HALFSPACE DEPTHS FOR SCATTER, CONCENTRATION AND SHAPE MATRICES

By Davy Paindaveine* and Germain Van Bever†

Université libre de Bruxelles

We propose halfspace depth concepts for scatter, concentration and shape matrices. For scatter matrices, our concept is similar to those from Chen, Gao and Ren (2017) and Zhang (2002). Rather than focusing, as in these earlier works, on deepest scatter matrices, we thoroughly investigate the properties of the proposed depth and of the corresponding depth regions. We do so under minimal assumptions and, in particular, we do not restrict to elliptical distributions nor to absolutely continuous distributions. Interestingly, fully understanding scatter halfspace depth requires considering different geometries/topologies on the space of scatter matrices. We also discuss, in the spirit of Zuo and Serfling (2000), the structural properties a scatter depth should satisfy, and investigate whether or not these are met by scatter halfspace depth. Companion concepts of depth for concentration matrices and shape matrices are also proposed and studied. We show the practical relevance of the depth concepts considered in a real-data example from finance.

1. Introduction. Statistical depth measures the centrality of a given location in $\mathbb{R}^k$ with respect to a sample of k-variate observations, or, more generally, with respect to a probability measure P over $\mathbb{R}^k$. The most famous depths include the halfspace depth (Tukey, 1975), the simplicial depth (Liu, 1990), the spatial depth (Vardi and Zhang, 2000) and the projection depth (Zuo, 2003). In the last decade, depth has also known much success in functional data analysis, where it measures the centrality of a function with respect to a sample of functional data. Some instances are the band depth (López-Pintado and Romo, 2009), the functional halfspace depth (Claeskens

*Research is supported by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy), the Crédit de Recherche J.0113.16 of the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique, and a grant from the National Bank of Belgium.

†Research is supported by the FC84444 grant of the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique.

MSC 2010 subject classifications: Primary 62H20; secondary 62G35

Keywords and phrases: curved parameter spaces, elliptical distributions, robustness, scatter matrices, shape matrices, statistical depth

et al., 2014) and the functional spatial depth (Chakraborty and Chaudhuri, 2014). The large variety of available depths made it necessary to introduce an axiomatic approach identifying the most desirable properties of a depth function; see Zuo and Serfling (2000) in the multivariate case and Nieto-Reyes and Battey (2016) in the functional one.

Statistical depth provides a center-outward ordering of the observations that allows to tackle in a robust and nonparametric way a broad range of inference problems; see Liu, Parelius and Singh (1999). For most depths, the deepest point is a robust location functional that extends the univariate median to the multivariate or functional setups; see, in particular, Cardot, Cénac and Godichon-Baggioni (2017) for a recent work on the functional spatial median. Beyond the median, depth plays a key role in the classical problem of defining multivariate quantiles; see, e.g., Hallin, Paindaveine and Šiman (2010) or Serfling (2010). In line with this, the collections of locations in $\mathbb{R}^k$ whose depth does not exceed a given level are sometimes called *quantile regions*; see, e.g., He and Einmahl (2017) in a multivariate extreme value theory framework. In the functional case, the quantiles in Chaudhuri (1996) may be seen as those associated with functional spatial depth; see Chakraborty and Chaudhuri (2014). Both in the multivariate and functional cases, supervised classification and outlier detection are standard applications of depth; we refer, e.g., to Cuevas, Febrero and Fraiman (2007), Paindaveine and Van Bever (2015), Dang and Serfling (2010), Hubert, Rousseeuw and Segaert (2015) and to the references therein.

In Mizera (2002), statistical depth was extended to a virtually arbitrary parametric framework. In a generic parametric model indexed by an ℓ-dimensional parameter ϑ, the resulting *tangent depth* $D_{P_n}(\vartheta_0)$ measures how appropriate a parameter value ϑ_0 is, with respect to the empirical measure P_n of a sample of k-variate observations $X_1, \ldots, X_n$ at hand, as one could alternatively do based on the likelihood $L_{P_n}(\vartheta_0)$. Unlike the MLE of ϑ, the depth-based estimator maximizing $D_{P_n}(\vartheta)$ is robust under mild conditions; see Section 4 of Mizera (2002). The construction, that for linear regression provides the Rousseeuw and Hubert (1999) depth, proved useful in various contexts. However, tangent depth requires evaluating the halfspace depth of a given location in $\mathbb{R}^\ell$, hence can only deal with low-dimensional parameters. In particular, tangent depth cannot cope with covariance or scatter matrix parameters ($\ell = k(k+1)/2$), unless k is as small as 2 or 3.

The crucial role played by scatter matrices in multivariate statistics, however, makes it highly desirable to have a satisfactory depth for such parameters, as phrased by Serfling (2004), that calls for an extension of the Mizera and Müller (2004) location-scale depth concept into a location-

scatter one. While computational issues prevent from basing this extension on tangent depth, a more ad hoc approach such as the one proposed in Zhang (2002) is suitable. Recently, another concept of scatter depth, that is very close in spirit to the one from Zhang (2002), was introduced in Chen, Gao and Ren (2017). Both proposals dominate tangent depth in the sense that, for k-variate observations, they rely on projection pursuit in $\mathbb{R}^k$ rather than in $\mathbb{R}^{k(k+1)/2}$, which allowed Chen, Gao and Ren (2017) to consider their depth even in high dimensions, under, e.g., sparsity assumptions. Both works, however, mainly focus on asymptotic, robustness and/or minimax convergence properties of the sample deepest scatter matrix. The properties of these scatter depths thus remain largely unknown, which severely affects the interpretation of the sample concepts.

In the present work, we consider a concept of halfspace depth for scatter matrices that is close to the Zhang (2002) and Chen, Gao and Ren (2017) ones. Unlike these previous works, however, we thoroughly study the properties of the scatter depth and of the corresponding depth regions. We do so under minimal assumptions and, in particular, we do not restrict to elliptical distributions nor to absolutely continuous distributions. Interestingly, fully understanding scatter halfspace depth requires considering different geometries/topologies on the space of scatter matrices. Like Donoho and Gasko (1992) and Rousseeuw and Ruts (1999) did for location halfspace depth, we study continuity and quasi-concavity properties of scatter halfspace depth, as well as the boundedness, convexity and compacity properties of the corresponding depth regions. Existence of a deepest halfspace scatter matrix, which is not guaranteed a priori, is also investigated. We further discuss, in the spirit of Zuo and Serfling (2000), the structural properties a scatter depth should satisfy and we investigate whether or not these are met by scatter halfspace depth. Moreover, companion concepts of depth for concentration matrices and shape matrices are proposed and studied. To the best of our knowledge, our results are the first providing structural and topological properties of depth regions outside the classical location framework. Throughout, numerical results illustrate our theoretical findings. Finally, we show the practical relevance of the depth concepts considered in a real-data example from finance.

The outline of the paper is as follows. In Section 2, we define scatter halfspace depth and investigate its affine invariance and uniform consistency. We also obtain explicit expressions of this depth for two distributions we will use as running examples in the paper. In Section 3, we derive the properties of scatter halfspace depth and scatter halfspace depth regions when considering the Frobenius topology on the space of scatter matrices, whereas we

do the same for the geodesic topology in Section 4. In Section 5, we identify the desirable properties a generic scatter depth should satisfy and investigate whether or not these are met by scatter halfspace depth. In Sections 6 and 7, we extend this depth to concentration and shape matrices, respectively. In Section 8, we treat a real-data example from finance. Final comments and perspectives for future work are provided in Section 9. Proofs and further numerical results are provided in the supplemental article Paindaveine and Van Bever (2017).

Before proceeding, we list here, for the sake of convenience, some notation to be used throughout. The collection of $k \times k$ matrices, $k \times k$ invertible matrices, and $k \times k$ symmetric matrices will be denoted as $\mathcal{M}_k$, GL_k, and $\mathcal{S}_k$, respectively (all matrices in this paper are real matrices). The identity matrix in $\mathcal{M}_k$ will be denoted as I_k. For any $A \in \mathcal{M}_k$, $\mathrm{diag}(A)$ will stand for the k-vector collecting the diagonal entries of A, whereas, for any k-vector v, $\mathrm{diag}(v)$ will stand for the diagonal matrix such that $\mathrm{diag}(\mathrm{diag}(v)) = v$. For $p \geq 2$ square matrices $A_1, \ldots, A_p$, $\mathrm{diag}(A_1, \ldots, A_p)$ will stand for the block-diagonal matrix with diagonal blocks $A_1, \ldots, A_p$. Any matrix A in $\mathcal{S}_k$ can be diagonalized into $A = O\,\mathrm{diag}(\lambda_1(A), \ldots, \lambda_k(A))\,O'$, where $\lambda_1(A) \geq \ldots \geq \lambda_k(A)$ are the eigenvalues of A and where the columns of the $k \times k$ orthogonal matrix $O = (v_1(A), \ldots, v_k(A))$ are corresponding unit eigenvectors (as usual, eigenvectors, and possibly eigenvalues, are only partly identified, but this will not play a role in the sequel). The spectral interval of A is $\mathrm{Sp}(A) := [\lambda_k(A), \lambda_1(A)]$. For any mapping $f : \mathbb{R} \to \mathbb{R}$, we let $f(A) = O\,\mathrm{diag}(f(\lambda_1(A)), \ldots, f(\lambda_k(A)))\,O'$. If Σ is a *scatter matrix*, in the sense that Σ belongs to the collection $\mathcal{P}_k$ of symmetric and positive definite $k \times k$ matrices, then this defines $\log(\Sigma)$ and Σ^t for any $t \in \mathbb{R}$. In particular, $\Sigma^{1/2}$ is the unique $A \in \mathcal{P}_k$ such that $\Sigma = AA'$, and $\Sigma^{-1/2}$ is the inverse of this symmetric and positive definite square root. Throughout, T will denote a location functional, that is, a function mapping a probability measure P to a real k-vector T_P. A location functional T is affine-equivariant if $T_{P_{A,b}} = AT_P + b$ for any $A \in GL_k$ and $b \in \mathbb{R}^k$, where the probability measure $P_{A,b}$ is the distribution of $AX + b$ when X has distribution P. A much weaker equivariance concept is centro-equivariance, for which $T_{P_{A,b}} = AT_P + b$ is imposed for $A = -I_k$ and $b = 0$ only. For a probability measure P over $\mathbb{R}^k$ and a location functional T, we will let $\alpha_{P,T} := \min(s_{P,T}, 1 - s_{P,T})$, where $s_{P,T} := \sup_{u \in \mathcal{S}^{k-1}} P[\{x \in \mathbb{R}^k : u'(x - T_P) = 0\}]$ involves the unit sphere $\mathcal{S}^{k-1} := \{x \in \mathbb{R}^k : \|x\|^2 = x'x = 1\}$ of $\mathbb{R}^k$. We will say that P is *smooth at* $\theta (\in \mathbb{R}^k)$ if the P-probability of any hyperplane of $\mathbb{R}^k$ containing θ is zero and that it is *smooth* if it is smooth at any θ. Finally, $\overset{\mathcal{D}}{=}$ will denote equality in distribution.

2. Scatter halfspace depth. We start by recalling the classical concept of location halfspace depth. To do so, let P be a probability measure over $\mathbb{R}^k$ and X be a random k-vector with distribution P, which allows us throughout to write $P[X \in B]$ instead of $P[B]$ for any k-Borel set B. The *location halfspace depth* of $\theta(\in \mathbb{R}^k)$ with respect to P is then

$$HD_P^{\text{loc}}(\theta) := \inf_{u \in \mathcal{S}^{k-1}} P[u'(X-\theta) \geq 0].$$

The corresponding *depth regions* $R_P^{\text{loc}}(\alpha) := \{\theta \in \mathbb{R}^k : HD_P^{\text{loc}}(\theta) \geq \alpha\}$ form a nested family of closed convex subsets of $\mathbb{R}^k$. The innermost depth region, namely $M_P^{\text{loc}} := \{\theta \in \mathbb{R}^k : HD_P^{\text{loc}}(\theta) = \max_{\eta \in \mathbb{R}^k} HD_P^{\text{loc}}(\eta)\}$ (the maximum always exists; see, e.g., Proposition 7 in Rousseeuw and Ruts, 1999), is a set-valued location functional. When a unique representative of M_P^{loc} is needed, it is customary to consider the *Tukey median* θ_P of P, that is defined as the barycenter of M_P^{loc}. The Tukey median has maximal depth (which follows from the convexity of M_P^{loc}) and is an affine-equivariant location functional.

In this paper, for a location functional T, we define the *T-scatter halfspace depth* of $\Sigma(\in \mathcal{P}_k)$ with respect to P as

$$HD_{P,T}^{\text{sc}}(\Sigma) := \inf_{u \in \mathcal{S}^{k-1}} \min\big(P\big[|u'(X-T_P)| \leq \sqrt{u'\Sigma u}\,\big], \\ P\big[|u'(X-T_P)| \geq \sqrt{u'\Sigma u}\,\big]\big). \tag{2.1}$$

This extends to a probability measure with arbitrary location the centered matrix depth concept from Chen, Gao and Ren (2017). If P is smooth, then the depth in (2.1) is also equivalent to the (Tukey version of) the dispersion depth introduced in Zhang (2002), but for the fact that the latter, in the spirit of projection depth, involves centering through a univariate location functional (both Zhang (2002) and Chen, Gao and Ren (2017) also propose bypassing centering through a pairwise difference approach that will be discussed in Section 9). While they were not considered in these prior works, it is of interest to introduce the corresponding depth regions

$$R_{P,T}^{\text{sc}}(\alpha) := \big\{\Sigma \in \mathcal{P}_k : HD_{P,T}^{\text{sc}}(\Sigma) \geq \alpha\big\}, \quad \alpha \geq 0. \tag{2.2}$$

We will refer to $R_{P,T}^{\text{sc}}(\alpha)$ as the *order-α (T-scatter halfspace) depth region of P*. Obviously, one always has $R_{P,T}^{\text{sc}}(0) = \mathcal{P}_k$. Clearly, the concepts in (2.1)-(2.2) give practitioners the flexibility to freely choose the location functional T; numerical results below, however, will focus on the depth $HD_P^{\text{sc}}(\Sigma)$ and on the depth regions $R_P^{\text{sc}}(\alpha)$ based on the Tukey median θ_P, that is the natural location functional whenever halfspace depth objects are considered.

To get a grasp of the scatter depth $HD_P^{\mathrm{sc}}(\Sigma)$, it is helpful to start with the univariate case $k = 1$. There, the location halfspace deepest region is the "median interval" $M_P^{\mathrm{loc}} = \arg\max_{\theta\in\mathbb{R}} \min(P[X \le \theta], P[X \ge \theta])$ and the Tukey median θ_P, that is, the midpoint of M_P^{loc}, is the usual representative of the univariate median. The scatter halfspace deepest region is then the median interval $M_P^{\mathrm{sc}} := \arg\max_{\Sigma\in\mathbb{R}_0^+} \min(P[(X-\theta_P)^2 \le \Sigma], P[(X-\theta_P)^2 \ge \Sigma])$ of $(X-\theta_P)^2$; call it the *median squared deviation* interval $\mathcal{I}_{\mathrm{MSD}}[X]$ (or $\mathcal{I}_{\mathrm{MSD}}[P]$) of $X \sim P$. Below, parallel to what is done for the median, $\mathrm{MSD}[X]$ (or $\mathrm{MSD}[P]$) will denote the midpoint of this MSD interval. In particular, if $\mathcal{I}_{\mathrm{MSD}}[P]$ is a singleton, then scatter halfspace depth is uniquely maximized at $\Sigma = \mathrm{MSD}[P] = (\mathrm{MAD}[P])^2$, where $\mathrm{MAD}[P]$ denotes the median absolute deviation of P. Obviously, the depth regions $R_P^{\mathrm{sc}}(\alpha)$ form a family of nested intervals, $[\Sigma_\alpha^-, \Sigma_\alpha^+]$ say, included in $\mathcal{P}_1 = \mathbb{R}_0^+$. It is easy to check that, if P is symmetric about zero with an invertible cumulative distribution function F and if T is centro-equivariant, then

$$HD_P^{\mathrm{sc}}(\Sigma) = HD_{P,T}^{\mathrm{sc}}(\Sigma) = 2\min\left(F(\sqrt{\Sigma}) - \tfrac{1}{2}, 1 - F(\sqrt{\Sigma})\right) \quad \text{and} \tag{2.3}$$

$$R_P^{\mathrm{sc}}(\alpha) = R_{P,T}^{\mathrm{sc}}(\alpha) = \left[(F^{-1}(\tfrac{1}{2}+\tfrac{\alpha}{2}))^2, (F^{-1}(1-\tfrac{\alpha}{2}))^2\right]. \tag{2.4}$$

This is compatible with the fact that the maximal value of $\Sigma \mapsto HD_P^{\mathrm{sc}}(\Sigma)$ (that is equal to $1/2$) is achieved at $\Sigma = (\mathrm{MAD}[P])^2$ only.

For $k > 1$, elliptical distributions provide an important particular case. We will say that $P = P^X$ is k-variate elliptical with location $\theta(\in \mathbb{R}^k)$ and scatter $\Sigma(\in \mathcal{P}_k)$ if and only if $X \overset{\mathcal{D}}{=} \theta + \Sigma^{1/2}Z$, where $Z = (Z_1, \ldots, Z_k)'$ is (i) spherically symmetric about the origin of $\mathbb{R}^k$ (that is, $OZ \overset{\mathcal{D}}{=} Z$ for any $k \times k$ orthogonal matrix O) and is (ii) standardized in such a way that $\mathrm{MSD}[Z_1] = 1$ (one then has $T_P = \theta$ for any affine-equivariant location functional T). Denoting by Φ the cumulative distribution function of the standard normal, the k-variate normal distribution with location zero and scatter I_k is then the distribution of $X := W/b$, where $b := \Phi^{-1}(\frac{3}{4})$ and W is a standard normal random k-vector. In this Gaussian case, we obtain

$$\begin{aligned} HD_{P,T}^{\mathrm{sc}}(\Sigma) &= \inf_{u\in\mathcal{S}^{k-1}} \min\left(P\left[|u'X| \le \sqrt{u'\Sigma u}\,\right], P\left[|u'X| \ge \sqrt{u'\Sigma u}\,\right]\right) \\ &= 2\min\left(\Phi\big(b\lambda_k^{1/2}(\Sigma)\big) - \tfrac{1}{2}, 1 - \Phi\big(b\lambda_1^{1/2}(\Sigma)\big)\right). \end{aligned} \tag{2.5}$$

One can check directly that $HD_{P,T}^{\mathrm{sc}}(\Sigma) \le HD_{P,T}^{\mathrm{sc}}(I_k) = 1/2$, with equality if and only if Σ coincides with the "true" scatter matrix I_k (we refer to Theorem 5.1 for a more general result). Also, Σ belongs to the depth region $R_{P,T}^{\mathrm{sc}}(\alpha)$ if and only if $\mathrm{Sp}(\Sigma) \subset [(\frac{1}{b}\Phi^{-1}(\frac{1}{2}+\frac{\alpha}{2}))^2, (\frac{1}{b}\Phi^{-1}(1-\frac{\alpha}{2}))^2]$.

Provided that the location functional used is affine-equivariant, extension to an arbitrary multinormal is based on the following affine-invariance result, which ensures in particular that scatter halfspace depth will not be affected by possible changes in the marginal measurement units (a similar result is stated in Zhang (2002) for the dispersion depth concept considered there).

THEOREM 2.1. *Let T be an affine-equivariant location functional. Then, (i) scatter halfspace depth is affine-invariant in the sense that, for any probability measure P over $\mathbb{R}^k$, $\Sigma \in \mathcal{P}_k$, $A \in GL_k$ and $b \in \mathbb{R}^k$, we have $HD^{\mathrm{sc}}_{P_{A,b},T}(A\Sigma A') = HD^{\mathrm{sc}}_{P,T}(\Sigma)$, where $P_{A,b}$ is as defined on page 4. Consequently, (ii) the regions $R^{\mathrm{sc}}_{P,T}(\alpha)$ are affine-equivariant, in the sense that, for any probability measure P over $\mathbb{R}^k$, $\alpha \geq 0$, $A \in GL_k$ and $b \in \mathbb{R}^k$, we have $R^{\mathrm{sc}}_{P_{A,b},T}(\alpha) = AR^{\mathrm{sc}}_{P,T}(\alpha)A'$.*

This result readily entails that if P is the k-variate normal with location θ_0 and scatter Σ_0, then, provided that T is affine-equivariant,

$$HD^{\mathrm{sc}}_{P,T}(\Sigma) = 2\min\Big(\Phi\big(b\lambda_k^{1/2}(\Sigma_0^{-1}\Sigma)\big) - \tfrac{1}{2}, 1 - \Phi\big(b\lambda_1^{1/2}(\Sigma_0^{-1}\Sigma)\big)\Big) \tag{2.6}$$

and $R^{\mathrm{sc}}_{P,T}(\alpha)$ is the collection of scatter matrices Σ for which $\mathrm{Sp}(\Sigma_0^{-1}\Sigma) \subset [(\frac{1}{b}\Phi^{-1}(\frac{1}{2}+\frac{\alpha}{2}))^2, (\frac{1}{b}\Phi^{-1}(1-\frac{\alpha}{2}))^2]$. For a non-Gaussian elliptical probability measure P with location θ_0 and scatter Σ_0, it is easy to show that $HD^{\mathrm{sc}}_{P,T}(\Sigma)$ will still depend on Σ only through $\lambda_1(\Sigma_0^{-1}\Sigma)$ and $\lambda_k(\Sigma_0^{-1}\Sigma)$.

As already mentioned, we also intend to consider non-elliptical probability measures. A running non-elliptical example will be the one for which P is the distribution of a random vector $X = (X_1, \ldots, X_k)'$ with independent Cauchy marginals. If T is centro-equivariant, then

$$HD^{\mathrm{sc}}_{P,T}(\Sigma) = 2\min\Big(\Psi\big(1/\max_s \sqrt{s'\Sigma^{-1}s}\,\big) - \tfrac{1}{2}, 1 - \Psi\big(\sqrt{\max(\mathrm{diag}(\Sigma))}\big)\Big), \tag{2.7}$$

where Ψ is the Cauchy cumulative distribution function and where the maximum in s is over all sign vectors $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$; see the supplemental article Paindaveine and Van Bever (2017) for a proof. For $k = 1$, this simplifies to $HD^{\mathrm{sc}}_{P,T}(\Sigma) = 2\min\big(\Psi\big(\sqrt{\Sigma}\,\big) - \frac{1}{2}, 1 - \Psi\big(\sqrt{\Sigma}\big)\big)$, which agrees with (2.3). For $k = 2$, we obtain

$$HD^{\mathrm{sc}}_{P,T}(\Sigma) = 2\min\Big(\Psi\big(\sqrt{\det(\Sigma)/s_\Sigma}\,\big) - \tfrac{1}{2}, 1 - \Psi\big(\sqrt{\max(\Sigma_{11}, \Sigma_{22})}\big)\Big),$$

where we let $s_\Sigma := \Sigma_{11} + \Sigma_{22} + 2|\Sigma_{12}|$. For a general k, a scatter matrix Σ belongs to $R^{\mathrm{sc}}_{P,T}(\alpha)$ if and only if $1/(s'\Sigma^{-1}s) \geq (\Psi^{-1}(\frac{1}{2}+\frac{\alpha}{2}))^2$ for all

$s \in \{-1,1\}^k$ and $\Sigma_{\ell\ell} \leq (\Psi^{-1}(1-\frac{\alpha}{2}))^2$ for all $\ell = 1,\ldots,k$. The problem of identifying the scatter matrix achieving maximal depth, if any (existence is not guaranteed), will be considered in Section 4. Figure 1 plots scatter halfspace depth regions in the Gaussian and independent Cauchy cases above. Examples involving distributions that are not absolutely continuous with respect to the Lebesgue measure will be considered in the next sections.

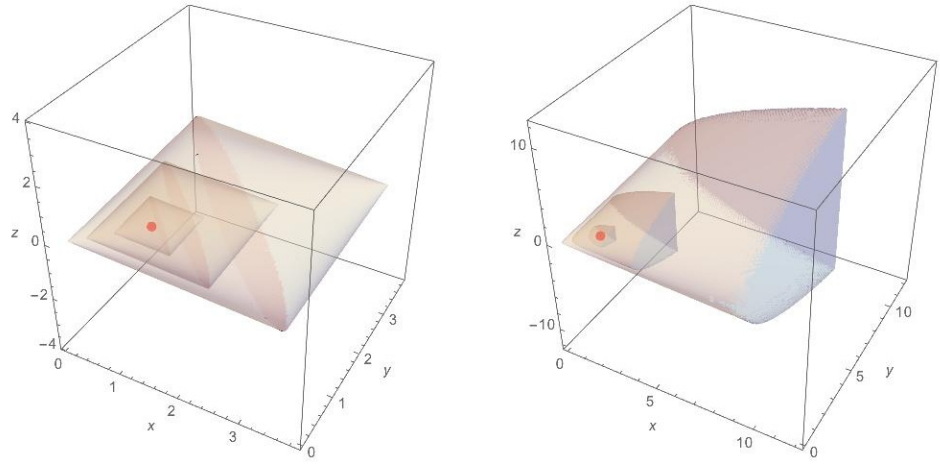

FIG 1. *Level sets of order $\alpha = .2, .3$ and .4, for any centro-symmetric T, of $(x,y,z) \mapsto HD^{\text{sc}}_{P,T}(\Sigma_{x,y,z})$, where $HD^{\text{sc}}_{P,T}(\Sigma_{x,y,z})$ is the T-scatter halfspace depth of $\Sigma_{x,y,z} = \left(\begin{smallmatrix} x & z \\ z & y \end{smallmatrix}\right)$ with respect to two probability measures P, namely the bivariate multinormal distribution with location zero and scatter I_2 (left) and the bivariate distribution with independent Cauchy marginals (right). The red points are those associated with I_2 (left) and $\sqrt{2}I_2$ (right), which are the corresponding deepest scatter matrices (see Sections 4 and 5).*

In the supplemental article Paindaveine and Van Bever (2017), we validate through a Monte Carlo exercise the expressions for $HD^{\text{sc}}_{P,T}(\Sigma)$ obtained in (2.6)-(2.7) above. Such a numerical validation is justified by the following uniform consistency result; see (6.2) and (6.6) in Donoho and Gasko (1992) for the corresponding location halfspace depth result, and Proposition 2.2(ii) in Zhang (2002) for the dispersion depth concept considered there.

THEOREM 2.2. *Let P be a smooth probability measure over $\mathbb{R}^k$ and T be a location functional. Let P_n denote the empirical probability measure associated with a random sample of size n from P and assume that $T_{P_n} \to T_P$ almost surely as $n \to \infty$. Then $\sup_{\Sigma\in\mathcal{P}_k} |HD^{\text{sc}}_{P_n,T}(\Sigma) - HD^{\text{sc}}_{P,T}(\Sigma)| \to 0$ almost surely as $n \to \infty$.*

This result applies in particular to the scatter halfspace depth $HD^{\mathrm{sc}}_P(\Sigma)$, as the Tukey median is strongly consistent without any assumption on P (for completeness, we show this in the supplemental article Paindaveine and Van Bever, 2017). Inspection of the proof of Theorem 2.2 reveals that the smoothness assumption is only needed to control the estimation of T_P, hence is superfluous when a constant location functional is used. This is relevant when the location is fixed, as in Chen, Gao and Ren (2017).

3. Frobenius topology. Our investigation of the further structural properties of the scatter halfspace depth $HD^{\mathrm{sc}}_{P,T}(\Sigma)$ and of the corresponding depth regions $R^{\mathrm{sc}}_{P,T}(\alpha)$ depends on the topology that is considered on $\mathcal{P}_k$. In this section, we focus on the topology induced by the *Frobenius metric space* $(\mathcal{P}_k, d_F)$, where $d_F(\Sigma_a, \Sigma_b) = \|\Sigma_b - \Sigma_a\|_F$ is the distance on $\mathcal{P}_k$ that is inherited from the Frobenius norm $\|A\|_F = \sqrt{\mathrm{tr}[AA']}$ on $\mathcal{M}_k$. The resulting *Frobenius topology* (or simply *F-topology*), generated by the F-balls $B_F(\Sigma_0, r) := \{\Sigma \in \mathcal{P}_k : d_F(\Sigma, \Sigma_0) < r\}$ with center Σ_0 and radius r, gives a precise meaning to what we call below *F-continuous* functions on $\mathcal{P}_k$, *F-open/F-closed* subsets of $\mathcal{P}_k$, etc. We then have the following result.

THEOREM 3.1. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is upper F-semicontinuous on $\mathcal{P}_k$, so that (ii) the depth region $R^{\mathrm{sc}}_{P,T}(\alpha)$ is F-closed for any $\alpha \geq 0$. (iii) If P is smooth at T_P, then $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is F-continuous on $\mathcal{P}_k$.*

For location halfspace depth, the corresponding result was derived in Lemma 6.1 of Donoho and Gasko (1992), where the metric on $\mathbb{R}^k$ is the Euclidean one. The similarity between the location and scatter halfspace depths also extends to the boundedness of depth regions, in the sense that, like location halfspace depth (Proposition 5 in Rousseeuw and Ruts, 1999), the order-α scatter halfspace depth region is bounded if and only if $\alpha > 0$.

THEOREM 3.2. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, for any $\alpha > 0$, $R^{\mathrm{sc}}_{P,T}(\alpha)$ is F-bounded (that is, it is included, for some $r > 0$, in the F-ball $B_F(I_k, r)$).*

This shows that, for any probability measure P, $HD^{\mathrm{sc}}_{P,T}(\Sigma)$ goes to zero as $\|\Sigma\|_F \to \infty$. Since $\|\Sigma\|_F \geq \lambda_1(\Sigma)$, this means that explosion of Σ (that is, $\lambda_1(\Sigma) \to \infty$) leads to arbitrarily small depth, which is confirmed in the multinormal case in (2.5). In this Gaussian case, however, implosion of Σ (that is, $\lambda_k(\Sigma) \to 0$) also provides arbitrarily small depth, but this is not captured by the general result in Theorem 3.2 (similar comments

can be given for the independent Cauchy example in (2.7)). Irrespective of the topology adopted (so that the F-topology is not to be blamed for this behavior), it is actually possible to have implosion without depth going to zero. We show this by considering the following example. Let $P = (1-s)P_1 + sP_2$, where $s \in (\frac{1}{2}, 1)$, P_1 is the bivariate standard normal and P_2 is the distribution of $\binom{0}{Z}$, where Z is univariate standard normal. Then, it can be showed that, for $\Sigma_n := \left(\begin{smallmatrix} 1/n & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and any centro-equivariant T, we have $HD^{\text{sc}}_{P,T}(\Sigma_n) \to 1 - s > 0$ as $n \to \infty$.

In the metric space $(\mathcal{P}_k, d_F)$, any bounded set is also *totally bounded*, that is, can be covered, for any $\varepsilon > 0$, by finitely many balls of the form $B_F(\Sigma, \varepsilon)$. Theorems 3.1-3.2 thus show that, for any $\alpha > 0$, $R^{\text{sc}}_{P,T}(\alpha)$ is both F-closed and totally F-bounded. However, since $(\mathcal{P}_k, d_F)$ is not complete, there is no guarantee that these regions are F-compact. Actually, these regions may fail to be F-compact, as we show through the example from the previous paragraph. For any $\alpha \in (0, 1-s)$, the scatter matrix Σ_n belongs to $R^{\text{sc}}_{P,T}(\alpha)$ for n large enough. However, the sequence (Σ_n) F-converges to $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, that does not belong to $R^{\text{sc}}_{P,T}(\alpha)$ (since it does not even belong to $\mathcal{P}_2$). Since this will also hold for any subsequence of (Σ_n), we conclude that, for $\alpha \in (0, 1-s)$, $R^{\text{sc}}_{P,T}(\alpha)$ is not F-compact in this example. This provides a first discrepancy between location and scatter halfspace depths, since location halfspace depth regions associated with a positive order α are always compact.

The lack of compacity of scatter halfspace depth regions may allow for probability measures for which no halfspace deepest scatter exists. This is actually the case in the bivariate mixture example above. There, letting $e_1 = (1, 0)'$ and assuming again that T is centro-equivariant, any $\Sigma \in \mathcal{P}_2$ indeed satisfies $HD^{\text{sc}}_{P,T}(\Sigma) \leq P[|e_1'X| \geq \sqrt{e_1'\Sigma e_1}] = P[|X_1| \geq \sqrt{\Sigma_{11}}] = (1-s)P[|Z| \geq \sqrt{\Sigma_{11}}] < 1 - s = \sup_{\Sigma \in \mathcal{P}_2} HD^{\text{sc}}_{P,T}(\Sigma)$, where the last equality follows from the fact that we identified a sequence (Σ_n) such that $HD^{\text{sc}}_{P,T}(\Sigma_n) \to 1 - s$. This is again in sharp contrast with the location case, for which a halfspace deepest location always exists; see, e.g., Propositions 5 and 7 in Rousseeuw and Ruts (1999). Identifying sufficient conditions under which a halfspace deepest scatter exists requires considering another topology, namely the *geodesic topology* considered in Section 4 below.

The next result states that scatter halfspace depth is a quasi-concave function, which ensures convexity of the corresponding depth regions; we refer to Proposition 1 (and to its corollary) in Rousseeuw and Ruts (1999) for the corresponding results on location halfspace depth.

THEOREM 3.3. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Sigma \mapsto HD^{\text{sc}}_{P,T}(\Sigma)$ is quasi-concave, in the sense that,*

for any $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ *and* $t \in [0,1]$, $HD^{sc}_{P,T}(\Sigma_t) \geq \min(HD^{sc}_{P,T}(\Sigma_a), HD^{sc}_{P,T}(\Sigma_b))$, *where we let* $\Sigma_t := (1-t)\Sigma_a + t\Sigma_b$; *(ii) for any* $\alpha \geq 0$, $R^{sc}_{P,T}(\alpha)$ *is convex.*

Strictly speaking, Theorem 3.3 is not directly related to the F-topology considered on $\mathcal{P}_k$. Yet we state the result in this section due to the link between the linear paths $t \mapsto \Sigma_t = (1-t)\Sigma_a + t\Sigma_b$ it involves and the "flat" nature of the F-topology (this link will become clearer below when we will compare with what occurs for the geodesic topology). Illustration of Theorem 3.3 will be provided in Figure 2 below, as well as in the supplemental article Paindaveine and Van Bever (2017).

4. Geodesic topology. Equipped with the inner product $<A, B> = \text{tr}[A'B]$, $\mathcal{M}_k$ is a Hilbert space. The resulting norm and distance are the Frobenius ones considered in the previous section. As an open set in $\mathcal{S}_k$, the parameter space $\mathcal{P}_k$ of interest is a differentiable manifold of dimension $k(k+1)/2$. The corresponding tangent space at Σ, which is isomorphic (via translation) to $\mathcal{S}_k$, can be equipped with the inner product $<A, B> = \text{tr}[\Sigma^{-1}A\Sigma^{-1}B]$. This leads to considering $\mathcal{P}_k$ as a Riemannian manifold, with the metric at Σ given by the differential $ds = \|\Sigma^{-1/2} d\Sigma\, \Sigma^{-1/2}\|_F$; see, e.g., Bhatia (2007). The length of a path $\gamma : [0,1] \to \mathcal{P}_k$ is then given by

$$L(\gamma) = \int_0^1 \left\| \gamma^{-1/2}(t) \frac{d\gamma(t)}{dt} \gamma^{-1/2}(t) \right\|_F dt.$$

The resulting geodesic distance between $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ is defined as

$$d_g(\Sigma_a, \Sigma_b) := \inf \left\{ L(\gamma) : \gamma \in \mathcal{G}(\Sigma_a, \Sigma_b) \right\} = \| \log(\Sigma_a^{-1/2} \Sigma_b \Sigma_a^{-1/2}) \|_F, \tag{4.1}$$

where $\mathcal{G}(\Sigma_a, \Sigma_b)$ denotes the collection of paths γ from $\gamma(0) = \Sigma_a$ to $\gamma(1) = \Sigma_b$ (the second equality in (4.1) is Theorem 6.1.6 in Bhatia, 2007). It directly follows from the definition of $d_g(\Sigma_a, \Sigma_b)$ that the geodesic distance satisfies the triangle inequality. Theorem 6.1.6 in Bhatia (2007) also states that all paths γ achieving the infimum in (4.1) provide the same geodesic $\{\gamma(t) : t \in [0,1]\}$ joining Σ_a and Σ_b, and that this geodesic can be parametrized as

$$\gamma(t) = \tilde{\Sigma}_t := \Sigma_a^{1/2} \left(\Sigma_a^{-1/2} \Sigma_b \Sigma_a^{-1/2} \right)^t \Sigma_a^{1/2}, \qquad t \in [0,1]. \tag{4.2}$$

By using the explicit formula in (4.1), it is easy to check that this particular parametrization of this unique geodesic is natural in the sense that $d_g(\Sigma_a, \tilde{\Sigma}_t) = t d_g(\Sigma_a, \Sigma_b)$ for any $t \in [0,1]$.

Below, we consider the natural topology associated with the metric space $(\mathcal{P}_k, d_g)$, that is, the topology whose open sets are generated by geodesic balls

of the form $B_g(\Sigma_0, r) := \{\Sigma \in \mathcal{P}_k : d_g(\Sigma, \Sigma_0) < r\}$. This topology — call it the *geodesic topology*, or simply *g-topology* — defines subsets of $\mathcal{P}_k$ that are g-open, g-closed, g-compact, and functions that are g-semicontinuous, g-continuous, etc. We will say that a subset R of $\mathcal{P}_k$ is g-bounded if and only if $R \subset B_g(I_k, r)$ for some $r > 0$ (we can safely restrict to balls centered at I_k since the triangle inequality guarantees that R is included in a finite-radius g-ball centered at I_k if and only if it is included in a finite-radius g-ball centered at an arbitrary $\Sigma_0 \in \mathcal{P}_k$). A g-bounded subset of $\mathcal{P}_k$ is also totally g-bounded, still in the sense that, for any $\varepsilon > 0$, it can be covered by finitely many balls of the form $B_g(\Sigma, \varepsilon)$; for completeness, we prove this in Lemma S.2.6 from the supplemental article Paindaveine and Van Bever (2017). Since $(\mathcal{P}_k, d_g)$ is complete (see, e.g., Proposition 10 in Bhatia and Holbrook, 2006), a g-bounded and g-closed subset of $\mathcal{P}_k$ is then g-compact.

We omit the proof of the next result as it follows along the exact same lines as the proof of Theorem 3.1, once it is seen that a sequence (Σ_n) converging to Σ_0 in $(\mathcal{P}_k, d_g)$ also converges to Σ_0 in $(\mathcal{P}_k, d_F)$.

THEOREM 4.1. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Sigma \mapsto HD^{\text{sc}}_{P,T}(\Sigma)$ is upper g-semicontinuous on $\mathcal{P}_k$, so that (ii) the depth region $R^{\text{sc}}_{P,T}(\alpha)$ is g-closed for any $\alpha \geq 0$. (iii) If P is smooth at T_P, then $\Sigma \mapsto HD^{\text{sc}}_{P,T}(\Sigma)$ is g-continuous on $\mathcal{P}_k$.*

The following result uses the notation $s_{P,T} := \sup_{u \in \mathcal{S}^{k-1}} P[u'(X - T_P) = 0]$ and $\alpha_{P,T} := \min(s_{P,T}, 1 - s_{P,T})$ defined in the introduction.

THEOREM 4.2. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, for any $\alpha > \alpha_{P,T}$, $R^{\text{sc}}_{P,T}(\alpha)$ is g-bounded, hence g-compact (if $s_{P,T} \geq 1/2$, then this result is trivial in the sense that $R^{\text{sc}}_{P,T}(\alpha)$ is empty for any $\alpha > \alpha_{P,T}$). In particular, if P is smooth at T_P, then $R^{\text{sc}}_{P,T}(\alpha)$ is g-compact for any $\alpha > 0$.*

This result complements Theorem 3.2 by showing that implosion always leads to a depth that is smaller than or equal to $\alpha_{P,T}$. In particular, in the multinormal and independent Cauchy examples in Section 2, this shows that both explosion and implosion lead to arbitrarily small depth, whereas Theorem 3.2 was predicting this collapsing for explosion only. Therefore, while the behavior of $HD^{\text{sc}}_{P,T}(\Sigma)$ under implosion/explosion of Σ is independent of the topology adopted, the use of the g-topology provides a better understanding of this behavior than the F-topology.

It is not possible to improve the result in Theorem 4.2, in the sense that $R^{\text{sc}}_{P,T}(\alpha_{P,T})$ may fail to be g-bounded. For instance, consider the probability measure P over $\mathbb{R}^2$ putting probability mass $1/6$ on each of the six

points $(0, \pm 1/2)$ and $(\pm 2, \pm 2)$, and let T be a centro-equivariant location functional. Clearly, $\alpha_{P,T} = s_{P,T} = 1/3$. Now, letting $\Sigma_n := \left(\begin{smallmatrix} 1/n & 0 \\ 0 & 1 \end{smallmatrix}\right)$, we have $P[|u'\Sigma_n^{-1/2}X| \leq 1] \geq 1/3$ and $P[|u'\Sigma_n^{-1/2}X| \geq 1] \geq 1/3$ for any $u \in \mathcal{S}^1$ (here, X is a random vector with distribution P), which entails that

$$HD^{\mathrm{sc}}_{P,T}(\Sigma_n) = \inf_{u\in\mathcal{S}^1} \min\left(P\left[|u'X| \leq \sqrt{u'\Sigma_n u}\right], P\left[|u'X| \geq \sqrt{u'\Sigma_n u}\right]\right)$$

$$= \inf_{u\in\mathcal{S}^1} \min\left(P[|u'\Sigma_n^{-1/2}X| \leq 1], P[|u'\Sigma_n^{-1/2}X| \leq 1]\right) \geq \frac{1}{3} = \alpha_{P,T},$$

so that $\Sigma_n \in R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ for any n. Since $d_g(\Sigma_n, I_2) \to \infty$, $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ is indeed g-unbounded.

An important benefit of working with the g-topology is that, unlike the F-topology, it allows to show that, under mild assumptions, a halfspace deepest scatter does exist. More precisely, we have the following result.

THEOREM 4.3. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Assume that $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ is non-empty. Then, $\alpha_{*P,T} := \sup_{\Sigma\in\mathcal{P}_k} HD^{\mathrm{sc}}_{P,T}(\Sigma) = HD^{\mathrm{sc}}_{P,T}(\Sigma_*)$ for some $\Sigma_* \in \mathcal{P}_k$.*

In particular, this result shows that for any probability measure P that is smooth at T_P, there exists a halfspace deepest scatter Σ_*. For the k-variate multinormal distribution with location zero and scatter I_k (and any centro-equivariant T), we already stated in Section 2 that $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is uniquely maximized at $\Sigma_* = I_k$, with a corresponding maximal depth equal to $1/2$. The next result identifies the halfspace deepest scatter (and the corresponding maximal depth) in the independent Cauchy case.

THEOREM 4.4. *Let P be the k-variate probability measure with independent Cauchy marginals and let T be a centro-equivariant location functional. Then, $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is uniquely maximized at $\Sigma_* = \sqrt{k}I_k$, and the corresponding maximal depth is $HD^{\mathrm{sc}}_{P,T}(\Sigma_*) = \frac{2}{\pi}\arctan\left(k^{-1/4}\right)$.*

For $k = 1$, the Cauchy distribution in this result is symmetric (hence, elliptical) about zero, which is compatible with the maximal depth being equal to $1/2$ there (Theorem 5.1 below shows that the maximal depth for absolutely continuous elliptical distributions is always equal to $1/2$). For larger values of k, however, this provides an example where the maximal depth is strictly smaller than $1/2$. Interestingly, this maximal depth goes (monotonically) to zero as $k \to \infty$. Note that, for the same distribution, location halfspace depth has, irrespective of k, maximal value $1/2$ (this follows, e.g., from Lemma 1 and Theorem 1 in Rousseeuw and Struyf, 2004).

In general, the halfspace deepest scatter Σ_* is not unique. This is typically the case for empirical probability measures P_n (note that the existence of a halfspace deepest scatter in the empirical case readily follows from the fact that $HD^{\mathrm{sc}}_{P_n,T}(\Sigma)$ takes its values in $\{\ell/n : \ell = 0, 1, \ldots, n\}$). For several purposes, it is needed to identify a unique representative of the halfspace deepest scatters, that would play a similar role for scatter as the one played by the Tukey median for location. To this end, one may consider here a *center of mass*, that is, a scatter matrix of the form

$$\Sigma_{P,T} := \arg\min_{\Sigma\in\mathcal{P}_k} \int_{R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T})} d_g^2(m,\Sigma)\, dm, \tag{4.3}$$

where dm is a mass distribution on $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T})$ with total mass one (the natural choice being the uniform over $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T})$). This is a suitable solution if $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T})$ is g-bounded (hence, g-compact), since Cartan (1929) showed that, in a simply connected manifold with non-positive curvature (as $\mathcal{P}_k$), every compact set has a unique center of mass; see also Proposition 60 in Berger (2003). Convexity of $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T})$ then ensures that $\Sigma_{P,T}$ has maximal depth. Like for location, this choice of $\Sigma_{P,T}$ as a representative of the deepest scatters guarantees affine equivariance (in the sense that $\Sigma_{P_{A,b},T} = A\Sigma_{P,T}A'$ for any $A \in GL_k$ and any $b \in \mathbb{R}^k$), provided that T itself is affine-equivariant. An alternative approach is to consider the scatter matrix $\Sigma_{P,T}$ whose vectorized form $\mathrm{vec}\,\Sigma_{P,T}$ is the barycenter of $\mathrm{vec}\, R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T})$. While this is a more practical solution for scatter matrices, the non-flat nature of some of the parameter spaces in Section 7 will require the more involved, manifold-type, approach in (4.3).

As a final comment related to Theorem 4.3, note that if $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ is empty, then it may actually be so that no halfspace deepest scatter does exist. An example is provided by the bivariate mixture distribution P in Section 3. There, we saw that, for any centro-equivariant T, no halfspace deepest scatter does exist, which is compatible with the fact that, for any Σ, $HD^{\mathrm{sc}}_{P,T}(\Sigma) < 1 - s = \alpha_{P,T}$, so that $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ is empty.

5. An axiomatic approach for scatter depth. Building on the properties derived in Liu (1990) for simplicial depth, Zuo and Serfling (2000) introduced an axiomatic approach suggesting that a generic location depth $D^{\mathrm{loc}}_P(\,\cdot\,) : \mathbb{R}^k \to [0,1]$ should satisfy the following properties: (P1) affine invariance, (P2) maximality at the symmetry center (if any), (P3) monotonicity relative to any deepest point, and (P4) vanishing at infinity. Without entering into details, these properties are to be understood as follows: (P1) means that $D^{\mathrm{loc}}_{P_{A,b}}(A\theta + b) = D^{\mathrm{loc}}_P(\theta)$ for any $A \in GL_k$ and $b \in \mathbb{R}^k$,

where $P_{A,b}$ is as defined on page 4; (P2) states that if P is symmetric (in some sense), then the symmetry center should maximize $D^{\text{loc}}_P(\cdot)$; according to (P3), $D^{\text{loc}}_P(\cdot)$ should be monotone non-increasing along any halfline originating from any P-deepest point; finally, (P4) states that as θ exits any compact set in $\mathbb{R}^k$, its depth should converge to zero. There is now an almost universal agreement in the literature that (P1)-(P4) are the natural desirable properties for location depths.

In view of this, one may wonder what are the desirable properties for a scatter depth. Inspired by (P1)-(P4), we argue that a generic scatter depth $D^{\text{sc}}_P(\cdot) : \mathcal{P}_k \to [0,1]$ should satisfy the following properties, all involving an (unless otherwise specified) arbitrary probability measure P over $\mathbb{R}^k$:

(Q1) *Affine invariance*: for any $A \in GL_k$ and $b \in \mathbb{R}^k$, $D^{\text{sc}}_{P_{A,b}}(A\Sigma A') = D^{\text{sc}}_P(\Sigma)$, where $P_{A,b}$ is still as defined on page 4;

(Q2) *Fisher consistency under ellipticity*: if P is elliptically symmetric with location θ_0 and scatter Σ_0, then $D^{\text{sc}}_P(\Sigma_0) \geq D^{\text{sc}}_P(\Sigma)$ for any $\Sigma \in \mathcal{P}_k$;

(Q3) *Monotonicity relative to any deepest scatter*: if Σ_a maximizes $D^{\text{sc}}_P(\cdot)$, then, for any $\Sigma_b \in \mathcal{P}_k$, $t \mapsto D^{\text{sc}}_P((1-t)\Sigma_a + t\Sigma_b)$ is monotone non-increasing over $[0,1]$;

(Q4) *Vanishing at the boundary of the parameter space*: if (Σ_n) F-converges to the boundary of $\mathcal{P}_k$ (in the sense that either $d_F(\Sigma_n, \Sigma) \to 0$ for some $\Sigma \in \mathcal{S}_k \setminus \mathcal{P}_k$ or $d_F(\Sigma_n, I_k) \to \infty$), then $D^{\text{sc}}_P(\Sigma_n) \to 0$.

While (Q1) and (Q3) are the natural scatter counterparts of (P1) and (P3), respectively, some comments are in order for (Q2) and (Q4). We start with (Q2). In essence, (P2) requires that, whenever an indisputable location center exists (as it is the case for symmetric distributions), this location should be flagged as most central by the location depth at hand. A similar reasoning leads to (Q2): we argue that, for an elliptical probability measure, the "true" value of the scatter parameter is indisputable, and (Q2) then imposes that the scatter depth at hand should identify this true scatter value as the (or at least, as a) deepest one. One might actually strengthen (Q2) by replacing the elliptical model there by a broader model in which the true scatter would still be clearly defined. In such a case, of course, the larger the model for which scatter depth satisfies (Q2), the better (a possibility, that we do not explore here, is to consider the union of the elliptical model and the *independent component model*; see Ilmonen and Paindaveine, 2011 and the references therein). This is parallel to what happens in (P2): the weaker the symmetry assumption under which (P2) is satisfied, the better (for instance, having (P2) satisfied with angular symmetry is better than having it satisfied with central symmetry only); see Zuo and Serfling (2000).

We then turn to (Q4), whose location counterpart (P4) is typically read by saying that the depth/centrality $D_P^{\text{loc}}(\theta_n)$ goes to zero when the point θ_n goes to the boundary of the *sample space*. In the spirit of parametric depth (Mizera, 2002; Mizera and Müller, 2004), however, it is more appropriate to look at θ_n as a candidate location fit and to consider that (P4) imposes that the appropriateness $D_P^{\text{loc}}(\theta_n)$ of this fit goes to zero as θ_n goes to the boundary of the *parameter space*. For location, the confounding between the sample space and parameter space (both are $\mathbb{R}^k$) allows for both interpretations. For scatter, however, there is no such confounding (the sample space is $\mathbb{R}^k$ and the parameter space is $\mathcal{P}_k$), and we argue (Q4) above is the natural scatter version of (P4): whenever Σ_n goes to the boundary of the parameter space $\mathcal{P}_k$, scatter depth should flag it as an arbitrarily poor candidate fit.

Theorem 2.1 states that scatter halfspace depth satisfies (Q1) as soon as it is based on an affine-equivariant T. Scatter halfspace depth satisfies (Q3) as well: if Σ_a maximizes $HD_{P,T}^{\text{sc}}(\,\cdot\,)$, then Theorem 3.3 indeed readily implies that $HD_{P,T}^{\text{sc}}((1-t)\Sigma_a + t\Sigma_b) \geq \min(HD_{P,T}^{\text{sc}}(\Sigma_a), HD_{P,T}^{\text{sc}}(\Sigma_b)) = HD_{P,T}^{\text{sc}}(\Sigma_b)$ for any $\Sigma_b \in \mathcal{P}_k$ and $t \in [0,1]$. The next Fisher consistency result shows that, provided that T is affine-equivariant, (Q2) is also met.

THEOREM 5.1. *Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and scatter Σ_0, and let T be an affine-equivariant location functional. Then, (i) $HD_{P,T}^{\text{sc}}(\Sigma_0) \geq HD_{P,T}^{\text{sc}}(\Sigma)$ for any $\Sigma \in \mathcal{P}_k$, and the equality holds if and only if $\text{Sp}(\Sigma_0^{-1}\Sigma) \subset \mathcal{I}_{\text{MSD}}[Z_1]$, where $Z = (Z_1, \ldots, Z_k)' \overset{\mathcal{D}}{=} \Sigma_0^{-1/2}(X - \theta_0)$; (ii) in particular, if $\mathcal{I}_{\text{MSD}}[Z_1]$ is a singleton (equivalently, if $\mathcal{I}_{\text{MSD}}[Z_1] = \{1\}$), then $\Sigma \mapsto HD_{P,T}^{\text{sc}}(\Sigma)$ is uniquely maximized at Σ_0.*

While (Q1)-(Q3) are satisfied by scatter halfspace depth without any assumption on P, (Q4) is not, as the mixture example considered in Section 3 shows (since the sequence (Σ_n) considered there has limiting depth $1-s > 0$). However, Theorem 3.2 reveals that (Q4) may fail only when $d_F(\Sigma_n, \Sigma) \to 0$ for some $\Sigma \in \mathcal{S}_k \setminus \mathcal{P}_k$. More importantly, Theorem 4.2 implies that T-scatter halfspace depth will satisfy (Q4) at any P that is smooth at T_P.

In a generic parametric depth setup, (Q3) would require that the parameter space is convex. If the parameter space rather is a non-flat Riemannian manifold, then it is natural to replace the "linear" monotonicity property (Q3) with a "geodesic" one. In the context of scatter depth, this would lead to replacing (Q3) with

($\widetilde{\text{Q3}}$) *Geodesic monotonicity relative to any deepest scatter*: if Σ_a maximizes $D_P^{\text{sc}}(\,\cdot\,)$, then, for any $\Sigma_b \in \mathcal{P}_k$, $t \mapsto D_P^{\text{sc}}(\tilde{\Sigma}_t)$ is monotone nonincreasing over $[0,1]$ along the geodesic path $\tilde{\Sigma}_t$ from Σ_a to Σ_b in (4.2).

We refer to Section 7 for a parametric framework where (Q3) cannot be considered and where ($\widetilde{\text{Q3}}$) needs to be adopted instead. For scatter, however, the hybrid nature of $\mathcal{P}_k$, which is both flat (as a convex subset of the vector space $\mathcal{S}_k$) and curved (as a Riemannian manifold with non-positive curvature), allows to consider both (Q3) and ($\widetilde{\text{Q3}}$). Just like (Q3) follows from quasi-concavity of the mapping $\Sigma \mapsto HD^{\text{sc}}_{P,T}(\Sigma)$, ($\widetilde{\text{Q3}}$) would follow from the same mapping being *geodesic quasi-concave*, in the sense that $HD_{P,T}(\tilde{\Sigma}_t) \geq \min(HD_{P,T}(\Sigma_a), HD_{P,T}(\Sigma_b))$ along the geodesic path $\tilde{\Sigma}_t$ from Σ_a to Σ_b. Geodesic quasi-concavity would actually imply that scatter halfspace depth regions are *geodesic convex*, in the sense that, for any $\Sigma_a, \Sigma_b \in R^{\text{sc}}_{P,T}(\alpha)$, the geodesic from Σ_a to Σ_b is contained in $R^{\text{sc}}_{P,T}(\alpha)$. We refer to Dümbgen and Tyler (2016) for an application of geodesic convex functions to inference on (high-dimensional) scatter matrices.

Theorem 3.3 shows that $\Sigma \mapsto HD^{\text{sc}}_{P,T}(\Sigma)$ is quasi-concave for any P. A natural question is then whether or not this extends to geodesic quasi-concavity. The answer is positive at any k-variate elliptical probability measure and at the k-variate probability measure with independent Cauchy marginals.

THEOREM 5.2. *Let P be an elliptical probability measure over $\mathbb{R}^k$ or the k-variate probability measure with independent Cauchy marginals, and let T be an affine-equivariant location functional. Then, (i) $\Sigma \mapsto HD^{\text{sc}}_{P,T}(\Sigma)$ is geodesic quasi-concave, so that (ii) $R^{\text{sc}}_{P,T}(\alpha)$ is geodesic convex for any $\alpha \geq 0$.*

We close this section with a numerical illustration of the quasi-concavity results in Theorems 3.3 and 5.2 and with an example showing that geodesic quasi-concavity may actually fail to hold. Figure 2 provides, for three bivariate probability measures P, the plots of $t \mapsto HD^{\text{sc}}_P(\Sigma_t)$ and $t \mapsto HD^{\text{sc}}_P(\tilde{\Sigma}_t)$, where $\Sigma_t = (1-t)\Sigma_a + t\Sigma_b$ is the linear path from $\Sigma_a = I_2$ to $\Sigma_b = \text{diag}(0.001, 20)$ and where $\tilde{\Sigma}_t = \Sigma_a^{1/2}(\Sigma_a^{-1/2}\Sigma_b\Sigma_a^{-1/2})^t\Sigma_a^{1/2}$ is the corresponding geodesic path. The three distributions considered are (i) the bivariate normal with location zero and scatter I_2, (ii) the bivariate distribution with independent Cauchy marginals, and (iii) the empirical distribution associated with a random sample of size $n = 200$ from the bivariate mixture distribution $P = \frac{1}{2}P_1 + \frac{1}{4}P_2 + \frac{1}{4}P_3$, where P_1 is the standard normal, P_2 is the normal with mean $(0, 4)'$ and covariance matrix $\frac{1}{10}I_2$, and P_3 is the normal with mean $(0, -4)'$ and covariance matrix $\frac{1}{10}I_2$. Figure 2 illustrates that (linear) quasi-concavity of scatter halfspace depth always holds, but that geodesic quasi-concavity may fail to hold. Despite this counterexample, extensive numerical experiments led us to think that geodesic quasi-concavity is the rule rather than the exception.

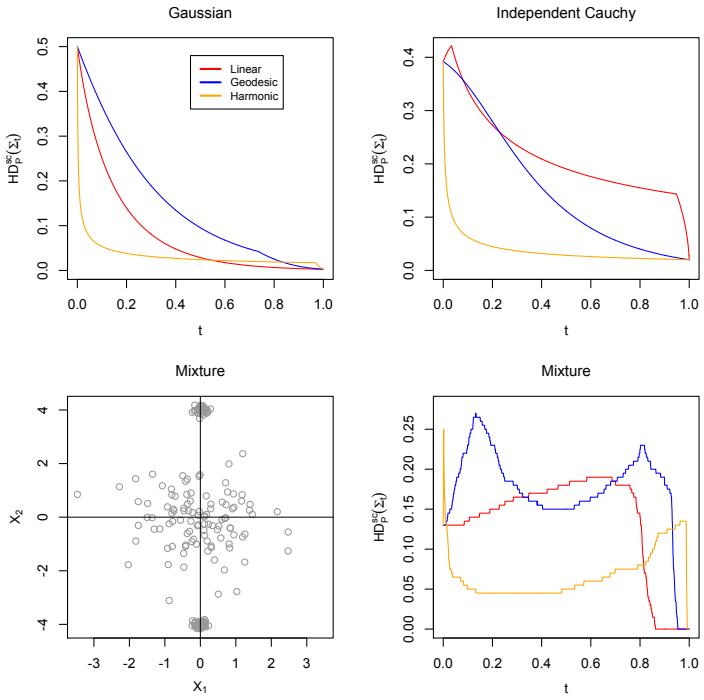

FIG 2. *Plots, for various bivariate probability measures P, of the scatter halfspace depth function $\Sigma \mapsto HD_P^{\mathrm{sc}}(\Sigma)$ along the linear path $\Sigma_t = (1-t)\Sigma_a + t\Sigma_b$ (red), the geodesic path $\tilde{\Sigma}_t = \Sigma_a^{1/2}(\Sigma_a^{-1/2}\Sigma_b\Sigma_a^{-1/2})^t\Sigma_a^{1/2}$ (blue), and the harmonic path $\Sigma_t^* = ((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1})^{-1}$ (orange), from $\Sigma_a = I_2$ to $\Sigma_b = \mathrm{diag}(0.001, 20)$; harmonic paths are introduced in Section 6. The probability measures considered are the bivariate normal with location zero and scatter I_2 (top left), the bivariate distribution with independent Cauchy marginals (top right), and the empirical probability measure associated with a random sample of size $n = 200$ from the bivariate mixture distribution described in Section 5 (bottom right). The scatter plot of the sample used in the mixture case is provided in the bottom left panel.*

6. Concentration halfspace depth. In various setups, the parameter of interest is the *concentration matrix* $\Gamma := \Sigma^{-1}$ rather than the scatter matrix Σ. For instance, in Gaussian graphical models, the (i,j)-entry of Γ is zero if and only if the ith and jth marginals are conditionally independent given all other marginals. It may then be useful to define a depth for inverse scatter matrices. The scatter halfspace depth in (2.1) naturally leads to defining the *T-concentration halfspace depth* of Γ with respect to P as

$$HD_{P,T}^{\mathrm{conc}}(\Gamma) := HD_{P,T}^{\mathrm{sc}}(\Gamma^{-1})$$

and the corresponding *T-concentration halfspace depth regions* as $R_{P,T}^{\mathrm{conc}}(\alpha) := \{\Gamma \in \mathcal{P}_k : HD_{P,T}^{\mathrm{conc}}(\Gamma) \geq \alpha\}$, $\alpha \geq 0$. As indicated by an anonymous ref-

eree, the definition of T-concentration halfspace depth alternatively results, through the use of "innovated transformation" (see, e.g., Hall and Jin, 2010, Fan, Jin and Yao, 2013, or Fan and Lv, 2016), from the concept of (an affine-invariant) T-scatter halfspace depth.

Concentration halfspace depth and concentration halfspace depth regions inherit the properties of their scatter antecedents, sometimes with subtle modifications. The former is affine-invariant and the latter are affine-equivariant as soon as they are based on an affine-equivariant T. Concentration halfspace depth is upper F- and g-semicontinuous for any probability measure P (so that the regions $R^{\text{conc}}_{P,T}(\alpha)$ are F- and g-closed) and F- and g-continuous if P is smooth at T_P. While the regions $R^{\text{conc}}_{P,T}(\alpha)$ are still g-bounded (hence also, F-bounded) for $\alpha > \alpha_{P,T}$, the outer regions $R^{\text{conc}}_{P,T}(\alpha)$, $\alpha \leq \alpha_{P,T}$, here may fail to be F-bounded (this is because implosion of Σ, under which scatter halfspace depth may fail to go below $\alpha_{P,T}$, is associated with explosion of Σ^{-1}). Finally, uniform consistency and existence of a concentration halfspace deepest matrix are guaranteed under the same conditions on P and T as for scatter halfspace depth.

Quasi-concavity of concentration halfspace depth and convexity of the corresponding regions require more comments. The linear path $t \mapsto (1-t)\Gamma_a + t\Gamma_b$ between the concentration matrices $\Gamma_a = \Sigma_a^{-1}$ and $\Gamma_b = \Sigma_b^{-1}$ determines a *harmonic path* $t \mapsto \Sigma_t^* := ((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1})^{-1}$ between the corresponding scatter matrices Σ_a and Σ_b. In line with the definitions adopted in the previous sections, we will say that $f : \mathcal{P}_k \to \mathbb{R}$ is *harmonic quasi-concave* if $f(\Sigma_t^*) \geq \min(f(\Sigma_a), f(\Sigma_b))$ for any $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ and $t \in [0,1]$, and that a subset R of $\mathcal{P}_k$ is *harmonic convex* if $\Sigma_a, \Sigma_b \in R$ implies that $\Sigma_t^* \in R$ for any $t \in [0,1]$. Clearly, concentration halfspace depth is quasi-concave if and only if scatter halfspace depth is harmonic quasi-concave, which turns out to be the case in the elliptical and independent Cauchy cases. We thus have the following result.

THEOREM 6.1. *Let P be an elliptical probability measure over $\mathbb{R}^k$ or the k-variate probability measure with independent Cauchy marginals, and let T be an affine-equivariant location functional. Then, (i) $\Gamma \mapsto HD^{\text{conc}}_{P,T}(\Gamma)$ is quasi-concave, so that (ii) $R^{\text{conc}}_{P,T}(\alpha)$ is convex for any $\alpha \geq 0$.*

However, concentration halfspace depth may fail to be quasi-concave, since, as we show by considering the mixture example in Figure 2, scatter halfspace depth may fail to be harmonic quasi-concave. The figure, that also plots scatter halfspace depth along harmonic paths, confirms that, while scatter halfspace depth is harmonic quasi-concave for the Gaussian and independent Cauchy examples there, it is not in the mixture example. In

this mixture example, thus, concentration halfspace depth fails to be quasi-concave and the corresponding depth regions fail to be convex. This is not a problem per se — recall that famous (location) depth functions, like, e.g., the simplicial depth from Liu (1990), may provide non-convex depth regions.

For completeness, we present the following result which shows that some form of quasi-concavity for concentration halfspace depth survives.

THEOREM 6.2. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Gamma \mapsto HD^{\text{conc}}_{P,T}(\Gamma)$ is harmonic quasi-concave, so that (ii) $R^{\text{conc}}_{P,T}(\alpha)$ is harmonic convex for any $\alpha \geq 0$.*

Since concentration halfspace depth is harmonic quasi-concave if and only if scatter halfspace depth is quasi-concave, the result is a direct corollary of Theorem 3.3. Quasi-concavity and harmonic quasi-concavity clearly are dual concepts, relative to scatter and concentration halfspace depths (which justifies the * notation in the path Σ_t^*, dual to Σ_t). Interestingly, $\Gamma \mapsto HD^{\text{conc}}_{P,T}(\Gamma)$ is geodesic quasi-concave if and only if $\Sigma \mapsto HD^{\text{sc}}_{P,T}(\Sigma)$ is, so that concentration halfspace depth regions are geodesic convex if and only if scatter halfspace depth regions are.

7. Shape halfspace depth. In many multivariate statistics problems (PCA, CCA, sphericity testing, etc.), it is sufficient to know the scatter matrix Σ up to a positive scalar factor. In PCA, for instance, all scatter matrices of the form $c\Sigma$, $c > 0$, indeed provide the same unit eigenvectors $v_\ell(c\Sigma)$, $\ell = 1, \ldots, k$, hence the same principal components. Moreover, when it comes to deciding how many principal components to work with, a common practice is to look at the proportions of explained variance $\sum_{\ell=1}^m \lambda_\ell(c\Sigma)/\sum_{\ell=1}^k \lambda_\ell(c\Sigma)$, $m = 1, \ldots, k-1$, which do not depend on c either. In PCA, thus, the parameter of interest is a *shape matrix*, that is, a normalized version, V say, of the scatter matrix Σ.

The generic way to normalize a scatter matrix Σ into a shape matrix V is based on a scale functional S, that is, on a mapping $S : \mathcal{P}_k \to \mathbb{R}_0^+$ satisfying (i) $S(I_k) = 1$ and (ii) $S(c\Sigma) = cS(\Sigma)$ for any $c > 0$ and $\Sigma \in \mathcal{P}_k$. In this paper, we will further assume that (iii) if $\Sigma_1, \Sigma_2 \in \mathcal{P}_k$ satisfy $\Sigma_2 \geq \Sigma_1$ (in the sense that $\Sigma_2 - \Sigma_1$ is positive semidefinite), then $S(\Sigma_2) \geq S(\Sigma_1)$. Such a scale functional leads to factorizing $\Sigma(\in \mathcal{P}_k)$ into $\Sigma = \sigma_S^2 V_S$, where $\sigma_S^2 := S(\Sigma)$ is the *scale* of Σ and $V_S := \Sigma/S(\Sigma)$ is its *shape matrix* (in the sequel, we will drop the subscript S in V_S to avoid overloading the notation). The resulting collection of shape matrices V will be denoted as $\mathcal{P}_k^S$. Note that the constraint $S(I_k) = 1$ ensures that, irrespective of the scale functional S adopted, I_k is a shape matrix. Common scale functionals satisfying (i)-(iii)

are (a) $S_{\mathrm{tr}}(\Sigma) = (\mathrm{tr}\,\Sigma)/k$, (b) $S_{\det}(\Sigma) = (\det\Sigma)^{1/k}$, (c) $S^*_{\mathrm{tr}}(\Sigma) = k/(\mathrm{tr}\,\Sigma^{-1})$, and (d) $S_{11}(\Sigma) = \Sigma_{11}$; we refer to Paindaveine and Van Bever (2014) for references where the scale functionals (a)-(d) are used. The corresponding shape matrices V are then normalized in such a way that (a) $\mathrm{tr}[V] = k$, (b) $\det V = 1$, (c) $\mathrm{tr}[V^{-1}] = k$, or (d) $V_{11} = 1$.

In this section, we propose a concept of halfspace depth for shape matrices. More precisely, for a probability measure P over $\mathbb{R}^k$, we define the (S,T)-*shape halfspace depth* of $V(\in \mathcal{P}^S_k)$ with respect to P as

$$HD^{\mathrm{sh},S}_{P,T}(V) := \sup_{\sigma^2>0} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V), \tag{7.1}$$

where $HD^{\mathrm{sc}}_{P,T}(\sigma^2 V)$ is the T-scatter halfspace depth of $\sigma^2 V$ with respect to P. The corresponding depth regions are defined as

$$R^{\mathrm{sh},S}_{P,T}(\alpha) := \{V \in \mathcal{P}^S_k : HD^{\mathrm{sh},S}_{P,T}(V) \geq \alpha\}$$

(alike scatter, we will drop the index T in $HD^{\mathrm{sh},S}_{P,T}(V)$ and $R^{\mathrm{sh},S}_{P,T}(\alpha)$ whenever T is the Tukey median). The halfspace deepest shape (if any) is obtained by maximizing the "profile depth" in (7.1), in the same way a profile likelihood approach would be based on the maximization of a (shape) profile likelihood of the form $L^{\mathrm{sh}}_V = \sup_{\sigma^2>0} L_{\sigma^2 V}$. To the best of our knowledge, such a profile depth construction has never been considered in the literature.

We start the study of shape halfspace depth by considering our running, Gaussian and independent Cauchy, examples. For the k-variate normal with location θ_0 and scatter Σ_0 (hence, with S-shape matrix $V_0 = \Sigma_0/S(\Sigma_0)$),

$$\sigma^2 \mapsto HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) = 2\min\left(\Phi\left(\frac{b\sigma\lambda_k^{1/2}(V_0^{-1}V)}{\sqrt{S(\Sigma_0)}}\right) - \frac{1}{2}, 1 - \Phi\left(\frac{b\sigma\lambda_1^{1/2}(V_0^{-1}V)}{\sqrt{S(\Sigma_0)}}\right)\right)$$

(see (2.6)) will be uniquely maximized at the σ^2-value for which both arguments of the minimum are equal. It follows that

$$HD^{\mathrm{sh},S}_{P,T}(V) = 2\Phi\big(c(V_0^{-1}V)\lambda_k^{1/2}(V_0^{-1}V)\big) - 1,$$

where $c(\Upsilon)$ is the unique solution of $\Phi\big(c(\Upsilon)\lambda_k^{1/2}(\Upsilon)\big) - \frac{1}{2} = 1 - \Phi\big(c(\Upsilon)\lambda_1^{1/2}(\Upsilon)\big)$. At the k-variate distribution with independent Cauchy marginals, we still have that (with the same notation as in (2.7))

$$HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) = 2\min\left(\Psi\big(\sigma/\max_s (s'V^{-1}s)^{1/2}\big) - \tfrac{1}{2}, 1 - \Psi\big(\sigma\sqrt{\max(\mathrm{diag}(V))}\big)\right)$$

is maximized for fixed V when both arguments of the minimum are equal, that is, when $\sigma^2 = \big(\max_s(s'V^{-1}s)/\max(\mathrm{diag}(V))\big)^{1/2}$. Therefore,

$$\begin{aligned} HD^{\mathrm{sh},S}_{P,T}(V) &= 2\,\Psi\Big(\big(\max_s(s'V^{-1}s)\max(\mathrm{diag}(V))\big)^{-1/4}\Big) - 1 \\ &= \frac{2}{\pi}\arctan\Big(\big(\max_s(s'V^{-1}s)\max(\mathrm{diag}(V))\big)^{-1/4}\Big). \end{aligned}$$

Figure 3 draws, for six probability measures P and any affine-equivariant T, contour plots of $(V_{11}, V_{12}) \mapsto HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$, where $HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$ is the shape halfspace depth of $V = \left(\begin{smallmatrix} V_{11} & V_{12} \\ V_{12} & 2-V_{11} \end{smallmatrix}\right)$ with respect to P. Letting $\Sigma_A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $\Sigma_B = \left(\begin{smallmatrix} 4 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\Sigma_C = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, the probability measures P considered are those associated (i) with the bivariate normal distributions with location zero and scatter Σ_A, Σ_B and Σ_C, and (ii) with the distributions of $\Sigma_A^{1/2}Z$, $\Sigma_B^{1/2}Z$ and $\Sigma_C^{1/2}Z$, where Z has independent Cauchy marginals. Note that the maximal depth is larger in the Gaussian cases than in the Cauchy ones, that depth monotonically decreases along any ray originating from the deepest shape matrix and that it goes to zero if and only if the shape matrix converges to the boundary of the parameter space. Shape halfspace depth contours are smooth in the Gaussian cases but not in the Cauchy ones.

In both the Gaussian and independent Cauchy examples above, the supremum in (7.1) is a maximum. For empirical probability measures P_n, this will always be the case since $HD^{\mathrm{sc}}_{P_n,T}(\sigma^2 V)$ then takes its values in $\{\ell/n : \ell = 0, 1, \ldots, n\}$. The following result implies in particular that a sufficient condition for this supremum to be a maximum is that P is smooth at T_P (which is the case in both our running examples above).

THEOREM 7.1. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Fix $V \in \mathcal{P}^S_k$ such that $cV \in R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ for some $c > 0$. Then, $HD^{\mathrm{sh},S}_{P,T}(V) = HD^{\mathrm{sc}}_{P,T}(\sigma^2_V V)$ for some $\sigma^2_V > 0$.*

The following affine-invariance/equivariance and uniform consistency results are easily obtained from their scatter antecedents.

THEOREM 7.2. *Let T be an affine-equivariant location functional. Then, (i) shape halfspace depth is affine-invariant in the sense that, for any probability measure P over $\mathbb{R}^k$, $V \in \mathcal{P}^S_k$, $A \in GL_k$ and $b \in \mathbb{R}^k$, we have $HD^{\mathrm{sh},S}_{P_{A,b},T}(AVA'/S(AVA')) = HD^{\mathrm{sh},S}_{P,T}(V)$, where $P_{A,b}$ is as defined on page 4. Consequently, (ii) shape halfspace depth regions are affine-equivariant, in the sense that $R^{\mathrm{sh},S}_{P_{A,b},T}(\alpha) = \{AVA'/S(AVA') : V \in R^{\mathrm{sh},S}_{P,T}(\alpha)\}$ for any probability measure P over $\mathbb{R}^k$, $\alpha \geq 0$, $A \in GL_k$ and $b \in \mathbb{R}^k$.*

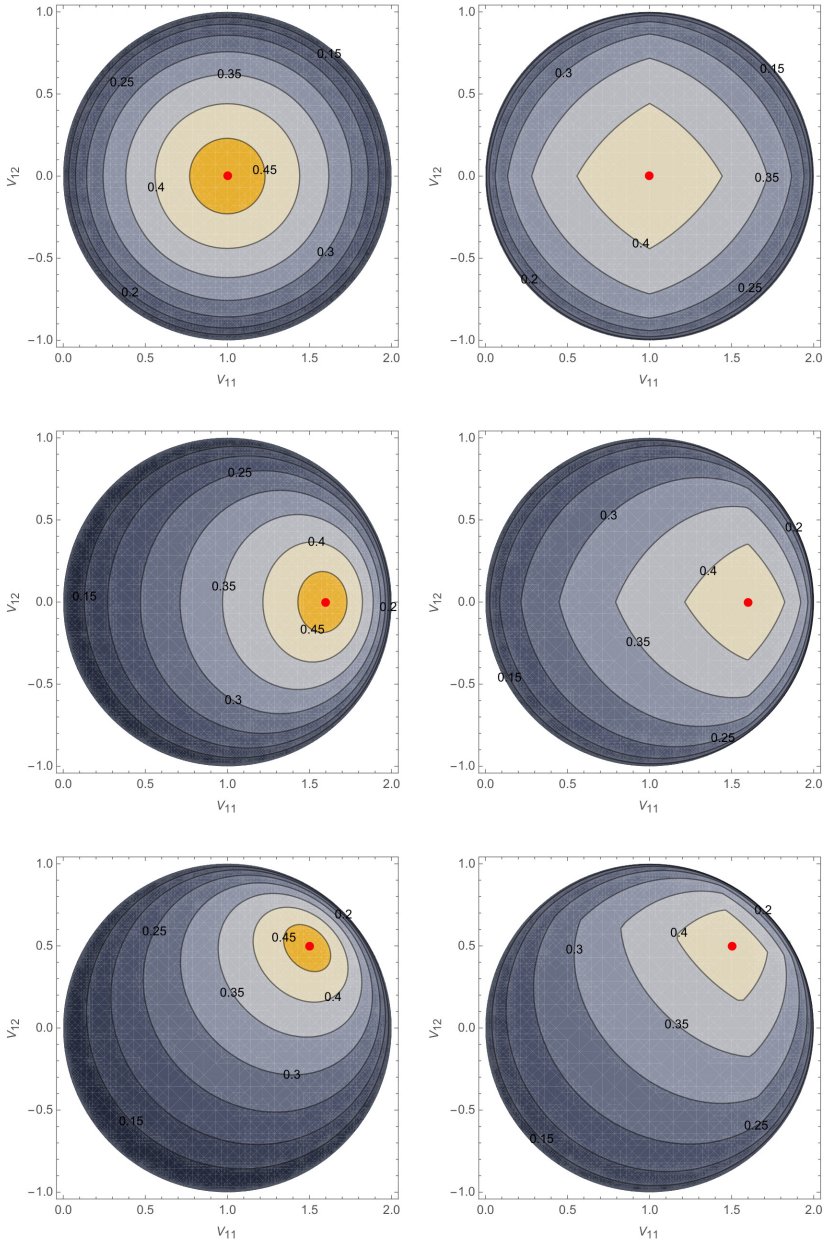

FIG 3. *Contour plots of* $(V_{11}, V_{12}) \mapsto HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$*, for several bivariate probability measures* P *and an arbitrary affine-equivariant location functional* T*, where* $HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$ *is the shape halfspace depth, with respect to* P*, of* $V = \left(\begin{smallmatrix} V_{11} & V_{12} \\ V_{12} & 2-V_{11} \end{smallmatrix}\right)$*. Letting* $\Sigma_A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$*,* $\Sigma_B = \left(\begin{smallmatrix} 4 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ *and* $\Sigma_C = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 1 \end{smallmatrix}\right)$*, the probability measures* P *considered are those associated (i) with the bivariate normal distributions with location zero and scatter* Σ_A*,* Σ_B *and* Σ_C *(top, middle and bottom left), and (ii) with the distributions of* $\Sigma_A^{1/2}Z$*,* $\Sigma_B^{1/2}Z$ *and* $\Sigma_C^{1/2}Z$*, where* Z *has mutually independent Cauchy marginals (top, middle and bottom right). In each case, the "true"* S_{tr}*-shape matrix is marked in red.*

THEOREM 7.3. *Let P be a smooth probability measure over $\mathbb{R}^k$ and T be a location functional. Let P_n denote the empirical probability measure associated with a random sample of size n from P and assume that $T_{P_n} \to T_P$ almost surely as $n \to \infty$. Then $\sup_{V \in \mathcal{P}_k^S} |HD_{P_n,T}^{\mathrm{sh},S}(V) - HD_{P,T}^{\mathrm{sh},S}(V)| \to 0$ almost surely as $n \to \infty$.*

Shape halfspace depth inherits the F- and g-continuity properties of scatter halfspace depth (Theorems 3.1 and 4.1, respectively), at least for a smooth P. More precisely, we have the following result.

THEOREM 7.4. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $V \mapsto HD_{P,T}^{\mathrm{sh},S}(V)$ is upper F- and g-semicontinuous on $R_{P,T}^{\mathrm{sh},S}(\alpha_{P,T})$, so that (ii) for any $\alpha \geq \alpha_{P,T}$, the depth region $R_{P,T}^{\mathrm{sh},S}(\alpha)$ is F- and g-closed. (iii) If P is smooth at T_P, then $V \mapsto HD_{P,T}^{\mathrm{sh},S}(V)$ is F- and g-continuous.*

The g-boundedness part of the following result will play a key role when proving the existence of a halfspace deepest shape.

THEOREM 7.5. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, for any $\alpha > \alpha_{P,T}$, $R_{P,T}^{\mathrm{sh},S}(\alpha)$ is F- and g-bounded, hence g-compact. If $s_{P,T} \geq 1/2$, then this result is trivial in the sense that $R_{P,T}^{\mathrm{sh},S}(\alpha)$ is empty for $\alpha > \alpha_{P,T}$.*

Comparing with the scatter result in Theorem 3.2, the shape result for F-boundedness requires the additional condition $\alpha > \alpha_{P,T}$ (for g-boundedness, this condition was already required in Theorem 4.2). This condition is actually necessary for scale functionals S for which implosion of a shape matrix V cannot be obtained without explosion, as it is the case, e.g., for S_{det} (the product of the eigenvalues of an S_{det}-shape matrix being equal to one, the smallest eigenvalue of V cannot go to zero without the largest going to infinity). We illustrate this on the bivariate discrete example discussed below Theorem 4.2, still with an arbitrary centro-equivariant T. The sequence of scatter matrices $\Sigma_n = \mathrm{diag}(\frac{1}{n}, 1)$ there defines a sequence of S_{det}-shape matrices $V_n = \mathrm{diag}(\frac{1}{\sqrt{n}}, \sqrt{n})$, that is neither F- nor g-bounded. Since $HD_{P,T}^{\mathrm{sh},S_{\mathrm{det}}}(V_n) \geq HD_{P,T}^{\mathrm{sc}}(\Sigma_n) \geq 1/3 = \alpha_{P,T}$ for any n, we conclude that $R_{P,T}^{\mathrm{sh},S_{\mathrm{det}}}(\alpha_{P,T})$ is both F- and g-unbounded. Note also that F-boundedness of $R_{P,T}^{\mathrm{sh},S}(\alpha)$ depends on S. In particular, it is easy to check that the condition $\alpha > \alpha_{P,T}$ for F-boundedness is not needed for the scale functional S_{tr}^* (that is, $R_{P,T}^{\mathrm{sh},S_{\mathrm{tr}}^*}(\alpha)$ is F-bounded for any $\alpha > 0$). Finally, one

trivially has that all $R^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(\alpha)$'s are F-bounded since the corresponding collection of shape matrices, $\mathcal{P}^{S_{\mathrm{tr}}}_k$, itself is F-bounded. Unlike F-boundedness, g-boundedness results are homogeneous in S, which further suggests that the g-topology is the most appropriate one to study scatter/shape depths.

As announced, the g-part of Theorem 7.5 allows to show that a halfspace deepest shape exists under mild conditions. More precisely, we have the following result.

THEOREM 7.6. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Assume that $R^{\mathrm{sh},S}_{P,T}(\alpha_{P,T})$ is non-empty. Then, $\alpha^S_{*P,T} := \sup_{V\in\mathcal{P}^S_k} HD^{\mathrm{sh},S}_{P,T}(V) = HD^{\mathrm{sh},S}_{P,T}(V_*)$ for some $V_* \in \mathcal{P}^S_k$.*

Alike scatter, a sufficient condition for the existence of a halfspace deepest shape is thus that P is smooth at T_P. In particular, a halfspace deepest shape exists in the Gaussian and independent Cauchy examples. In the k-variate independent Cauchy case, it readily follows from Theorem 4.4 that, irrespective of the centro-equivariant T used, $HD^{\mathrm{sh},S}_{P,T}(V)$ is uniquely maximized at $V_* = I_k$, with corresponding maximal depth $\frac{2}{\pi}\arctan\left(k^{-1/4}\right)$. The next Fisher-consistency result states that, in the elliptical case, the halfspace deepest shape coincides with the "true" shape matrix.

THEOREM 7.7. *Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and scatter Σ_0, hence with S-shape matrix $V_0 = \Sigma_0/S(\Sigma_0)$, and let T be an affine-equivariant location functional. Then, (i) $HD^{\mathrm{sh},S}_{P,T}(V_0) \geq HD^{\mathrm{sh},S}_{P,T}(V)$ for any $V \in \mathcal{P}^S_k$; (ii) if $\mathcal{I}_{\mathrm{MSD}}[Z_1]$ is a singleton (equivalently, if $\mathcal{I}_{\mathrm{MSD}}[Z_1] = \{1\}$), where $Z = (Z_1,\ldots,Z_k)' \overset{\mathcal{D}}{=} \Sigma_0^{-1/2}(X-\theta_0)$, then $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$ is uniquely maximized at V_0.*

We conclude this section by considering quasi-concavity properties of shape halfspace depth and convexity properties of the corresponding depth regions. It should be noted that, for some scale functionals S, the collection $\mathcal{P}^S_k$ of S-shape matrices is not convex; for instance, neither $\mathcal{P}^{S_{\det}}_k$ nor $\mathcal{P}^{S^*_{\mathrm{tr}}}_k$ is convex, so that it does not make sense to investigate whether or not $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$ is quasi-concave for these scale functionals. It does, however, for S_{tr} and S_{11}, and we have the following result.

THEOREM 7.8. *Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Fix $S = S_{\mathrm{tr}}$ or $S = S_{11}$. Then, (i) $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$ is quasi-concave, that is, for any $V_a, V_b \in \mathcal{P}^S_k$ and $t \in [0,1]$, $HD^{\mathrm{sh},S}_{P,T}(V_t) \geq \min(HD^{\mathrm{sh},S}_{P,T}(V_a), HD^{\mathrm{sh},S}_{P,T}(V_b))$, where we let $V_t := (1-t)V_a + tV_b$; (ii) for any $\alpha \geq 0$, $R^{\mathrm{sh},S}_{P,T}(\alpha)$ is convex.*

As mentioned above, neither $\mathcal{P}_k^{S_{\text{det}}}$ nor $\mathcal{P}_k^{S_{\text{tr}}^*}$ are convex in the usual sense (unlike for S_{tr} and S_{11}, thus, a unique halfspace deepest shape could not be defined through barycenters but would rather require a center-of-mass approach as in (4.3)). However, $\mathcal{P}_k^{S_{\text{det}}}$ is geodesic convex, which justifies studying the possible geodesic convexity of $R^{\text{sh}}_{P,S_{\text{det}}}(\alpha)$ (this provides a parametric framework for which the shape version of (Q3) in Section 5 cannot be considered and for which it is needed to adopt the corresponding Property ($\widetilde{\text{Q3}}$) instead). Similarly, $\mathcal{P}_k^{S_{\text{tr}}^*}$ is harmonic convex, so that it makes sense to investigate the harmonic convexity of $R^{\text{sh},S_{\text{tr}}^*}_{P,T}(\alpha)$. We have the following results.

THEOREM 7.9. *Let T be an affine-equivariant location functional and P be an arbitrary probability measure over $\mathbb{R}^k$ for which T-scatter halfspace depth is geodesic quasi-concave. Then, (i) $V \mapsto HD^{\text{sh},S_{\text{det}}}_{P,T}(V)$ is geodesic quasi-concave, so that (ii) $R^{\text{sh},S_{\text{det}}}_{P,T}(\alpha)$ is geodesic convex for any $\alpha \geq 0$.*

THEOREM 7.10. *Let T be an affine-equivariant location functional and P be an arbitrary probability measure over $\mathbb{R}^k$ for which T-scatter halfspace depth is harmonic quasi-concave. Then, (i) $V \mapsto HD^{\text{sh},S_{\text{tr}}^*}_{P,T}(V)$ is harmonic quasi-concave, so that (ii) $R^{\text{sh},S_{\text{tr}}^*}_{P,T}(\alpha)$ is harmonic convex for any $\alpha \geq 0$.*

An illustration of Theorems 7.8-7.10 is provided in the supplemental article Paindaveine and Van Bever (2017).

8. A real-data application. In this section, we analyze the returns of the Nasdaq Composite and S&P500 indices from February 1st, 2015 to February 1st, 2017. During that period, for each trading day and for each index, we collected returns every 5 minutes (that is, the difference between the index at a given time and 5 minutes earlier, when available), resulting in usually 78 bivariate observations per day. Due to some missing values, the exact number of returns per day varies, and only days with at least 70 observations were considered. The resulting dataset comprises a total of 38489 bivariate returns distributed over $D = 478$ trading days.

The goal of this analysis is to determine which days, during the two-year period, exhibit an atypical behavior. In line with the fact that the main focus in finance is on volatily, atypicality here will refer to deviations from the "global" behavior either in scatter (i.e., returns do not follow the global dispersion pattern) or in scale only (i.e., returns show a usual shape but their overall size is different). Atypical days will be detected by comparing intraday estimates of scatter and shape with a global version.

Below, $\hat{\Sigma}_{\text{full}}$ will denote the minimum covariance determinant (MCD) scatter estimate on the empirical distribution P_{full} of the returns over the two-year period, and $\hat{V}_{\text{full}}$ will stand for the resulting shape estimate $\hat{V}_{\text{full}} = \hat{\Sigma}_{\text{full}}/S_{\text{det}}(\hat{\Sigma}_{\text{full}})$. For any $d = 1, \ldots, D$, $\hat{\Sigma}_d$ and $\hat{V}_d$ will denote the corresponding estimates on the empirical distribution P_d on day d.

The rationale behind the choice of MCD rather than standard covariance as an estimation method for scatter/shape is twofold. First, the former will naturally deal with outliers inherently arising in the data (the first few returns after an overnight or weekend break are famously more volatile and their importance should be downweighted in the estimation procedure). Second, as hinted above, the global estimate will provide a baseline to measure the atypicality of any given day, which will be done, among others, using its intraday depth. It would be natural to use halfspace deepest scatter/shape matrices on P_{full} as global estimates for scatter/shape. While locating the exact maxima is a non-trivial task, the MCD shape estimator has already a high depth value ($HD^{\text{sh},S_{\text{det}}}_{P_{\text{full}}}(\hat{V}_{\text{full}}) = 0.481$), which makes it a very good proxy for the halfspace deepest shape. For the same reason, the scaled MCD estimator $\bar{\Sigma}_{\text{full}} = \sigma^2_{\text{full}} V_{\text{full}}$ with $\sigma^2_{\text{full}} = \text{argmax}_{\sigma^2} HD^{\text{sc}}_{P_{\text{full}}}(\sigma^2 V_{\text{full}})$ (that, obviously, satisfies $HD^{\text{sc}}_{P_{\text{full}}}(\bar{\Sigma}_{\text{full}}) = 0.481$) is similarly an excellent proxy for the halfspace deepest scatter. In contrast, the shape estimate associated with the standard covariance matrix (resp., the deepest scaled version of the covariance matrix) has a global shape (resp., scatter) depth of only 0.426.

For each day, the following measures of (a)typicality (three for scatter, three for shape) are computed: (i) the scatter depth $HD^{\text{sc}}_{P_d}(\bar{\Sigma}_{\text{full}})$ of $\bar{\Sigma}_{\text{full}}$ in day d, (ii) the shape depth $HD^{\text{sh},S_{\text{det}}}_{P_d}(\hat{V}_{\text{full}})$ of $\hat{V}_{\text{full}}$ in day d, (iii) the scatter Frobenius distance $d_F(\hat{\Sigma}_d, \hat{\Sigma}_{\text{full}})$, (iv) the shape Frobenius distance $d_F(\hat{V}_d, \hat{V}_{\text{full}})$, (v) the scatter geodesic distance $d_g(\hat{\Sigma}_d, \hat{\Sigma}_{\text{full}})$, and (vi) the shape geodesic distance $d_g(\hat{V}_d, \hat{V}_{\text{full}})$. Of course, low depths or high distances point to atypical days. Practitioners might be tempted to base the distances in (iii)-(vi) on standard covariance estimates, which would actually provide poorer performances in the present outlier detection exercise (due to the masking effect resulting from using a non-robust global dispersion measure as a baseline). Here, we rather use MCD-based estimates to ensure a fair comparison with the depth-based methods in (i)-(ii).

Figure 4 provides the plots of the quantities in (i)-(vi) above as a function of d, $d = 1, \ldots, D$. Major events affecting the returns during the two years are marked there. They are (1) the Black Monday on August 24th, 2015 (orange) when world stock markets went down substantially, (2) the crude oil crisis on January 20th, 2016 (dark blue) when oil barrel prices fell sharply, (3) the Brexit vote aftermath on June 24th, 2016 (dark green), (4) the end

of the low volatility period on September 13th, 2016 (red), (5) the Donald Trump election on November 9th, 2016 (purple), and (6) the announcement and aftermath of the federal rate hikes on December 14th, 2016 (teal).

Detecting atypical events was achieved by flagging outliers in either collections of scatter or shape depth values. This was conducted by constructing box-and-whisker plots of those collections and marking events with depth value below 1.5 IQR of the first quartile. This procedure flagged events (1), (2) and (6) as outlying in scatter and 21 days — including events (1), (2), (3) and (5) — as atypical in shape. Most of the resulting 22 outlying days can be associated (that is, are temporally close) to one of the events (1)-(6) above. For example, 9 days are flagged within the period extending from January 20th, 2016 to February 9th, 2016, during which continuous slump in oil prices rocked the marked strongly, with biggest loss for S&P 500 index on February 9th. Remarkably, out of the 22 flagged outliers, only two (namely October 1st, 2015 and December 14th, 2015) could not be associated with major events. Event (4), although not deemed outlying, was added to mark the end of the low volatility period.

Events (1) and (2) are noticeably singled out by all outlyingness measures, displaying low depth values and high Frobenius and geodesic distances, but the four remaining events tell a very different story. In particular, event (6) exhibits a low scatter depth but a relatively high shape depth, which means that this day shows a shape pattern that is in line with the global one but is very atypical in scale (that is, in volatility size). Quite remarkably, the four distances considered fail to flag this day as an atypical one. A similar behavior appears throughout the two-month period spanning July, August and early September 2016 (between events (3) and (4)), during which the markets have seen a historical streak of small volatility. This period presents widely varying scatter depth values together with stable and high shape depth values, which is perfectly in line with what has been seen on the markets, where only the volatility of the indices was low in days that were otherwise typical. Again, the four distance plots are blind to this relative behavior of scatter and shape in the period.

Events (3) to (5) are picked up by depth measures and scatter distances, though more markedly by the former. This is particularly so for event (3), which sticks out sharply in both depths. The fact that the scatter depth is even lower than the shape depth suggests that event (3) is atypical not only in shape but also in scale. Interestingly, distance measures fully miss the shape outlyingness of this event. Actually, shape distances do not assign large values to any of the events (3) to (6) and, from March 2016 onwards, these distances stay in the same range — particularly so for the Frobenius

ones in (iv). In contrast, the better ability of shape depth to spot outlyingness may be of particular importance in cases where one wants to discard the overall volatility size to rather focus on the shape structure of the returns.

To summarize, the detection of atypical patterns in the dispersion of intraday returns can more efficiently be performed with scatter/shape depths than on the basis of distance measures. Arguably, the fact that the proposed depths use all observations and not a sole estimate of scatter/shape allows to detect deviations from global behaviors more sharply. As showed above, comparing scatter and shape depth values provides a tool that permits the distinction between shape and scale outliers.

9. Final comments and perspectives. In this work, we thoroughly investigated the structural properties of a concept of scatter halfspace depth linked to those proposed in Zhang (2002) and Chen, Gao and Ren (2017). While we tried doing so under minimal assumptions, alternative scatter halfspace depth concepts may actually require even weaker assumptions, but they typically would make the computational burden heavier in the sample case. As an example, one might alternatively define the scatter halfspace depth of $\Sigma(\in \mathcal{P}_k)$ with respect to P as

$$HD_P^{\mathrm{sc,alt}}(\Sigma) = \sup_{\theta\in\mathbb{R}^k} HD_{P,\theta}^{\mathrm{sc}}(\Sigma), \tag{9.1}$$

where $HD_{P,\theta}^{\mathrm{sc}}(\Sigma)$ is the scatter halfspace depth associated with the constant location functional at θ. This alternative scatter depth concept satisfies a uniform consistency result such as the one in Theorem 2.2 without any condition on P, whereas the scatter halfspace depth $HD_{P,T}^{\mathrm{sc}}(\Sigma)$ in (2.1) requires that P is smooth (see Theorem 2.2). In the sample case, however, evaluation of $HD_{P_n}^{\mathrm{sc,alt}}(\Sigma)$ is computationally much more involved than $HD_{P_n,T}^{\mathrm{sc}}(\Sigma)$. Alternative concentration and shape halfspace depth concepts may be defined along the same lines and will show the same advantages/disadvantages. compared to those proposed in this paper.

Another possible concept of scatter halfspace depth bypasses the need to choose a location functional T by exploiting a pairwise difference approach; see Zhang (2002) and Chen, Gao and Ren (2017). In our notation, the resulting scatter depth of Σ with respect to $P = P^X$ is

$$HD_P^{\mathrm{sc,U}}(\Sigma) = HD_{P^{X-\tilde{X}},0}^{\mathrm{sc}}(\Sigma), \tag{9.2}$$

where $\tilde{X}$ is an independent copy of X and where 0 denotes the origin of $\mathbb{R}^k$. On one hand, the sample version of (9.2) is a U-statistic of order two, which will increase the computational burden compared to the sample version

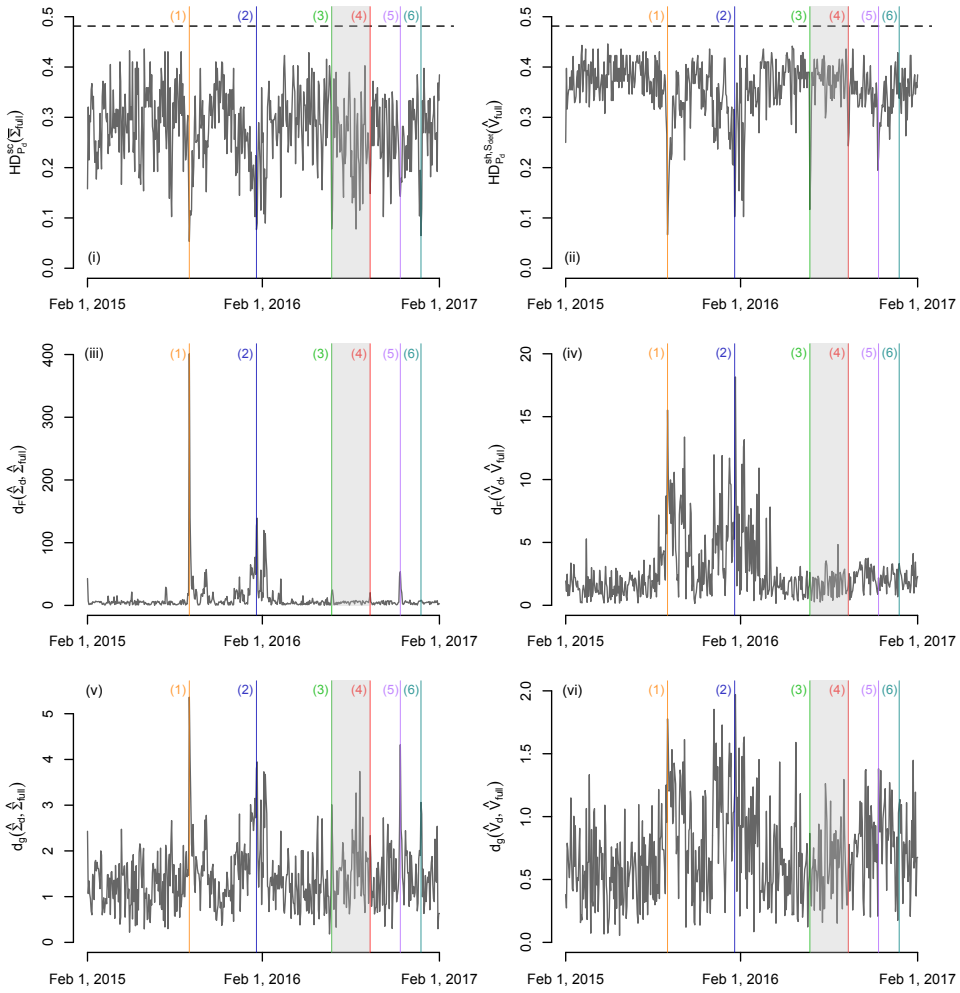

FIG 4. *Plots of (i)* $HD^{\mathrm{sc}}_{P_d}(\bar{\Sigma}_{\mathrm{full}})$*, (ii)* $HD^{\mathrm{sh},S_{\mathrm{det}}}_{P_d}(\hat{V}_{\mathrm{full}})$*, (iii)* $d_F(\hat{\Sigma}_d, \hat{\Sigma}_{\mathrm{full}})$*, (iv)* $d_F(\hat{V}_d, \hat{V}_{\mathrm{full}})$*, (v)* $d_g(\hat{\Sigma}_d, \hat{\Sigma}_{\mathrm{full}})$ *and (vi)* $d_g(\hat{V}_d, \hat{V}_{\mathrm{full}})$*, as a function of* d*, for the MCD scatter and shape estimates described in Section 8. The horizontal dotted lines in (i)-(ii) correspond to the global depths* $HD^{\mathrm{sc}}_{P_{\mathrm{full}}}(\bar{\Sigma}_{\mathrm{full}})$ *and* $HD^{\mathrm{sh},S_{\mathrm{det}}}_{P_{\mathrm{full}}}(\hat{V}_{\mathrm{full}})$*, respectively. All depths make use of the Tukey median as a location functional. Vertical lines mark the six events listed in Section 8.*

of (2.1). On the other hand, uniform consistency results for (9.2) (which here follow from Glivenko-Cantelli results for U-processes, such as the one in Corollary 3.3 from Arcones and Giné, 1993) will again hold without any assumption on P, which is due to the fact that, as already mentioned, the smoothness assumption in Theorem 2.2 is superfluous when a constant lo-

cation functional T is used. At first sight, thus, the pros and cons for (9.2) are parallel to those for (9.1), that is, weaker distributional assumptions are obtained at the expense of computational ease. However, (9.2) suffers from a major disadvantage: it does not provide Fisher consistency at the elliptical model (see (Q2) in Section 5). This results from the fact that if $P = P^X$ is elliptical with location θ and scatter Σ, then $P^{X-\tilde{X}}$ is elliptical with location 0 and scatter $c_P\Sigma$, where the scalar factor c_P depends on the type of elliptical distribution: for multinormal and Cauchy elliptical distributions, e.g., $c_P = 2$ and 4, respectively, so that if one replaces $X - \tilde{X}$ with $(X - \tilde{X})/\sqrt{2}$ to achieve Fisher consistency at the multinormal, then Fisher consistency will still not hold at the Cauchy. Actually, the maximizer of $HD_P^{\mathrm{sc,U}}(\Sigma)$ is useless as a measure of scatter for the original probability measure P, as its interpretation requires knowing which type of elliptical distribution P is. This disqualifies the pairwise difference scatter depth, as well as the companion concentration depth. Note, however, that the corresponding shape depth will not suffer from this Fisher consistency problem since the normalization of scatter matrices into shape matrices will get rid of the scalar factor c_P.

As both previous paragraphs suggest and as it is often the case with statistical depth, computational aspects are key for the application of the proposed depths. Evaluating (good approximations of) the scatter halfspace depth $HD^{\mathrm{sc}}_{P_n,T}(\Sigma)$ of a given Σ can of course be done for very small dimensions $k = 2$ or 3 by simply sampling the unit sphere $\mathcal{S}^{k-1}$. Even for such small dimensions, however, computing the halfspace deepest scatter is nontrivial: while scatter halfspace depth relies on a low-dimensional (that is, k-dimensional) projection-pursuit approach, identifying the halfspace deepest scatter indeed requires exploring the collection of scatter matrices $\mathcal{P}_k$, that is of higher dimension, namely of dimension $k(k+1)/2$. Fortunately, the fixed-location scatter halfspace depth — hence, also its T-version proposed in this paper, after appropriate centering of the observations — can be computed in higher dimensions through the algorithm proposed in Chen, Gao and Ren (2017), where the authors performed simulations requiring to compute the deepest scatter matrix for dimensions and sample sizes as large as 10 and 2000, respectively. Their implementation of this algorithm is available as an R package at `https://github.com/ChenMengjie/DepthDescent`.

As pointed by an anonymous referee, the concept of scatter halfspace depth also makes sense when the parameter space is the compactification of $\mathcal{P}_k$, that is, is the collection $\overline{\mathcal{P}}_k$ of $k \times k$ symmetric positive *semi*-definite matrices. Interestingly, it is actually easier to investigate the properties of scatter halfspace depth over $\overline{\mathcal{P}}_k$ than over $\mathcal{P}_k$. The F-continuity and F-boundedness results in Theorems 3.1-3.2 extend, mutatis mutandis, to $\overline{\mathcal{P}}_k$.

Unlike $(\mathcal{P}_k, d_F)$, the metric space $(\overline{\mathcal{P}}_k, d_F)$ is complete, so that the regions $R^{\mathrm{sc}}_{P,T}(\alpha)$ are then F-compact for any $\alpha > 0$. Consequently, a trivial adaptation of the proof of Theorem 4.3 allows to show that there *always* exists a halfspace deepest scatter matrix in $\overline{\mathcal{P}}_k$. It is fortunate that these neat results can be established by considering the F-distance only, as the geodesic distance, that is unbounded on $\overline{\mathcal{P}}_k \times \overline{\mathcal{P}}_k$, could not have been considered here. Of course, in many applications, $\mathcal{P}_k$ remains the natural parameter space since many multivariate statistics procedures will require inverting scatter matrices. In such applications, it will be of little help to practitioners that the deepest halfspace scatter matrix belongs to $\overline{\mathcal{P}}_k \setminus \mathcal{P}_k$, which explains why our detailed investigation focusing on $\mathcal{P}_k$ is of key importance.

Perspectives for future research are rich and diverse. The proposed halfspace depth concepts for scatter, concentration and shape can be extended to other scatter functionals of interest. In particular, halfspace depths that are relevant for PCA could result from the "profile depth" approach in Section 7. For instance, the T-*"first principal direction" halfspace depth* of $\beta (\in \mathcal{S}^{k-1})$ with respect to the probability measure P over $\mathbb{R}^k$ can be defined as

$$HD^{1^{\mathrm{st}}\mathrm{pd}}_{P,T}(\beta) = \sup_{\Sigma \in \mathcal{P}_{k,1,\beta}} HD^{\mathrm{sc}}_{P,T}(\Sigma), \quad \text{with } \mathcal{P}_{k,1,\beta} := \{\Sigma \in \mathcal{P}_k : \Sigma\beta = \lambda_1(\Sigma)\beta\}.$$

The halfspace deepest first principal direction is a promising robust estimator of the true underlying first principal direction, at least under ellipticity. Obviously, the depth of any other principal direction, or the depth of any eigenvalue, can be defined accordingly. Another direction of research is to explore inferential applications of the proposed depths. Clearly, point estimation is to be based on halfspace deepest scatter, concentration or shape matrices; Chen, Gao and Ren (2017) partly studied this already for scatter in high dimensions. Hypothesis testing is also of primary interest. In particular, a natural test for $\mathcal{H}_0 : \Sigma = \Sigma_0$, where $\Sigma_0 \in \mathcal{P}_k$ is fixed, would reject the null for small values of $HD^{\mathrm{sc}}_{P_n,T}(\Sigma_0)$. For shape matrices, a test of sphericity would similarly reject the null for small values of $HD^{\mathrm{sh},S}_{P_n,T}(I_k)$. These topics, obviously, are beyond the scope of the present work.

Acknowledgements. We would like to thank the Editor Edward I. George, the Associate Editor and three anonymous referees for their insightful comments and suggestions, that led to a substantial improvement of a previous version of this work.

SUPPLEMENTARY MATERIAL

Supplement to "Halfspace Depths for Scatter, Concentration and Shape Matrices"

(doi: completed by the typesetter; .pdf). In this supplement, we conduct a Monte Carlo exercise validating the explicit scatter halfspace depth expressions obtained in the Gaussian and independent Cauchy examples. We also provide illustrations of Theorem 3.3 and Theorems 7.8-7.10. Finally, we prove all theorems stated in this paper.

REFERENCES

Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542.

Berger, M. (2003). *A Panoramic View of Riemannian Geometry*. Springer-Verlag, Berlin Heidelberg.

Bhatia, R. (2007). *Positive definite matrices. Princeton Series in Applied Mathematics*. Princeton University Press, Princeton, NJ.

Bhatia, R. and Holbrook, J. (2006). Riemannian geometry and matrix geometric means. *Linear Algebra Appl.* **413** 594–618.

Cardot, H., Cénac, P. and Godichon-Baggioni, A. (2017). Online estimation of the geometric median in Hilbert spaces : Non asymptotic confidence balls. *Ann. Statist.* **45** 591–614.

Cartan, E. (1929). Groupes simples clos et ouverts et géometrie riemannienne. *J. Math. Pures Appl.* **8** 1–33.

Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231.

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

Chen, M., Gao, C. and Ren, Z. (2017). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.,* to appear.

Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.* **109** 411–423.

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

Dang, X. and Serfling, R. J. (2010). Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties. *J. Statist. Plann. Inference* **140** 198–213.

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

Dümbgen, L. and Tyler, D. E. (2016). Geodesic convexity and regularized scatter estimators. ArXiv:1607.05455v2.

Fan, Y., Jin, J. and Yao, Z. (2013). Optimal classification in sparse Gaussian graphic model. *Ann. Statist.* **41** 2537–2571.

Fan, Y. and Lv, J. (2016). Innovated scalable efficient estimation in ultra-large gaussian graphical models. *Ann. Statist.* **44** 2098–2126.

Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732.

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

He, Y. and Einmahl, J. H. J. (2017). Estimation of extreme depth-based quantile regions. *J. R. Stat. Soc. Ser. B* **79** 449–461.

Hubert, M., Rousseeuw, P. J. and Segaert, P. (2015). Multivariate functional outlier detection. *Stat. Methods Appl.* **24** 177–202.

Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39** 2448–2476.

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. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858.

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

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

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

Nieto-Reyes, A. and Battey, H. (2016). A topologically valid definition of depth for functional data. *Statist. Sci.* **31** 61–79.

Paindaveine, D. and Van Bever, G. (2014). Inference on the shape of elliptical distributions based on the MCD. *J. Multivariate Anal.* **129** 125-144.

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

Paindaveine, D. and Van Bever, G. (2017). Supplement to "Halfspace depths for scatter, concentration and shape matrices". Manuscript under consideration.

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

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

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

Serfling, R. J. (2004). Some perspectives on location and scale depth functions. *J. Amer. Statist. Assoc.* **99** 970–973.

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* 523–531. Canad. Math. Congress, Montreal, Que.

Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426.

Zhang, J. (2002). Some extensions of Tukey's depth function. *J. Multivariate Anal.* **82** 134–165.

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

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

Université libre de Bruxelles
ECARES and Département de Mathématique
Avenue Roosevelt, 50, ECARES, CP114/04
B-1050, Brussels
Belgium
E-mail: dpaindav@ulb.ac.be
URL: http://homepages.ulb.ac.be/~dpaindav

Université libre de Bruxelles
ECARES, Avenue Roosevelt, 50, CP114/04
B-1050, Brussels
Belgium
E-mail: gvbever@ulb.ac.be
URL: https://sites.google.com/site/germainvanbever