ON DEPTH AND DEEP POINTS: A CALCULUS[1]

BY IVAN MIZERA

Comenius University and University of Alberta

For a general definition of depth in data analysis a differential-like calculus is constructed in which the location case (the framework of Tukey's median) plays a fundamental role similar to that of linear functions in the mathematical analysis. As an application, a lower bound for maximal regression depth is proved in the general multidimensional case—as conjectured by Rousseeuw and Hubert and others. This lower bound is demonstrated to have an impact on the breakdown point of the maximum depth estimator.

1. Introduction and outline. The notion of depth provides a route to possible analogs of the sample median and quantiles—beyond the univariate location model. Unlike other approaches, depth substantially elaborates on the order structure of the sample space.

In the univariate location case, the first relevant observations can be traced in Hotelling (1929) and Chamberlin (1933). For multivariate location, the proposal of Tukey (1975) was developed by Donoho and Gasko (1992); the germ of the idea appeared already in Hodges (1955). A breakthrough broadening the understanding of depth was the invention of regression depth by Rousseeuw and Hubert (1999a); see also Rousseeuw and Hubert (1999b) and Hubert, Rousseeuw and Van Aelst (1999); the precursors here were Edgeworth (1888) and Daniels (1954). The interrelations between location and regression were indicated by Hill (1960) and Carrizosa (1996).

In pursuit of the regression version, Rousseeuw and Hubert (1999a) isolated the crucial general essence of halfspace depth: its connection to admissibility in a certain data-analytic sense. In the present paper, we further develop their idea: a general definition of depth is formalized in the framework of vector optimization. Several examples aim at convincing the reader that this way of thinking about depth opens a transparent route to depth-based analysis in various statistical models. Furthermore, the vector-optimization approach leads to a sort of differential calculus, where the simplest, multivariate location depth plays the role of the prototype—similarly as linear structures do in the analysis of nonlinear ones. To

illustrate the power of these techniques, we show how they can be successfully used for dealing with the so-called centerpoint problem; in particular, we settle certain standing conjectures from the literature. Finally, we give some statistical implications of centerpoint considerations: results about the bias and breakdown properties of maximum depth estimators in a general setting.

The range of questions opened is broad and difficult to cover in one paper. At the moment, our objective is to lay necessary theoretical and mathematical foundations; the detailed study of depth and depth-based procedures in concrete statistical models is left for the forthcoming work. For applications and aspects of depth, the reader may also consult Liu, Parelius and Singh (1999), Caplin and Nalebuff (1988, 1991a, b) and other references given or not given in this paper. The relevant (largely asymptotic) theory can be found in He and Wang (1997), Nolan (1992, 1999), He and Portnoy (1998) and Bai and He (1999). Our approach to depth is limited to what is called halfspace or Tukey's depth; for other brands of depth in multivariate location see Liu, Parelius and Singh (1999) and Liu and Singh (1993), and in linear regression Rousseeuw and Hubert (1999a).

In Section 2, the definitions of global, local and tangent depth are introduced, based on vector optimization formalism—with weak optimality as the basic notion. General inequalities between, as well as convexity-based criteria for the equality of sorts of depth are given. A few statistical models (covering nonetheless all instances studied in the literature so far) are introduced as examples.

Section 3 extends tangent depth to the general measure-theoretic setting—covering both finite-sample and population distributions. The equality to intuitive definitions, given in the literature for several models, is demonstrated with examples. The centerpoint questions are expounded in Section 3.3; the most important result is Theorem 3.3. Its less general version Theorem 3.4 admits a shorter and more intuitive proof and extends also to multivariate regression.

Statistical implications of centerpoint considerations are treated in Section 4. Particularly, in Section 4.1 we show that the bias sets of maximum depth estimators are contained in upper level sets of depth—this fact giving another statistical interpretation for depth contours. In Section 4.2, we give lower bounds for the total variation and contamination breakdown points and apply them to regression depth, with the help of Theorem 5.25—the latter also implies the existence of a fit with maximal regression depth in the general situation, stated in Theorem 3.5.

Mathematical details are collected in Section 5, which, besides the proofs of the results, contains also a wealth of the supporting material and adds a lot of illuminating details to the—albeit in principle self-contained—presentation in Sections 2, 3 and 4. The proofs of all theorems and propositions from Section 2 can be found in Section 5.1, in their order of appearance. Theorem 3.2 is proved at the beginning of Section 5.4, followed later by the proof of Proposition 3.1; Theorem 3.4 in Section 5.7; Section 5.8 contains proofs of both theorems from Section 4; and finally, Theorem 5.25, whose direct consequence is Theorem 3.5,

is formulated and proved in Section 5.9. The notation and techniques developed in Section 5 are also used in the Appendix, devoted to the full proof of Theorem 3.3.

Some of the results have been announced in Mizera (1998) and Portnoy and Mizera (1999).

2. Depth in data analysis.

2.1. *Preliminaries.* We denote by $\mathbb{R}^p$ the p-dimensional Euclidean space, by $\|\cdot\|$ its Euclidean norm, and by 0 the point with zero norm. If $\vartheta, u \in \mathbb{R}^p$, $u \neq 0$, we write $\mathsf{H}_{\vartheta,u}$ for the set $\{x \in \mathbb{R}^p : u^{\mathrm{T}}(x - \vartheta) \geq 0\}$, the closed halfspace whose boundary contains ϑ and is orthogonal to the vector u pointing into the halfspace. Obviously, $\mathsf{H}_{\vartheta,u} = \mathsf{H}_{\vartheta,cu}$ for any $c > 0$; hence we often pick u from the sphere $\mathbb{S}^{p-1} = \{x \in \mathbb{R}^p : \|x\| = 1\}$. If $\vartheta = 0$, we abbreviate $\mathsf{H}_{0,u}$ to H_u.

We denote by A^{c} the complement of A—with respect to a basic set which should be clear from the context. For instance, if A is a subset of the indexing set $\{1, 2, \ldots, n\}$, A^{c} denotes the set $\{1, 2, \ldots, n\} \smallsetminus A$.

2.2. *Generalities.* A typical data-analytic model consists of a collection Z of observations $z_1, z_2, \ldots, z_n$, whose values lie in a sample space Z. For these data, a fit is sought: an element ϑ of a parameter space Θ.

To give each observation its impact on the result, a criterial function F_i is attached to every observation z_i; the lower the value of F_i at ϑ, the better ϑ fits z_i. Fits that yield optimal values for all F_i—that is, uniformly best solutions—occur only for rare data configurations. Classical strategies thus consider trade-offs, giving each point its share via a compound criterial function— the sum of all F_i, say.

The approach involving depth could be characterized as elaborating on a "degree of data-analytic admissibility." The general definition of depth, given by Rousseeuw and Hubert (1999a), says that "the depth of ϑ is the smallest number of observations that would need to be removed to make ϑ a nonfit." The word "nonfit" means a parameter value inadmissible from a data-analytic view, a parameter value with zero depth. In what follows, we develop this idea in the framework of vector, multi-objective optimization, in the setting which employs criterial functions. Such an approach provides a guide how to define depth in various statistical models— via a natural transition from classical techniques. It also lays a firm technical foundation for the study of depth and related notions by methods of differential calculus.

2.3. *Global depth.* Let n stand for card Z and N for the index set $\{1, 2, \ldots, n\}$; if $A \subseteq N$, we write $\#A$ for card A divided by n.

Suppose that a criterial function $F_i = F_{z_i}$ acting from Θ to $[0, \infty)$ is attached to every observation z_i from Z; note that the notation implies that criterial functions for two equal observations coincide. A parameter value $\vartheta \in \tilde{\Theta} \subseteq \Theta$ will be called

weakly optimal in $\tilde{\Theta}$ with respect to $A \subseteq N$, if $A \neq \varnothing$ and there is no $\tilde{\vartheta} \in \tilde{\Theta}$ such that $F_{z_i}(\tilde{\vartheta}) < F_{z_i}(\vartheta)$ for all $i \in A$. We define the global depth of $\vartheta \in \Theta$ to be

$$(1) \quad d_G(\vartheta, \mathcal{Z}) = \min \#\{A \subseteq N : \vartheta \text{ is not weakly optimal in } \Theta \text{ with respect to } A^c\}.$$

If ϑ is not weakly optimal with respect to the full collection of F_i's, then the minimal set A in (1) is empty and $d_G(\vartheta, \mathcal{Z}) = 0$. On the other hand, if ϑ is weakly optimal with respect to any subset of N, then $d_G(\vartheta, \mathcal{Z}) = 1$; by definition, ϑ is not weakly optimal with respect to the empty set of criterial functions.

If we choose certain other criterial functions $\tilde{F}_{z_i}$ to be attached to any data point z_i, then such a choice may lead to the same depth—as long as $F_{z_i}(\vartheta) < F_{z_i}(\tilde{\vartheta})$ if and only if $\tilde{F}_{z_i}(\vartheta) < \tilde{F}_{z_i}(\tilde{\vartheta})$. Obviously, depth depends essentially only on the order induced by the criterial functions; nevertheless, the specific form of criterial functions often allows for better technical handling than abstract order notions.

To achieve immediate compatibility with the population case, we define depth as minimal *proportion*—and not *number* of observations (as common in the literature). Our depth thus has values $0, 1/n, 2/n, \ldots, 1$ instead of $0, 1, 2, \ldots, n$—a minor detail ignored in the sequel, where we may speak about the equality of depths even if it may actually hold up to a multiplication by n.

The word "admissibility" we used in the introduction suggests the relationship to those notions from statistical theory related to loss and risk functions under a probabilistic model for the data. While from the decision-theoretic aspect the relationship is really close, we have to raise several cautions. First, "admissibility" considered here is a different, data-analytic one: "residual admissibility." An example: any multivariate location estimator contained in the convex hull of the observations—the sample mean, for instance—is weakly optimal (and thus "residual admissible"). Compare this with the well-known fact from statistical theory: the sample mean in dimensions beyond three is inadmissible with respect to quadratic loss function under a normal sampling distribution.

Second, "admissibility" in decision theory means what the vector optimization literature calls Pareto optimality: there is no solution which performs *strictly better in one* criterion and *better or equally well in others*. But, for the definition of depth, we do not use Pareto, but weak optimality (also known as "weak Pareto optimality," "Slater optimality," "weak efficiency"): there is no solution performing *strictly better in all* criteria. The reason is that in order to have a consistent definition of depth, an omission of a criterial function should not create optimality, only possibly destroy it. This is a property of weak, but not Pareto optimality.

To prevent misunderstandings in the sequel, we therefore avoid the word "admissibility"; we also speak about "criterial" rather than "loss" functions. On the other hand, we do not adopt the word "nonfit" either; as given in Rousseeuw and Hubert (1999a), we consider it semantically on a more general and, to an extent, intuitive level than our rigorous "not weakly optimal fit."

2.4. *Examples.*   We now introduce several statistical models to be analyzed in the sequel.

EXAMPLE 1 (Multivariate location).   In the location model, $\mathsf{Z} = \Theta = \mathbb{R}^p$. Natural criterial functions are those based on the distance $\|\vartheta - z\|$ of ϑ from z; in fact, any increasing function of this distance results in the same depth. For technical convenience, we choose $F_z(\vartheta) = \frac{1}{2}\|\vartheta - z\|^2$.

It is not hard to see that ϑ is weakly optimal if and only if it lies in the convex hull of the data points. Applying the definition of the global depth to this model, we obtain the definition of location depth used by Rousseeuw and Hubert (1999a): $d_G(\vartheta, \mathcal{Z})$ is the minimal proportion of points whose removal makes ϑ lie outside the convex hull of the remaining ones. This definition is equivalent [see (2.6)] to the halfspace definition of Tukey (1975) and Donoho and Gasko (1992).

EXAMPLE 2 (Linear regression).   In the regression model, an observation $z_i = (x_i^{\mathrm{T}}, y_i)^{\mathrm{T}}$ consists of a response y_i and a vector of covariates x_i. The sample space is $\mathsf{Z} = \mathsf{X} \times \mathbb{R}$, where X is a subset of $\mathbb{R}^p$; the parameter space is $\Theta = \mathbb{R}^p$. Most models have $x_i = (1, w_i^{\mathrm{T}})^{\mathrm{T}}$ and $\mathsf{X} = \{1\} \times \mathbb{R}^{p-1}$ accordingly—regression with intercept. When w_i runs over $\mathbb{R}$ we speak about the simple linear regression; when it runs over $\mathbb{R}^{p-1}$, about multiple linear regression. The general model covers also cases when $x_i = g(w_i)$ and w_i is a lesser-dimensional covariate; for instance, quadratic regression, when $x_i = (1, w_i, w_i^2)^{\mathrm{T}}$.

Natural criterial functions in regression depend on the residuals. Any increasing function of the absolute residuals leads to the same result [see (2.6)], regression depth of Rousseeuw and Hubert (1999a). We choose $F_z(\vartheta) = \frac{1}{2}(y - \vartheta^{\mathrm{T}}x)^2$; another possibility is $F_z(\vartheta) = |y - \vartheta^{\mathrm{T}}x|$.

EXAMPLE 3 (General, nonlinear regression).   Nonlinear regression is a generalization of the linear one; however, the usual notation is slightly different. The observations are $z_i = (w_i^{\mathrm{T}}, y_i)^{\mathrm{T}}$, drawn from $\mathsf{Z} = \mathsf{W} \times \mathbb{R}$; the functional dependence is given by a regression function $f(\vartheta, w)$, the regression being linear if $f(\vartheta, w) = \vartheta^{\mathrm{T}}g(w)$.

Similarly to linear regression, we choose $F_z(\vartheta) = \frac{1}{2}(y - f(\vartheta, w))^2$. Simple nonlinear (linearizable) models were considered in Rousseeuw and Hubert (1999a) and Van Aelst, Rousseeuw, Hubert and Struyf (2001).

EXAMPLE 4 (Multivariate linear regression).   Multivariate regression is a generalization of Example 2 in another direction. The functional dependence remains linear, but the response y_i is allowed to be multi-dimensional. An observation $z_i = (x_i^{\mathrm{T}}, y_i^{\mathrm{T}})^{\mathrm{T}}$ belongs to $\mathsf{Z} = \mathsf{X} \times \mathbb{R}^m$, where X is a subset of $\mathbb{R}^k$. For notational convenience, we consider the parameter Θ to be a $k \times m$ matrix lying in $\Theta = \mathbb{R}^p = \mathbb{R}^{km}$ (interpreting it also as a vector, if necessary). The criterial functions are $F_z(\Theta) = \frac{1}{2}\|y - \Theta^{\mathrm{T}}x\|^2$.

Our attention to this model was turned by the work of Bern and Eppstein (2000), who gave a geometric definition of multivariate regression depth.

EXAMPLE 5 (Orthogonal regression).   In orthogonal regression, the observations z_i are points from $\mathbb{R}^p$ and fits are k-dimensional affine subspaces in $\mathbb{R}^p$. We consider only the simple (traditional) case when $k = p - 1$ and fits are hyperplanes, affine subspaces of codimension 1; this case parallels classical regression, where one of the variables is interpreted as the response and other as covariates. A hyperplane is parametrized by $\vartheta = (s, \beta^{\mathrm{T}})^{\mathrm{T}}$, where $\beta \in \mathbb{S}^{p-1}$ is a unit vector orthogonal to the hyperplane and $s\beta$ is the intersection of the hyperplane with the linear space generated by β. The resulting parameter space is $\mathbb{R} \times \mathbb{S}^{p-1}$. We found this parametrization convenient, despite the lack of identification: (s, β^{T}) and $(-s, -\beta)$ represent the same hyperplane.

The fact that the regression is orthogonal is expressed by the choice of the criterial functions. They are based on the orthogonal distances of observations to the fitted hyperplane: $F_z(\vartheta) = \frac{1}{2}(\beta^{\mathrm{T}} z - s)^2$.

2.5. *Local depth.*   Weak optimality can be effectively studied in a way akin to the classical approach of the differential calculus to extrema. The first step is a transition to local notions. We define the local depth of $\vartheta \in \Theta$ to be

(2)
$$d_{\mathrm{loc}}(\vartheta, \mathcal{Z}) = \min \#\{A \subseteq N : \vartheta \text{ is weakly optimal in}$$
$$\text{no neighborhood of } \vartheta \text{ w.r.t. } A^{\mathrm{c}}\},$$

where "in no neighborhood of ϑ" means "in no open $\tilde{\Theta} \subseteq \Theta$ containing ϑ." Since Θ is itself a neighborhood of ϑ, the global depth never exceeds the local one; the following theorem gives a sufficient condition for their equality. For the definition of quasiconvexity, see Section 5.1.

THEOREM 2.1.   *For any $\vartheta \in \Theta$, $d_{\mathrm{G}}(\vartheta, \mathcal{Z}) \leq d_{\mathrm{loc}}(\vartheta, \mathcal{Z})$. If Θ is an open convex subset of $\mathbb{R}^p$ and all $F_i = F_{z_i}$, attached to the data points $z_i \in \mathcal{Z}$, are strictly quasiconvex (in particular, convex), then $d_{\mathrm{G}}(\vartheta, \mathcal{Z}) = d_{\mathrm{loc}}(\vartheta, \mathcal{Z})$.*

2.6. *Tangent depth.*   In the calculus methodology of handling extremes, the crucial second step is the use of derivatives. Suppose that Θ is a p-dimensional manifold. Given a function F from Θ to $\mathbb{R}$, we denote by $\nabla F(\vartheta)$ the derivative (gradient) of F at ϑ: a linear functional from the tangent space of Θ at ϑ to $\mathbb{R}$, representing the local linear approximation of F at ϑ. Since all tangent spaces are isomorphic to $\mathbb{R}^p$, we identify in the usual fashion $\nabla F(\vartheta)$ with a vector in $\mathbb{R}^p$, the vector of partial derivatives of F at ϑ.

(The reader not comfortable with this language may think about derivatives in a less sophisticated way: in most models, the parameter space is actually $\mathbb{R}^p$ and

gradients are the vectors of the partial derivatives taken in the elementary way. Example 5, however, shows the need for the advanced formalism.)

The differential approach to vector optimization dates back to Frisch (1966). As in the classical calculus, we deal here with first-order necessary and second-order sufficient conditions; for Pareto optima those were developed by Smale (1973, 1975a, b) and Wan (1975, 1978). Their first-order, necessary condition for a Pareto optimum turns out to be actually the same as the one required for a weak optimum; see Section 5.1 for more details. It provides also a sufficient condition when the criterial functions possess a certain degree of convexity; and since this is all we need for the development of depth theory, we do not introduce any second-order sufficient conditions for weak optima in this paper.

If S is a subset S of $\mathbb{R}^p$, we say that S surrounds $\vartheta \in \mathbb{R}^p$ (ϑ is surrounded by S) whenever ϑ lies in the convex hull of S.

PROPOSITION 2.2. *If $S \subseteq \mathbb{R}^p$ surrounds $\vartheta \in \mathbb{R}^p$, then it contains a finite subset with at most $p + 1$ elements that also surrounds ϑ. The following are equivalent*:

(i) $\vartheta \in \mathbb{R}^p$ *is surrounded by S*;
(ii) *there is no open halfspace in $\mathbb{R}^p$ which contains S and has ϑ on its boundary*;
(iii) *if S is finite and $S = \{\eta_1, \eta_2, \ldots, \eta_k\}$, there are nonnegative $\lambda_1, \lambda_2, \ldots, \lambda_k$, not all equal to zero and such that $\sum \lambda_i \eta_i = \vartheta$.*

It is not hard to see that if ϑ is locally weakly optimal, then the origin 0 must be surrounded by $\nabla F_i(\vartheta)$; in other words, once all $\nabla F_i(\vartheta)$ are contained in an open halfspace with 0 on its boundary, then ϑ is not locally weakly optimal. Hence, we have a necessary condition for weak optimality which leads to the following definition. Suppose that Φ is a function from $\Theta \times \mathsf{Z} \to \mathbb{R}^p$. Writing its values for $\vartheta \in \Theta$ and $z \in \mathsf{Z}$ as $\Phi_\vartheta(z)$, we define the tangent depth of $\vartheta \in \Theta$ to be

$$(3) \qquad d_{\mathrm{T}}^\Phi(\vartheta, \mathsf{Z}) = \min_{u \neq 0} \#\{i : u^{\mathrm{T}} \Phi_\vartheta(z_i) \geq 0\} = \min_{\|u\|=1} \#\{i : \Phi_\vartheta(z_i) \in \mathsf{H}_u\}.$$

The minimum is taken over all closed halfspaces with 0 on the boundary—we may therefore take it over $u \neq 0$ or over $\|u\| = 1$, whichever is more convenient. The connection between local and tangent depth is established by setting $\Phi_\vartheta(z) = \nabla F_z(\vartheta)$; nevertheless, the general definition opens room for general Φ, not necessarily coming from the vector-optimization problem (an analogy could be general estimating equations not necessarily arising from the maximization of a likelihood). In our examples, we omit the superscript Φ whenever its form is clear from the context.

A parameter ϑ surrounded by gradients may not yet be a weak local optimum. Nevertheless, it often is, for instance, when all F_i are convex. For the definition of pseudoconvexity, see Section 5.1.

THEOREM 2.3.   *Suppose that all $F_i = F_{z_i}$, attached to the data points $z_i \in \mathcal{Z}$, are differentiable and let $\Phi_{\vartheta}(z_i) = \nabla F_{z_i}(\vartheta)$. For any $\vartheta \in \Theta$, $d_{\mathrm{loc}}(\vartheta, \mathcal{Z}) \leq d_{\mathrm{T}}^{\Phi}(\vartheta, \mathcal{Z})$. If Θ is an open convex subset of $\mathbb{R}^p$ and all F_i are pseudoconvex* (in particular, convex), *then $d_{\mathrm{T}}^{\Phi}(\vartheta, \mathcal{Z}) = d_{\mathrm{loc}}(\vartheta, \mathcal{Z}) = d_{\mathrm{G}}(\vartheta, \mathcal{Z})$.*

The proofs of Theorems 2.1 and 2.3 given in Section 5.1 reveal that neither strict quasiconvexity nor pseudoconvexity is a minimal sufficient condition required for the equality of depths.

2.7. *Examples of tangent depth.*   Now we are ready to reconsider our examples and show that they cover all instances of halfspace depth studied in the literature so far.

EXAMPLE 1.   The criterial functions F_z are convex and

$$(4) \qquad\qquad \Phi_{\vartheta}(z) = \nabla F_z(\vartheta) = \vartheta - z.$$

Therefore,

$$d_{\mathrm{G}}(\vartheta, \mathcal{Z}) = d_{\mathrm{T}}(\vartheta, \mathcal{Z}) = \min_{u \neq 0} \#\{i : u^{\mathrm{T}}(\vartheta - z_i) \geq 0\}$$

$$= \min_{u \neq 0} \#\{i : u^{\mathrm{T}}(z_i - \vartheta) \geq 0\} = \min_{\|u\|=1} \#\{i : z_i \in \mathsf{H}_{\vartheta,u}\}.$$

This shows the equality to the original halfspace definition of Tukey (1975) and Donoho and Gasko (1992).

EXAMPLE 4.   Preserving the matrix dimension of Θ, we obtain that

$$(5) \qquad\qquad \Phi_{\vartheta}(z) = \nabla F_z(\Theta) = -x(y - \Theta^{\mathrm{T}}x)^{\mathrm{T}} = xx^{\mathrm{T}}\Theta - xy^{\mathrm{T}}.$$

We denote by $U \cdot V$ the inner product of two matrices considered as vectors: $U \cdot V = \mathrm{tr}(U^{\mathrm{T}}V) = \mathrm{tr}(UV^{\mathrm{T}}) = \mathrm{tr}(V^{\mathrm{T}}U) = \mathrm{tr}(VU^{\mathrm{T}})$. Since F_z is convex for any z, we have

$$d_{\mathrm{G}}(\Theta, \mathcal{Z}) = d_{\mathrm{loc}}(\Theta, \mathcal{Z}) = d_{\mathrm{T}}^{\Phi}(\Theta, \mathcal{Z})$$

$$= \min_{U \neq \mathsf{O}} \#\{i : U \cdot (x_i x_i^{\mathrm{T}}\Theta - x_i y_i^{\mathrm{T}}) \geq 0\}$$

$$(6) \qquad\qquad = \min_{U \neq \mathsf{O}} \#\{i : \mathrm{tr}((x_i x_i^{\mathrm{T}}\Theta - x_i y_i^{\mathrm{T}})^{\mathrm{T}}U) \geq 0\}$$

$$= \min_{U \neq \mathsf{O}} \#\{i : -\mathrm{tr}((y_i - \Theta^{\mathrm{T}}x_i)x_i^{\mathrm{T}}U) \geq 0\}$$

$$= \min_{U \neq \mathsf{O}} \#\{i : -(x_i^{\mathrm{T}}U)(y_i - \Theta^{\mathrm{T}}x_i) \geq 0\},$$

with U running over all $k \times m$ matrices not identically equal to zero.

A question arises about the relationship of not weakly optimal Θ and nonfits as defined by Bern and Eppstein (2000). If the two notions are equivalent, that is, pick up identical sets of Θ, then the definition of multivariate regression depth given above is equivalent to that of Bern and Eppstein (2000). Despite certain positive evidence (D. Eppstein, personal communication), we now believe that the implication holds only in one direction: our definition minimizes over the larger set, thus the multivariate regression depth defined here does not exceed that of Bern and Eppstein (2000); if Θ is weakly optimal, it cannot be a nonfit of Bern and Eppstein (2000), but there are nonfits which are not weakly optimal. Note that the inequality has the right direction to ensure that our centerpoint-lower bounds on the multivariate regression depth are applicable also to that of Bern and Eppstein (2000).

EXAMPLE 2. This model is a special case of Example 4; using (6) yields

$$(7) \quad d_{\mathrm{G}}(\vartheta, \mathcal{Z}) = d_{\mathrm{loc}}(\vartheta, \mathcal{Z}) = d_{\mathrm{T}}(\vartheta, \mathcal{Z}) = \min_{u \neq 0} \#\{i : u^{\mathrm{T}}(x_i x_i^{\mathrm{T}} \vartheta - x_i y_i) \geq 0\}.$$

Obviously, the criterial functions remain convex. We may reexpress (7) in many ways:

$$(8) \qquad d_{\mathrm{T}}(\vartheta, \mathcal{Z}) = \min_{u \neq 0} \#\{i : -(u^{\mathrm{T}}x_i)\,\mathrm{sgn}(y_i - \vartheta^{\mathrm{T}}x_i) \geq 0\}$$

$$(9) \qquad = \min_{u \neq 0} \#\{i : \mathrm{sgn}(u^{\mathrm{T}}x_i)\,\mathrm{sgn}(y_i - \vartheta^{\mathrm{T}}x_i) \geq 0\};$$

in all these formulas, we are free to include or drop the minus sign and also to restrict the domain of minimization from $u \neq 0$ to $\|u\| = 1$. Incidentally, (8) can be interpreted as arising from the criterial functions of $|y - \vartheta^{\mathrm{T}}x|$; but not all reexpressions should arise in a similar way.

To see that equations (7)–(9) yield the regression depth whose geometric definition was given by Rousseeuw and Hubert (1999a), note first that the minimized function is piecewise constant; hence the equivalent result is obtained by a minimization over a dense set of directions u. Fix this set to be the set S of all $u \in \mathbb{S}^{p-1}$ such that $u^{\mathrm{T}}x_i \neq 0$ for all x_i. When $x_i = (1, w_i)^{\mathrm{T}}$, the complement of S contains only finitely many hyperplanes of lower dimension; the general case is treated similarly below. The expression (9) then equals the minimal proportion over S of observations such that

$$(10) \qquad \begin{array}{l} \text{either } (u^{\mathrm{T}}x_i > 0 \text{ and } \mathrm{sgn}(y_i - \vartheta^{\mathrm{T}}x_i) \geq 0) \\[4pt] \text{or } (u^{\mathrm{T}}x_i < 0 \text{ and } \mathrm{sgn}(y_i - \vartheta^{\mathrm{T}}x_i) \leq 0). \end{array}$$

It is perhaps interesting to mention an equivalent possibility that recently appeared in Adrover, Maronna and Yohai (2000): take the minimal proportion of observations such that

$$\mathrm{sgn}\left(\frac{y_i - \vartheta^{\mathrm{T}}x_i}{u^{\mathrm{T}}x_i}\right) \geq 0.$$
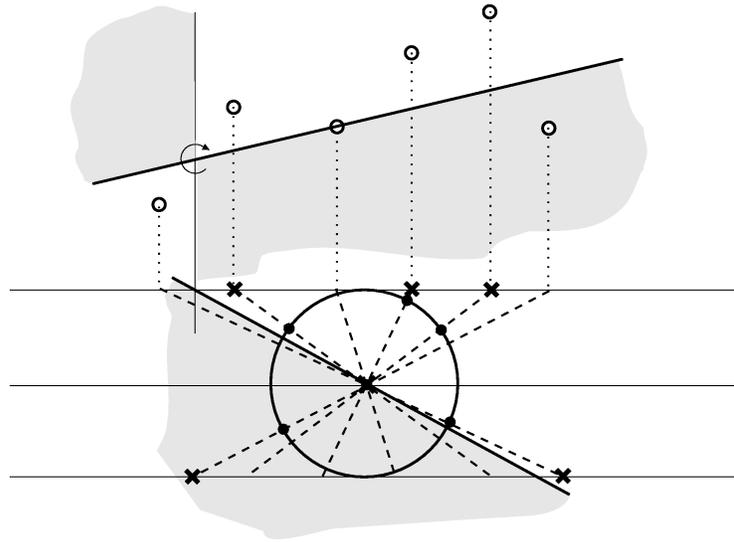
FIG. 1.   *Regression depth as location depth.*

The set of observations satisfying (10) corresponds to the shaded area in Figure 1. All observations lying on the solid line are included, and no observation is expected to lie on the vertical line. The set given by (10) contains all observations that are met by a line during its rotation to the vertical position, in the sense indicated in Figure 1.

Regression without the intercept (when $x_i = w_i$) behaves similarly; we only have to pay separate attention to points with $x = 0$. They have no influence on the resulting fit, but increase simultaneously the depth of all fits; in fact, we may remove them all without destroying the weak optimality of any fit.

Figure 1 illustrates how tangent depth may be evaluated in simple regression (and it is not hard to see how the same method works in general). We evaluate the depth of the (solid) regression line $\vartheta = (\alpha, \beta)^{\mathrm{T}}$, with respect to the (small empty circles) observations (w_i, y_i); recall that $x_i = (1, w_i)^{\mathrm{T}}$. We start by projecting the observations onto the covariate space $\{1\} \times \mathbb{R}^{p-1}$ (indicated by dotted lines). Then we project them further (along dashed lines) to the unit sphere—that is, we plot their normalized directions (solid circles), in the positive or negative halfspace according to the sign of the residual; if the residual is zero, as happens for one point at Figure 1, the corresponding projection goes to the origin (slightly obscured). The desired result is obtained as the location depth of the origin with respect to these projections (a minimizing halfspace is indicated by the shaded area). Points represented by crosses show how the same method would equivalently work for $F_z(\vartheta) = |y - \vartheta^{\mathrm{T}} x|$, when (8) replaces (7).

EXAMPLE 3.   All previous examples involved convex criterial functions. General nonlinear regression models provide exceptions—yet in many cases local

and tangent, or global and local depth are equal. The more thorough analysis of various nonlinear regression models is beyond the scope of this paper.

EXAMPLE 5. Also in this example, we limit our analysis to the simple two-dimensional case, where the observations are $z_i = (x_i, y_i)^{\mathrm{T}}$. For $\vartheta \in \mathbb{R} \times \mathbb{S}^1$, consider local coordinates $(s, t)^{\mathrm{T}} = \varphi(\vartheta)$ with the inverse $\varphi^{-1}(s, t) = (s, \beta(t)^{\mathrm{T}})^{\mathrm{T}} = (s, (-\sin t, \cos t)^{\mathrm{T}})^{\mathrm{T}}$. Representing $\nabla F_z(\vartheta)$ by the partial derivatives in these local coordinates, we obtain

$$\nabla F_z(\vartheta) = -(\beta^{\mathrm{T}} z - s)(1, z^{\mathrm{T}} \Gamma \beta)^{\mathrm{T}},$$

where Γ denotes the matrix of the clockwise rotation by 90 degrees (in $\mathbb{R}^2$); note that

$$-\Gamma\beta = (-\cos t, -\sin t)^{\mathrm{T}} = \frac{\partial}{\partial t}\beta(t).$$

Let ϑ_0 be the line represented by local coordinates $(0, 0)$, the line $y = 0$ in $\mathbb{R}^2$. Since $\varphi^{-1}(0, 0) = (0, \beta(0)^{\mathrm{T}})^{\mathrm{T}} = (0, (0, 1)^{\mathrm{T}})^{\mathrm{T}}$, we have, for $z = (x, y)^{\mathrm{T}}$,

$$(11) \qquad \nabla F_z(\vartheta_0) = -y(1, x)^{\mathrm{T}} = (-y, -yx)^{\mathrm{T}}.$$

The evaluation of the tangent depth of any ϑ can be reduced to this canonical case by the appropriate rotation and translation (which leave all orthogonal distances unchanged). In the simple linear regression model, the same line is represented as $y = 0x + 0$; equation (5) gives that $\nabla F_z(\vartheta) = -y(1, w)^{\mathrm{T}}$. Recall that w in simple linear regression corresponds in this example to x; in other words, we obtained the same expression in both cases. Therefore, the orthogonal regression tangent depth of ϑ is the regression depth of the line $y = 0$ after a Euclidean change of coordinates that carries the line represented by ϑ to the line $y = 0$. We obtained a geometric definition of the orthogonal regression tangent depth (Figure 2, left): (i) given a line orthogonal to the line represented by ϑ, take the minimum of the proportion of observations in the shaded area, and of those in its complement (counting those lying on the line represented by ϑ in both cases); (ii) minimize over all orthogonal lines that do not contain any of observations.
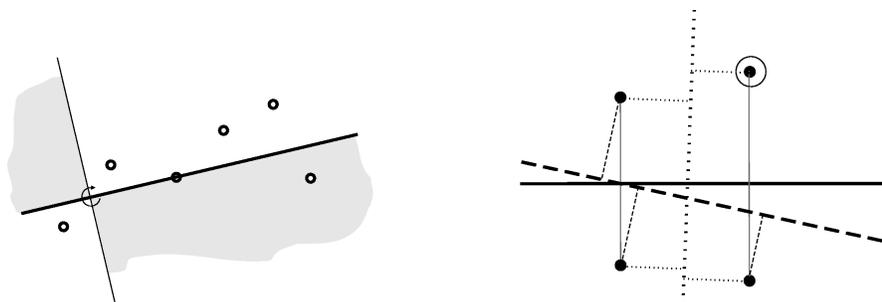

FIG. 2. *Depth in orthogonal regression.*

Figure 2 (right) shows that the orthogonal regression model is interesting also from another aspect: each of the depth notions introduced above may result in a different value. Consider the fit represented by the the solid line. It is easily seen that its tangent depth is 2. After removing the circled point, the orthogonal residuals of the dashed line uniformly supersede those of the solid one—therefore the local depth of the solid line is at most 1; on the other hand, the fit represented by the solid line is locally weakly optimal—hence its local depth is at least 1. Finally, the orthogonal residuals of the dotted line uniformly supersede those of the solid one; hence the fit is not globally weakly optimal and its global depth is 0.

3. Tangent depth in a measure-theoretic setting.

3.1. *Preliminaries.* By "a measure on $\mathbb{X}$" we always understand a measure defined on the appropriate Borel σ-algebra of a separable metrizable space $\mathbb{X}$; examples of $\mathbb{X}$ are $\mathbb{R}^p$, $\mathbb{S}^{p-1}$, the p-dimensional unit ball $\mathbb{D}^p = \{x \in \mathbb{R}^p : \|x\| \leq 1\}$ or the projective plane $\mathbb{RP}^{p-1}$ (arising from $\mathbb{S}^{p-1}$ by identifying the antipodal points). Halfspaces and similar geometric constructions are thus measurable; we implicitly assume the measurability of any set or function under consideration. With only the exception of Lebesgue measure, all measures we work with are "subprobability" measures—bounded measures with total mass not exceeding 1.

We denote by $f \circ g$ the composition mapping assigning the value $f(g(x))$ to x. Particularly, if P is a measure and g a function on $\mathbb{X}$, then $P \circ g^{-1}$ is another measure with the same total mass, assigning $P\{x : g(x) \in E\}$ to any E.

3.2. *Tangent depth in probability fields.* For statistical considerations, we need an analog of depth defined for population distributions—we need a model for sampling, or target for asymptotics, or both. The following general definition of the tangent depth extends that given in Section 2. It stems from the fact that formula (3) can be easily rewritten to depend on the empirical probability supported by the gradient points. The analogy ("plug-in") principle suggests then to replace the empirical probability by a general one.

Let $\mathbb{X}$ be any space where halfspaces can be reasonably defined: $\mathbb{R}^p$, but also $\mathbb{S}^{p-1}$ and $\mathbb{D}^p$ whose halfspaces are formed by the intersection of those from $\mathbb{R}^p$ with $\mathbb{S}^{p-1}$ or $\mathbb{D}^p$, respectively. We will later add also the cosmic space $\bar{\mathbb{R}}^p$ [see (5.3)]. For any measure Q on $\mathbb{X}$, we define

$$d(Q) = \inf_{u \neq 0} Q(\mathsf{H}_u) = \inf_{\|u\|=1} Q(\mathsf{H}_u).$$

Note that a minimizing halfspace may not exist, and therefore the use of inf instead of min is essential. The infimum, on the other hand, should not be necessarily taken over the full set of directions.

PROPOSITION 3.1. *Let $\mathbb{X}$ be $\mathbb{R}^p$, $\mathbb{S}^{p-1}$, $\mathbb{D}^p$ or $\bar{\mathbb{R}}^p$. For any measure Q on $\mathbb{X}$, $d(Q) = \inf_{u \in S} Q(\mathsf{H}_u)$ whenever S is a dense subset of $\mathbb{S}^{p-1}$.*

An important instance of a set *S* appearing in Proposition 3.1 is given by the following theorem.

THEOREM 3.2. *Let $\mathbb{X}$ be $\mathbb{R}^p, \mathbb{S}^{p-1}, \mathbb{D}^p$ or $\bar{\mathbb{R}}^p$. For any probability measure Q on $\mathbb{X}$, $\mu(S(Q)) = 1$, where μ is the uniform distribution on $\mathbb{S}^{p-1}$ and $S(Q)$ is:*

for $\mathbb{X} = \mathbb{R}^p, \mathbb{D}^p$ or $\bar{\mathbb{R}}^p$, the set of all $u \in \mathbb{S}^{p-1}$ such that $Q(\partial \mathsf{H}_u \smallsetminus \{0\}) = 0$;
for $\mathbb{X} = \mathbb{S}^{p-1}$, the set of all $u \in \mathbb{S}^{p-1}$ such that $Q(\partial \mathsf{H}_u) = 0$.

Let Θ be a metrizable topological space. We define a probability field on $\mathbb{X}$ indexed by Θ to be a function $\{P_\vartheta : \vartheta \in \Theta\}$ assigning a probability measure P_ϑ on $\mathbb{X}$ to each $\vartheta \in \Theta$. An important instance is when Θ is a p-dimensional manifold, $\mathbb{X} = \mathbb{R}^p$ is isomorphic to the tangent space of Θ at each of its points ϑ, and $\Phi_\vartheta(Z) = \nabla F_Z(\vartheta)$; in this case we speak of a gradient probability field. The motivation for the name is: if F is a differentiable function on Θ, then its gradient $\nabla F(\vartheta)$ defines a vector field, a mapping that assigns a vector from the tangent space of Θ at ϑ to each $\vartheta \in \Theta$. A collection of n differentiable functions produces n gradients; these n points support an empirical probability.

Given a probability field, then $d(P_\vartheta)$ naturally yields a depth function on Θ. To proceed more in analogy with the definition of tangent depth in 2.6, let us consider again a function Φ from $\Theta \times \mathsf{Z} \to \mathbb{X}$, where $\mathbb{X}$ is now any space like $\mathbb{R}^p$ or $\mathbb{S}^{p-1}$, where the notion of halfspace makes sense. Let Φ_ϑ^{-1} stand for the preimage set-valued mapping corresponding to the function $\Phi_\vartheta(\cdot)$: $\Phi_\vartheta^{-1}(E) = \{z : \Phi_\vartheta(z) \in E\}$. If P is a probability on Z, Φ gives birth to a probability field $P_\vartheta = P \circ \Phi_\vartheta^{-1}$. We define the tangent depth of ϑ to be

$$d_{\mathrm{T}}^\Phi(\vartheta, P) = d(P_\vartheta) = d(P \circ \Phi_\vartheta^{-1}) = \inf_{u \neq 0} P\big(\Phi_\vartheta^{-1}(\mathsf{H}_u)\big).$$

Again, the definition depends on Φ and P. When P is a distribution of a random variable Z, we abuse the notation and write $d_{\mathrm{T}}^\Phi(\vartheta, Z)$ instead of the precise $d_{\mathrm{T}}^\Phi(\vartheta, \mathcal{L}(Z))$, having $P_\vartheta(E) = \mathbb{P}[\Phi(Z, \vartheta) \in E]$ for any E and

$$d_{\mathrm{T}}^\Phi(\vartheta, Z) = \inf_{u \neq 0} \mathbb{P}\big[Z \in \Phi_\vartheta^{-1}(\mathsf{H}_u)\big] = \inf_{u \neq 0} \mathbb{P}[\Phi_\vartheta(Z) \in \mathsf{H}_u].$$

The symbol d_{T} is now used in several formally different ways, which nonetheless all express the same concept: $d_{\mathrm{T}}^\Phi(\vartheta, \mathcal{Z})$ is equal to $d_{\mathrm{T}}^\Phi(\vartheta, Z) = d_{\mathrm{T}}^\Phi(\vartheta, \mathcal{L}(Z))$, whenever $\mathcal{L}(Z)$ is the empirical distribution of $\mathcal{Z}$.

All general definitions of depth that occurred in the literature so far are special instances of our general definition of tangent depth. We do not introduce analogous extensions for global and local depth in this paper. Such a development might be relatively straightforward, but formal subtleties needed for analogs of Theorems 2.1 and 2.3 would divert us from the main theme.

EXAMPLE 1 (General version). Recall (4): in the multivariate location setting, $\Phi_\vartheta(z) = \nabla F_z(\vartheta) = \vartheta - z$. If Z is a Z-valued random variable with distribution P, then

$$\begin{aligned}
d_T^\Phi(\vartheta, Z) &= \inf_{u \neq 0} \mathbb{P}[u^T(\vartheta - Z) \geq 0] = \inf_{u \neq 0} \mathbb{P}[Z \in \mathsf{H}_{\vartheta,u}] \\
&= \inf_{u \neq 0} P_\vartheta(\mathsf{H}_u) \\
&= \inf_{u \neq 0} P(\mathsf{H}_{\vartheta,u}),
\end{aligned}$$

where $P_\vartheta = \mathcal{L}(\vartheta - Z)$. The analogy with the finite-sample case is straightforward.

EXAMPLE 2 (General version). In the linear regression setting, $\Phi_\vartheta(z) = \Phi_\vartheta(x, y) = \nabla F_z(\vartheta) = xx^T\vartheta - xy$. Thus, $P_\vartheta = \mathcal{L}(XX^T\vartheta - XY)$. For a random element $Z = (X, Y)$, we obtain straightforward analogs of (7), (8) and (9):

$$\begin{aligned}
d_T^\Phi(\vartheta, Z) &= \inf_{u \neq 0} P_\vartheta(\mathsf{H}_u) = \inf_{u \neq 0} \mathbb{P}[u^T(XX^T\vartheta - XY) \geq 0] \\
&= \inf_{\|u\|=1} \mathbb{P}[-u^T X \operatorname{sgn}(Y - X^T\vartheta) \geq 0] \\
&= \inf_{u \in \mathbb{S}^{p-1}} \mathbb{P}[\operatorname{sgn}(u^T X) \operatorname{sgn}(Y - X^T\vartheta) \geq 0].
\end{aligned}$$

To establish the equivalence to the geometric definition of Rousseeuw and Hubert (1999a), we have to invoke Proposition 3.1. Consider the set S of directions u such that $\mathbb{P}[X^T u = 0]$. Theorem 3.2 says that such a set S is dense in $\mathbb{S}^{p-1}$, if $\mathbb{P}[X = 0] = 0$. Once this holds, $d_T(\vartheta, Z)$ is equal to the infimum of

$$\mathbb{P}[u^T X > 0 \text{ and } \operatorname{sgn}(Y - \vartheta^T X) \geq 0 \text{ or } u^T X < 0 \text{ and } \operatorname{sgn}(Y - \vartheta^T X) \leq 0],$$

over u satisfying $\mathbb{P}[u^T X] = 0$, as in the finite-sample case. When the regression is with intercept, the first coordinate of X is equal to 1, hence $\mathbb{P}[X \neq 0] = 1$. The regression without intercept can be treated similarly: we just decompose the distribution of Z to the part with $X = 0$ and the rest (see Section 5 for a more formal treatment).

EXAMPLE 4 (General version). Also for this example, the general version is a straightforward extension of the finite-sample case: $\Phi_\vartheta(z) = \nabla F_z(\Theta) = xx^T\Theta - xy^T$, $P_\vartheta = \mathcal{L}(XX^T\Theta - XY^T)$ and

$$\begin{aligned}
d_T^\Phi(\Theta, \mathcal{Z}) &= \inf_{U \neq \mathsf{O}} \mathbb{P}[U \cdot (XX^T\Theta - XY^T) \geq 0] \\
&= \inf_{U \neq \mathsf{O}} \mathbb{P}[\operatorname{tr}((XX^T\Theta - XY^T)^T U) \geq 0] \\
&= \inf_{U \neq \mathsf{O}} \mathbb{P}[-(X^T U)(Y - \Theta^T X) \geq 0],
\end{aligned}$$

a direct analog of (6).

3.3. *Existence of centerpoints.* If the parametric space is a manifold, then we may speak about its dimension (equal to the dimension of its tangent space); from the statistical point of view, it is the number of independent parameters in the model. For instance, when $\Theta = \mathbb{R}^p$, then its dimension is simply p. A centerpoint is a parameter whose (tangent) depth is not less than $1/(1 + \dim \Theta)$.

EXAMPLE 1. In the multivariate location model, the existence of a centerpoint for any probability P on $\mathbb{R}^p$ is a well-known mathematical result, proved by Neumann (1945) for $p = 2$ and by Rado (1946) in general. Birch (1959) gave an alternative proof, similar to that by Donoho and Gasko (1992), who rediscovered the result for statistics. A very short and elementary proof for finite-sample point configurations can be found on page 66 of Edelsbrunner (1987).

EXAMPLE 2. The centerpoint problem in the linear regression setting was raised by Rousseeuw and Hubert (1999a), who also established the important special case: they proved the existence of a centerpoint for $p = 2$, for simple regression. Their ingenious geometric construction based on the ham-and-sandwich theorem seems to be, however, not extendable to a higher-dimensional case, where they conjectured the same general lower bound $1/(p + 1)$ for the maximal depth: separately for all finite-sample data—Conjecture 1(a) in Rousseeuw and Hubert (1999a), Conjecture 1 in Rousseeuw and Hubert (1999b)—and for any absolutely continuous population distribution—Conjecture 1(b) in Rousseeuw and Hubert (1999a). The following theorem establishes the lower bound for maximal depth for *any* random variable $Z = (X, Y)$, settling thus both conjectures of Rousseeuw and Hubert (1999b).

THEOREM 3.3. *Let* $\Phi_\vartheta(z) = \Phi_\vartheta(x, y) = -x(y - x^{\mathrm{T}}\vartheta)$ *for any* $\vartheta \in \Theta = \mathbb{R}^p$ *and any* (x, y) *from* $\mathsf{X} \times \mathbb{R}$, *where* $\mathsf{X} \subseteq \mathbb{R}^p$. *For every probability P on* $\mathsf{X} \times \mathbb{R}$,

$$(12) \qquad \sup_{\vartheta \in \Theta} d_{\mathrm{T}}^{\Phi}(\vartheta, P) \geq \frac{1}{1 + \dim \Theta}.$$

The proof of Theorem 3.3 in its full generality is long and employs technical constructions specific for the univariate regression setting. An easier, restricted, but also more general version of Theorem 3.3 is obtained under an additional assumption involving the identification index

$$\Delta(X) = \sup_{\vartheta \neq 0} \mathbb{P}[X^{\mathrm{T}}\vartheta = 0],$$

the quantity well known from breakdown and consistency considerations in linear regression. Note that the definition of $\Delta(X)$ does not depend on whether ϑ runs over $\mathbb{R}^k$ or $\bar{\mathbb{R}}^k$.

EXAMPLE 4 (Centerpoints).   Amenta, Bern, Eppstein and Teng (2000) gave another proof of Conjecture 1(a), a more geometric one. Later, Bern and Eppstein (2000) introduced (in the finite-sample setting) a multivariate generalization of regression depth, and conjectured a lower bound on the maximal depth for their definition. Motivated by this development, we formulated Theorem 3.4 in the revised version to cover the multivariate case as well. Theorem 3.4 relies on the general Theorem 5.19 and Proposition 5.21.

THEOREM 3.4.   *Let* $\Phi_\vartheta(z) = \Phi_\vartheta(x, y) = -x(y^\mathrm{T} - x^\mathrm{T}\vartheta)$ *for any* $\vartheta \in \Theta = \mathbb{R}^p = \mathbb{R}^{mk}$ *and any* (x, y) *from* $\mathsf{X} \times \mathbb{R}^m$, *where* $\mathsf{X} \subseteq \mathbb{R}^p$. *For every probability* P *on* $\mathsf{X} \times \mathbb{R}$ *the following holds*: *if* $\Delta(X) < 1/(1 + p)$ *for any random variable* $Z = (X, Y)$ *whose distribution is* P, *then there is* $\dot{\Theta} \in \mathbb{R}^p$ *such that*

$$d_\mathrm{T}^\Phi(\dot{\Theta}, P) = \sup_{\Theta \in \Theta} d_\mathrm{T}^\Phi(\Theta, P) \geq \frac{1}{1 + \dim \Theta}.$$

The additional assumption introduced by Theorem 3.4 is probably more restrictive from the mathematical than the statistical aspect: it says that the regressors are in (fairly) general position, a condition satisfied in many real situations. In simple linear regression, for instance, more than one third of the observations should lie in a covariate point to make the assumption invalid. Note also that if the distribution of X is absolutely continuous, then $\Delta(X) = 0$; thus Theorem 3.4 immediately proves the original Conjecture 1(b) of Rousseeuw and Hubert (1999a).

The only missing link between Theorem 3.4 and Theorem 3.3 is our inability to handle the smoothing approximation successfully. For the details, see Section 5 and the Appendix; here we only mention that the whole question can be reduced to a problem whether a regression probability field generated by P can be approximated by smooth enough (absolute continuity is enough but is not necessary) regression probability fields generated by P_ν, in a way that

(13) $$\limsup_{\nu \to \infty} \sup_{\vartheta \in \mathbb{R}^p} d_\mathrm{T}(\vartheta, P_\nu) \leq \sup_{\vartheta \in \mathbb{R}^p} d_\mathrm{T}(\vartheta, P).$$

If yes, then the conclusion of Theorem 3.4 holds for P even if the assumption on $\Delta(X)$ is not satisfied. Thus, if the existence of such an approximation for general P would yield a theorem more general than Theorem 3.3. We do not know an answer to this question, though a positive one seems plausible; we would like to remark only that we are aware of counterexamples showing that mere weak convergence of P_ν to P does not imply (13).

On the other hand, however, it seems that the proper treatment of the smoothing trick would not illuminate any new feature of the problem and is only a technical necessity. To bypass it, we had to undertake a painstaking way expounded in the Appendix, where the proof of Theorem 3.3 follows basically the same scheme

as that of Theorem 3.4, but on a considerably higher technical level. Realizing the importance of linear regression in statistics, we believe that it is essential to have Theorem 3.3 rigorously proved in the maximal possible generality—in fact, Theorem 3.4 does not cover even all possible finite-sample cases. Also, our Theorem 3.3 may be of independent mathematical interest. Unfortunately, the technique does not extend to multivariate regression depth, where the complete solution of the centerpoint problem remains open.

Note that Theorem 3.3 guarantees only a bound on the maximal depth. The existence of a centerpoint then follows from the existence of a point with maximal depth (in $\mathbb{R}^p$). This fact is not obvious, though in many cases comes trivially: for instance, if P is the empirical distribution of a finite sample. It also follows from Theorem 3.4 under the assumption $\Delta(X) < 1/(p+1)$. For (univariate) linear regression, the general existence of the deepest parameter is established in the following theorem.

THEOREM 3.5. *Let* $\Phi_\vartheta(z) = \Phi_\vartheta(x, y) = -x(y - x^{\mathrm{T}}\vartheta)$ *for any* $\vartheta \in \Theta = \mathbb{R}^p$ *and any* (x, y) *from* $\mathsf{X} \times \mathbb{R}$, *where* $\mathsf{X} \subseteq \mathbb{R}^p$. *For any probability* P *on* $\mathbb{R} \times \mathsf{X}$, *there is* $\dot{\vartheta} \in \mathbb{R}^p$ *such that*

$$d_{\mathrm{T}}^\Phi(\dot{\vartheta}, P) = \sup_{\vartheta \in \mathbb{R}^p} d_{\mathrm{T}}^\Phi(\vartheta, P).$$

The proof is a direct consequence of Theorem 5.25, formulated and proved in Section 5. Again, we do not know whether the same general result—although plausible—holds for the multivariate regression depth.

4. Bias, breakdown point and maximum depth estimators.

4.1. *Bias sets of maximum depth estimators and depth contours.* One of the reasons for the study of centerpoints is their implications on the bias and breakdown of maximum depth estimators. For location and regression, these aspects were illustrated by Donoho and Gasko (1992), Rousseeuw and Hubert (1999a) and Van Aelst and Rousseeuw (2000). Here we give a view from a more general perspective.

We formalize *an estimator* as a mapping $\mathcal{T}$ assigning one or more parameters from Θ to any probability P with values in Z (many robust estimators yield non-unique results for certain datasets, thus we have to work on this level of generality). The setting in which data are represented by probabilities encompasses a variety of situations and is also perfectly relevant for finite-sample data (see below). If $\mathcal{E}$ is a set of probabilities, then $\mathcal{T}(\mathcal{E})$ denotes the set of all possible values of $\mathcal{T}$ under $P \in \mathcal{E}$. We are interested in the behavior of $\mathcal{T}(\mathcal{E})$ when $\mathcal{E}$ is a neighborhood of P.

Neighborhoods are constructed with the help of a distance π defined on the space of probabilities on Z. As a rule, this distance depends only on the

laws of random variables under consideration. It may be a metric, but not necessarily; all we need is that for any random variable Z, the balls $\mathsf{B}_\pi(P, \varepsilon) = \{\tilde{P} : \pi(P, \tilde{P}) < \varepsilon\}$ decrease with ε and shrink to $\{P\}$ for $\varepsilon = 0$. A frequent choice for π is the total variation metric: $\pi(P, \tilde{P}) = \inf|P(A) - \tilde{P}(A)|$, inf taken over all (measurable) A. Another popular (and nonmetric) choice is the contamination distance γ: $\gamma(P, \tilde{P}) \leq \varepsilon$ if $\tilde{P} = (1 - \varepsilon)P + \varepsilon Q$ for some probability Q on Z. The inequality $\gamma(P, \tilde{P}) \leq \upsilon(P, \tilde{P})$ implies that $\mathsf{B}_\gamma(P, \varepsilon) \subseteq \mathsf{B}_\upsilon(P, \varepsilon)$ for all P and any $\varepsilon > 0$. There may be good reasons for adopting other distances, as well as to reject certain ones—for a thorough discussion, see Davies (1993).

Maximum depth estimators, defined in particular statistical models, can be viewed as generalizations of the sample median. Given a function Φ from $\Theta \times \mathsf{Z}$ to $\mathbb{R}^p$, we call $\mathcal{T}$ a *maximum depth estimator* if

$$\mathcal{T}(P) = \{\dot{\vartheta} \in \Theta : d_\mathrm{T}^\Phi(\dot{\vartheta}, P) \geq d_\mathrm{T}^\Phi(\vartheta, P) \text{ for all } \vartheta \in \Theta\}.$$

The following theorem shows that bias sets of any maximum depth estimator are closely related to the upper level sets of depth. This gives another interpretation for depth contours in data analysis—in addition to those found in Liu, Parelius and Singh (1999).

THEOREM 4.1.   *Let Φ_ϑ be, for any $\vartheta \in \Theta$, a (measurable) function from Z to $\mathbb{R}^p$ and let $\mathcal{T}$ be a maximum depth estimator. For any $\varepsilon \geq 0$ and any P, the inclusion*

$$(14) \qquad\qquad \mathcal{T}\big(\mathsf{B}_\pi(P, \varepsilon)\big) \subseteq \big\{\vartheta : d_\mathrm{T}^\Phi(\vartheta, P) \geq \delta\big\}$$

holds with:

(i) $\delta = \eta - \varepsilon$, *if $d_\mathrm{T}(\dot{\vartheta}, \tilde{P}) \geq \eta$ for all $\dot{\vartheta} \in \mathcal{T}(\tilde{P})$ and all $\tilde{P}$;*
(ii) $\delta = \eta - 2\varepsilon$, *if $d_\mathrm{T}(\dot{\vartheta}, P) \geq \eta$ for all $\dot{\vartheta} \in \mathcal{T}(P)$ and $\pi = \upsilon$;*
(iii) $\delta = \eta(1 - \varepsilon) - \varepsilon$, *if $d_\mathrm{T}(\dot{\vartheta}, P) \geq \eta$ for all $\dot{\vartheta} \in \mathcal{T}(P)$ and $\pi = \gamma$.*

We believe that in many concrete cases the inclusion (14) is actually equality. The attainable uniform bound required by (i) may be lower than the depth of the maximum depth estimator—it means that (ii) or (iii) often gives sharper bounds.

4.2. *Breakdown points of maximum depth estimators.*   An important indicator of the bias behavior is the breakdown point, which says at which ε the estimator "breaks down": the bias sets start to be unacceptably rich, that is, unbounded (or containing all parameter points or at least some unacceptable ones). In this paper, we limit our analysis to parameter spaces equal to $\mathbb{R}^p$ or similar—in other words, endowed with a structure of "boundedness." Knowing once what "bounded" means (in $\mathbb{R}^p$: a bounded set is contained in a ball with finite perimeter), we define the breakdown point of $\mathcal{T}$ at P to be

$$\varepsilon_\pi^*(\mathcal{T}, P) = \inf\{\varepsilon > 0 : \mathcal{T}\big(\mathsf{B}_\pi(P, \varepsilon)\big) \text{ is not bounded in } \Theta\}.$$

Note that when we restrict our attention to those P whose laws are empirical distributions of the n-tuples—provided that the same restriction is applied to all probabilities appearing in the definition of $\mathsf{B}_v(P, \varepsilon)$—then for $\pi = v$ we obtain nothing but the popular finite-sample Donoho–Huber replacement breakdown point. [Thus, it is not "another" breakdown theory which we study here, but the one perfectly relevant also for the finite-sample case. Rigorously, it is not hard to see that the finite-sample Donoho–Huber replacement breakdown point is always bounded from below by $\varepsilon_v^*(P)$ evaluated at P whose law is the empirical distribution of a dataset Z; equality actually holds in all but some artificial cases.]

Let $\mathcal{T}$ be a maximum depth estimator. A quantity influencing the breakdown point of $\mathcal{T}$ is "the depth of infinity"

$$d_\mathrm{T}(\infty, P) = \inf_{A \text{ bounded}} \sup\{d_\mathrm{T}^\Phi(\vartheta, P) : \vartheta \in A^\mathsf{c}\},$$

where inf is taken over all bounded subsets A of Θ. It is not hard to see that when $\Theta = \mathbb{R}^p$, $d_\mathrm{T}^\Phi(\infty, P)$ is equal to the supremum of $\limsup_{v \to \infty} d_\mathrm{T}^\Phi(\vartheta_v, P)$ taken over all sequences such that $\|\vartheta_v\| \to \infty$.

THEOREM 4.2. *Let Φ_ϑ be, for any $\vartheta \in \Theta$, a (measurable) function from Z to $\mathbb{R}^p$ and let $\mathcal{T}$ be a maximum depth estimator.*

(i) *If $d_\mathrm{T}(\dot{\vartheta}) \geq \eta$ for all $\dot{\vartheta} \in \mathcal{T}(\tilde{P})$ and all $\tilde{P}$, then $\varepsilon_\pi^*(\mathcal{T}, P) \geq \eta - d_\mathrm{T}^\Phi(\infty, P)$ (for π equal either to v or γ).*

(ii) *If $d_\mathrm{T}(\dot{\vartheta}) \geq \eta$ for all $\dot{\vartheta} \in \mathcal{T}(P)$, then*

$$\varepsilon_v^*(\mathcal{T}, P) \geq \tfrac{1}{2}\big(\eta - d_\mathrm{T}^\Phi(\infty, P)\big)$$

and

$$\varepsilon_\gamma^*(\mathcal{T}, P) \geq \frac{\eta - d_\mathrm{T}^\Phi(\infty, P)}{1 + \eta}.$$

EXAMPLE 1 (Breakdown point).   In the location setting, $d_\mathrm{T}(\infty, P) = 0$ for any P. Thus, Theorem 4.2(i) gives that both $\varepsilon_v^*(\mathcal{T}, P)$ and $\varepsilon_\gamma^*(\mathcal{T}, P)$ are bounded by $1/(p + 1)$ from below, $p = \dim \Theta$. If the depth of the Tukey median $\mathcal{T}$ is $1/2$, as often happens for symmetric distributions, then by Theorem 4.2(ii), $\varepsilon_v^*(\mathcal{T}, P) \geq 1/4$ and $\varepsilon_\gamma^*(\mathcal{T}, P) \geq 1/3$, regardless of the dimension.

EXAMPLE 2 (Breakdown point).   Let $Z = (X, Y)$ be a random variable whose distribution is P. The theory developed in Section 5 yields upper bounds for $d_\mathrm{T}(\infty, P)$. The cruder bound $d_\mathrm{T}(\infty, P) \leq \Delta(X)$ follows from Proposition 5.21(ii) and Proposition 5.12(i); Theorem 4.2(i) then yields the general lower bound $1/(1 + p) - \Delta(X)$ for the breakdown point of the maximum depth estimator. For instance, if Z corresponds to the finite-sample data n-tuple in general position,

then $\Delta(X) = (p-1)/n$ and the finite sample breakdown point is not less than $n/(p+1) - p + 1$.

Theorem 4.2(ii) yields the dimension-free bound for the maximum depth estimator: if its depth is $1/2$ then its breakdown point is $(1-2\Delta(X))/4$ for the total variation and $(1-2\Delta(X))/3$ for the contamination breakdown point. Van Aelst and Rousseeuw (2000) and Van Aelst, Rousseeuw, Hubert and Struyf (2001) considered the situation when "the model holds": there is a parameter ϑ_0 such that the conditional distribution of Z given X is symmetric about $\ell(\vartheta_0)$. In such a case, the depth of ϑ_0 is $\delta_0 + (1-\delta_0)/2$ with $\delta_0 = \mathbb{P}[X^T\vartheta_0 = Y]$. Theorem 5.25 then yields the bound

$$d_T(\infty, P) \leq \Delta_0 + \tfrac{1}{2}\big(\Delta(X) - \Delta_0\big),$$

where $\Delta_0 = \sup_{\vartheta \neq 0} \mathbb{P}[X^T\vartheta_0 = Y \text{ and } X^T\vartheta = 0] \leq \delta_0$; by Theorem 4.2(ii), we obtain

$$\varepsilon_\upsilon^*(\mathcal{T}, P) \geq \tfrac{1}{4}\big(1 + \delta_0 - \Delta(X) - \Delta_0\big) \geq \tfrac{1}{4}\big(1 - \Delta(X)\big).$$

For the contamination breakdown, Theorem 4.2(iii) gives that

$$\varepsilon_\gamma^*(\mathcal{T}, P) \geq \frac{1 + \delta_0 - \Delta(X) - \Delta_0}{3 + \delta_0}.$$

We believe that these bounds are in general sharp; note that if Z has an absolutely continuous distribution, they reduce to $1/4$ and $1/3$, respectively.

5. Mathematical details.

5.1. *Weak optimality, convexity.* In this subsection we prove all the results from Section 2. Let F be a real function F defined on a convex domain Θ. Following the terminology of Ponstein (1967), we call F quasiconvex if all upper-level sets are convex [$F(x) \leq \min\{F(a), F(b)\}$ for any x lying on the line connecting a and b]. We call F strictly quasiconvex if the following holds: given any a, b with $F(a) < F(b)$, $F(ta + (1-t)b) < F(b)$ for every $t \in (0, 1)$. We call F pseudoconvex if given any a, b with $F(a) < F(b)$, there exists $c > 0$ and $\tau \in (0, 1]$ such that $F(ta + (1-t)b) \leq F(b) - ct$ for every $t \in [0, \tau]$.

PROPOSITION 5.1.   (i) *If F is convex, then it is pseudoconvex.*

 (ii) *If F is pseudoconvex and continuous, then it is strictly quasiconvex.*

(iii) *If F is pseudoconvex, differentiable and $\nabla F(\vartheta) = 0$, then ϑ is its point of global minimum.*

 (iv) *If F is strictly quasiconvex or pseudoconvex, then it is quasiconvex.*

PROOF.   See Ortega and Rheinboldt [(1970), pages 102–105] (in a slightly different terminology there).   □

PROOF OF THEOREM 2.1.   Suppose that we removed enough observations to destroy local weak optimality. Then ϑ cannot be globally weakly optimal; hence we would need not more observations to destroy the global optimality. This proves the general inequality.

To prove the equality under the additional hypothesis, note that if ϑ is not a weak optimum in Θ with respect to A^{c}, then there is $\tilde{\vartheta} \in \Theta$ such that $F_i(\tilde{\vartheta}) < F_i(\vartheta)$ for all $i \in A^{\mathrm{c}}$. The strict quasiconvexity of all F_i implies then that ϑ cannot be weakly optimal in any open neighborhood of ϑ, with respect to the same A^{c}; this proves the converse inequality and hence equality. $\square$

PROOF OF PROPOSITION 2.2.   The first part of the proposition is just a slightly different statement of the Carathéodory theorem. The equivalent conditions follow from direct combinations of known convexity properties: note that (ii) is the consequence of the Minkowski separation theorem. $\square$

The next theorem gives a sufficient condition for weak optimality when criterial functions are quasiconvex; compare with the necessary and sufficient conditions for Pareto optimality given on page 71 of Smale (1975a).

THEOREM 5.2.   *Suppose that $F_1, F_2, \ldots, F_k$ are differentiable at ϑ. If ϑ is locally weakly optimal with respect to $F_1, F_2, \ldots, F_k$, then 0 is surrounded by $\{\nabla F_1(\vartheta), \nabla F_2(\vartheta), \ldots, \nabla F_k(\vartheta)\}$; that is,*

(S): $\sum \lambda_i \nabla F_i(\vartheta) = 0$ *for some nonnegative and not all equal to zero constants $\lambda_1, \lambda_2, \ldots, \lambda_k$.*

Conversely, ϑ is a weak local optimum, if (S) holds, all F_i are quasiconvex and

(M):   *any F_i with nonzero λ_i in (S) is locally nondecreasing in any direction u such that $u^{\mathrm{T}} \nabla F_i(\vartheta) = 0$.*

PROOF.   If (S) does not hold, then all $\nabla F_i(\vartheta)$ are contained in an open halfspace with 0 on its boundary, by Proposition 2.2. Hence, there is a direction u such that $u^{\mathrm{T}} \nabla F_i(\vartheta) < 0$ for all i; therefore, all F_i are locally decreasing in the direction of u and hence ϑ cannot be weakly optimal.

To prove the converse, we will show that under (S) and (M) weak optimality cannot be violated in any direction—this is sufficient in view of quasiconvexity of criterial functions. Fix a direction u. If there is F_i such that $u^{\mathrm{T}} \nabla F_i(\vartheta) < 0$, then this F_i is locally increasing in the direction of u and the desired conclusion holds. If there is no such F_i, then all $\nabla F_i(\vartheta)$ lie in a subspace $\{x : x^{\mathrm{T}} u = 0\}$. By (S), there is at least one nonzero λ_i; by (M), the corresponding F_i is locally nondecreasing in the direction u and we obtain the same conclusion again. $\square$

PROPOSITION 5.3.   *Suppose that F is differentiable at ϑ. If F is pseudoconvex, then F is nondecreasing in any direction u such that $u^{\mathrm{T}} \nabla F(\vartheta) = 0$.*

PROOF.   If $u^{\mathrm{T}} \nabla F(\vartheta) = 0$, then the directional derivative of F in the direction u is 0. The proposition follows from Proposition 5.1(iii) and the fact that the restriction of a pseudoconvex F on the line in the direction of u passing through ϑ is again pseudoconvex.   □

PROOF OF THEOREM 2.3.   If there is a halfspace H_u containing a proportion $d_{\mathrm{T}}(\vartheta, \mathcal{Z})$ of the gradients, then after removing the observations corresponding to the gradients from H_u the remaining ones do not surround 0, due to Proposition 2.2(ii). Theorem 5.2 then implies that ϑ cannot be locally weakly optimal. This proves the inequality in the general situation.

To prove the converse inequality under the additional hypothesis, suppose that we have removed a proportion $d_{\mathrm{loc}}(\vartheta, \mathcal{Z})$ of the observations so that ϑ is not anymore a local weak optimum with respect to the remaining ones. By Theorem 5.2, this means that then either (S) or (M) must be violated. It must be (S), since (M) is implied for pseudoconvex F_i by Proposition 5.3. The remaining observations therefore do not surround 0; but then they lie in an open halfspace, due to Proposition 2.2(ii). Hence $d_{\mathrm{T}}(\vartheta, \mathcal{Z}) \leq d_{\mathrm{loc}}(\vartheta, \mathcal{Z})$.   □

5.2. *Atomic decomposition of measures.*   For the proof of Theorem 3.2, we need a combinatorial technique generalizing the principle that a bounded measure may possess only a countable number of atoms. The same technique is used for the proof of Lemma A.2.1 in the Appendix; unfortunately, we do not know about any reference in the literature which would allow us to skip the following technical development.

We introduce the following notation: $\mathbb{X}$ is the basic set (a measurable space, e.g., $\mathbb{S}^{p-1}$, $\mathbb{R}^p$ or any other; we do not require any topological assumptions) and $\mathcal{A}$ is an atomic system: a set of (measurable) subsets of $\mathbb{X}$ that can be written as a union

$$(15) \qquad\qquad \mathcal{A} = \mathcal{A}^0 \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \cdots \cup \mathcal{A}^r$$

such that $\mathcal{A}^0$ consists of a single element (often $\varnothing$, but not always) and the following property holds: if $A \in \mathcal{A}^i$, $E \in \mathcal{A}^j$ and $A \subseteq E$, then $i \leq j$; and if $A \subset E$, then $i < j$ (that is, equality of i and j implies the equality of A and E). It is not hard to see that any atomic system has the property of the intersection system introduced by Balek and Mizera (1997): the intersection of any two distinct elements from $\mathcal{A}^i$ belongs to $\mathcal{A}^j$ with $j < i$.

A canonical example is $\mathcal{A} = \mathcal{A}^1$ consisting of all singletons; here $\mathcal{A}^0 = \{\varnothing\}$. For this example, our theory reduces to the well-known considerations concerning atoms. This example could be iterated to higher cardinalities, but we use two other instances instead: the system of linear subspaces and the system of affine subspaces of $\mathbb{R}^p$ (or $\bar{\mathbb{R}}^p$). In both of them, $\mathcal{A}^i$ consist of subspaces with dimension equal to i; for linear subspaces $\mathcal{A}^0 = \{0\}$, while for affine ones $\mathcal{A}^0 = \{\varnothing\}$.

If Q is a measure and E is a (measurable) set, we denote by $Q \llcorner E$ the restriction ("trace") of Q to E: the measure satisfying $(Q \llcorner E)(A) = Q(E \cap A)$. We say that Q is supported by A if $Q \llcorner A = Q$; that is, $Q(E) = Q(E \cap A)$ for every E.

PROPOSITION 5.4. *Suppose that $\mathcal{A}$ is an atomic system on $\mathbb{X}$. Any finite measure Q can be written as the sum*

$$
\text{(16)} \qquad Q = \sum_{i=0}^{r} \sum_{A \in \mathcal{A}_Q^i} Q^A + Q^\omega,
$$

where sets $\mathcal{A}_Q^i$ are at most countable (they may be empty), every measure Q^A is supported by A and the measure Q^j given by

$$
\text{(17)} \qquad Q^j = \sum_{i=0}^{j} \sum_{A \in \mathcal{A}_Q^i} Q^A
$$

satisfies $Q(A) - Q^j(A) = 0$ for any $A \in \mathcal{A}^j$.

PROOF. In the proof, we use the following corollary of the Hahn–Saks–Vitali theorem—see Doob [(1994), Theorems III.10 and IX.10, pages 30 and 155]: if Q_ν is a countable system of measures such that the sum $\sum_\nu Q_\nu(E)$ converges for all (measurable) E, then this sum defines a measure $Q(E)$. We write $Q_1 \le Q_2$, if $Q_1(E) \le Q_2(E)$ for all E.

We define the decomposition (16) inductively. In the initial step, take the only element A of $\mathcal{A}^0$; if $Q(A) > 0$, then set $\mathcal{A}_Q^0 = \{A\}$ and $Q^A = Q \llcorner A$. Otherwise (e.g., if $A = \varnothing$), set $\mathcal{A}_Q^0 = \varnothing$. In each case, we have that $Q^0(E) \le Q(E)$ for any E.

Suppose now that for all $i < j$, we have already defined the sets $\mathcal{A}_Q^i$, as well as measures Q^A for every $A \in \mathcal{A}_Q^i$; the sets $\mathcal{A}_Q^i$ are at most countable, the measure Q^{j-1}, given by (17), satisfies the inequality $Q^{j-1} \le Q$, and finally $Q(E) - Q^{j-1}(E) = 0$ for any $A \in \mathcal{A}^i$ and any $i < j$. Since $Q^{j-1} \le Q$, $Q - Q^{j-1}$ is a measure; denote it by $\tilde{Q}$. Let $\mathcal{A}_Q^j$ be the (possibly empty) collection of all $A \in \mathcal{A}^j$ such that $\tilde{Q}(A) > 0$. For every $A \in \mathcal{A}_Q^j$, let $Q^A = (\tilde{Q} \llcorner A)$. The intersection of any two distinct elements $\mathcal{A}^j$ belongs to $\mathcal{A}^i$ with $i < j$; therefore this intersection has measure $\tilde{Q}$ equal to 0. It follows that $\mathcal{A}_Q^j$ is at most countable and for any E,

$$
\text{(18)} \qquad \sum_{A \in \mathcal{A}_Q^j} Q^A(E) = \sum_{A \in \mathcal{A}_Q^i} \tilde{Q}(E \cap A) = \tilde{Q}\left(E \cap \bigcup_{A \in \mathcal{A}_Q^i} A \right)
$$

$$
\le \tilde{Q}(E) = Q(E) - Q^{j-1}(E);
$$

hence Q^j, defined by (17), satisfies the inequality $Q^j \le Q$. The way we constructed all measures Q^A so far implies that

$$
\tilde{Q} - \sum_{A \in \mathcal{A}_Q^j} Q^A = Q - Q^{j-1} - \sum_{A \in \mathcal{A}_Q^j} Q^A = Q - Q^j.
$$

If $i < j$ and E belongs to $\mathcal{A}^i$, then $\tilde{Q}(E)$ is equal to 0 as well as all $Q^A(E)$, due to the inequality $Q^A \leq Q$ holding for any $A \in \mathcal{A}^j_Q$. The same holds if E is from $\mathcal{A}^j_Q \smallsetminus \mathcal{A}^j$. Finally, if E is from $\mathcal{A}^j_Q$, then

$$(19) \qquad \tilde{Q}(E) = Q^E(E) \leq \sum_{A \in \mathcal{A}^j_Q} Q^A(E).$$

This yields, together with (18), that $Q(E) - Q^j(E) = 0$, concluding the induction step.

After repeating the construction for $j = 1, 2, \ldots, r$, we finally set $Q^\omega = Q - Q^r$. $\square$

We call the system of measures appearing in (16) an atomic decomposition of a (finite) measure Q with respect to an atomic system $\mathcal{A}$. We denote by $\mathcal{A}_Q$ the union of sets $\mathcal{A}^i_Q$ for $i = 0, 1, 2, \ldots, r$—the set of atoms. Finally, the measure Q^ω is a nonatomic part of Q with respect to $\mathcal{A}$, since $Q^\omega(A) = 0$ for any $A \in \mathcal{A}$.

PROPOSITION 5.5. *Suppose that $\mathcal{A}$ is an atomic system on $\mathbb{X}$ and Q is a finite measure. If $E \in \mathcal{A}$ and $Q(E) > 0$, then E contains a member from $\mathcal{A}_Q$.*

PROOF. Since $Q^\omega(E) = 0$, Proposition 5.4 implies that there is $\tilde{A} \in \mathcal{A}_Q$ such that $Q^{\tilde{A}}(E) > 0$; we may pick it from $\mathcal{A}^j_Q$ with smallest possible j, so that for all $A \in \mathcal{A}^i_Q$ with $i < j$, $Q^A(E) = 0$. Now, $0 < Q^{\tilde{A}}(E) = Q^{\tilde{A}}(\tilde{A} \cap E)$; thus $\tilde{A} \cap E$ must belong to $\mathcal{A}^i_Q$ with $i \geq j$. Since $\tilde{A} \cap E \subseteq \tilde{A}$, it follows that $\tilde{A} \cap E = \tilde{A}$; that is, $\tilde{A} \subseteq E$. $\square$

5.3. *Cosmic spaces.* Hereafter, the letter ν is reserved for passages to infinity along positive integers: all limits involving ν are with respect to $\nu \to \infty$.

A cosmic space $\bar{\mathbb{R}}^p$ is a compactification of $\mathbb{R}^p$ (a compact topological space containing $\mathbb{R}^p$ as an open dense subset), which is formed by adding a point with infinite distance from the origin in each direction. The name is due to Rockafellar and Wets (1998); we refer there for more details and background. Rigorously spoken, $\bar{\mathbb{R}}^p$ is the union of $\mathbb{R}^p$ and a homeomorphic copy $\partial\mathbb{R}^p$ of $\mathbb{S}^{p-1}$: it is convenient to write elements of $\partial\mathbb{R}^p$ as $\omega(u)$, where $u \in \mathbb{S}^{p-1}$ and ω is a one-to-one mapping of $\mathbb{S}^{p-1}$ to $\partial\mathbb{S}^{p-1}$.

The space $\bar{\mathbb{R}}^p$ can be conveniently described in polar coordinates. The norm $\|x\|$ extends the Euclidean norm on $\mathbb{R}^p$; it is equal to ∞ for $x \in \partial\mathbb{R}^p$. We define the direction of x to be

$$(\!(x)\!) = \begin{cases} 0, & \text{if } x = 0, \\ x/\|x\|, & \text{if } x \neq 0 \text{ and } x \in \mathbb{R}^p, \\ u, & \text{if } x \in \partial\mathbb{R}^p, x = \omega(u) \text{ for } u \in \mathbb{S}^{p-1}. \end{cases}$$

The direction defined in this way satisfies the following equalities for all $x, y \in \mathbb{R}^p$:

$$(20) \qquad \operatorname{sgn}(x^{\mathsf{T}} y) = \operatorname{sgn}(((x))^{\mathsf{T}} y) = \operatorname{sgn}(x^{\mathsf{T}} ((y))) = \operatorname{sgn}(((x))^{\mathsf{T}} ((y))).$$

In the vein of (20), we define for any $x, y \in \bar{\mathbb{R}}^p$,

$$(21) \qquad \operatorname{sgn}(x \cdot y) = \operatorname{sgn}(((x))^{\mathsf{T}} ((y))).$$

In what follows, we actually never need the full definition of the inner product, just its sign; nevertheless, for better readability we often write $x \cdot y \geq 0$ instead of $\operatorname{sgn}(x \cdot y) \geq 0$.

Norm and direction characterize the convergence in $\bar{\mathbb{R}}^p : x_\nu \to x$ if and only if $\|x_\nu\| \to \|x\|$, and either $x = 0$ or $((x_\nu)) \to ((x))$; note that the latter may not hold if $x = 0$. It is not hard to see that the convergence defined in this way extends the standard one in $\mathbb{R}^p$ and is consistent with the topological properties of $\bar{\mathbb{R}}^p$. Particularly, $\partial \mathbb{R}^p$ is homeomorphic to $\mathbb{S}^{p-1}$; $\mathbb{R}^p$ is open and dense in $\bar{\mathbb{R}}^p$; and $\bar{\mathbb{R}}^p$ is compact and homeomorphic to the ball $\mathbb{D}^p$.

For later convenience, we extend the mapping ω from $\mathbb{S}^{p-1}$ to $\bar{\mathbb{R}}^p$. We set $\omega(0) = 0$; for $x \neq 0$, let $\omega(x) = \omega(((x)))$, where the right-hand side refers to the original definition of ω. Albeit the extended ω is no longer one–one, it satisfies the following identity for all $x \in \bar{\mathbb{R}}^p$:

$$(22) \qquad ((\omega(x))) = ((x)).$$

We write $x = -y$ if and only if $((x)) = -((y))$. For all $x, y \in \bar{\mathbb{R}}^p$,

$$(23) \qquad \operatorname{sgn}((-x) \cdot y) = \operatorname{sgn}(x \cdot (-y)) = -\operatorname{sgn}(x \cdot y).$$

PROPOSITION 5.6. *Let $x_\nu, y_\nu \in \mathbb{R}^p$. If $x_\nu \to x \in \partial \mathbb{R}^p$ and $\|y_\nu\| < \infty$, then $x_\nu + y_\nu \to x$.*

PROOF. A straightforward verification: just note that $\|x_\nu + y_\nu\| \geq \|x_\nu\| - \|y_\nu\| \to \infty$ and then that

$$((x_\nu + y_\nu)) = \frac{x_\nu + y_\nu}{\|x_\nu + y_\nu\|} = \frac{\frac{x_\nu}{\|x_\nu\|} + \frac{y_\nu}{\|x_\nu\|}}{\left\| \frac{x_\nu}{\|x_\nu\|} + \frac{y_\nu}{\|x_\nu\|} \right\|} \to \frac{((x))}{\|((x))\|} = ((x));$$

these two facts together give the desired convergence. $\square$

Obviously, any measure on $\mathbb{R}^p$ may be extended to $\bar{\mathbb{R}}^p$ and, since $\mathbb{R}^p \subset \bar{\mathbb{R}}^p$, any theorem about measures on $\bar{\mathbb{R}}^p$ holds for measures on $\mathbb{R}^p$.

5.4. *General properties of depth.* In this subsection, $\mathbb{X}$ may be $\mathbb{R}^p$, $\mathbb{D}^p$, $\mathbb{S}^{p-1}$, or $\bar{\mathbb{R}}^p$; we do all the proofs for $\mathbb{X} = \bar{\mathbb{R}}^p$, the other cases being entirely similar. If Q is a measure on $\mathbb{X}$, we write $S(Q)$ for the set of all $u \in \mathbb{S}^{p-1}$ such that $Q(\partial \mathsf{H}_u \smallsetminus \{0\}) = 0$, and h_Q for the function from $\mathbb{S}^{p-1}$ to $\mathbb{R}$ such that $h_Q(u) = Q(\mathsf{H}_u)$.

PROOF OF THEOREM 3.2. Let, for $j = 0, 1, 2, \ldots, p-1$, the sets $\mathcal{A}^j$ consist of all subsets of $\mathbb{X}$ of the form

$$\{x \in \bar{\mathbb{R}}^p : x^\mathsf{T} u_1 = 0, x^\mathsf{T} u_2 = 0, \ldots, x^\mathsf{T} u_{p-i} = 0\} \smallsetminus \{0\},$$

where $u_1, u_2, \ldots, u_{p-i}$ are linearly independent vectors from $\mathbb{S}^{p-1}$, for $i \le j$. The union $\mathcal{A}$ of all $\mathcal{A}^j$ is an atomic system consisting of all linear subspaces of $\bar{\mathbb{R}}^p$ with codimension at least 1 with the origin $\{0\}$ removed. We apply Proposition 5.5 to the measure $Q \llcorner (\bar{\mathbb{R}}^p \smallsetminus \{0\})$: any $E \in \mathcal{A}$ contains A from a countable system $\mathcal{A}_Q$.

For any $A \in \mathcal{A}$, let $A^\perp$ denote the set of all $u \in \mathbb{S}^{p-1}$ such that $u \cdot x = 0$ for all x in A. Any such $A^\perp$ has codimension at least 1, hence its dimension is less than the full dimension of $\mathbb{S}^{p-1}$—and therefore its measure μ is 0, as well as the measure μ of any countable union of such sets, in particular of the union $\mathcal{A}_Q^\perp$ of all $A^\perp$ such that $A \in \mathcal{A}_Q$. If $A \subseteq E$, then $A^\perp \supseteq E^\perp$; thus by Proposition 5.5, $\mathcal{A}_Q^\perp$ contains all u such that $Q(\{x \in \bar{\mathbb{R}}^p : x \cdot u = 0\} \smallsetminus \{0\}) > 0$. Since $\mu(\mathcal{A}_Q^\perp) = 0$, we obtain that $\mu(S(Q)) = 1$; therefore $S(Q)$ is dense in $\mathbb{S}^{p-1}$. □

Weak convergence for subprobability measures is understood in the sense of Billingsley (1968, 1971): we write $Q_\nu \rightharpoonup Q$, if the integrals of all bounded continuous functions with respect to Q_ν converge to those with respect to Q. Weak convergence implies that the total mass of Q_ν converges to Q; our sequences are always formed by measures with the same total mass, hence it is not hard to see that all relevant theorems hold (trivially if the total mass is zero, and by division of the total mass otherwise—switching to "conditional probabilities").

PROPOSITION 5.7. *Suppose that Q_ν, Q are measures on $\bar{\mathbb{R}}^p$ and $Q_\nu \rightharpoonup Q$. If $u_\nu, u \in \mathbb{S}^{p-1}$ and $u_\nu \to u$, then*

$$\limsup_{\nu \to \infty} Q_\nu(\mathsf{H}_{u_\nu}) \le Q(\mathsf{H}_u). \tag{24}$$

If $Q(\partial \mathsf{H}_u) = 0$, then $Q_\nu(\mathsf{H}_{u_\nu}) \to Q(\mathsf{H}_u)$.

PROOF. Note first the following. Let $u_\nu, u \in \mathbb{S}^{p-1}$, $u_\nu \to u$ and $x_\nu, x \in \bar{\mathbb{R}}^p$, $x_\nu \to x$. If $u_\nu^\mathsf{T}((x_\nu)) \ge 0$ for infinitely many ν, then $u^\mathsf{T}((x)) \ge 0$; this holds trivially if $x = 0$ and follows from the fact that $((x_\nu)) \to ((x))$ otherwise. Conversely, if $u^\mathsf{T}((x)) > 0$, then $u_\nu^\mathsf{T}((x_\nu)) > 0$ for all but a finite number of ν, since in this case $x \ne 0$ and thus $((x_\nu)) \to ((x))$.

We start the proof by invoking the Skorokhod representation: it yields the existence of random vectors ξ_ν, ξ, defined on the common probability space $(\Omega, \mathscr{S}, \mathbb{P})$, such that $\mathcal{L}(\xi_\nu) = Q_\nu$, $\mathcal{L}(\xi) = Q$ and $\xi_\nu \to \xi$ almost surely. With the help of (21), we obtain

$$
\begin{aligned}
\limsup_{\nu \to \infty} Q_\nu(\mathsf{H}_{u_\nu}) = \limsup_{\nu \to \infty} \mathbb{P}[u_\nu \cdot \xi_\nu \geq 0] &= \limsup_{\nu \to \infty} \mathbb{P}[u_\nu^{\mathrm{T}}(\!(\xi_\nu)\!) \geq 0] \\
&\leq \mathbb{P}[u^{\mathrm{T}}(\!(\xi)\!) \geq 0] = \mathbb{P}[u \cdot \xi \geq 0] = Q(\mathsf{H}_u);
\end{aligned}
$$
(25)

this proves (24). If $Q(\partial\mathsf{H}_u) = 0$, then

$$
\begin{aligned}
Q(\mathsf{H}_u) = \mathbb{P}[u \cdot \xi > 0] = \mathbb{P}[u^{\mathrm{T}}(\!(\xi)\!) > 0] &\leq \liminf_{\nu \to \infty} \mathbb{P}[u_\nu^{\mathrm{T}}(\!(\xi_\nu)\!) > 0] \\
&\leq \liminf_{\nu \to \infty} \mathbb{P}[u_\nu^{\mathrm{T}}(\!(\xi_\nu)\!) \geq 0] = \liminf_{\nu \to \infty} Q_\nu(\mathsf{H}_{u_\nu}).
\end{aligned}
$$
(26)

The convergence follows from the combination of (25) and (26). □

PROPOSITION 5.8. *Let $\Phi_\vartheta(z)$ be a function from $\Theta \times \mathsf{Z}$ to $\mathbb{R}^p$; let Q_ν, Q be measures on Z.*

(i) *If, for $\vartheta_\nu \to \vartheta$,*

$$(\!(\Phi_{\vartheta_\nu}(z_\nu))\!) \to (\!(\Phi_\vartheta(z))\!) \qquad whenever\ z_\nu \to z$$
(27)

for almost all z with respect to Q (in particular, when Φ is jointly continuous in ϑ and z), then $Q_\nu \rightharpoonup Q$ and $\vartheta_\nu \to \vartheta$ imply that

$$\limsup_{\nu \to \infty} Q_\nu \circ \Phi_{\vartheta_\nu}^{-1}(\mathsf{H}_u) \leq Q \circ \Phi_\vartheta^{-1}(\mathsf{H}_u).$$
(28)

If (28) holds only with $z_\nu = z$ for all ν and Q-almost all z (that holds, e.g., if Φ is continuous only in ϑ), then (28) still holds with $Q_\nu = Q$ for all ν.

(ii) *If $\Phi_\vartheta(z)$, as a function of ϑ, is continuous at ϑ for Q-almost all z, then $Q \circ \Phi_{\vartheta_\nu}^{-1} \rightharpoonup Q \circ \Phi_\vartheta^{-1}$.*

PROOF. The proof is completely analogous to that of Proposition 5.7; the second part is an application of the continuous mapping theorem from the theory of weak convergence. □

PROPOSITION 5.9. *For any measure Q on $\bar{\mathbb{R}}^p$, a function $h_Q(u) = Q(\mathsf{H}_u)$ is upper semicontinuous on $\mathbb{S}^{p-1}$, and continuous at every $u \in S(Q)$, the set of all $u \in \mathbb{S}^{p-1}$ such that $Q(\partial\mathsf{H}_u \smallsetminus \{0\}) = 0$.*

PROOF. Upper semicontinuity follows from Proposition 5.7, setting $Q_\nu = Q$ for all ν. To prove continuity, we apply the same proposition to measures defined by $Q_\nu(E) = Q(E) = Q(E \smallsetminus \{0\})$ for all E and ν. □

PROOF OF PROPOSITION 3.1.   The proposition directly follows from Proposition 5.9. In a more detailed way, if $h_Q(u) < d(Q) + \varepsilon$, then there is a sequence $u_\nu \in S$ such that $u_\nu \to u$; upper semicontinuity yields that

$$\inf_{u \in S} h_Q(u) \le \limsup_{\nu \to \infty} h_Q(u_\nu) \le h_Q(u) \le d(Q) + \varepsilon,$$

and since ε was arbitrary, (ii) follows (the converse inequality is obvious). $\square$

PROPOSITION 5.10.   *For any measure Q on $\bar{\mathbb{R}}^p$, $Q(\{0\}) \le d(Q) \le (Q(\{0\}) + Q(\bar{\mathbb{R}}^p))/2$.*

PROOF.   The first inequality is obvious. The second one follows from Proposition 3.1 via Theorem 3.2, which implies that the set of those u for which $Q(\mathsf{H}_u) + Q(\mathsf{H}_{-u}) = 1 + Q(\{0\})$ is dense in $\mathbb{S}^{p-1}$. $\square$

PROPOSITION 5.11.   *Let Q_ν, Q be measures on $\bar{\mathbb{R}}^p$ such that $Q_\nu \rightharpoonup Q$. If $Q(\partial \mathsf{H}_u) = 0$ for all $u \in \mathbb{S}^{p-1}$, then $Q_\nu(\mathsf{H}_u) \to Q(\mathsf{H}_u)$ uniformly in $u \in \mathbb{S}^{p-1}$.*

PROOF.   We topologize the space of all closed proper halfspaces in a natural way: $\mathsf{H}_{u_\nu} \to \mathsf{H}_u$ if and only if $u_\nu \to u$. Clearly, this topology makes the space of all halfspaces homeomorphic to $\mathbb{S}^{p-1}$, hence compact and metrizable. If $\mathsf{H}_{u_\nu} \to \mathsf{H}_u$ and $Q_\nu \rightharpoonup Q$, then $Q_\nu(\mathsf{H}_{u_\nu}) \to Q(\mathsf{H}_u)$, by Proposition 5.7. Finally, we have also that $Q(\mathsf{H}_{u_\nu}) \to Q(\mathsf{H}_u)$, again by Proposition 5.7. We verified all assumptions of Proposition 2.2 from page 6 of Bickel and Millar (1992); the proposition follows. $\square$

PROPOSITION 5.12.   *For any measures Q_ν, Q on $\bar{\mathbb{R}}^p$, the following implications hold*:

(i) *if $\limsup_{\nu \to \infty} Q_\nu(\mathsf{H}_u) \le Q(\mathsf{H}_u)$ for any $u \in \mathbb{S}^{p-1}$, then $\limsup_{\nu \to \infty} d(Q_\nu) \le d(Q)$*;

(ii) *if $Q_\nu \rightharpoonup Q$ and $Q_\nu(\mathsf{H}_u) \to Q(\mathsf{H}_u)$ uniformly in $u \in \mathbb{S}^{p-1}$, then $d(Q_\nu) \to d(Q)$.*

PROOF.   In (i), the assumption asserts that for fixed u, $Q(\mathsf{H}_u)$ as a function of Q is upper semicontinuous (with respect to the weak topology). Depth function is thus the infimum of upper semicontinuous functions and therefore also upper semicontinuous. The assumption in (ii) says that the functions $h_{Q_\nu}(u) = Q_\nu(\mathsf{H}_u)$ converge uniformly to the function $h_Q(u) = Q(\mathsf{H}_u)$; the convergence of infima follows. $\square$

5.5. *Measures concentrated in closed halfspaces.* Hereafter, we denote by G_u the open halfspace that is the interior of H_u. Taking complements with respect to $\mathbb{X}$ (the correct choice of $\mathbb{X}$ being hopefully clear from the context), we have that $\mathsf{G}_u = \mathsf{H}^{\mathrm{c}}_{-u}$.

We say that w lies on the *arc* connecting u and v from $\mathbb{S}^{p-1}$ if one of the following three possibilities holds: either $w = u = v$; or $u = -v$ (and w is arbitrary); or w lies between u and v on the shorter part of the circumference created by the intersection of $\mathbb{S}^{p-1}$ and the linear subspace generated by u and v.

PROPOSITION 5.13. *Suppose that Q is a measure on $\bar{\mathbb{R}}^p$ such that $Q(\mathsf{G}_u) = 0$ for some $u \in \mathbb{S}^{p-1}$. For every $v_1, v_2 \in \mathbb{S}^{p-1}$, if $v_1 \neq u$ lies on the arc connecting u and v_2, then $Q(\mathsf{H}_{v_1}) \leq Q(\mathsf{H}_{v_2})$.*

PROOF. Note first that the assumptions imply that $v_2 \neq u$; if $v_2 = -u$, then $Q(\mathsf{H}_{v_2}) = Q(\bar{\mathbb{R}}^p)$ and the proposition holds trivially—hence we will assume that $v_2 \neq -u$ as well. In such a case, u and v_2 are not collinear and

$$\tilde{u} = \frac{v_2 - u(u^{\mathrm{T}} v_2)}{\|v_2 - u(u^{\mathrm{T}} v_2)\|}$$

is well defined. It is straightforward to verify that (see Figure 3): $u^{\mathrm{T}}\tilde{u} = 0$, hence u is orthogonal to $\tilde{u}$ and lies in $\mathsf{H}_{\tilde{u}}$; $\tilde{u}$ lies in the linear space spanned by u and v_2 and therefore on the same circumference as u, v_2 and thus also v_1; finally, $v_2^{\mathrm{T}}\tilde{u} \geq 0$, hence v_2 lies in the halfspace $\mathsf{H}_{\tilde{u}}$. Since u_1 lies on the arc connecting u and v_2, it also lies in $\mathsf{H}_{\tilde{u}}$. We express v_1 and v_2 in polar coordinates: for $i = 1, 2$,

$$v_i = u(u^{\mathrm{T}} v_i) + \tilde{u}(\tilde{u}^{\mathrm{T}} v_i) = u \cos\varphi_i + \tilde{u} \sin\varphi_i,$$

where the angles φ_i are from $(0, \pi)$. The condition that v_1 lies on the arc connecting u and v_2 means that $\varphi_1 < \varphi_2$ and hence $\cot\varphi_1 > \cot\varphi_2$.
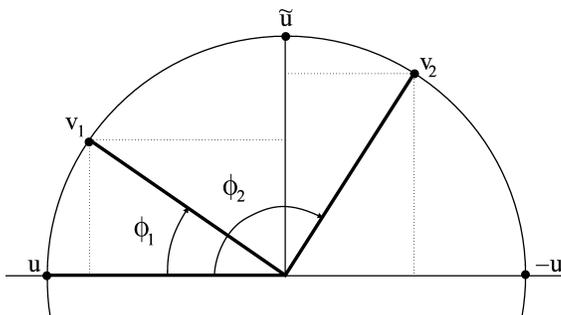

FIG. 3. *An illustration for the proof of Proposition* 5.13.

Now it is sufficient to show that $H_{v_1} \smallsetminus H_{v_2} \subseteq H^c_{-u}$, since then $Q(H_{v_1} \smallsetminus H_{v_2}) \leq Q(H^c_{-u}) = 0$ and by subadditivity,

$$Q(H_{v_1}) \leq Q(H_{v_1} \smallsetminus H_{v_2}) + Q(H_{v_2}).$$

Suppose that $x \in H_{v_1} \smallsetminus H_{v_2}$. Then

$$(29) \qquad ((x))^T v_1 = ((x))^T u \cos\varphi_1 + ((x))^T \tilde{u} \sin\varphi_1 \geq 0,$$

$$(30) \qquad ((x))^T v_2 = ((x))^T u \cos\varphi_2 + ((x))^T \tilde{u} \sin\varphi_2 < 0.$$

Dividing (29) and (30) by $\sin\varphi_1$ and $\sin\varphi_2$, respectively (we excluded the possibility that either one of them is zero at the beginning), and combining both inequalities, we obtain

$$((x))^T u \, (\cot\varphi_1 - \cot\varphi_2) > 0.$$

Hence $((x))^T u > 0$, that is, $x \in H^c_{-u}$. $\square$

PROPOSITION 5.14. *If Q is a measure on $\bar{\mathbb{R}}^p$ such that $Q(G_u) = 0$ for some $u \in \mathbb{S}^{p-1}$, then $d(Q) = d(Q \llcorner \partial H_u)$.*

PROOF.   We have to prove that $d(Q) \leq d(Q \llcorner \partial H_u)$; the reverse inequality is trivial. Fix $\varepsilon > 0$. Consider $v \in \partial H_u \cap \mathbb{S}^{p-1}$. For any $w \in \mathbb{S}^{p-1}$,

$$Q(H_w) = Q(H_w \cap G_u) + Q(H_w \cap \partial H_u) + Q(H_w \cap G_{-u})$$
$$= Q(H_w \cap \partial H_u) + Q(H_w \cap G_{-u}),$$

since $Q(G_u) = 0$. If w approaches u along the arc connecting u and v, then $Q(H_w \cap G_{-u}) \to 0$. Therefore, there is $w \in \mathbb{S}^{p-1}$ such that

$$Q(H_w) \leq Q(H_w \cap \partial H_u) + \varepsilon = Q(H_v \cap \partial H_u) + \varepsilon = (Q \llcorner \partial H_u)(H_v) + \varepsilon.$$

It follows that $d(Q) \leq d(Q \llcorner \partial H_u) + \varepsilon$. Since ε was arbitrary, the desired inequality follows. $\square$

5.6. *Skeletons, centrality, topology.* Throughout this subsection, Q is a measure on $\bar{\mathbb{R}}^p$ or $\mathbb{S}^{p-1}$, $h_Q(u) = Q(H_u)$ is a function on $\mathbb{S}^{p-1}$, λ a nonnegative constant and μ denotes the uniform distribution on $\mathbb{S}^{p-1}$. For any λ, we define a λ-skeleton of Q to be

$$s_\lambda(Q) = \int_{\mathbb{S}^{p-1}} u\bigl(h_Q(u) \wedge \lambda\bigr) \, d\mu(u).$$

Clearly, any skeleton assigns to Q a vector in a unit ball $\mathbb{D}^p$. Skeletons give vector-field approximations to probability fields; the inspiration for the notion came from Birch (1959).

PROPOSITION 5.15.   *If* $\lambda = \lambda(Q) \geq 0$ *depends continuously on* Q, *with respect to the weak topology, and* $Q(\{0\}) = 0$, *then* $s_{\lambda(\cdot)}(\cdot)$ *is continuous at* Q.

PROOF.   Suppose that $Q_\nu \rightharpoonup Q$ and $Q(\{0\}) = 0$. By Proposition 5.7 and Theorem 3.2, $h_{Q_\nu}(u) \to h_Q(u)$ for μ-almost all u. If $\lambda(Q_\nu) \to \lambda(Q)$, then

$$u\big(h_{Q_\nu}(u) \wedge \lambda(Q_\nu)\big) \to u\big(h_Q(u) \wedge \lambda(Q)\big)$$

for μ-almost all u. The proof is concluded by the application of the Lebesgue dominated convergence theorem.   □

We call Q λ-central, if the level set $\{u \in \mathbb{S}^{p-1} : h_Q(u) \leq \lambda\}$ surrounds 0. Note that λ-centrality of Q means also its $\tilde{\lambda}$-centrality whenever $\lambda \leq \tilde{\lambda}$. The following proposition expresses the basic Helly–Carathéodory principle.

PROPOSITION 5.16.   *If* $\lambda > d(Q)$ *and* $s_\lambda(Q) = 0$, *then* Q *is* λ-*central and* $\lambda \geq 1/(p+1)$.

PROOF.   Since $\lambda > d(Q)$, Proposition 3.1 and Theorem 3.2 imply the existence of $u \in S(Q)$ such that $h_Q(u) < \lambda$. By Proposition 5.9, the level set $E_\lambda = \{u \in \mathbb{S}^{p-1} : h_Q(u) \leq \lambda\}$ contains some neighborhood of u; hence $\mu(E_\lambda) > 0$. Since $s_\lambda(Q) = 0$, the set E_λ cannot be contained in any open halfspace H_u^c; hence E_λ surrounds 0, by Proposition 2.2(ii), and Q is λ-central.

Proposition 2.2(iii) then yields the existence of a set $V \subseteq E_\lambda$ surrounding 0 with cardinality at most $m + 1$. If a set V surrounds 0, then the system $\{\mathsf{H}_v\}_{v \in V}$ covers $\bar{\mathbb{R}}^p$ and

$$1 \leq \sum_{v \in V} P_\vartheta(\mathsf{H}_v) \leq (p+1)\lambda. \qquad \qquad □$$

The following proposition specifies how skeletons behave on the compactifying boundary.

PROPOSITION 5.17.   *Let* Q *be a probability measure on* $\bar{\mathbb{R}}^p$ *such that* $Q(\mathsf{H}_{-u}) = 1$ *for some* $u \in \mathbb{S}^{p-1}$. *For any* $\lambda \geq 0$, $s_\lambda(Q) \in \mathsf{H}_{-u}$.

PROOF.   Given $u \in \mathbb{S}^{p-1}$, we write v^* for the reflection of $v \in \mathbb{S}^{p-1}$ about $\partial\mathsf{H}_u$: $v^* = v - 2u(u^\mathsf{T}v)$. This reflection places points from the open hemispheres $\mathsf{G}_u \cap \mathbb{S}^{p-1}$ and $\mathsf{G}_{-u} \cap \mathbb{S}^{p-1}$ into one–one correspondence.

Suppose that $v \in \mathsf{G}_u \cap \mathbb{S}^{p-1}$, $v \neq u$. Let $\tilde{v} = v - u(u^\mathsf{T}v)$; note that $\tilde{v}$ is orthogonal to u and thus lies in $\partial\mathsf{H}_u$. We will decompose both v and v^* to a sum of vectors collinear with u and $\tilde{v}$. First, $u^\mathsf{T}v^* = -u^\mathsf{T}v$, hence the corresponding components in the direction of u are of the same magnitude. Second, since

$$\tilde{v}^\mathsf{T}v^* = \tilde{v}^\mathsf{T}v = 1 - (u^\mathsf{T}v)^2 \geq 0,$$

v and v^* lie in $\mathsf{H}_{\tilde{v}}$, and v lies on the arc connecting u and v^*: all three are collinear and if $v \in \mathsf{G}_u$, that is, $u^\mathsf{T}v > 0$, then $u^\mathsf{T}v^* < 0$. Proposition 5.13 implies that $Q(\mathsf{H}_v) \leq Q(\mathsf{H}_{v^*})$; therefore,

$$(31) \qquad\qquad h_Q(v) \wedge \lambda \leq h_Q(v^*) \wedge \lambda$$

for all $\lambda \geq 0$. Computing the λ-skeleton, we obtain

$$(32) \qquad \begin{aligned} s_\lambda(Q) &= \int_{\mathbb{S}^{p-1}} v\big(h_Q(v) \wedge \lambda\big) \, d\mu(v) \\ &= \int_{\mathsf{G}_u \cup \mathsf{G}_{-u}} v\big(h_Q(v) \wedge \lambda\big) \, d\mu(v) + \int_{\partial \mathsf{H}_u} v\big(h_Q(v) \wedge \lambda\big) \, d\mu(v). \end{aligned}$$

The second integral results in a vector lying in $\partial \mathsf{H}_u$. The first integral in (32) can be further orthogonally decomposed to

$$(33) \qquad \begin{aligned} & u \int_{\mathsf{G}_u} (u^\mathsf{T}v)\big(\big(h_Q(v) \wedge \lambda\big) - \big(h_Q(v^*) \wedge \lambda\big)\big) \, d\mu(v) \\ & \qquad + \int_{\mathsf{G}_u} \tilde{v} \, \frac{\tilde{v}^\mathsf{T}v}{\|\tilde{v}\|^2} \big(\big(h_Q(v) \wedge \lambda\big) + \big(h_Q(v^*) \wedge \lambda\big)\big) \, d\mu(v). \end{aligned}$$

The second integral in (33) results again in $\partial \mathsf{H}_u$. The first one results in H_{-u}, due to (31); since $\partial \mathsf{H}_u \subseteq \mathsf{H}_{-u}$, we obtain by (32) and (33) that $s_\lambda(Q) \in \mathsf{H}_{-u}$. $\quad \square$

A point where the vector field is equal to 0 is its critical point. Critical points of continuous vector fields are effectively hunted by the Kronecker index theory. Let S, T be topological spaces. Continuous mappings f, g from S to T are called homotopic, if one can be continuously transformed to another: there is a continuous mapping π from $S \times [0, 1]$ to T such that $\pi(x, 0) = f(x)$ and $\pi(x, 1) = g(x)$. A continuous mapping from $\mathbb{S}^{p-1}$ to $\mathbb{S}^{p-1}$ is homotopic to a constant mapping if and only if it can be extended to a continuous mapping of $\mathbb{D}^p$ to $\mathbb{S}^{p-1}$. Homotopy is an equivalence relation: if f is homotopic to a constant and f and g are homotopic, then g is homotopic to a constant. A topological space T is called contractible if the identity mapping from T to T is homotopic to a constant. A more thorough picture of homotopic equivalence of continuous mappings of spheres gives the theory of degree: a mapping is homotopic to a constant if and only if its topological degree is 0. On the sphere, neither the identity mapping $f(x) = x$ (degree 1) nor the antipodal mapping $f(x) = -x$ (degree -1) is homotopic to a constant. If a restriction to the boundary ∂U of a region U of a vector field is nonsingular, its degree is called an index of a vector field over U. The following proposition is a reformulation of the well-known principle from the theory of vector fields, saying that a continuous vector field on ∂U with a nonzero index possesses a critical point inside U.

PROPOSITION 5.18. *Let ξ be a continuous vector field on $\mathbb{R}^p$. If $\mathbb{R}^p$ contains a subset S homeomorphic to $\mathbb{S}^{p-1}$ and such that the restriction of ξ to S is not homotopic to a constant, then ξ has a critical point in the region bounded by S (possibly lying on its boundary).*

PROOF [see Dodson and Parker (1997), page 25]. Instead of $\mathbb{R}^p$, we might take any contractible topological space Θ, due to the pathwise connectedness of $\mathbb{S}^{p-1}$; see, for instance, Exercise 19(ii), page 26 of Rotman (1988). $\square$

The following theorem is the core of our general approach to centerpoint hunting, in the vein of Birch (1959) and Donoho and Gasko (1992). The general theorem given below covers the simplest case when the parameter space is $\Theta = \mathbb{R}^p$ (Examples 1, 2, and 4 showed the practical importance of this case). More sophisticated parameter spaces can be analyzed similarly, only the adequate topological engine would be different.

THEOREM 5.19. *If a probability field $\{P_\vartheta : \vartheta \in \bar{\mathbb{R}}^p\}$ on $\bar{\mathbb{R}}^p$ is:*

(i) *continuous in the weak topology at all $\vartheta \in \bar{\mathbb{R}}^p : P_{\vartheta_\nu} \rightharpoonup P_\vartheta$ whenever $\vartheta_\nu \to \vartheta$,*

(ii) *$P_\vartheta(\mathsf{H}_{-\vartheta}) = 0$ for all $\vartheta \in \partial\mathbb{R}^p$,*

(iii) *$P_\vartheta(\{0\}) = 0$ for all $\vartheta \in \mathbb{R}^p$,*

then there exists a point $\dot\vartheta \in \mathbb{R}^p$ such that $d(P_{\dot\vartheta}) \geq 1/(p+1)$.

PROOF. For fixed $\lambda > 0$, the continuity of the probability field and the fact that $P_\vartheta(\{0\}) = 0$ for all ϑ imply, via Proposition 5.15, the continuity of the λ-skeleton $s_\lambda(P_\vartheta)$ as a vector field dependent on ϑ. Assumption (ii) implies, via Proposition 5.17, that $s_\lambda(P_\vartheta) \in \mathsf{H}_\vartheta$ for any $\vartheta \in \partial\mathbb{R}^p$. The Poincaré–Bohl theorem [see Dodson and Parker (1997), page 19] says that if f and g are two continuous functions from $\mathbb{S}^{p-1}$ to $\mathbb{S}^{p-1}$ not pointing to opposite directions at any point, then f and g are homotopic. This holds in our case; the skeleton $s_\lambda(P_\vartheta)$ never points in the direction opposite to ϑ, hence its degree on $\partial\mathbb{R}^p$ is equal to the degree of the identity mapping on $\mathbb{S}^{p-1}$. The latter is equal to 1; therefore the λ-skeleton cannot be homotopic to a constant. By Proposition 5.18, such a λ-skeleton has a critical point in $\mathbb{R}^p$. By Proposition 5.16, if $\lambda > \sup\{d(P_\vartheta) : \vartheta \in \bar{\mathbb{R}}^p\}$, then λ cannot be less than $1/(p+1)$. Thus, $\sup\{d(P_\vartheta) : \vartheta \in \bar{\mathbb{R}}^p\} \geq 1/(p+1)$. Proposition 5.12 and compactness of $\bar{\mathbb{R}}^p$ yield a deepest parameter, a point $\dot\vartheta \in \bar{\mathbb{R}}^p$ such that $d(P_{\dot\vartheta}) = \sup\{d(P_\vartheta) : \vartheta \in \bar{\mathbb{R}}^p\}$. Apparently, $d(P_{\dot\vartheta}) \geq 1/(p+1)$; finally, $\dot\vartheta$ cannot lie in $\partial\mathbb{R}^p$, since from (ii) follows that $d(P_\vartheta) = 0$ for any $\vartheta \in \partial\mathbb{R}^p$. $\square$

5.7. *Applications*: *regression probability fields.* Two principal techniques to overcome technical difficulties in the application of Theorem 5.19 are compactification and smoothing. Recall that $\bar{\mathbb{X}}$ is a compactification of $\mathbb{X}$ if it is compact and $\mathbb{X}$ is dense in $\bar{\mathbb{X}}$; we denote $\bar{\mathbb{X}} \setminus \mathbb{X}$ by $\partial\mathbb{X}$. We say that a probability field $\{\bar{P}_\vartheta : \vartheta \in \bar{\Theta}\}$ on $\bar{\mathbb{X}}$ is a compactification of a probability field $\{P_\vartheta : \vartheta \in \Theta\}$ on $\mathbb{X}$, if $\bar{\mathbb{X}}$ is a compactification of $\mathbb{X}$, the indexing set $\bar{\Theta}$ is a compactification of Θ, and $\bar{P}_\vartheta = P_\vartheta$ for all $\vartheta \in \Theta$, in the sense that $\bar{P}_\vartheta \llcorner \mathbb{X} = P_\vartheta$ and $\bar{P}_\vartheta(\partial\mathbb{X}) = 0$.

For models with $\Theta = \mathbb{R}^p$, the useful compactifications are those with $\bar{\Theta} = \bar{\mathbb{R}}^p$— for these, Theorem 5.19 yields the existence of a centerpoint, in the original $\Theta = \mathbb{R}^p$ (so that we may eventually forget about the compactification trick). The appropriate compactified probability field can be found by a natural process of continuous extension. As far as the assumptions of Theorem 5.19 are concerned, note first that (ii) implies (iii) also for all $x \in \partial\mathbb{R}^p$; for verifying (i), the following scheme may be helpful.

PROPOSITION 5.20. *Suppose that Θ is an open, dense subset in $\bar{\Theta}$. The probability field $\{\bar{P}_\vartheta : \vartheta \in \bar{\Theta}\}$ is continuous with respect to weak topology, if $\vartheta_\nu \to \vartheta$ implies weak convergence of $\bar{P}_{\vartheta_\nu}$ to $\bar{P}_\vartheta$ in the following three cases*:

(a) $\vartheta_\nu \in \Theta$, $\vartheta \in \Theta$;
(b) $\vartheta_\nu \in \Theta$, $\vartheta \in \bar{\Theta} \setminus \Theta$;
(c) $\vartheta_\nu \in \bar{\Theta} \setminus \Theta$, $\vartheta \in \bar{\Theta} \setminus \Theta$.

PROOF. The proof is a straightforward manipulation with subsequences. □

To check (a) in Proposition 5.20, we verify the continuity of the original probability field—setting $\bar{P}_\vartheta = P_\vartheta$ for $\vartheta \in \Theta$ and extending the measures $\bar{P}_\vartheta$, if needed. If we define $\bar{P}_\vartheta$ for $\vartheta \in \partial\mathbb{R}^p$ by continuous extension, then (b) comes automatically. Finally, checking (c) is often easy.

Another problem in Theorem 5.19 is its assumption (iii), usually not holding when the underlying distribution has atoms. A technique to overcome this difficulty is smoothing: we add a small perturbation with absolutely continuous distribution. However, it may be not that easy to backtrack the smoothing approximation successfully; unless we are able to control the depth on the compactifying boundary (for instance).

Let us illustrate all this technology on the simplest example.

EXAMPLE 1 (Centerpoints). Let P be a probability on $\mathbb{R}^p$; let Z be a random variable whose distribution is P. The function $\Phi_\vartheta(z) = \vartheta - z$ is continuous, jointly in ϑ and z, hence the gradient probability field $P_\vartheta = \mathcal{L}(\vartheta - Z)$ is continuous at any $\vartheta \in \mathbb{R}^p$, by Proposition 5.8—we checked assumption (a) of Proposition 5.20. If $\vartheta \in \partial\mathbb{R}^p$, then Proposition 5.6 yields that $\vartheta_\nu - z$ converges to ϑ; we extend the definition of Φ setting $\Phi_\vartheta(z) = \vartheta$ for $\vartheta \in \partial\mathbb{R}^p$. By Proposition 5.6, this extension

preserves the continuity of Φ, jointly in ϑ and z; therefore, assumptions (b) and (c) are checked as well and Proposition 5.20 yields assumption (i) of Theorem 5.19.

Note that for $\vartheta \in \partial \mathbb{R}^p$, the probability P_ϑ is concentrated in ϑ, therefore $P_\vartheta(\mathsf{H}_{-\vartheta}) = 0$; this checks assumption (ii) of Theorem 5.19. At this point, (ii) holds without any continuity assumption on Z; this implies, in particular, that $d(P_\vartheta) = 0$ for any $\vartheta \in \partial \mathbb{R}^p$.

The only remaining assumption of Theorem 5.19 is (iii). If it holds—for instance, when the distribution of Z is absolutely continuous—then we are able to prove the existence of a centerpoint in $\mathbb{R}^p$. However, assumption (iii) does not hold when the distribution of Z has atoms with nonzero probability. To overcome this difficulty, we consider a sequence of compactified probability fields arising from random variables $Z + Z_\nu$, with the following properties: $Z_\nu \to 0$ and hence the distributions P_ν of $Z + Z_\nu$ converge weakly to P; the distributions of all Z_ν, and hence of $Z + Z_\nu$, are absolutely continuous. All the perturbed probability fields satisfy (i), (ii) and (iii); Theorem 5.19 yields a sequence of centerpoints $\dot{\vartheta}_\nu \in \mathbb{R}^p$ such that $d_{\mathrm{T}}(\dot{\vartheta}_\nu, Z + Z_\nu) \geq 1/(p+1)$.

Such a sequence has a subsequence converging to $\dot{\vartheta} \in \bar{\mathbb{R}}^p$. The continuity of Φ implies, via Proposition 5.8, that $\limsup_{\nu \to \infty} P_\nu \circ \Phi_{\dot{\vartheta}_\nu}^{-1}(\mathsf{H}_u) \leq P \circ \Phi_{\dot{\vartheta}}^{-1}(\mathsf{H}_u)$ (we abuse the subsequence notation in the usual way). Proposition 5.12 then yields that $d_{\mathrm{T}}(\dot{\vartheta}, P) \geq 1/(p+1)$. Finally, $\dot{\vartheta}$ cannot lie in $\partial \mathbb{R}^p$ since all $\vartheta \in \partial \mathbb{R}^p$ have $d_{\mathrm{T}}(\vartheta, Z) = 0$; recall that this was shown without the help of the absolute continuity assumption.

In regression models, we follow essentially the same scheme. The required technical facts are summarized in the following proposition.

PROPOSITION 5.21. *Let $Z = (X, Y)$ be a random variable with values in $\mathsf{X} \times \mathbb{R}^m$ where $\mathsf{X} \subseteq \mathbb{R}^k$. If $\{P_\Theta : \Theta \in \mathbb{R}^p\}$ is a probability field such that $P_\Theta = \mathcal{L}(X X^{\mathrm{T}} \Theta - X Y^{\mathrm{T}})$ for any $\Theta \in \mathbb{R}^p = \mathbb{R}^{km}$, then it has a compactification, a probability field $\{P_\Theta : \Theta \in \bar{\mathbb{R}}^p\}$ on $\bar{\mathbb{R}}^p$, with the following properties*:

(i) *the probability field $\{P_\Theta : \Theta \in \bar{\mathbb{R}}^p\}$ is continuous (in weak topology) at every $\Theta \in \mathbb{R}^p$ and at every $\Theta \in \partial \mathbb{R}^p$ such that $\mathbb{P}[X \cdot \Theta = 0]$*;
(ii) *$\limsup_{\nu \to \infty} P_{\Theta_\nu} \leq P_\Theta$, whenever $\Theta_\nu \in \mathbb{R}^p$ and $\Theta_\nu \to \Theta \in \bar{\mathbb{R}}^p$*;
(iii) *for $\Theta \in \partial \mathbb{R}^p$, $P(\mathsf{H}_\Theta) = 1$ and $d(P_\Theta) = P_\Theta(\{0\}) = P_\Theta(\mathsf{H}_{-\Theta}) = \mathbb{P}[X \cdot \Theta = 0]$*.

If, moreover, $\Delta(X) = 0$, then:

(iv) *$\{P_\Theta : \Theta \in \bar{\mathbb{R}}^p\}$ is continuous (in weak topology) at all $\Theta \in \bar{\mathbb{R}}^p$*;
(v) *for every $\Theta \in \partial \mathbb{R}^p$, $d(P_\Theta) = P(\mathsf{H}_{-\Theta}) = 0$*.

The values of the compactification at $\vartheta \in \partial \mathbb{R}^p$ are given by $P_\Theta = \mathcal{L}(\omega(X X^{\mathrm{T}}(\!(\Theta)\!)))$.

Before proving Proposition 5.21, we show how it can be applied in the proof of Theorem 3.4. Note that Proposition 5.21 and Theorem 5.19 imply Conjecture 1(b) of Rousseeuw and Hubert (1999a) immediately. If the distribution of Z is absolutely continuous (with respect to the appropriate Lebesgue measure), then $\Delta(X) = 0$, assumption (iii) of Theorem 5.19 holds, and Proposition 5.21(iv) and (v) implies assumptions (i) and (ii), respectively.

PROOF OF THEOREM 3.4.   If $\Delta(X) > 0$, then the compactified probability field may not be continuous with respect to weak topology at points from $\partial\mathbb{R}^p$. We consider a sequence of compactified probability fields generated by random variables $Z + Z_\nu = (X + X_\nu, Y + Y_\nu)$ where $Z_\nu \to 0$. All Z_ν, and thus all $Z + Z_\nu$, have absolutely continuous distributions. Hence, $\Delta(X + X_\nu) = 0$ and all probability fields generated by perturbed random variables satisfy all assumptions of Theorem 5.19. We obtain a sequence of centerpoints $\dot{\vartheta}_\nu \in \mathbb{R}^p$, $d_{\mathrm{T}}(\dot{\vartheta}_\nu, Z + Z_\nu) \geq 1/(p+1)$; take its subsequence with a limit $\dot{\vartheta} \in \bar{\mathbb{R}}^p$. From Proposition 5.21(ii) follows, via Proposition 5.12(i), that $d_{\mathrm{T}}(\dot{\vartheta}, Z) \geq 1/(p+1)$. The final step is to show that $\dot{\vartheta}$ cannot lie in $\partial\mathbb{R}^p$. By Proposition 5.21(iii), $d_{\mathrm{T}}(\vartheta, Z) = \mathbb{P}[X \cdot \vartheta = 0] \leq \Delta(X)$ for $\vartheta \in \partial\mathbb{R}^p$. Therefore, the desired result follows—if $\Delta(X) < 1/(p+1)$.

By Proposition 5.21(ii), 5.21(i) and 5.12, tangent depth is upper semicontinuous on $\bar{\mathbb{R}}^p$, which is compact; this proves the existence of the deepest parameter $\dot{\vartheta}$ in $\bar{\mathbb{R}}^p$. Proposition 5.21(iii) then implies that $\dot{\vartheta}$ cannot lie in $\partial\mathbb{R}^p$.   □

PROOF OF PROPOSITION 5.21.   The continuity of the probability field at all $\Theta \in \mathbb{R}^p$ follows, via Proposition 5.8(ii), from the joint continuity (in x, y and Θ) of the function $\Phi_\Theta(x, y) = xx^{\mathrm{T}}\Theta - xy^{\mathrm{T}}$. In the same way, (ii) follows from Proposition 5.8(ii) for $\Theta \in \mathbb{R}^p$.

Suppose now that $\Theta_\nu \to \Theta$, $\Theta_\nu \in \mathbb{R}^p$, $\Theta \in \partial\mathbb{R}^p$. Since $\|\Theta_\nu\| \to \infty$, we may assume that $\Theta_\nu \neq 0$; in view of the fact that also $((\Theta_\nu)) \to ((\Theta))$, we obtain that

$$(34) \qquad ((x_\nu x_\nu^{\mathrm{T}}\Theta_\nu - x_\nu y_\nu^{\mathrm{T}})) = ((x_\nu x_\nu^{\mathrm{T}}((\Theta_\nu)) - x_\nu y_\nu^{\mathrm{T}}\|\Theta_\nu\|^{-1})) \to ((xx^{\mathrm{T}}((\Theta)))).$$

This suggests the form of the compactified probability field: $P_\Theta = \mathcal{L}(\omega(XX^{\mathrm{T}}((\Theta))))$. Under this definition, Proposition 5.8(i), in view of (34), completes the proof of (ii). To complete the proof of part (i), we first need to show that

$$(35) \qquad\qquad xx^{\mathrm{T}}\Theta_\nu - xy^{\mathrm{T}} \to \omega(xx^{\mathrm{T}}((\Theta)))$$

for almost all (x, y) with respect to $P = \mathcal{L}(X, Y)$. Suppose that $\mathbb{P}[X^{\mathrm{T}}((\Theta)) = 0] = 0$; then $X \neq 0$ almost surely, and therefore $XX^{\mathrm{T}}((\Theta)) \neq 0$ almost surely. Hence, for P-almost all x (and y), $\|xx^{\mathrm{T}}((\Theta_\nu))\| \to \|xx^{\mathrm{T}}((\Theta))\| > 0$ and therefore

$$(36) \qquad\qquad \|xx^{\mathrm{T}}\Theta_\nu\| = \|\Theta_\nu\|\|xx^{\mathrm{T}}((\Theta_\nu))\| \to \infty$$

and

$$(37) \qquad ((xx^{\mathrm{T}}\Theta_\nu)) = \frac{xx^{\mathrm{T}}((\Theta_\nu))}{\|xx^{\mathrm{T}}((\Theta_\nu))\|}$$

$$\longrightarrow \quad \frac{xx^{\mathrm{T}}((\Theta))}{\|xx^{\mathrm{T}}((\Theta))\|} = ((xx^{\mathrm{T}}((\Theta)))) = ((\omega(xx^{\mathrm{T}}((\Theta))))).$$

By Proposition 5.6, (36) and (37) imply (35). Finally, if $\Theta_\nu, \Theta \in \partial\mathbb{R}^p$ and $\Theta_\nu \to \Theta$, then

$$(38) \quad ((\omega(xx^{\mathrm{T}}((\Theta_\nu))))) = ((xx^{\mathrm{T}}((\Theta_\nu)))) \quad \longrightarrow \quad ((xx^{\mathrm{T}}((\Theta)))) = ((\omega(xx^{\mathrm{T}}((\Theta))))),$$

completing the proof of (iii) completely.

Let $\Theta \in \partial\mathbb{R}^p$. By (22), (21) and (20),

$$P_\Theta(\mathsf{H}_\Theta) = \mathbb{P}[\Theta \cdot \omega(XX^{\mathrm{T}}((\Theta))) \geq 0] = \mathbb{P}[((\Theta)) \cdot ((XX^{\mathrm{T}}((\Theta)))) \geq 0]$$

$$(39) \qquad = \mathbb{P}[((\Theta)) \cdot (XX^{\mathrm{T}}((\Theta))) \geq 0] = \mathbb{P}[\mathrm{tr}(((\Theta))^{\mathrm{T}}XX^{\mathrm{T}}((\Theta))) \geq 0]$$

$$= \mathbb{P}[X^{\mathrm{T}}((\Theta))((\Theta))^{\mathrm{T}}X \geq 0] = \mathbb{P}[\|((\Theta))^{\mathrm{T}}X\|^2 \geq 0] = 1.$$

Similarly, for all $U \in \mathbb{S}^{p-1}$,

$$P_\Theta(\mathsf{H}_U) = \mathbb{P}[U \cdot \omega(XX^{\mathrm{T}}((\Theta))) \geq 0] = \mathbb{P}[\mathrm{tr}(U^{\mathrm{T}}XX^{\mathrm{T}}((\Theta))) \geq 0]$$

$$(40) \qquad = \mathbb{P}[X^{\mathrm{T}}((\Theta))U^{\mathrm{T}}X \geq 0] \geq \mathbb{P}[X^{\mathrm{T}}((\Theta)) = 0] = \mathbb{P}[X \cdot \Theta = 0];$$

$X \cdot \Theta = 0$ means that $X^{\mathrm{T}}\vartheta_j = 0$ for $j = 1, 2, \dots, m$, where ϑ_j is the jth column of Θ. By (23),

$$P_\Theta(\mathsf{H}_{-\Theta}) = \mathbb{P}[-((\Theta)) \cdot (XX^{\mathrm{T}}((\Theta))) \geq 0]$$

$$(41) \qquad = \mathbb{P}[-\mathrm{tr}(((\Theta))^{\mathrm{T}}XX^{\mathrm{T}}((\Theta))) \geq 0] = \mathbb{P}[-X^{\mathrm{T}}((\Theta))((\Theta))^{\mathrm{T}}X \geq 0]$$

$$= \mathbb{P}[\|((\Theta))^{\mathrm{T}}X\|^2 \leq 0] = \mathbb{P}[((\Theta))^{\mathrm{T}}X = 0] = \mathbb{P}[X \cdot \Theta = 0].$$

Since (40) and (41) hold for any $\Theta \in \partial\mathbb{R}^p$, we obtain the proof of (iii):

$$d(P_\Theta) = P_\Theta(\mathsf{H}_{-\Theta}) = \mathbb{P}[((\Theta))^{\mathrm{T}}X = 0] = \mathbb{P}[((\Theta))^{\mathrm{T}}XX^{\mathrm{T}} = 0]$$

$$(42) \qquad = \mathbb{P}[((\Theta))^{\mathrm{T}}XX^{\mathrm{T}} = 0] = \mathbb{P}[XX^{\mathrm{T}}((\Theta)) = 0] = P_\Theta(\{0\}).$$

The rest of the proposition is easy: just note that if $\Delta(X) = 0$, then $\mathbb{P}[X \cdot \Theta = 0] = 0$ for all $\Theta \in \bar{\mathbb{R}}^p$. $\square$

5.8. *Bias and breakdown.* In this subsection, the function Φ involved in the definition of tangent depth is supposed to be an arbitrary function from $\Theta \times \mathsf{Z}$ to $\mathbb{R}^p$, measurable as a function from Z to $\mathbb{R}^p$ for any $\vartheta \in \Theta$, and we suppress the dependence on it in the notation. As in Section 4, υ stands for the variation metric and γ for the contamination distance.

PROPOSITION 5.22.   *Suppose that $\pi(P, \tilde{P}) \le \varepsilon$. Then*:

(i) $|d_{\mathrm{T}}(\vartheta, P) - d_{\mathrm{T}}(\vartheta, \tilde{P})| \le \varepsilon$ *if* $\pi = \upsilon$;

(ii) $(1 - \varepsilon)d_{\mathrm{T}}(\vartheta, P) \le d_{\mathrm{T}}(\vartheta, \tilde{P}) \le d_{\mathrm{T}}(\vartheta, P) + \varepsilon$ *if* $\pi = \gamma$.

PROOF.   The assumption $\upsilon(P, \tilde{P}) \le \varepsilon$ implies that

$$\left| P\big(\Phi_{\vartheta}^{-1}(\mathsf{H}_u)\big) - \tilde{P}\big(\Phi_{\vartheta}^{-1}(\mathsf{H}_u)\big) \right| \le \varepsilon$$

for any $u \ne 0$ and any $\vartheta \in \Theta$. This implies (i). The second inequality in (ii) follows from (i) and the inequality $\upsilon(P, \tilde{P}) \le \gamma(P, \tilde{P}) \le \varepsilon$. To see the first inequality in (ii), just note that

$$(1 - \varepsilon)P\big(\Phi_{\vartheta}^{-1}(\mathsf{H}_u)\big) \le \tilde{P}\big(\Phi_{\vartheta}^{-1}(\mathsf{H}_u)\big)$$

for any any $u \ne 0$ and any $\vartheta \in \Theta$.   □

PROOF OF THEOREM 4.1.   To see (i), note first that $\tilde{\vartheta}$ can be from $\mathcal{T}(\tilde{P})$ only if $d_{\mathrm{T}}(\tilde{\vartheta}, \tilde{P}) \ge \eta$; but then, by Proposition 5.22, $d_{\mathrm{T}}(\tilde{\vartheta}, P) \ge \eta - \varepsilon$, proving (i). Let $\dot{\vartheta}$ be from $\mathcal{T}(P)$. Any $\tilde{\vartheta}$ can be from $\mathcal{T}(\tilde{P})$ only if $d_{\mathrm{T}}(\dot{\vartheta}, \tilde{P}) \le d_{\mathrm{T}}(\tilde{\vartheta}, \tilde{P})$. We obtain that

$$(43) \qquad \eta \le d_{\mathrm{T}}(\dot{\vartheta}, P) \le d_{\mathrm{T}}(\dot{\vartheta}, \tilde{P}) + \varepsilon \le d_{\mathrm{T}}(\tilde{\vartheta}, \tilde{P}) + \varepsilon \le d_{\mathrm{T}}(\tilde{\vartheta}, P) + 2\varepsilon,$$

by the repeated application of Proposition 5.22(i). This proves (ii). Finally, the proof of (iii) is analogous, only now Proposition 5.22(ii) gives

$$(44) \quad \eta(1 - \varepsilon) \le (1 - \varepsilon)d_{\mathrm{T}}(\dot{\vartheta}, P) \le d_{\mathrm{T}}(\dot{\vartheta}, \tilde{P}) \le d_{\mathrm{T}}(\tilde{\vartheta}, \tilde{P}) \le d_{\mathrm{T}}(\tilde{\vartheta}, P) + \varepsilon$$

instead of (43).   □

Suppose now that $\Theta = \mathbb{R}^p$ (all reasoning is valid for any locally compact space).

PROPOSITION 5.23.   *If* $\vartheta_n \in \Theta$, $\vartheta_n \to \infty$, *then* $\limsup_{n \to \infty} d_{\mathrm{T}}(\vartheta_n, P) \le d_{\mathrm{T}}(\infty, P)$.

PROOF.   A straightforward consequence of the definition of $d_{\mathrm{T}}(\infty, P)$.   □

PROOF OF THEOREM 4.2.   Fix $\varepsilon > \varepsilon_{\pi}^*(\mathcal{T}, P)$. Then $\mathcal{T}(\mathsf{B}_\pi(P, \varepsilon))$ is not bounded and hence there is a sequence $\tilde{\vartheta}_n \in \mathcal{T}(\tilde{P}_n)$ such that $d_{\mathrm{T}}(\tilde{\vartheta}_n, \tilde{P}_n) \ge \eta$. By Propositions 5.22 and 5.23, $\varepsilon \ge \eta - d_{\mathrm{T}}(\infty, P)$. Since ε was arbitrary, (i) follows.

The proof of (ii) starts in the same way and we arrive at an observation similar to (43), with $\tilde{\vartheta}$ replaced by $\tilde{\vartheta}_n$. Proposition 5.23 then yields

$$\eta \le \limsup_{n \to \infty} d_{\mathrm{T}}(\tilde{\vartheta}_n, P) + 2\varepsilon \le d_{\mathrm{T}}(\infty, P) + 2\varepsilon$$

which proves (ii). Finally, (iii) comes also in an analogous way, only (44) is used instead of (43), giving

$$(1 - \varepsilon)\eta \leq \limsup_{n \to \infty} d_{\mathrm{T}}(\tilde{\vartheta}_n, P) + \varepsilon \leq d_{\mathrm{T}}(\infty, P) + \varepsilon,$$

implying (iii). $\square$

5.9. *A refined analysis of the regression model.*   In this subsection, $Z = (X, Y)$ is a random variable with values in $\mathsf{X} \times \mathbb{R}$ and $\{P_\vartheta : \vartheta \in \bar{\mathbb{R}}^p\}$ is a (linear, not multivariate) regression probability field: $P_\vartheta = \mathcal{L}(X X^{\mathrm{T}} \vartheta - X Y^{\mathrm{T}})$ for $\vartheta \in \mathbb{R}^p$.

The geometric definition of the regression depth shows that it is independent of the parametrization (this principle holds in greater generality, but we will not develop this theme here). The following notion comes from the theory of nonlinear regression. Given a regression function $f(x, \vartheta)$, the solution locus of $\vartheta \in \Theta$ is defined to be

$$\ell(\vartheta) = \{(y, x) : y = f(x, \vartheta), \ x \in \mathsf{X}\}.$$

In linear regression, $\ell(\vartheta) = \{(y, x) : y = x^{\mathrm{T}} \vartheta, x \in \mathsf{X}\}$. A closer investigation of graphs of fitted surfaces reveals how this definition can be extended to ϑ from $\partial \mathbb{R}^p$: exploring the limits of sets $\ell(\vartheta)$ in terms of set convergence, we arrive at the definition

$$\ell(\vartheta) = \{(y, x) : x \cdot \vartheta = 0, \ x \in \mathsf{X}\}.$$

This indicates a statistical interpretation for the parameter points from the cosmic extension: they correspond to "vertical" regression fits. Since $x \cdot \vartheta = 0$ if and only if $x \cdot (-\vartheta) = 0$, we have that $\ell(\vartheta) = \ell(-\vartheta)$; that is, each vertical fit may be interpreted as occurring once in upwards and once in downwards orientation. Therefore, a projective plane might be more appropriate compactification than the cosmic one from the statistical point of view; however, the latter has more convenient topological properties. Note that by Proposition 5.21(iii), $d(P_{-\vartheta}) = \mathbb{P}[X \cdot (-\vartheta) = 0] = \mathbb{P}[X \cdot \vartheta = 0] = d(P_{-\vartheta})$ for any $\vartheta \in \partial \mathbb{R}^p$, hence the cosmic compactification is consistent from the depth point of view.

For the proof of Theorem 5.25, the following technical result is required.

PROPOSITION 5.24.   *If* $\vartheta_\nu \to \vartheta \in \partial \mathbb{R}^p$, *then* $P_{\vartheta_\nu}(\mathsf{G}_{-\vartheta}) \to 0$.

PROOF.   Just note that, for any $\vartheta \in \mathbb{R}^p$ and any $u \in \bar{\mathbb{R}}^p$,

$$\begin{aligned}
(45) \qquad \mathbb{P}[u \cdot (X X^{\mathrm{T}} \vartheta - X Y) = 0] &= \mathbb{P}[-((u))^{\mathrm{T}} (X X^{\mathrm{T}} \vartheta - X Y) = 0] \\
&\geq \mathbb{P}[((u))^{\mathrm{T}} X = 0] = \mathbb{P}[X \cdot u = 0]
\end{aligned}$$

and therefore

$$\limsup_{\nu \to \infty} \mathbb{P}[(-\vartheta) \cdot (XX^{\mathrm{T}}\vartheta_\nu - XY) > 0]$$

$$= \limsup_{\nu \to \infty}\big(\mathbb{P}[(-\vartheta) \cdot (XX^{\mathrm{T}}\vartheta_\nu - XY) \geq 0] - \mathbb{P}[(-\vartheta) \cdot (XX^{\mathrm{T}}\vartheta_\nu - XY) = 0]\big)$$

$$\leq \limsup_{\nu \to \infty} \mathbb{P}[(-\vartheta) \cdot (XX^{\mathrm{T}}\vartheta_\nu - XY) \geq 0] - \mathbb{P}[X \cdot (-\vartheta) = 0]$$

$$\leq \mathbb{P}[X \cdot \vartheta = 0] - \mathbb{P}[X \cdot \vartheta = 0] = 0. \qquad \square$$

Assume the same setting as in the definition of tangent depth. For any $E \subseteq \mathsf{Z}$, we define

$$d_{\mathrm{T}}^{\Phi}(\vartheta, P \llcorner E) = d\big((P \llcorner E) \circ \Phi_{\vartheta}^{-1}\big) = \inf_{u \neq 0} P\big(\Phi_{\vartheta}^{-1}(\mathsf{H}_u) \cap E\big) = d\big(P_{\vartheta} \llcorner \Phi(E, \vartheta)\big),$$

where $\Phi(E, \vartheta) = \{\Phi(z, \vartheta) : z \in E\}$ and $P_{\vartheta} = P \circ \Phi_{\vartheta}^{-1}$. If P is the distribution of a random variable Z, then we write $d_{\mathrm{T}}^{\Phi}(\vartheta, Z \llcorner E)$ for $d_{\mathrm{T}}^{\Phi}(\vartheta, \mathcal{L}(Z) \llcorner E)$; in such a case,

$$d_{\mathrm{T}}^{\Phi}(\vartheta, Z \llcorner E) = \inf_{u \neq 0} \mathbb{P}\big[Z \in \Phi_{\vartheta}^{-1}(\mathsf{H}_u) \cap E\big] = \inf_{u \neq 0} \mathbb{P}[\Phi_{\vartheta}(Z) \in \mathsf{H}_u \text{ and } Z \in E].$$

THEOREM 5.25. *Suppose that $Z = (X, Y)$ is a random variable with values in $\mathsf{X} \times \mathbb{R} \subseteq \mathbb{R}^{p+1}$, and $\Phi_{\vartheta}(z) = \Phi_{\vartheta}(y, x) = xx^{\mathrm{T}}\vartheta - xy$ for any $\vartheta \in \Theta = \mathbb{R}^p$. If ϑ_ν is a sequence from $\mathbb{R}^p$ such that $\vartheta_\nu \to \vartheta \in \partial\mathbb{R}^p$ for $\nu \to \infty$, then*

$$(46) \qquad \limsup_{\nu \to \infty} d_{\mathrm{T}}^{\Phi}(\vartheta_\nu, Z) \leq \sup_{\eta \in \mathbb{R}^p} d_{\mathrm{T}}^{\Phi}\big(\eta, Z \llcorner \ell(\vartheta)\big).$$

PROOF.   We prove first that

$$(47) \qquad \limsup_{\nu \to \infty} d(P_{\vartheta_\nu}) \leq \limsup_{\nu \to \infty} d(P_{\vartheta_\nu} \llcorner \partial\mathsf{H}_{\vartheta})$$

(which actually means equality since the reverse is obvious) and then the equality

$$(48) \qquad \limsup_{\nu \to \infty} d(P_{\vartheta_\nu} \llcorner \partial\mathsf{H}_{\vartheta}) = \limsup_{\nu \to \infty} d_{\mathrm{T}}\big(\vartheta_\nu, Z \llcorner \ell(\vartheta)\big).$$

Fix $\varepsilon > 0$. By Proposition 5.24, there is ν_0 such that $P_{\vartheta_\nu}(\mathsf{G}_{-\vartheta}) < \varepsilon$ for $\nu > \nu_0$; thus by Proposition 5.14,

$$d(P_{\vartheta_\nu}) \leq d(P_{\vartheta_\nu} \llcorner \mathsf{H}_{\vartheta}) + \varepsilon = d(P_{\vartheta_\nu} \llcorner \partial\mathsf{H}_{\vartheta}) + \varepsilon.$$

Since ε is arbitrary, (47) follows. The verification of (48) is rather straightforward. Note first that

$$d(P_{\vartheta_\nu} \llcorner \partial\mathsf{H}_{\vartheta}) = \inf_{u \neq 0} \mathbb{P}[u^{\mathrm{T}}(XX^{\mathrm{T}}\vartheta_\nu - XY) \geq 0 \text{ and } ((\vartheta))^{\mathrm{T}}(XX^{\mathrm{T}}\vartheta_\nu - XY) = 0]$$

and

$$d_{\mathrm{T}}(\vartheta_\nu, Z\llcorner\ell(\vartheta)) = \inf_{u\neq 0} \mathbb{P}[u^{\mathrm{T}}(XX^{\mathrm{T}}\vartheta_\nu - XY) \geq 0 \text{ and } ((\vartheta))^{\mathrm{T}}X = 0].$$

While in the general case only the obvious inequality given by (45) holds, the equality (48) holds when $\vartheta_\nu \to \vartheta \in \partial\mathbb{R}^p$. Let $\zeta_\nu = XX^{\mathrm{T}}\vartheta_\nu - XY$. We obtain that

$$\begin{aligned}
\limsup_{\nu\to\infty} &\, d(P_{\vartheta_\nu}\llcorner\partial\mathsf{H}_\vartheta)\\
&= \limsup_{\nu\to\infty} \inf_{u\neq 0} \mathbb{P}[u^{\mathrm{T}}\zeta_\nu \geq 0 \text{ and } ((\vartheta))^{\mathrm{T}}(XX^{\mathrm{T}}\vartheta_\nu - XY) = 0]\\
&= \limsup_{\nu\to\infty} \inf_{u\neq 0} \big(\mathbb{P}[u^{\mathrm{T}}\zeta_\nu \geq 0 \text{ and } ((\vartheta))^{\mathrm{T}}X = 0]\\
&\qquad\qquad\qquad + \mathbb{P}[u^{\mathrm{T}}\zeta_\nu \geq 0 \text{ and } ((\vartheta))^{\mathrm{T}}X \neq 0 \text{ and } X^{\mathrm{T}}\vartheta_\nu - Y = 0]\big)\\
&\leq \limsup_{\nu\to\infty} \Big(\inf_{u\neq 0} \mathbb{P}[u^{\mathrm{T}}\zeta_\nu \geq 0 \text{ and } ((\vartheta))^{\mathrm{T}}X = 0]\\
&\qquad\qquad\qquad + \mathbb{P}[((\vartheta))^{\mathrm{T}}X \neq 0 \text{ and } X^{\mathrm{T}}\vartheta_\nu = Y]\Big)\\
&\leq \limsup_{\nu\to\infty} d_{\mathrm{T}}(\vartheta_\nu, Z\llcorner\ell(\vartheta))\\
&\qquad + \limsup_{\nu\to\infty} \mathbb{P}\big[((\vartheta))^{\mathrm{T}}X \neq 0 \text{ and } X^{\mathrm{T}}((\vartheta_\nu)) = Y\|\vartheta_\nu\|^{-1}\big]\\
&\leq \limsup_{\nu\to\infty} d_{\mathrm{T}}(\vartheta_\nu, Z\llcorner\ell(\vartheta)) + \mathbb{P}[((\vartheta))^{\mathrm{T}}X \neq 0 \text{ and } X^{\mathrm{T}}((\vartheta)) = 0],
\end{aligned}$$

yielding the desired result, since the last probability is equal to zero. □

In the linear regression with $\Theta = \mathbb{R}^p$, Theorem 5.25 gives that

$$(49) \qquad\qquad d_{\mathrm{T}}(\infty, Z) \leq \sup_{\vartheta\in\mathbb{R}^p} \sup_{\eta\in\mathbb{R}^p} d_{\mathrm{T}}^\Phi(\eta, Z\llcorner\ell(\vartheta)).$$

As already mentioned, the bound $d_{\mathrm{T}}(\infty, Z) \leq \Delta(X)$ follows from Propositions 5.21(ii) and 5.12(i). If "the model holds," that is, there is a parameter ϑ_0 such that the conditional distribution of Z given X is symmetric about $\ell(\vartheta_0)$, then the depth of ϑ_0 is $\delta_0 + (1 - \delta_0)/2$ with $\delta_0 = \mathbb{P}[X^{\mathrm{T}}\vartheta_0 = Y]$; by Theorem 5.25,

$$d_{\mathrm{T}}(\infty, Z) \leq \Delta_0 + \tfrac{1}{2}\big(\Delta(X) - \Delta_0\big),$$

where $\Delta_0 = \sup_{\vartheta\neq 0} \mathbb{P}[X^{\mathrm{T}}\vartheta_0 = Y \text{ and } X^{\mathrm{T}}\vartheta = 0] \leq \delta_0$.

APPENDIX: PROOF OF THEOREM 3.3 IN FULL GENERALITY

A.1. Additional facts from measure theory. We write cl E for the topological closure of a set E; all other notation introduced throughout the paper is kept. In what follows, Q_ν, Q_ν^i, Q^i and Q are all (bounded) measures on a separable metric space $\mathbb{X}$.

LEMMA A.1.1.   *Suppose that $Q_\nu \rightharpoonup Q$. If E_1, E_2 are closed sets such that $Q_\nu(E_1 \cup E_2) = Q(E_1 \cup E_2)$ and $Q_\nu(E_1 \cap E_2) = Q(E_1 \cap E_2) = 0$ for all ν, then $Q_\nu(E_i) \to Q(E_i)$ for $i = 1, 2$.*

PROOF.   Since E_i are closed, we have for $i = 1, 2$,

(50) $$\limsup_{\nu \to \infty} Q_\nu(E_i) \le Q(E_i).$$

Using the fact that $Q_\nu(E_2) = Q_\nu(E_1 \cup E_2) - Q_\nu(E_1)$ for all ν and also for Q in place of Q_ν, we obtain, in view of (50) for E_2, that

$$Q(E_1) \le \liminf_{\nu \to \infty} Q_\nu(E_1);$$

together with (50) for E_1 this yields the convergence for E_1. The proof for E_2 is symmetric.   □

LEMMA A.1.2.   *Suppose that $Q_\nu = \sum_{i=1}^{\infty} Q_\nu^i$ and $Q = \sum_{i=1}^{\infty} Q^i$.*

 (i) *If $Q_\nu^i \rightharpoonup Q^i$ for all i, then $Q_\nu \rightharpoonup Q$.*
 (ii) *If $\mathcal{E}$ is a system of sets such that $Q_\nu^i(E) \to Q^i(E)$ uniformly for all $E \in \mathcal{E}$, then $Q_\nu(E) \to Q(E)$ uniformly for all E.*

PROOF.   We prove (ii) first. Fix $\varepsilon > 0$. Choose k such that

(51) $$\sum_{i=k+1}^{\infty} Q^i(\mathbb{X}) \le \frac{\varepsilon}{4}.$$

Choose $\nu_0 = \nu(\varepsilon)$ such that for $\nu > \nu_0$,

(52) $$|Q_\nu^i(\mathbb{E}) - Q^i(\mathbb{E})| \le \frac{\varepsilon}{4k}$$

for all $i = 1, 2, \ldots, k$, all $E \in \mathcal{E}$ and also $E = \mathbb{X}$. Note that

(53)
$$\sum_{i=1}^{k} Q_\nu^i(\mathbb{X}) \ge \sum_{i=1}^{k} Q^i(\mathbb{X}) - \sum_{i=1}^{k} |Q_\nu^i(\mathbb{X}) - Q^i(\mathbb{X})|$$
$$\ge 1 - \frac{\varepsilon}{4} - \frac{k\varepsilon}{4k} = 1 - \frac{\varepsilon}{2}.$$

Combining (51), (52) and (53) yields for the given $\varepsilon > 0$: there is ν_0 such that for all $\nu > \nu_0$ and all $E \in \mathcal{E}$,

$$|Q_\nu(E) - Q(E)| \le \left| \sum_{i=k+1}^{\infty} Q_\nu^i(E) \right| + \left| \sum_{i=1}^{k} Q_\nu^i(E) - Q^i(E) \right| + \left| \sum_{i=k+1}^{\infty} Q^i(E) \right|$$

$$\le \sum_{i=k+1}^{\infty} Q_\nu^i(\mathbb{X}) + \sum_{i=1}^{k} |Q_\nu^i(E) - Q^i(E)| + \sum_{i=k+1}^{\infty} Q^i(\mathbb{X})$$

$$\le \frac{\varepsilon}{4} + \frac{k\varepsilon}{4k} + \frac{\varepsilon}{2} = \varepsilon.$$

Since ε was arbitrary, the uniform continuity is proved.

The proof of (i) goes along the same lines, only in (53) we take E to be just a single Q-continuity set; it is clear that then it is also a Q^i-continuity set for all i. $\square$

A.2. Additional geometric facts. We need a lemma similar to Theorem 3.2, but specific for the regression setting.

LEMMA A.2.1. *For any probability Q on $\mathsf{X} \times \mathbb{R}$, the set $\{\vartheta : Q(\{(x, y) : y = \vartheta^{\mathsf{T}} x\}) > 0\}$ has zero Lebesgue measure in $\Theta = \mathbb{R}^p$.*

PROOF. Consider an atomic system $\mathcal{A}$ in $\mathsf{Z} = \mathbb{R} \times \mathsf{X}$, the union of $\mathcal{A}^j$ for $j = 0, 1, 2, \ldots, k$, defined in the following way. The single element of $\mathcal{A}^0$ is $\varnothing$; $\mathcal{A}^1$ contains all points $(x, y) \in \mathsf{Z}$, that is, all sets consisting of all $(x, x^{\mathsf{T}} \vartheta) \in \mathsf{Z}$ such that ϑ belongs to a codimension 1 affine subspace of $\mathbb{R}^p$; $\mathcal{A}^2$ contains all sets consisting of all $(x, x^{\mathsf{T}} \vartheta) \in \mathsf{Z}$ such that ϑ belongs to a codimension 2 affine subspace of $\mathbb{R}^p$ (lines in the simple linear regression); generally, $\mathcal{A}^j$ contains all sets composed from all $(x, x^{\mathsf{T}} \vartheta) \in \mathsf{Z}$ such that ϑ belongs to a codimension j affine subspace of $\mathbb{R}^p$. It is straightforward to verify that $\mathcal{A}$ is an atomic system. By Propositions 5.4 and 5.5, any $E \in \mathcal{A}$ such that $Q(A) > 0$ contains an A from a countable system $\mathcal{A}_Q$; there is a similar dual structure, the set of all ϑ such that $(x, x^{\mathsf{T}} \vartheta) \in E$ is contained in the set of all ϑ with $(x, x^{\mathsf{T}} \vartheta) \in A$; and any such set has codimension at least 1, thus has zero Lebesgue measure in $\mathbb{R}^p$, as does any countable union of them. $\square$

The rest of this subsection is oriented toward the proof of Lemma A.2.4, which gives a special condition for the uniform convergence of Q_ν to Q on halfspaces, applicable to regression probability fields. We keep the same halfspace notation H_u and G_u for the corresponding hemispheres $\mathsf{H}_u \cap \mathbb{S}^{p-1}$ and $\mathsf{G}_u \cap \mathbb{S}^{p-1}$. The collection of all finite intersections of closed hemispheres in $\mathbb{S}^{p-1}$ is denoted by $\mathcal{H}^{p-1}$. Recall that a system of sets is called a convergence-determining class if $Q_\nu(E) \to Q(E)$ for all its elements E implies that $Q_\nu \rightharpoonup Q$.

LEMMA A.2.2. *For any p, $\mathcal{H}^{p-1}$ is a convergence-determining class on $\mathbb{S}^{p-1}$.*

PROOF. The lemma follows from Billingsley [(1968), Corollary 1, page 14]: $\mathcal{H}^{p-1}$ is closed under the formation of finite intersections and given any $x \in \mathbb{S}^{p-1}$ and any $\varepsilon > 0$, there is a finite intersection E of hemispheres such that x lies in the interior of E and E is contained in the ball with center x and radius ε. $\square$

Suppose that $\mathcal{A}$ is an atomic system composed from all linear (proper) subspaces of $\mathbb{R}^p$ intersected with $\mathbb{S}^{p-1}$. That is, $\mathcal{A}^0$ contains only $\varnothing$, $\mathcal{A}^1$ only antipodal pairs of points, $\mathcal{A}^2$ circles, and so forth.

LEMMA A.2.3. *Let H be from $\mathcal{H}^{p-1}$. If $A \in \mathcal{A}^j$, then there is a closed set $A_H \subseteq A$ such that $(A \cap H) \cup A_H = A$, and $(A \cap H) \cap A_H = H \cap A_H$ is contained in a (possibly empty) union of a finite number of elements from $\mathcal{A}^i$, $i < j$.*

PROOF. We define A_H to be $\mathrm{cl}(A \cap H^c)$. Since A is closed, $A_H \subseteq A_H$; thus $(A \cap H) \cup A_H \subseteq A$. Conversely, the properties of closure give that

$$(A \cap H) \cup A_H \supseteq (A \cap H) \cup (A \cap H^c) = A \cap (H \cup H^c) = A;$$

hence $(A \cap H) \cup A_H = A$.

The rest of the proposition is proved by induction with respect to the number of hemispheres whose intersection gives H. Suppose first that H is itself a closed hemisphere. If $A \subseteq H$, then $A \cap H^c = \varnothing$ and thus also $A_H = H \cap A_H = \varnothing$. Otherwise, again by the properties of closure,

$$H \cap A_H \subseteq H \cap A \cap \mathrm{cl}(H^c) = A \cap H \cap H^c = A \cap \partial H;$$

the properties of the atomic system imply that the last intersection belongs to the $\mathcal{A}^i$ with $i < j$ ($A \cap \partial H$ is a subset of H, but it is not equal to H, since A is not a subset of H).

Suppose now that the proposition holds for H_1 and H_2; let $H = H_1 \cap H_2$. By the properties of closure,

$$H \cap A_H = H \cap \mathrm{cl}(A \cap (H_1 \cap H_2)^c) = H \cap \left(\mathrm{cl}(A \cap H_1^c) \cup \mathrm{cl}(A \cap H_2^c)\right)$$

$$(54) \qquad = \left(H \cap \mathrm{cl}(A \cap H_1^c)\right) \cup \left(H \cap \mathrm{cl}(A \cap H_2^c)\right)$$

$$= \left(H_1 \cap \mathrm{cl}(A \cap H_1^c)\right) \cup \left(H_2 \cap \mathrm{cl}(A \cap H_2^c)\right).$$

The induction assumption implies that any of the two last terms in (54) are contained in the union of a finite number of elements of $\mathcal{A}^i$ with $i < j$. Therefore the same is true for $A \cap A_H$. $\square$

We call a set $E \subseteq \mathbb{S}^{p-1}$ symmetric if $E = -E = \{-x : x \in E\}$. Let π be the factorization mapping from $\mathbb{S}^{p-1}$ to the projective plane $\mathbb{RP}^{p-1}$, the mapping identifying antipodal points: $\pi(-x) = \pi(x)$ for all $x \in \mathbb{S}^{p-1}$. Any symmetric set is a preimage of a subset of $\mathbb{RP}^{p-1}$ under π.

LEMMA A.2.4. *Let Q_ν and Q be probability measures on $\mathbb{S}^{p-1}$ such that $Q_\nu(E) = Q(E)$ for all symmetric sets E and all ν. If $Q_\nu \rightharpoonup Q$, then $Q_\nu(\mathsf{H}_u) \to Q(\mathsf{H}_u)$ uniformly in $u \in \mathbb{S}^{p-1}$.*

PROOF. Consider the atomic decomposition of Q_ν and Q, with the atomic system $\mathcal{A}$ used in Lemma A.2.3 above. Since all sets in $\mathcal{A}$ are symmetric, it follows that $\mathcal{A}^j_{Q_\nu} = \mathcal{A}^j_Q$ for all ν and all j; only measures Q_ν^A may differ.

We show first that $Q_\nu^A \rightharpoonup Q^A$ for all A. In view of Lemma A.2.3, it is sufficient to prove that $Q_\nu^A(E) \to Q^A(E)$ for all E which are intersections of finitely many

closed hemispheres. Suppose that $A \in \mathcal{A}_Q^j$ and let $E \in \mathcal{H}^{p-1}$. The set $A \cap E$ is closed, and by Lemma A.2.3 there is a closed set $A_E \subseteq A$ such that $(A \cap E) \cup A_E = A$. The intersection $(A \cap E) \cap A_E = E \cap A_E$ lies in the union of finitely many sets from $\mathcal{A}^i$, $i < j$; hence $Q^A((A \cap E) \cap A_E) = 0$, as well as $Q_\nu^A((A \cap E) \cap A_E) = 0$ for all ν. The application of Lemma A.1.1 then gives the desired convergence. The weak convergence of Q_ν^ω to Q^ω follows from Lemma A.1.2(i).

Fix $A \in \mathcal{A}^j$. Let H_u be a closed hemisphere in $\mathbb{S}^{p-1}$. If $A \subseteq \mathsf{H}_u$, then

(55) $$Q_\nu^A(\mathsf{H}_u) = Q_\nu^A(A) = Q^A(A) = Q^A(\mathsf{H}_u),$$

thus the convergence is clearly uniform in this case. If H_u does not contain A, then $A \cap \partial \mathsf{H}_u$ lies in $\mathcal{A}^i$, $i < j$, and therefore $Q^A(\partial \mathsf{H}_u) = 0$. The intersections with A of all hemispheres not containing A form a set of hemispheres which is isomorphic to the set of all hemispheres in $\mathbb{S}^{j-1}$. By the properties of the atomic decomposition given by Proposition 5.4, all these hemispheres are Q^A-continuity sets. Thus, we may apply Proposition 5.11 and obtain the uniform convergence for these hemispheres. Combining this convergence with that implied by (55), we have that $Q_\nu^A(\mathsf{H}_u) \to Q^A(\mathsf{H}_u)$ uniformly for all $u \in \mathbb{S}^{p-1}$. Since $Q^\omega(\partial \mathsf{H}_u) = 0$ for all u, Proposition 5.11 gives the same conclusion for the sequence Q_ν^ω. The proof of the lemma is then finished by the application of Lemma A.1.2(ii). $\square$

Lemma A.2.4 helps us to prove convergence of depth for sequences of measures on $\mathbb{S}^{p-1}$ with a constant marginal on $\mathbb{RP}^{p-1}$. Given a probability Q on $\mathbb{X}$ and a mapping f whose range is a subset $\mathbb{X}$, let $f^{-1}(Q)$ denote the set of all probabilities of the form $Q \circ f^{-1}$. For any probability Q on $\mathbb{S}^{p-1}$, the definitions h_Q, $S(Q)$, $d(Q)$ and $s_\lambda(Q)$ are extended in the natural way: hemispheres stand for halfspaces.

LEMMA A.2.5. *For any probability $\bar{Q}$ on $\mathbb{RP}^{p-1}$, $\pi^{-1}(\bar{Q})$ is a compact convex subset of the space of all probabilities on $\mathbb{S}^{p-1}$* (*with respect to the weak topology*). *The function $d(\cdot)$, as well as $s_{d(\cdot)+\varepsilon}(\cdot)$ for any $\varepsilon > 0$, is continuous on $\pi^{-1}(\bar{Q})$.*

PROOF. From the continuity of π it follows that $\pi^{-1}(\bar{Q})$ is closed; since it is a subset of the compact space of the probabilities on $\mathbb{S}^{p-1}$, it is itself compact. Convexity follows by straightforward verification involving symmetric sets.

Suppose that $Q_\nu \rightharpoonup Q$ and all Q_ν and Q belong to $\pi^{-1}(\bar{Q})$. Lemma A.2.4 gives that $h_{Q_\nu}(u) \to h_Q(u)$ uniformly for all $u \in \mathbb{S}^{p-1}$; the convergence of $d(Q_\nu)$ to $d(Q)$ follows. The continuity of the skeleton function follows then from Proposition 5.15. $\square$

LEMMA A.2.6. *Let Q be a probability on $\mathbb{R}^p$ such that $Q(\mathsf{H}_{-u}) = 1$ for some $u \in \mathbb{S}^{p-1}$. If $\lambda > d(Q)$ and $Q(\mathsf{G}_{-u}) > 0$, then $s_\lambda(Q) \in \mathsf{G}_{-u}$; in particular, $s_\lambda(Q) \neq 0$.*

PROOF.   The proof is a follow-up of that of Proposition 5.17 and uses the same notation. To prove that $s_\lambda(Q) \in \mathsf{G}_{-u}$, we have to show only that the first integral in (33) is nonzero. This is true if sharp inequality in (31) holds for some set of v's with positive measure μ.

Let u satisfy the assumptions of the lemma. We will prove that there is a neighborhood $\mathcal{U}(u, \varepsilon)$ of this u such that

(56) $h_Q(v) < h_Q(v^*)$ for every $v \in \mathcal{U}(u, \varepsilon)$

[recall that $v^* = v - 2u(u^{\mathrm{T}}v)$ is the reflection of v about $\partial \mathsf{H}_u$]. Given $\delta > 0$, we write H_u^δ for the δ-fattening of H_u: $\mathsf{H}_u^\delta = \{x \in \mathbb{R}^p : (\!(x)\!)^{\mathrm{T}}u \geq -\delta\}$. The proof of Proposition 5.17 showed, via Proposition 5.13, that

$$Q(\mathsf{H}_v) = Q(\mathsf{H}_v \smallsetminus \mathsf{H}_{v^*}) + Q(\mathsf{H}_v \cap \mathsf{H}_{v^*}) = Q(\mathsf{H}_v \cap \mathsf{H}_{v^*}).$$

Since, on the other hand,

$$Q(\mathsf{H}_{v^*}) = Q(\mathsf{H}_{v^*} \smallsetminus \mathsf{H}_v) + Q(\mathsf{H}_v \cap \mathsf{H}_{v^*}),$$

we obtain that

(57) $Q(\mathsf{H}_{v^*}) - Q(\mathsf{H}_v) = Q(\mathsf{H}_{v^*} \smallsetminus \mathsf{H}_v).$

For arbitrary $v \in \mathbb{S}^{p-1}$ the following holds: if $x \in \mathsf{H}_v$, then

(58) $(\!(x)\!)^{\mathrm{T}}u = (\!(x)\!)^{\mathrm{T}}(u - v) + (\!(x)\!)^{\mathrm{T}}u \geq (\!(x)\!)^{\mathrm{T}}(u - v) \geq -\|u - v\|$

and therefore $x \in \mathsf{H}_u^{\|u-v\|}$. Hence, $\mathsf{H}_v \subseteq \mathsf{H}_u^{\|u-v\|}$ and thus

(59) $\mathsf{H}_v^{\mathsf{c}} \supseteq (\mathsf{H}_u^{\|u-v\|})^{\mathsf{c}}.$

The reflection about $\partial \mathsf{H}_u$ carries v to v^* and u to $-u$; it is a Euclidean transformation, hence $\|u - v\| = \|(-u) - v^*\|$. Suppose that $x \in \mathsf{H}_{v^*}^{\mathsf{c}}$. Then $x^{\mathrm{T}}v^* < 0$; similarly as in (58), we obtain that

$$(\!(x)\!)^{\mathrm{T}}(-u) = (\!(x)\!)^{\mathrm{T}}\big((-u) - v^*\big) + (\!(x)\!)^{\mathrm{T}}v^* \leq (\!(x)\!)^{\mathrm{T}}\big((-u) - v^*\big)$$

$$\leq \|(-u) - v^*\| = \|u - v\|;$$

hence $x \in \mathsf{H}_u^{\|u-v\|}$. Thus,

(60) $\mathsf{H}_{v^*}^{\mathsf{c}} \supseteq (\mathsf{H}_u^{\|u-v\|})^{\mathsf{c}}.$

Combining (59) and (60), we obtain from (57) that

$$Q(\mathsf{H}_{v^*}) - Q(\mathsf{H}_v) \geq Q\big((\mathsf{H}_u^{\|u-v\|})^{\mathsf{c}}\big).$$

For $\|u - v\| \to 0$, the set $(\mathsf{H}_u^{\|u-v\|})^{\mathsf{c}}$ monotonically increases to $\mathsf{H}_u^{\mathsf{c}} = \mathsf{G}_{-u}$; since $Q(\mathsf{G}_u) > 0$, we may conclude that there is $\varepsilon > 0$ such that $Q(\mathsf{H}_{v^*}) - Q(\mathsf{H}_v) > 0$ whenever $\|u - v\| < \varepsilon$. This proves (56).

Suppose now that $\lambda > d(Q)$. There is v_2 such that $h_Q(v_2) < \lambda$. Take the arc connecting u and v_2; according to Proposition 5.13, h_Q is nonincreasing along

this arc when v approaches u. Therefore, we may choose $v_1 \neq u$ in $\mathcal{U}(u, \varepsilon)$ such that $h_Q(v_1) < \lambda$. By Proposition 5.9, h_Q is upper semicontinuous, hence there is a neighborhood $\mathcal{U}(v_1, \eta)$ such that $h_Q(v) < \lambda$ for all $v \in \mathcal{U}(v_1, \eta)$; we may choose this neighborhood to be contained in $\mathcal{U}(u, \varepsilon)$. In view of (56), we have proved that $h_Q(v) \wedge \lambda < h_Q(v^*) \wedge \lambda$ for all $v \in \mathcal{U}(v_1, \eta)$. This concludes the proof, since $\mu(\mathcal{U}(v_1, \eta)) > 0$. $\square$

Note that Proposition 3.1 implies that the set $\{v \in \mathbb{S}^{p-1} : h_Q(v) \leq \lambda\} \cap S$ is nonempty for any S dense in $\mathbb{S}^{p-1}$ and any $\lambda > d(Q)$.

A.3. Set-valued topology and spheric closures. The proof of Theorem 3.3 uses the theory of set-valued vector fields; for more background on this mathematical apparatus, see Rockafellar and Wets (1998) or Klein and Thompson (1984). A set-valued mapping $\mathcal{F}$ from a metric space S to a metric space T is any mapping from S to the set 2^T of all subsets of T. Any set-valued mapping has the graph $\mathrm{Gr}(\mathcal{F}) = \{(s, t) \in S \times T : t \in F(s)\}$. Conversely, any subset of $S \times T$ defines a set-valued mapping from S to T. The domain of a set-valued mapping $\mathcal{F}$ is the set $\mathrm{Dom}\,\mathcal{F} = \{s : \mathcal{F}(s) \neq \varnothing\}$, and the range is the set $\mathrm{Rng}\,\mathcal{F} = \{t : t \in \mathcal{F}(s)$ for some $s \in S\}$. The composition of set-valued mappings $\mathcal{F}_1$ from S_1 to S_2 and $\mathcal{F}_2$ from S_2 to S_3 is the set-valued mapping $\mathcal{F}$ from S_1 to S_3 such that

$$\mathcal{F}(s_3) = \{s_3 : s_3 \in \mathcal{F}_2(s_2) \text{ and } s_2 \in \mathcal{F}_1(s_1)\}.$$

A set-valued mapping $\mathcal{F}$ is called strongly outer semicontinuous at s, if for any open set $E \supseteq \mathcal{F}(s)$, any sequence $t_n \in \mathcal{F}(s_n)$ with $s_n \to s$ is eventually in E. The more widespread terminology is "upper semicontinuous," but we would like to avoid confusion with the upper semicontinuity used in the previous sections. A set-valued mapping $\mathcal{F}$ is single-valued at s, if $\mathcal{F}(s)$ is a singleton; in such a case, it is strongly outer semicontinuous if and only if it is continuous as the ordinary mapping in the usual sense. A closure of $\mathcal{F}$ is the set-valued function $\tilde{\mathcal{F}}$ such that $\mathrm{Gr}(\tilde{\mathcal{F}}) = \mathrm{cl}(\mathrm{Gr}(\mathcal{F}))$. A function with closed graph ($\tilde{\mathcal{F}} = \mathcal{F}$) is called closed—"outer semicontinuous" by Rockafellar and Wets (1998). A set-valued function $\mathcal{F}$ is closed-valued (compact-valued, convex-valued) if $\mathcal{F}(s)$ is closed (compact, convex) for all s, respectively. Note that each closed function is closed-valued, but the converse is not necessarily true.

LEMMA A.3.1. *Let S, T be metrizable spaces. If T is compact, then a set-valued mapping $\mathcal{F}$ from S to T is closed if and only if it is strongly outer semicontinuous and closed-valued.*

PROOF. See Klein and Thompson [(1984), Theorems 7.1.15 and 7.1.16, page 78], or Nikaido [(1968), Lemma 4.4, page 66]. $\square$

Let $S \subseteq \bar{S}$ and $\mathcal{F}$, $\bar{\mathcal{F}}$ be set-valued functions from S, $\bar{S}$, respectively, to T; $\bar{\mathcal{F}}$ is called an extension of $\mathcal{F}$, if $\bar{\mathcal{F}}(s) = \mathcal{F}(s)$ for all $s \in S$.

LEMMA A.3.2. *Let S, $\bar{S}$, T be metrizable spaces. Suppose that T is compact and S is a dense subset of $\bar{S}$.*

(i) If $\mathcal{F}$ is closed, with respect to the relative topology on $S \times T$ (particularly, if $\mathcal{F}$ is a single-valued mapping continuous on S), then its closure $\bar{\tilde{\mathcal{F}}}$ in $S \times T$ is a closed extension of $\mathcal{F}$ to $\bar{S}$.

(ii) If $\mathcal{F}$ is single-valued on S, with values given by an ordinary (single-valued) mapping f continuous on S, then the values of the closed extension $\bar{\tilde{\mathcal{F}}}$ at $s \in \bar{S} \smallsetminus S$ are exactly all limit values of all possible sequences $f(s_\nu)$ with $s_\nu \to s$.

PROOF. For (i), see Nikaido (1968), Theorem 4.7, page 72. To see (ii), note first that all limit points of sequences $f(s_\nu)$ with $s_\nu \to s$ are in $\mathrm{Gr}(\bar{\tilde{\mathcal{F}}})$, since the latter is closed. Suppose that $s_0 \in \bar{S} \smallsetminus S$ and $t_0 \in \bar{\tilde{\mathcal{F}}}(s_0)$. If there is an open neighborhood $\mathcal{U}$ of (s_0, t_0) in $\bar{S} \times T$ containing no point of $\mathrm{Gr}(\bar{\tilde{\mathcal{F}}})$ with $s \in S$ then $\mathrm{Gr}(\bar{\tilde{\mathcal{F}}}) \smallsetminus \mathcal{U}$ is a closed set containing $\mathrm{Gr}(\mathcal{F})$; this is a contradiction with the minimality of the closure. Hence there is a sequence $s_\nu \to s_0$ such that $f(s_\nu) \to t_0$. $\square$

Unlike the closure operation, we define the closed convex hull of the set-valued function $\mathcal{F}$ pointwise: $(\mathrm{cc}\,\mathcal{F})(s) = \mathrm{cl}(\mathrm{conv}(\mathcal{F}(s)))$ for all s.

LEMMA A.3.3. *Let $\mathcal{F}$ be a compact-valued set-valued mapping from S to T, a complete metrizable convex subspace of a linear topological space. The strong outer semicontinuity of $\mathcal{F}$ implies that of $\mathrm{cc}\,\mathcal{F}$.*

PROOF. See Nikaido [(1968), Theorem 4.8, page 72], or Borisovich, Gel'man, Myshkis and Obukhovskii [(1982), Theorem 1.3.21, page 138]. $\square$

Recall that the set of all probability measures on $\mathbb{S}^{p-1}$ with the weak topology is a compact, convex subset of the complete metrizable space of all positive continuous linear functionals on the space $\mathcal{C}(\mathbb{S}^{p-1})$ of continuous functions from $\mathbb{S}^{p-1}$ to $\mathbb{R}$. The latter is the convex cone in the locally convex topological linear space of all continuous linear functionals on $\mathcal{C}(\mathbb{S}^{p-1})$ with the weak*-topology; see [Tjur (1980), page 19]. Recall also the definition of $P \circ g^{-1}$ for a measure P and a function g on $\mathbb{X}$. This definition works well even when g is defined on a subset $\tilde{\mathbb{X}}$, resulting in $(P \llcorner \tilde{\mathbb{X}}) \circ g^{-1}$. We will use the notation $P \circ g^{-1}$ in this extended meaning; of course, if P is supported by $\tilde{\mathbb{X}}$, the result is the same.

Let $\{P_\vartheta : \vartheta \in \Theta\}$ be a probability field on $\mathbb{R}^p$. We define a new, set-valued mapping $\mathcal{P}(\vartheta)$ from $\mathrm{cl}\,\dot\Theta$ to the set of probability measures on $\mathbb{S}^{p-1}$—a set-valued probability field on $\mathbb{S}^{p-1}$, indexed by Θ. We start with the set $\dot\Theta$ of all ϑ such that $P_\vartheta(\{0\}) = 0$; we require that the original probability field $\{P_\vartheta : \vartheta \in \Theta\}$ be continuous (in the weak topology) on $\dot\Theta$, and also that $\dot\Theta$ be dense in Θ; these conditions give the construction its sense. The closure of $\dot\Theta$ could be understood

in Θ; however, it is more convenient to consider it in a compactification $\bar{\Theta}$ instead—if $\dot{\Theta}$ is dense in Θ, so is it in $\bar{\Theta}$.

We define $\mathcal{P}$ in three steps. Let κ be the mapping assigning to x from $\mathbb{R}^p$ its direction $((x))$; note that the mapping is continuous on $\mathbb{R}^p \smallsetminus \{0\}$. In the first step, we set $\mathcal{P}(\vartheta) = \{P_\vartheta \circ \kappa^{-1}\}$ for any $\vartheta \in \dot{\Theta}$; any such P_ϑ is supported by the complement of $\{0\}$, hence $\mathcal{P}$ is single-valued on $\dot{\Theta}$.

In the second step, we construct a strongly outer semicontinuous set-valued extension $\dot{\mathcal{P}}$ of $\mathcal{P}$ to cl Θ. By the continuous mapping theorem, $P_{\vartheta_\nu} \circ \kappa^{-1}$ converges weakly to $P_\vartheta \circ \kappa^{-1}$, whenever $\vartheta \in \dot{\Theta}$ and $P_{\vartheta_\nu} \rightharpoonup P_\vartheta$. Therefore, $\mathcal{P}$ is continuous on $\dot{\Theta}$. This is all we need to apply Lemma A.3.2(i): the closure $\dot{\mathcal{P}}$ of $\mathcal{P}$ is closed on cl Θ; then Lemma A.3.1 yields its strong outer semicontinuity. The closure is also an extension of $\mathcal{P}$, that is, $\dot{\mathcal{P}}(\vartheta)$ is single-valued and equal to $\dot{\mathcal{P}}(\vartheta)$ for all $\vartheta \in \dot{\Theta}$.

In the third step, we define $\mathcal{P}(\vartheta)$ to be the convex closure of $\dot{\mathcal{P}}$: for all $\vartheta \in \Theta$, $\mathcal{P}(\vartheta) = \mathrm{cc}\, \dot{\mathcal{P}}(\vartheta)$. The convexification leaves the singleton values for $\vartheta \in \dot{\Theta}$ unchanged, hence our notation remains consistent. Lemma A.3.3 implies that $\mathcal{P}(\vartheta)$ is strongly outer semicontinuous (on the whole Θ); obviously, it is also convex-valued.

We call the resulting set-valued probability field on $\mathbb{S}^{p-1}$ indexed by $\bar{\Theta}$ a spheric closure of the original field $\{P_\vartheta : \vartheta \in \mathbb{R}^p\}$. Spheric closures replace probability fields on $\mathbb{R}^p$ by those on $\mathbb{S}^{p-1}$, bypassing thus assumption (iii) of Theorem 5.19. For ϑ outside $\dot{\Theta}$, the spheric closure may be thought of as replacing the original probability by the set of probabilities representing all ways how the mass from the origin can be spread along $\mathbb{S}^{p-1}$. In the tangent depth setting, such a replacement can never raise depth. Recall that $h_Q(u) = Q(\mathsf{H}_u)$, for any measure Q on $\mathbb{S}^{p-1}$ (where H_u denotes hemispheres) or on $\mathbb{R}^p$ or $\bar{\mathbb{R}}^p$ (where H_u denotes halfspaces).

LEMMA A.3.4. *Let Φ be a function from $\Theta \times \mathsf{Z}$ to $\mathbb{R}^p$ continuous in the variable ϑ for Q-almost all z. Let P be a probability on Z and let $\mathcal{P}$ be a spheric closure of the probability field $\{P_\vartheta : \vartheta \in \Theta\}$ such that $P_\vartheta = P \circ \Phi_\vartheta^{-1}$. For any $\vartheta \in \Theta$ and any $Q \in \mathcal{P}(\vartheta)$, $h_Q(u) \leq h_{P_\vartheta}(u)$ for all $u \in \mathbb{S}^{p-1}$; in particular, $d(Q) \leq d(\vartheta, P)$.*

PROOF.   If Q is the single element of $\mathcal{P}(\vartheta)$ for $\vartheta \in \dot{\Theta}$, then the lemma trivially holds. Fix $\vartheta \in \Theta \smallsetminus \dot{\Theta}$. We will show that any $Q \in \mathcal{P}(\vartheta)$ is of the form

$$(61) \qquad\qquad Q = \widehat{P} \circ \Phi_\vartheta^{-1} \circ \kappa^{-1} + \tilde{Q}$$

where $\widehat{P} = P \llcorner (\mathsf{Z} \smallsetminus \Phi_\vartheta^{-1}(\{0\}))$ and

$$(62) \qquad \tilde{Q}(\mathbb{S}^{p-1}) = 1 - \widehat{P}(\mathbb{S}^{p-1}) = P\big(\Phi_\vartheta^{-1}(\{0\})\big) = P_\vartheta(\{0\}).$$

Observe that the set of Q satisfying (61) is convex and closed under weak convergence; hence it is sufficient to prove (61) for all $Q \in \dot{\mathcal{P}}(\vartheta)$.

Fix $Q \in \dot{\mathcal{P}}(\vartheta)$. According to Lemma A.3.2(ii), there is a sequence $\vartheta_\nu \to \vartheta$ such that

(63) $$P_{\vartheta_\nu} \circ \kappa^{-1} = P \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1} \rightharpoonup Q.$$

The decomposition

$$P = \widehat{P} + \dot{P} = P \llcorner \left(\mathsf{Z} \smallsetminus \Phi_\vartheta^{-1}(\{0\}) \right) + P \llcorner \Phi_\vartheta^{-1}(\{0\})$$

(the dependence of ϑ is suppressed in the notation) leads to the decomposition for P_ϑ

(64) $$P \circ \Phi_\vartheta^{-1} = \widehat{P} \circ \Phi_\vartheta^{-1} + \dot{P} \circ \Phi_\vartheta^{-1}$$

and also for $P_{\vartheta_\nu} \circ \kappa^{-1}$:

(65) $$P \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1} = \widehat{P} \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1} + \dot{P} \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1}.$$

By the continuous mapping theorem,

(66) $$\widehat{P} \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1} \rightharpoonup \widehat{P} \circ \Phi_\vartheta^{-1} \circ \kappa^{-1}.$$

From (63) and (66) follows that there is $\tilde{Q}$ such that

(67) $$\dot{P} \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1} \rightharpoonup \tilde{Q}.$$

The decomposition (61) follows from (63), (65), (66) and (67); the identity (62) from (67) and the fact that

$$\dot{P} \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1}(\mathbb{S}^{p-1}) = 1 - \widehat{P} \circ \Phi_{\vartheta_\nu}^{-1} \circ \kappa^{-1}(\mathbb{S}^{p-1})$$

$$= 1 - \left(1 - P\left(\Phi_\vartheta^{-1}(\{0\})\right) \right).$$

The decomposition (61) yields that

$$Q(\mathsf{H}_u) = \widehat{P} \circ \Phi_\vartheta^{-1} \circ \kappa^{-1}(\mathsf{H}_u) + \tilde{Q}(\mathsf{H}_u) \leq \widehat{P} \circ \Phi_\vartheta^{-1}(\mathsf{H}_u) + P\left(\Phi_\vartheta^{-1}(\{0\})\right)$$

$$= P \circ \Phi_\vartheta^{-1}(\mathsf{H}_u)$$

where H_u stands for a hemisphere or halfspace, whichever is appropriate. $\square$

The appropriate topological tool for spheric closures uses an approach similar to that of Cellina (1969), with a slightly more powerful approximation theorem.

LEMMA A.3.5. *Suppose that S is a compact, convex subset of a convex metrizable subspace of a locally convex topological vector space. Let $\mathcal{F}$ be a set-valued vector field which is a composition of a strongly outer semicontinuous convex-valued mapping F from $\mathbb{D}^p$ to S and a single-valued mapping given by a continuous function f from S to $\mathbb{D}^p$. Suppose that $\mathcal{F}(\vartheta) \subseteq \mathsf{G}_{g(\vartheta)}$ for all $\vartheta \in \mathbb{S}^{p-1}$, where g is a continuous function from $\mathbb{S}^{p-1}$ to $\mathbb{S}^{p-1}$. If g is not homotopic to a constant, then there is a point ϑ in the interior of $\mathbb{D}^p$ such that $0 \in \mathcal{F}(\vartheta)$.*

PROOF. First, we show that there is $\eta > 0$ such that

$$(68) \qquad u^{\mathrm{T}} g(\vartheta) \geq \eta \qquad \text{whenever } u \in \mathcal{F}(\vartheta) \text{ and } \vartheta \in \mathbb{S}^{p-1}.$$

Assume the contrary. Then there are u_ν, ϑ_ν such that $u_\nu \in \mathcal{F}(\vartheta_\nu)$ and $u_\nu^{\mathrm{T}} g(\vartheta_\nu) \to 0$. From the compactness of $\mathbb{S}^{p-1}$ and $\mathbb{D}^p$ follows that we may assume, passing to a subsequence if necessary, that $\vartheta_\nu \to \vartheta \in \mathbb{S}^{p-1}$ and $u_\nu \to u \in \mathbb{D}^p$. The composition $\mathcal{F}$ of the strongly outer semicontinuous mapping F and single-valued continuous, hence also strongly outer semicontinuous mapping f (we abuse slightly the notation) is itself strongly outer semicontinuous; see Klein and Thompson [(1984), Theorem 7.3.11 on page 87], or Nikaido [(1968), Theorem 4.6 on page 71]. We have that $u_\nu^{\mathrm{T}} g(\vartheta_\nu) \to u^{\mathrm{T}} g(\vartheta) = 0$, a contradiction.

Now, we invoke the approximation theorem: Proposition 1.1.10 on page 70 of Borisovich, Gel'man, Myshkis and Obukhovskii (1980). For any $\varepsilon > 0$, there exists a single-valued ε-approximation of F: a continuous function f_ε from $\mathbb{D}^p$ to S with the following property: for every $\vartheta \in \mathbb{D}^p$ there are $\vartheta_\varepsilon \in \mathbb{D}^p$ and $x_\varepsilon \in F(\vartheta_\varepsilon)$ such that $\|\vartheta - \vartheta_\varepsilon\| < \varepsilon$ and $d(f_\varepsilon(\vartheta), x_\varepsilon) < \varepsilon$, where d is the metric on S and the values of f_ε lie in the convex closure of S (which is S itself).

Using the uniform continuity of f (S is compact), we can extend this approximation to the composition of F and f, and construct a sequence f_ν of functions from $\mathbb{D}^p$ to S such that for every $\vartheta \in \mathbb{D}^p$ there are $\vartheta_\nu \in \mathbb{D}^p$ and $u_\nu \in f(F(\vartheta)) = \mathcal{F}(\vartheta)$ satisfying

$$(69) \qquad \|\vartheta - \vartheta_\nu\| < \frac{1}{\nu} \quad \text{and} \quad \|f(f_\nu(\vartheta)) - u_\nu\| < \frac{1}{\nu}$$

and, moreover,

$$(70) \qquad \|g(\vartheta) - g(\vartheta_\nu)\| < \frac{1}{\nu},$$

using also the uniform continuity of g.

Suppose that $\vartheta \in \mathbb{S}^{p-1}$. We will show that $u_\nu^{\mathrm{T}} g(\vartheta) > \eta$ for all ν sufficiently large, uniformly for $\vartheta \in \mathbb{S}^{p-1}$. Assume the contrary. Then there is a sequence ϑ^ν such that (abusing the notation again) $u_\nu^{\mathrm{T}} g(\vartheta_\nu^\nu) \leq \eta$. We may again suppose (by compactness and passing to a subsequence) that $u_\nu \to u$ and $\vartheta^\nu \to \vartheta$, hence $\vartheta_\nu^\nu \to \vartheta$ too. Then we obtain that $u^{\mathrm{T}} g(\vartheta) \leq \eta$, a contradiction with (68): the strong outer semicontinuity of $\mathcal{F}$ yields that $u \in \mathcal{F}(\vartheta)$, since $u_\nu \in \mathcal{F}(\vartheta_\nu^\nu)$.

Now, for ν sufficiently large: $(g(\vartheta_\nu))^{\mathrm{T}} u_\nu^{\mathrm{T}} > \eta$, and therefore also $(g(\vartheta_\nu))^{\mathrm{T}} \times f(f_\nu(\vartheta)) > \eta$, uniformly for $\vartheta \in \mathbb{S}^{p-1}$, due to (69) and (70). Therefore, $f(f_\nu(\vartheta))$ and $g(\vartheta)$ cannot point in opposite directions for any ϑ (if ν is sufficiently large); this means that $f \circ f_\nu$ and g are homotopic. Thus, if g is not homotopic to a constant, then neither is $f \circ f_\nu$; Proposition 5.18 yields a critical point $\vartheta^\nu \in \mathbb{D}^p$ such that $f(f_\nu(\vartheta)) = 0$. By the compactness of $\mathbb{D}^p$, there is $\vartheta \in \mathbb{D}^p$ such that (possibly for a subsequence) $\vartheta^\nu \to \vartheta$. Let ϑ_ν^ν and u_ν be the points satisfying (69); we have that $\vartheta_\nu^\nu \to \vartheta$ and $u_\nu \to 0$. Since $u_\nu \in \mathcal{F}(\vartheta_\nu^\nu)$, the

strong outer semicontinuity of $\mathcal{F}$ yields that $0 \in \mathcal{F}(\vartheta)$. Finally, the possibility that $\vartheta \in \mathbb{S}^{p-1}$ is ruled out by our assumptions. $\square$

As can be seen from the proof, the assumption that $\mathcal{F}(\vartheta) \subseteq \mathsf{G}_{g(\vartheta)}$ is not really necessary. We might formulate Lemma A.3.5 in the vein of Proposition 5.18: either there is a critical point on the boundary (and then there is nothing to prove), or $0 \notin \mathcal{F}(\vartheta)$ for all $\vartheta \in \mathbb{S}^{p-1}$; Lemma A.3.5 holds under this assumption as well, only the proof is more tedious.

A.4. Spheric closures of regression fields. Recall that π denotes the factorization mapping from $\mathbb{S}^{p-1}$ to the projective plane $\mathbb{RP}^{p-1}$ and $\pi^{-1}(\bar{Q})$ stands for the set of all probabilities of the form $\bar{Q} \circ \pi^{-1}$ for a given probability $\bar{Q}$ on $\mathbb{RP}^{p-1}$.

LEMMA A.4.1.   *Let $Z = (X, Y)$ be a random variable with values in $\mathsf{X} \times \mathbb{R}$ where $\mathsf{X} \subseteq \mathbb{R}^p$; let $\Theta = \mathbb{R}^p$. If $\{P_\vartheta : \vartheta \in \Theta\}$ is the regression probability field defined by $P_\vartheta = \mathcal{L}(XX^{\mathrm{T}}\vartheta - XY)$ for any $\vartheta \in \Theta$, then*:

(i) *$\{P_\vartheta : \vartheta \in \Theta\}$ has a spheric closure $\mathcal{P}(\vartheta)$, well-defined for all $\vartheta \in \bar{\Theta} = \bar{\mathbb{R}}^p$*;
(ii) *there is a probability $\bar{Q}$ on $\mathbb{RP}^{p-1}$ such that $Q \in \pi^{-1}(\bar{Q})$ for any $Q \in \mathcal{P}(\vartheta)$ and any $\vartheta \in \bar{\mathbb{R}}^p$*;
(iii) *for any $\vartheta \in \partial\mathbb{R}^p$, $Q(\mathsf{H}_\vartheta) = 1$ and $Q(\mathsf{G}_\vartheta) > 1 - \Delta(X)$ for any $Q \in \mathcal{P}(\vartheta)$.*

PROOF.   The proof of (i) consists merely in noting that in this particular setting, the density of $\dot{\Theta}$ in Θ follows from Lemma A.2.1; and the continuity of $\{P_\vartheta : \vartheta \in \Theta\}$ in any $\vartheta \in \Theta$, and thus in $\dot{\Theta}$, from Proposition 5.21(i).

For any $\vartheta \in \dot{\Theta}$, the measure $P_\vartheta \circ \kappa^{-1}$ belongs to $\pi^{-1}(\bar{Q})$ where $\bar{Q}$ is a fixed probability measure on $\mathbb{RP}^{p-1}$; just note that for any symmetric set E,

$$\mathbb{P}[((XX^{\mathrm{T}}\vartheta - XY)) \in E] = \mathbb{P}[((X)) \operatorname{sgn}(X^{\mathrm{T}}\vartheta - XY) \in E] = \mathbb{P}[((X)) \in E],$$

thus $(P_\vartheta \circ \kappa^{-1})(E)$ is the same for all $\vartheta \in \dot{\Theta}$. The rest of (ii) follows from the fact that $\pi^{-1}(\bar{Q})$ is convex and closed in the weak topology.

In particular, for any $\vartheta \in \bar{\mathbb{R}}^p$ and any $Q \in \mathcal{P}(\vartheta)$,

$$(71) \qquad\qquad Q(\partial\mathsf{H}_u) = \mathbb{P}[((X)) \in \mathsf{H}_u] = \mathbb{P}[u^{\mathrm{T}}X = 0].$$

Fix $\vartheta \in \partial\mathbb{R}^p$. If $Q \in \dot{\mathcal{P}}(\vartheta)$, then Lemma A.3.2(ii) yields a sequence $\vartheta_\nu \in \dot{\Theta}$ such that $\vartheta_\nu \to \vartheta$, and $Q_\nu = P_{\vartheta_\nu} \circ \kappa^{-1} \rightharpoonup Q$. By Lemma A.2.4,

$$Q(\mathsf{H}_{-((\vartheta))}) = \lim_{\nu \to \infty} Q_\nu(\mathsf{H}_{-((\vartheta))}) = \lim_{\nu \to \infty} P_{\vartheta_\nu}(\mathsf{H}_{-((\vartheta))})$$

$$= \lim_{\nu \to \infty} \mathbb{P}[-((\vartheta))^{\mathrm{T}}(XX^{\mathrm{T}}\vartheta_\nu - XY) \geq 0]$$

$$= \lim_{\nu \to \infty} \mathbb{P}\left[-((\vartheta))^{\mathrm{T}}\left(XX^{\mathrm{T}}((\vartheta_\nu)) - \frac{XY}{\|\vartheta_\nu\|}\right) \geq 0\right]$$

$$\leq \mathbb{P}[-((\vartheta))^{\mathrm{T}}XX^{\mathrm{T}}((\vartheta)) \geq 0] = \mathbb{P}[X^{\mathrm{T}}((\vartheta)) = 0]$$

(here H_u stands again both for a hemisphere or halfspace). By (71), $Q(\mathsf{G}_{-(\!(\vartheta)\!)}) = 0$ and hence $Q(\mathsf{H}_{(\!(\vartheta)\!)}) = 1$, for any $Q \in \dot{\mathcal{P}}(\vartheta)$. This conclusion is preserved under convex combinations and also under taking limits in the weak topology; thus, $Q(\mathsf{H}_{(\!(\vartheta)\!)}) = 1$ for all $Q \in \mathcal{P}(\vartheta)$. By (71) again, we obtain that $Q(\mathsf{G}_{(\!(\vartheta)\!)}) = 1 - Q(\partial \mathsf{H}_{(\!(\vartheta)\!)}) = 1 - \mathbb{P}[(\!(\vartheta)\!)^{\mathrm{T}} X = 0] > 1 - \Delta(X)$. $\square$

PROOF OF THEOREM 3.3.   Let $Z = (X, Y)$ be a random variable with values in $\mathsf{X} \times \mathbb{R}$ whose distribution is P. We proceed by induction with respect to p. For $p = 1$, the model reduces to the univariate location model and the theorem holds: the sample median has depth not less than $1/2$.

Suppose that (12) holds for all dimensions less than p. If $\Delta(X) = 1$, the problem is not identifiable; we can reparametrize it to a lesser dimensional one and use the induction assumption. The lower bound obtained in this way is more stringent than $1/(p + 1)$ and all our fits are extensions of the lower-dimensional ones; therefore, the inequality (12) holds.

Thus, suppose that $\Delta(X) < 1$ and assume the contrary to the inequality (12): there is $\varepsilon > 0$ such that

(72)
$$\sup_{\vartheta \in \mathbb{R}^p} d(\vartheta, P) + \varepsilon < \frac{1}{p + 1}.$$

We know that Lemma A.4.1(i) guarantees the existence of the spheric closure of the corresponding regression probability field, a strongly outer semicontinuous and convex-valued set-valued mapping $\mathcal{P}(\cdot)$ defined on $\bar{\mathbb{R}}^p$, whose range is, by Lemma A.4.1(ii), a subset of $\pi^{-1}(\bar{Q})$ for some fixed probability $\bar{Q}$ on $\mathbb{RP}^{p-1}$; note that $\pi^{-1}(\bar{Q})$ is a convex, closed (and hence compact, with respect to the weak topology) subset of the space of all probability measures on $\mathbb{S}^{p-1}$. Lemma A.2.5 says that $s_{d(\cdot)+\varepsilon}(\cdot)$, a (single-valued) function assigning every $Q \in \pi^{-1}(\bar{Q})$ a vector in the unit ball $\mathbb{D}^p$, is continuous (in the weak topology). Let $\mathcal{F}$ be the composition of $\mathcal{P}(\cdot)$ and $s_{d(\cdot)+\varepsilon}(\cdot)$: a set-valued mapping from $\bar{\mathbb{R}}^p$ to $\mathbb{D}^p$. Lemmas A.4.1 and A.2.6 guarantee that for $\vartheta \in \partial \mathbb{R}^p$, $\mathcal{F}(\vartheta) \subseteq \mathsf{G}_{(\!(\vartheta)\!)}$. Since $\bar{\mathbb{R}}^p$ and $\partial \mathbb{R}^p$ are topologically equivalent to $\mathbb{D}^p$ and $\mathbb{S}^{p-1}$, we may apply Lemma A.3.5, which yields the existence of $\dot{\vartheta} \in \mathbb{R}^p$ such that $0 \in \mathcal{F}(\dot{\vartheta})$; that is, there is $Q \in \mathcal{P}(\dot{\vartheta})$ such that $s_{d(Q)+\varepsilon}(Q) = 0$. By Proposition 5.16 and Lemma A.3.4, $1/(p+1) \leq d(Q) + \varepsilon \leq d(\vartheta, P) + \varepsilon$. We have obtained a contradiction of (72). $\square$

NOTE ADDED IN PROOF (Proof of Theorem 3.5).   We use formula (46). There are two possibilities: either all suprema, for various ϑ, at the right-hand side of (46) are strictly less than maximal depth (i.e., supremum of depth)—in such a case the theorem is proved; or, there is ϑ such that the right-hand side of (46) is equal to the maximal depth—then we consider the whole problem restricted to $\ell(\vartheta)$ and use induction with respect to the dimension of the parameter space as in the proof of Theorem 3.3. $\square$

REFERENCES

ADROVER, J., MARONNA, R. and YOHAI, V. (2000). Relationships between maximum depth and projection regression estimates. *J. Statist. Plann. Inference*. **105** 363–375.

AMENTA, N., BERN, M., EPPSTEIN, D. and TENG, S.-H. (2000). Regression depth and center points. *Discrete Comput. Geom.* **23** 305–323.

BAI, Z.-D. and HE, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Ann. Statist.* **27** 1616–1637.

BALEK, V. and MIZERA, I. (1997). The closeness of the range of a probability on a certain system of random events—an elementary proof. *Bull. Belg. Math. Soc. Simon Stevin* **4** 621–624.

BERN, M. and EPPSTEIN, D. (2000). Multivariate regression depth. In *Proceedings of the 16th Annual Symposium on Computational Geometry* 315–321. ACM Press, New York.

BICKEL, P. J. and MILLAR, P. W. (1992). Uniform convergence of probability measures on classes of functions. *Statist. Sinica* **2** 1–15.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BILLINGSLEY, P. (1971). *Weak Convergence of Measures*: *Applications in Probability*. SIAM, Philadelphia.

BIRCH, B. J. (1959). On 3N points in a plane. *Proc. Cambridge Philos. Soc.* **55** 289–293.

BORISOVICH, Y. G., GEL'MAN, B. D., MYSHKIS, A. D. and OBUKHOVSKII, V. V. (1980). Topological methods in the fixed-point theory of multi-valued maps. *Uspekhi Mat. Nauk* **35** 59–126. [Translation (1980) *Russian Math. Surveys* **35** 65–143.]

BORISOVICH, Y. G., GEL'MAN, B. D., MYSHKIS, A. D. and OBUKHOVSKII, V. V. (1982). Multivalued mappings. In *Mathematical Analysis* **19** 127–230. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatsii, Moscow (in Russian).

CAPLIN, A. and NALEBUFF, B. (1988). On 64%-majority rule. *Econometrica* **56** 787–814.

CAPLIN, A. and NALEBUFF, B. (1991a). Aggregation and social choice: A mean voter theorem. *Econometrica* **59** 1–23.

CAPLIN, A. and NALEBUFF, B. (1991b). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica* **59** 25–59.

CARRIZOSA, E. (1996). A characterization of halfspace depth. *J. Multivariate Anal.* **58** 21–26.

CELLINA, A. (1969). Approximation of set-valued functions and fixed point theorems. *Ann. Mat. Pura Appl.* **82** 17–24.

CHAMBERLIN, E. (1933). *The Theory of Monopolistic Competition*. Harvard Univ. Press.

DANIELS, H. E. (1954). A distribution-free test for regression parameters. *Ann. Math. Statist.* **25** 499–513.

DAVIES, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.* **21** 1843–1899.

DODSON, C. T. J. and PARKER, P. E. (1997). *A User's Guide to Algebraic Topology*. Kluwer, Dordrecht.

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

DOOB, J. L. (1994). *Measure Theory*. Springer, New York.

EDELSBRUNNER, H. (1987). *Algorithms in Combinatorial Geometry*. Springer, Berlin.

EDGEWORTH, F. Y. (1888). On a new method of reducing observations relating to several quantities. *Philosophical Magazine* **25** 184–191.

FRISCH, R. (1966). *Maxima and Minima*: *Theory and Economic Applications*. Reidel, Dordrecht.

HE, X. and PORTNOY, S. (1998). Asymptotics of the deepest line. In *Applied Statistical Science III*: *Papers in Honor of A. K. Md. E. Saleh* (S. E. Ahmed, M. Ahsanullah and B. K. Sinha, eds.) 71–81. Nova Science Publications, Commack, NY.

HE, X. and WANG, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495–504.

HILL, B. M. (1960). A relationship between Hodges' bivariate sign test and a non-parametric test of Daniels. *Ann. Math. Statist.* **31** 1190–1192.

HODGES, J. L., JR. (1955). A bivariate sign test. *Ann. Math. Statist.* **26** 523–527.

HOTELLING, H. (1929). Stability in competition. *Econom. J.* **39** 41–57.

HUBERT, M., ROUSSEEUW, P. J. and VAN AELST, S. (1999). Similarities between location depth and regression depth. In *Statistics in Genetics and in the Environmental Sciences* (L. Fernholz, S. Morgenthaler and W. Stahel, eds.) 159–172. Birkhäuser, Basel.

KLEIN, E. and THOMPSON, A. C. (1984). *Theory of Correspondences*. Wiley, New York.

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

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

MIZERA, I. (1998). On depth and deep points: A calculus (abstract). *IMS Bull.* **27** 207.

NEUMANN, B. H. (1945). On an invariant of plane regions and mass distributions. *J. London Math. Soc.* **20** 226–237.

NIKAIDO, H. (1968). *Convex Structures and Economic Theory*. Academic Press, New York.

NOLAN, D. (1992). Asymptotics for multivariate trimming. *Stochastic Process. Appl.* **42** 157–169.

NOLAN, D. (1999). On min-max majority and deepest points. *Statist. Probab. Lett.* **43** 325–333.

ORTEGA, J. M. and RHEINBOLDT, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York.

PONSTEIN, J. (1967). Seven kinds of convexity. *SIAM Rev.* **9** 115–119.

PORTNOY, S. and MIZERA, I. (1999). Comment on "Regression depth," by P. J. Rousseeuw and M. Hubert. *J. Amer. Statist. Assoc.* **94** 417–419.

RADO, R. (1946). A theorem on general measure. *J. London Math. Soc.* **21** 291–300.

ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis*. Springer, Berlin.

ROTMAN, J. J. (1988). *An Introduction to Algebraic Topology*. Springer, New York.

ROUSSEEUW, P. J. and HUBERT, M. (1999a). Regression depth. *J. Amer. Statist. Assoc.* **94** 388–402.

ROUSSEEUW, P. J. and HUBERT, M. (1999b). Depth in an arrangement of hyperplanes. *Discrete Comput. Geom.* **22** 167–176.

SMALE, S. (1973). Global analysis and economics I: Pareto optimum and a generalization of Morse theory. In *Dynamical Systems. Proc. Symposium Univ. Bahia, Salvador* (M. M. Peixoto, ed.) 531–544. Academic Press, New York.

SMALE, S. (1975a). Sufficient conditions for an optimum. *Dynamical Systems. Lecture Notes in Math.* **468** 287–292. Springer, Berlin.

SMALE, S. (1975b). Optimizing several functions. In *Manifolds—Tokyo 1973. Proc. International Conference on Manifolds and Related Topics in Topology* (A. Hattori, ed.) 69–75. Univ. Tokyo Press.

TJUR, T. (1980). *Probability Based on Radon Measures*. Wiley, New York.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proc. International Congress of Mathematicians* **2** 523–531. Canad. Math. Congress, Montreal.

VAN AELST, S. and ROUSSEEUW, P. J. (2000). Robustness of deepest regression. *J. Multivariate Anal.* **73** 82–106.

VAN AELST, S., ROUSSEEUW, P. J., HUBERT, M. and STRUYF, A. (2001). The deepest regression method. *J. Multivariate Anal.* **81** 138–166.

WAN, Y. H. (1975). On local Pareto optima. *J. Math. Econom.* **2** 35–42.

WAN, Y. H. (1978). On the structure and stability of local Pareto optima in a pure exchange economy. *J. Math. Econom.* **5** 255–274.

DEPARTMENT OF PROBABILITY
    AND STATISTICS
COMENIUS UNIVERSITY IN BRATISLAVA
MLYNSKÁ DOLINA
SK-84215 BRATISLAVA
SLOVAKIA

DEPARTMENT OF MATHEMATICAL
    AND STATISTICAL SCIENCES
UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA T6G 2G1
CANADA
E-MAIL: mizera@stat.ualberta.ca