Depth Functions on General Data Spaces, II. Formulation and Maximality, with Consideration of the Tukey, Projection, Spatial, and “Contour” Depths

Robert Serfling

Serfling & Thompson Statistical Consulting

May 2019

1Serfling & Thompson Statistical Consulting, 1921 Sparrows Point Drive, Plano, Texas 75023 USA. Email: rjserfling@gmail.com. Website: www.utdallas.edu/~serfling.
Abstract

The question of what the defining properties for depth functions should be is approached through axiomatic considerations, leading to characterizations of the key defining properties of monotonicity, invariance, nondegeneracy, and nullness at infinity. Also, the highly desired property of maximality of depth at a center of symmetry is studied under the central, angular, and halfspace notions of symmetry for probability distributions. It is seen that the center-locating and median-defining roles of depth functions hold without technical issues of dimensionality of the data space. Applying these results and also certain general perspectives on the critical evaluation of proposed depths and on the role of population versions, nondegeneracy issues with the halfspace and projection depths are interpreted and monotonicity issues with the spatial depth are clarified. This work is carried out in the setting of a general data space.

AMS 2000 Subject Classification: Primary 62G99 Secondary 62H99

Key words and phrases: Depth functions; Symmetry; Tukey halfspace depth; Projection depth; Spatial depth; Contour depth.
1 Introduction

In extending univariate quantile, order statistic, rank, and outlier detection methods to data sets in higher dimensional Euclidean and more complex spaces, one must compensate for lack of a natural order and a well-determined median-type notion of center. The “depth function approach” conceived by Tukey (1975) accomplishes this through a center-outward ordering, where “deeper” corresponds to more central with the deepest point a “center”. Given a data set $X_n$ of size $n$ in a data space $\mathcal{X}$, one identifies not only an $X_n$-based “center” in $\mathcal{X}$, but also, for any point $x \in \mathcal{X}$, the degree of its centrality with respect to $X_n$, yielding a center-outward ordering of points in $\mathcal{X}$ according to some formulated depth function $D(\cdot,X_n)$.

The intrinsic nature of a given depth function $D(\cdot,X_n)$ can be obscured, however, by the associated computational details and heuristics (indeed, a depth function might have various algorithmic implementations). In order to explore and better understand a depth function, it is helpful and sometimes simpler to introduce an associated “population” version $D(x,P)$, $x \in \mathcal{X}$, for $P$ an arbitrary probability distribution on $\mathcal{X}$. Here $D(x,P)$ measures the depth of $x$ relative to $P$, and a sample version $D(\cdot,X_n)$ is recovered by taking $P = \hat{P}_n$, the empirical distribution placing probability $n^{-1}$ on each observation in $X_n$, or alternatively by taking a smoothed version of $\hat{P}_n$.

As a key motif of the depth approach, the notion of “centrality” receives a stronger focus and a more explicit structural role than in the univariate setting. Since univariate scaled-deviation type outlyingness functions have long been extended to multivariate versions, however, and since centrality and outlyingness are actually equivalent through an inverse relationship, what is especially novel in the depth approach is the focus on centrality and the accomplishment of the ordering without necessarily specifying in advance a “center”, which instead is given by the deepest point of the depth function. Although technically equivalent algorithmically, determination of a center and identification of outliers are very different both conceptually and in application.

The Tukey (1975) depth function in particular is based on counts of sample points in halfspaces. Beginning with Liu (1988, 1990), many other versions of “depth” have also been formulated, not only for Euclidean but also for Banach or Hilbert spaces, reflecting various points of view. Comparison of points in a space $\mathcal{X}$ relative to their representativeness of the main body of data points must take into account not only the locations of the points but also, when present, other distinguishing features such as their shapes. Thus some depth functions are quite $\mathcal{X}$-specific in their formulations. Depth functions vary widely in structure, properties, robustness, computational ease, and asymptotic behavior. Because these aspects trade off against each other with differing subjective priorities, no depth exhibits overall dominance by
any common agreement. In practice, one chooses by evaluating candidates according to one's adopted criteria. See Serfling (2019) for further introductory background discussion of the depth approach.

A fundamental challenge is to specify the properties most desired for any proposed depth function. Here a number of key questions arise. What type of invariance under transformations of coordinates is desired? What kind of monotonicity as one moves outward from the center? Can nondegeneracy be assured? Should the depth be zero at infinity? If the underlying population satisfies some notion of symmetry about a center, is a “good” depth maximal at this center? Whereas a depth function generates a family of contours, does the converse hold in a practical sense? Are there intrinsic features of depth functions that hold across all types of data space $\mathcal{X}$?

Besides methodological treatment of the depth approach by Liu, Parelius and Singh (1999), theoretical developments on desired properties are given by Zuo and Serfling (2000a,b) and Serfling (2006) for Euclidean data and by Nieto-Reyes and Battey (2016) and Mosler and Polyakova (2018) for functional data.

Here we develop a number of technical results for depths on general $\mathcal{X}$ and apply them to enhance our understanding of the well-established Tukey (halfspace), projection, and spatial depths as well as the “contour” depth in general settings. Besides the above questions, our treatment also addresses two additional questions. Does unfavorable performance found for a population depth in some particular instance carry substantive implications for practical use of the sample version? Are some populations pathologically unsuitable for the formulation of depth functions? In this regard, we draw upon certain perspectives and assertions argued in detail in Serfling (2019) and stated here as

Selected Guidelines:

A [18, Assertion 2.2] The depth approach applies in practice to data sets $\mathcal{X}_n$ of finite size $n > 1$ in some data space $\mathcal{X}$ of dimension $d < \infty$, with $d < n$ preferred whenever feasible.

B [18, Assertion 2.4] Properties desired for a population depth function $D(\cdot, P)$ on a space $\mathcal{X}$ need not be satisfied for two classes of rather irrelevant $P$, those which either (i) are non-atomic with support a proper subset of $\mathcal{X}$ with empty interior, or (ii) place probability 1 on a proper subset $E$ of $\mathcal{X}$ that is dense in $\mathcal{X}$, i.e., for $P \in \mathcal{P}_1(\mathcal{X}) \cup \mathcal{P}_2(\mathcal{X})$, where

\[ \mathcal{P}_1(\mathcal{X}) = \{ P \in \mathcal{X} : P \text{ is non-atomic, supp } P \neq \mathcal{X}, (\text{supp } P) = \emptyset \}; \]

\[ \mathcal{P}_2(\mathcal{X}) = \{ P : \exists E = \text{essupp } P \text{ satisfying } E \neq \partial E \text{ and } E = \mathcal{X} \}, \]

and “essupp $P$” (essential supremum) is defined as any subset $E$ of supp $P$ for which $P(E) = 1$ and $P(E^c) = 0$, that is, for which $P(X \in E) = 1$ and $P(X \in E^c) = 0$ for $X$ having distribution $P$. 

2
The practical merits of a depth function should be evaluated solely on the basis of performance of the sample version \( D(\cdot, \mathbf{X}_n) \) as a statistical procedure. The population version \( D(\cdot, P) \) is relevant only for insights.

In Sect. 3, we provide a rather axiomatic approach toward formulation of key properties for defining depth (and outlyingness) functions. This leads to a formal definition of depth functions in terms of four key properties involving *monotonicity, invariance, nondegeneracy*, and *nullness at infinity*. Preliminary treatment of contour sets and notions of invariance is provided in Sect. 2.

In Sect. 4, we examine the classical *halfspace, projection, and spatial depth functions* with respect to the four defining properties. None of these depths fulfills all of the requirements perfectly, and we discuss their shortcomings precisely with a view to practical significance. We also consider "contour depth".

A further highly desired property for depth functions, having nearly the status of a defining property, is *maximality at a center of symmetry*. In Sect. 5, relevant results are developed for depths on general \( \mathcal{X} \), with respect to three leading nonparametric notions of symmetry – central, angular, and halfspace. Also, the halfspace, projection, and spatial depth functions are revisited with comparative discussion, and a comment on contour depth is provided.

Sect. 6 presents brief concluding remarks.

Proofs are provided in Appendix A and background technical material in Appendix B.

## 2 Preliminaries

### 2.1 Contours and contour sets of a function

Often quite useful in considering a real-valued function \( f \) on any space \( \mathcal{Y} \) are its contours and contour sets. Here we define these and examine their properties for \( f \) a *monotone* (entirely nondecreasing or entirely nonincreasing) function. Put \( \gamma_U = \sup_y f(y) \leq \infty \) and \( \gamma_L = \inf_y f(y) \geq -\infty \).

**Definition 2.1 (Contours and Contour Sets).** *(i)* The upper and lower contour sets (or upper and lower level sets) for a mapping \( f : \mathcal{Y} \mapsto \mathbb{R} \) are defined as

\[
U(f, t) = \{ y \in \mathcal{Y} : f(y) \geq t \} \quad \text{and} \quad L(f, t) = \{ y \in \mathcal{Y} : f(y) \leq t \},
\]

respectively, for levels \( t \in \mathbb{R} \).

*(ii)* For \( \gamma_L < \gamma_U \), the contour (or level set) at level \( t \) is defined as

\[
C(f, t) = \begin{cases} 
U(f, \gamma_U), & t = \gamma_U \text{ (if attained)}, \\
U(f, t) \cap L(f, t), & \gamma_L < t < \gamma_U, \\
L(f, \gamma_L), & t = \gamma_L \text{ (if attained)}.
\end{cases}
\]
For $t < \gamma_L$ or $t > \gamma_U$, or if $\gamma_L = \gamma_U$, no contour is defined.  \[ \square \]

For $\gamma_L < t < \gamma_U$, if each point of $C(f, t)$ is a limit point of both $U(f, t)$ and $L(f, t)$, then $C(f, t) = \partial U(f, t) = \partial L(f, t)$, the common boundary of the upper and lower contour sets. With increasing $t$, the upper (lower) level sets form a decreasing (an increasing) sequence of nested sets. For $f$ constant over an interval $a \leq t \leq b$, the contours is a band. Thus contours are of greatest interest when $f$ is strictly monotonic.

In treating depth and outlyingness functions, the case of $f$ monotone on rays from a center is of special interest. For $f$ nonincreasing on rays\(^1\) outward from a point $y_0$, which then is a maximal point of $f$, it is equivalent that the upper level sets $U(f, t)$ are all star-shaped\(^2\) about $y_0$. Further, $U(f, t)$ is bounded if and only if $t > \gamma_L = \inf_y f(y)$ ($\geq -\infty$). (If $t = \gamma_L$ is attained by $f$, then $U(f, t) = Y$ for $t \leq \gamma_L$, although $U(f, \gamma_L) = U(f, \gamma_L)$.) Analogous statements hold for $f$ nondecreasing on rays. The following lemma is straightforward to prove.

**Lemma 2.1.** For $f$ nonincreasing (nondecreasing) on rays from $y_0$, the upper (lower) level sets are star-shaped about $y_0$, and conversely, and $f$ is maximal (minimal) at $y_0$. Further, the upper (lower) contour sets at level $t$ are bounded if $\inf_y f(y) < t \leq \sup_y f(y)$ ($\inf_y f(y) \leq t < \sup_y f(y)$).

Star-shaped sets are simply connected\(^3\). Also, a set is convex\(^4\) if and only if it is star-shaped with respect to each of its points. If the upper (lower) sets of $f$ are in fact convex, then $f$ is called quasiconcave (quasiconvex), a relaxation of concavity (convexity) of $f$. If the epigraph\(^5\) $\text{Epi}(-f)$ ($\text{Epi}(f)$) is convex in $\mathcal{X} \times \mathbb{R}$, then $f$ is concave\(^6\) (convex).

### 2.2 Desired invariance properties of depth functions

Searching for important features and structure in a data set is leading focus in exploratory data analysis. Here it is essential to insist that such findings be invariant or equivariant under selected changes of coordinates, at least within

---

\(^1\)A function $f$ is nonincreasing (nondecreasing) on rays if, for any $y_0$ at which $f$ is maximal (minimal), $f$ decreases (increases) along the ray $\{y_0 + \alpha(y - y_0), 0 < \alpha \leq 1\}$ from $y_0$ to any other point $y$. Equivalently, $f(y) \leq (\geq) f(y_0 + \alpha(y - y_0))$, $\alpha \in [0, 1]$, for such $y_0$ and $y$.

\(^2\)A set $S$ is star-shaped (or “star-convex” or “radially convex”) if there exists $s_0 \in S$ such that the line segment $(1 - \alpha)s_0 + \alpha s$, $0 \leq \alpha \leq 1$, from $s_0$ to $s$ is in $S$ for each $s \in S$.

\(^3\)A set $S$ is simply connected if it is path-connected (any $x$ and $y$ in $S$ can be joined by a path in $S$) and any loop in $S$ can be contracted to a point in $S$.

\(^4\)A set $S$ is convex if any $x$ and $y$ in $S$ can be joined by a line segment lying wholly in $S$.

\(^5\)The epigraph of a function $f$ is $\text{Epi}(f) = \{(x, z) : x \in \mathcal{X}, z \in \mathbb{R}, z \geq f(x)\}$.

\(^6\)A function $f$ is concave (convex) if $f(\lambda x + (1 - \lambda)y) \geq (\leq) \lambda f(x) + (1 - \lambda)f(y)$, for all $x, y \in \mathcal{X}$ and $\lambda \in [0, 1]$.
some specified equivalence relation on the elements of description. Otherwise, interesting aspects that arise for consideration could be mere artifacts of the given coordinate system. In particular, this principle applies to data-based characterizations of “depth” or “outlyingness” of points in a data space $\mathcal{X}$. Here we formulate the relevant notion, *invariance within equivalence*\(^7\).

Denote the class of bounded real-valued $\mathcal{A}$-measurable mappings defined on $\mathcal{X}$ by $\mathcal{M}(\mathcal{X}) = \{\mathcal{A}$-measurable $m : \mathcal{X} \mapsto \mathbb{R}, \sup_x m(x) < \infty\}$. Suppose that, for certain purposes, an equivalence relation $E$ on $\mathcal{M}(\mathcal{X})$ is apropos, with equivalence classes generated as the orbits\(^8\) of a class $\mathcal{F}$ of transformations taking $\mathcal{M}(\mathcal{X})$ into itself. Then mappings $m$ in the same orbit of $\mathcal{M}(\mathcal{X})$ under $\mathcal{F}$ are considered $E$-equivalent.

Further, suppose that the mappings in $\mathcal{M}(\mathcal{X})$ should be invariant if a coordinate-changing transformation\(^9\) $g$ is applied to $\mathcal{X}$. That is, for a mapping $m^* \in \mathcal{M}(\mathcal{X})$, the role of $m^*$ on $\mathcal{X}$ is handed over to an associated mapping $m^*_g$ on $g\mathcal{X}$, with the desired invariance expressed by $m^*_g(gx) = m^*(x)$, $x \in \mathcal{X}$. (Here $m^*_g \in \mathcal{M}(\mathcal{X})$ and $g\mathcal{X} = \mathcal{X}$.) Moreover, this is to hold for each $g$ in a designated class $\mathcal{G}(\mathcal{X})$ (typically a group) of transformations on $\mathcal{X}$.

However, in the active presence of the equivalence $E$, one requires only that the function $m^*_g(gx)$, $x \in \mathcal{X}$, belong to the $E$-orbit of $m^*$ rather than be identical with $m^*$. This avoids undue stringency in characterizing “invariance”. A mapping $m^*$ satisfying this property is $\mathcal{G}(\mathcal{X})$-*invariant within $E$-equivalence*.

If the chosen mapping $m^*$ in $\mathcal{M}(\mathcal{X})$ is data-based, say $m^*(\mathbb{X}_n)$, then the invariance of $m^*$ with respect to $\mathcal{G}(\mathcal{X})$ in the standard sense means that $m^*(\mathbb{X}_n)$ depends on $\mathbb{X}_n$ only through a *maximal invariant statistic*\(^{10}\). However, for $m^*$ invariant only *within $E$-equivalence*, such a representation of $m^*(\mathbb{X}_n)$ no longer need hold (although it might).

### 3 Formal definitions of depth functions

Before introducing formal definitions of depth and outlyingness functions, let us explore intuitively their two key roles regarding points in $\mathcal{X}$, *center-outward ordering*, and *measurement of degree of centrality or outlyingness*. We propose for this purpose certain “axioms” characterizing the nature and meaning of depth and outlyingness functions. For a given distribution $P$ on $\mathcal{X}$ and given depth function $D(x, P)$ and associated outlyingness function $O(x, P)$, $x \in \mathcal{X}$,

---

\(^7\)A related topic, *invariant within equivalence coordinate systems*, is treated in Serfling (2015).

\(^8\)The orbit of $m$ relative to $\mathcal{F}$ is the set $\{fm : f \in \mathcal{F}\}$.

\(^9\)It is assumed that coordinate-changing transformations are “one-to-one” (injective) and “onto” (surjective) mappings, or “bijections”.

\(^{10}\)A *maximal invariant statistic* on $\mathbb{X}_n$ is any labeling of the orbits of $\mathcal{X}_n$ induced by the transformations $g\mathbb{X}_n$ for $g$ in $\mathcal{G}(\mathcal{X})$. See Lehmann and Romano (2005), Sect. 6.2.
these are:

(A1) Some point $\tilde{\theta}(P) \in X$ plays the role of “center” for the given $D(\cdot, P)$ or $O(\cdot, P)$.

(A2) For any $x \in X$, all points $x'$ on the ray joining $\tilde{\theta}(P)$ and $x$ are “at least as central as $x$”, or, equivalently, “not more outlying than $x$”.

(A3) For any $x$ and $x' \in X$, the relative centrality of $x$ and $x'$ is reflected by the ratio of $D(x, P)$ and $D(x', P)$ and the relative outlyingness by the ratio of $O(x, P)$ and $O(x', P)$.

(A4) Points $x$ in the support of $P$, except possibly on its boundary, have positive depth value, $D(x, P) > 0$.

(A5) Sufficiently central points have negligible outlyingness value, sufficiently outlying points have negligible depth value, with $O(x, P)$ equal to zero at the “center” and $D(x, P)$ equal to zero at “infinity”.

A key role of depth and outlyingness functions is to generate center-outward ordering. This is provided by (A1) and (A2), which yield via the discussion in Sect. 2.1 that ideally a depth function $D$ should be nonincreasing along any ray outward from any point $\tilde{\theta}(P)$ of maximality and, equivalently, have bounded upper contour sets star-shaped about $\tilde{\theta}(P)$. Likewise, ideally an outlyingness function $O$ should be nondecreasing on any ray outward from a point $\tilde{\theta}(P)$ of minimality and have bounded lower contour sets star-shaped about $\tilde{\theta}(P)$. These are highly desirable properties to be incorporated into the definitions of $D$ and $O$ functions.

The other direct role of depth and outlyingness functions is to measure the degree of “centrality” or of “outlyingness”, respectively. This requires taking account of the actual values of the $D$ or $O$ functions rather than just the monotonicity that determines a center-outward ordering and is accommodated by (A3), (A4), and (A5). Here (A5) reflects that one cannot attach a positive “degree of centrality” to points sufficiently far from the center, nor a positive “degree of outlyingness” to the center. Finally, (A4) reflects that points in the support of $P$ should have positive centrality value, just as they have positive density or probability mass, except possibly within a set of $P$-probability zero. This is a nondegeneracy condition. Indeed, sufficiently central points in the convex hull of supp $P$, for example the point of maximal depth, should also have positive depth value. These features too should be included in defining $D$ and $O$ functions.

Regarding (A3), in order for the function values of $D$ or $O$ to meaningfully measure centrality or outlyingness in a relative sense, the measures cannot be

\footnote{Alternatively, the desired monotonicity along rays may instead hold for a closely related subsidiary function in terms of which the depth or outlyingness function is defined (see Sect. 4.3).}
mere artifacts of the particular coordinate system being used. Rather, at least relatively, they should be invariant under typical changes of the coordinates, in the same spirit, for example, that the usual correlation measure in statistical analysis is invariant under location and scale changes in the data. This is implicit in (A3).

Invariance for $D$ and $O$ functions involves two aspects. (a) Firstly, note that multiplying $D$ or $O$ by a constant $c > 0$ does not change performance with respect to “axioms” (A1)-(A5). In the terminology of Sect. 2.2, $D$ and $O$ functions each satisfy the equivalence relation $E_0$ on $\mathcal{M}(\mathcal{X})$ corresponding to invariance under the group of scale transformations on $\mathcal{M}(\mathcal{X})$,

$$F_0(\mathcal{X}) = \{ f_c : f_cm = cm, \ m \in \mathcal{M}(\mathcal{X}), \ c > 0 \}.$$ 

Thus $D$ functions differing only by a multiplicative constant are $E_0$-equivalent, and likewise for $O$ functions. (b) Secondly, for a coordinate transformation $g$ from a designated relevant class $\mathcal{G}(\mathcal{X})$, the relative centrality or outlyingness of two points $x$ and $x'$ in $\mathcal{X}$ according to $D$ or $O$ and that of $gx$ and $gx'$ according to $D_g$ or $O_g$ induced on $g\mathcal{X}$ should agree precisely. Here the depth function $D_g$ on $g\mathcal{X}$ corresponding to the depth function $D(\cdot, P)$ on $\mathcal{X}$ is given by $D(\cdot, Pg^{-1})$, and $O_g$ is similarly defined. The desired invariance for $D$ (and similarly for $O$) is thus $\{ D(gx, Pg^{-1}), \ x \in \mathcal{X} \} = D(\cdot, P), \ g \in \mathcal{G}(\mathcal{X})$, except that actually this must be required merely within $E_0$-equivalence:

$$\{ D(gx, Pg^{-1}), \ x \in \mathcal{X} \} \text{ belongs to the } E_0\text{-orbit of } D(\cdot, P), \ g \in \mathcal{G}(\mathcal{X}).$$

The definitions of depth or outlyingness functions should include specification of a favorable $\mathcal{G}(\mathcal{X})$-invariance within $E_0$-equivalence property.

In practice, for a particular application context as represented by the type of data space $\mathcal{X}$ being considered, there is a preferred class $\mathcal{G}_0(\mathcal{X})$. A larger class would be too stringent and rule out some $D$ or $O$ functions simply on the basis of some irrelevant requirements, while a smaller class would permit a serious shortcoming from a practical standpoint. For example, the class of all affine transformations is preferred for $\mathcal{X} = \mathbb{R}^d$, but for $\mathcal{X}$ a Hilbert space of curves this class is too stringent because it includes heterogeneous scale changes of coordinates, a type of coordinate change not of practical interest for curves.

In general, for any $\mathcal{X}$ the preferred class $\mathcal{G}_0(\mathcal{X})$ should at least include the groups of translations, $\mathcal{G}_1(\mathcal{X}) = \{ g_b : g_bx = x + b, \ x \in \mathcal{X}, \ b \in \mathcal{X} \}$, scale changes, $\mathcal{G}_2(\mathcal{X}) = \{ g_c : g_cx = cx, \ x \in \mathcal{X}, \ c > 0 \}$, and reflections, $\mathcal{G}_3(\mathcal{X}) = \{ g_d : g_dx = 2d - x \text{ (reflection of } x \text{ about } d) , \ x \in \mathcal{X}, \ d \in \mathcal{X} \}$. Also, for $\mathcal{X}$ normed, $\mathcal{G}_0(\mathcal{X})$ should include rotations (linear isometries), $\mathcal{G}_4(\mathcal{X}) = \{ g : g \text{ is a rotation of } \mathcal{X} \text{ onto itself: } \|gx\| = \|x\|, \ x \in \mathcal{X} \}$. Thus, for the case of normed $\mathcal{X}$, minimally $\mathcal{G}_0(\mathcal{X})$ should include the class

$$\mathcal{G}_0(\mathcal{X}) = \{ g_{c,A,b} : g_{c,A,b}x = cAx + b, \ b \in \mathcal{X}, \ c \in \mathbb{R}, \ \|Ay\| = \|y\| \text{ for } y \in \mathcal{X} \}. \quad (1)$$
On the basis of the above considerations, *key formal properties desired for depth functions $D$ and outlyingness functions $O$, in general for any $X$ and $P \in \mathcal{P}_0(X)$, are as follows.* Put $\text{chs}(P) = \text{ch}(\text{supp } P)$, the *convex hull of the support of $P$*. Note that, for the empirical distribution $\hat{P}_n$ based on data set $X_n$, $\text{chs}(\hat{P}_n) = \text{ch}(X_n)$.

**(M) Monotonicity Property:** $D(\cdot, P)$ is nonincreasing ($O(\cdot, P)$ nondecreasing) on rays outward from a point $\tilde{\theta}(P)$ of maximality (minimality) depending on $P$ and belonging to $\text{chs}(P)$. Thus $D(\cdot, P)$ has bounded upper contour sets, and $O(\cdot, P)$ bounded lower contour sets, star-shaped about $\tilde{\theta}(P)$.

**(I) Invariance Property:** $D(\cdot, P)$ and $O(\cdot, P)$ are $\mathcal{G}(\mathcal{X})$-invariant within $E_0$-equivalence, for a suitable $\mathcal{G}(\mathcal{X})$.

**(N) Nondegeneracy Property:** $P(D(X, P) > 0) = 1$, $X$ distributed as $P$.

**(Z) Zero Property:** $\inf_{x \in X} D(x, P) = 0 = \inf_{x \in X} O(x, P)$.

Property (M) fulfills (A1) and (A2). When the maximum depth region is a set of points, a rule may be introduced to select a single point $\tilde{\theta}(P)$ as “center”. Property (I) fulfills (A3), Property (N) fulfills (A4), and Property (Z) fulfills (A5). With the substitution of $\hat{P}_n$ for $P$, these properties become defined and meaningful for the data-based versions $D(\cdot, X_n)$ and $O(\cdot, X_n)$.

Note that (N) implies the weaker *nondegeneracy condition*,

$$(n) \sup_{x \in \mathcal{X}} D(x, P) > \inf_{x \in \mathcal{X}} D(x, P),$$

that appropriately disallows depth and outlyingness functions constant over $x \in \mathcal{X}$, such as the depth $D(x, P) \equiv 0$, $x \in \mathcal{X}$, which serves no useful purpose although it satisfies (M) (I), and (Z). However, (n) is somewhat too weak, allowing for example depths which are zero everywhere except at some point $x_0 \in \text{chs}(P)$, which satisfy (M), (I), and (Z) but have limited practical utility, only identifying a center. It also allows sample depths that satisfy $D(X_i, X_n) = 0$, $i = 1, \ldots, n$. Such degeneracies are excluded by (N). For perspective, however, let us note the following important fact applying Guideline C in Sect. 1.

**Remark 3.1.** Failure of a depth function $D(\cdot, P)$ to satisfy Property (N) in an infinite-dimensional setting does not imply failure also for the associated data-based depth $D(\cdot, \hat{P}_n) = D(\cdot, X_n)$ in its finite-dimensional setting.  

For some depths not fulfilling Property (M), suitable alternative properties can be formulated, as with the spatial depth treated in Sect. 4.3. However, Property (M) is the most straightforward and most appealing monotonicity property and thus preferred.
As a minimal standard for depth functions on any vector space \( X \), \( \mathcal{G}(X) \) in Property (I) should include \( \mathcal{G}_1, \mathcal{G}_3, \) and \( \mathcal{G}^{(0)} \) for normed \( X \). For specific types of \( X \), the preferred class \( \mathcal{G}_0(X) \) includes additional invariance conditions. In characterizing a particular depth function, the largest feasible class within the preferred class \( \mathcal{G}_0(X) \), if not \( \mathcal{G}_0(X) \) itself, should be taken as the relevant \( \mathcal{G}(X) \). This facilitates judicious evaluation and comparison of competitive depth functions.

On the basis of the preceding considerations, let us formally define depth and outlyingness functions as follows.

**Definition 3.1 (Depth and Outlyingness Functions).** Associated with a data space \( X \), depth and outlyingness functions are \( \mathcal{A} \)-measurable mappings \( D(\cdot, \cdot) : X \times \mathcal{P}_0(X) \to \mathbb{R} \) and \( O(\cdot, \cdot) : X \times \mathcal{P}_0(X) \to \mathbb{R} \), such that \( D(\cdot, P) \) and \( O(\cdot, P) \) are well-defined and fulfill Properties (M), (I), (N), and (Z) for all \( P \in \mathcal{P}_0(X) \). For a data set \( \mathbb{X} \) in \( X \), data-based versions \( D(\cdot, \mathbb{X}) \) and \( O(\cdot, \mathbb{X}) \) are similarly defined.

Many data-based depth functions satisfy \( D(x, \mathbb{X}_n) = 0 \) for \( x \) outside \( \text{ch}(\mathbb{X}) \). However, when \( D(\cdot, \mathbb{X}_n) \) is viewed as an estimator of \( D(\cdot, P) \), this need not be imposed as a strict requirement. In cases when the dimensionality \( d \) of \( X \) exceeds the sample size \( n \), it can happen for some depth functions that \( D(x, \mathbb{X}_n) = 0 \) even within \( \text{ch}(\mathbb{X}) \) except for \( x \) in a set of \( P \)-probability 0. Thus some data-based depths \( D(x, \mathbb{X}_n) \) can equal 0 almost everywhere \([P]\) for some \( P \), the exceptional points \( x \) perhaps including the point of maximality. For such depths, this failure of Property (N) must be avoided by arranging that \( d < n \), or by smoothing the given depth. See Guideline A in Sect. 1.

Since the centrality and outlyingness of a point \( x \) increase and decrease together, depth functions generate associated outlyingness functions, and vice versa. A connection is conveniently expressed in terms of normalized versions taking values over \([0, 1]\) or \([0, 1)\). Then, through the relationships \( O = 1 - D \) and \( D = 1 - O \), the \( D \) and \( O \) functions are equivalent as inverses of each other. (For sup \( D(x) = M \), the equivalent version \( M^{-1}D \) takes values over \([0, 1]\), as does \( M^{-1}O \) for sup \( O(x) = M < \infty \). For sup \( O(x) = \infty \), the equivalent version \( O/(1 + O) \) takes values over \([0, 1)\).) Despite formal equivalence, \( D \) and \( O \) functions have different roles in practice, each having its own body of specialized procedures and results apropos to applications. On the other hand, for simplicity or for special emphasis, theoretical discussions may involve just one of these instead of both together and leave implicit the implications regarding the other.

One might consider representing \( D(\cdot, P) \) in Def. 3.1 by just a single member from its \( E_0 \)-orbit, for example by normalizing the depth so that \( \text{sup}_x D(x, P) = 1 \). However, this obscures the \( E_0 \)-equivalence, which is a key conceptual and operational element of the \( \mathcal{G}(X) \)-invariance property. Further, for depth func-
tions that are $G(\mathcal{X})$-invariant strictly, without invoking $E_0$-equivalence, such normalization is not needed. In any case, normalizing to achieve $\sup_x D(x, P) = 1$ is an unnecessary complication and also requires re-normalizing $D$ after any transformation. Unproductive constraints are to be avoided except for use in alternative representations of the depth function for certain purposes, as above in expressing equivalences with outlyingness functions.

**Remark 3.2.** Continuous versus discrete $P$. It can happen that a depth notion fulfills Property (M) for $P$ continuous but not for certain discrete $P$. An example is the simplicial depth on $\mathbb{R}^d$, as pointed out in Zuo and Serfling (2000a). However, such a shortcoming is not necessarily a practical concern, since continuous models have discrete approximations and vice versa. In the nonparametric setting, an assumption of continuity of $P$ for convenience is not seriously restrictive (although nevertheless we do not impose this here).

**Remark 3.3.** Equivariance of deepest points. Multiplication by $c > 0$ changes the values of a depth function but not the point(s) of maximality. That is, $E_0$-equivalent depth functions share the same maximal points in $\mathcal{X}$. It thus follows from the $G(\mathcal{X})$-invariance within $E_0$-equivalence that “deepest” points transform to deepest points under transformations in $G(\mathcal{X})$. That is, if $\tilde{\theta}(P)$ denotes the unique maximal point of $D(\cdot, P)$, then the corresponding maximal point $\tilde{\theta}(Pg^{-1})$ of $D(\cdot, Pg^{-1})$ satisfies $\tilde{\theta}(Pg^{-1}) = g\tilde{\theta}(P)$, and likewise sample versions satisfy $\tilde{\theta}(gX) = g\tilde{\theta}(X)$, $g \in G(\mathcal{X})$. These equations express that the functional $\tilde{\theta}(P)$ and the statistic $\tilde{\theta}(X)$ are equivariant with respect to $g \in G(\mathcal{X})$. With respect to particular transformation classes, sample deepest points thus satisfy:

\[
\begin{align*}
\tilde{\theta}(X + b) &= \tilde{\theta}(X) + b, \quad b \in \mathcal{X} \quad \text{(translation equivariance, under } G_1) \\
\tilde{\theta}(cX) &= c\tilde{\theta}(X), \quad c > 0 \in \mathbb{R} \quad \text{(scale equivariance, under } G_2) \\
\tilde{\theta}(2d - X) &= 2d - \tilde{\theta}(X), \quad d \in \mathcal{X} \quad \text{(reflection equivariance, under } G_3) \\
\tilde{\theta}(aX + b) &= a\tilde{\theta}(X) + b, \quad a \in \mathbb{R}, \quad b \in \mathcal{X} \quad \text{(under } G_1, G_2, \text{ and } G_3 \text{ together).}
\end{align*}
\]

These are minimal requirements for $\tilde{\theta}(X)$ to be a location measure, analogous to univariate location measures such as the mean and median.

The simple “depth approach” represented by Def. 3.1 is applied in practice by developing conceptually some technique for constructing a “depth” $D(x, P)$ over $x$ in $\mathcal{X}$. Each choice of data space $\mathcal{X}$ presents its unique challenge, and, for each, quite different notions of “centrality” are possible, leading to different $D$ functions whose merits are evaluated with high priority on invariance and robustness criteria. Thus Def. 3.1 provides merely an admissibility criterion for depth functions, whose appeal on other grounds must also be evaluated and compared.
4 The Tukey (halfspace), projection, spatial, and contour depth functions

In the setting of general $\mathcal{X}$, we examine quite interesting and quite different depth functions, the Tukey (halfspace), projection, spatial, and contour depth functions. These are revisited in Sect. 5 in illustrating maximality at center properties. Appendix B provides all needed technical preliminaries.

4.1 The Tukey (halfspace) depth

It is straightforward to formulate the Tukey halfspace depth, which is well-established for $\mathcal{X} = \mathbb{R}^d$, in a quite general setting. Recall that $\mathcal{H}(\mathcal{X})$ denotes the class of closed halfspaces in $\mathcal{X}$.

**Definition 4.1 (Tukey (Halfspace) Depth).** Let $\mathcal{X}$ be a topological vector space. For a distribution $P$ on $\mathcal{X}$, the associated Tukey (or halfspace) depth function is defined as $D_T(x, P) = \inf_{H \in \mathcal{H}(\mathcal{X})} \{P(H) : x \in H\}$, $x \in \mathcal{X}$, for each $x$ the infimum probability under $P$ of halfspaces containing $x$.

One can replace “$x \in H$” by ‘$x \in \partial H$” in the above. Equivalently, $D_T(x, P)$ has the representations

$$D_T(x, P) = \inf_{u \in S^*} P(u(X - x) \geq 0), \ x \in \mathcal{X},$$

$$= \inf_{u \in S^*} P(H[x,u]), \ x \in \mathcal{X},$$

$$= \inf_{u \in S^*} \min\{P(u(X) \leq u(x)), P(u(X) \geq u(x))\}, \ x \in \mathcal{X},$$

where $H[x,u]$ denotes the halfspace $\{x' \in \mathcal{X} : u(x' - x) \geq 0\}$ with $x$ on its boundary, $X$ is a random element having distribution $P$, and “$S^*$” may be the dual $\mathcal{X}^*$ or its closed unit ball $B^*$ or simply its closed unit sphere $\{u : u \in \mathcal{X}^*, \|u\| = 1\}$. We may select “$S^*$” as convenient in any particular context.

In the univariate setting $\mathcal{X} = \mathbb{R}$, the Tukey depth is $D^{(1)}_T(x, P) = \min\{P(X \leq x), P(X \geq x)\}, \ x \in \mathbb{R}$, the minimum tail probability cut off by the point $x$, with deepest point the usual median. Thus we have the further representation

$$D_T(x, P) = \inf_{u \in S^*} D^{(1)}_T(u(x), P_u(X)), \ x \in \mathcal{X},$$

where $P_u(X)$ denotes the distribution of $u(X)$. In this connection, the following two lemmas are useful and easily proved.

**Lemma 4.1.** The univariate Tukey depth satisfies $D^{(1)}_T(x, P) = 0$ if and only if $x$ is a continuity point of $P$ and $x \notin (\text{supp } P)^\circ$.  


Lemma 4.2. For \( x \in \mathcal{X} \), \( u \in \mathcal{X}^* \), and \( \mathcal{X} \) having distribution \( P \) on \( \mathcal{X} \), we have \( x \in \text{supp} \ P \) if and only if \( u(x) \in \text{supp} \ P_{u(X)} \).

It is technically advantageous when the infimum over a set is attained by an element of the set. For a broad class of spaces \( \mathcal{X} \), and for \( P \) such that all hyperplanes have \( P \)-probability zero, the Tukey depth may be given as the probability of a particular “minimal” halfspace \( H[x, u_x] \), for some \( u_x \in \mathcal{X}^* \). Then positivity of \( D_T(x, P) \) rests explicitly upon that of \( P(H[x, u_x]) \) or the associated univariate Tukey depth \( D^{(1)}_T(u_x(x), P_{u_x(X)}) \). Such minimal halfspaces are determined by applying the weak-* sequential compactness of \( \mathcal{B}^* \) that holds for any separable normed space \( \mathcal{X} \) and any reflexive Banach space \( \mathcal{X} \) and thus, in particular, for all Hilbert spaces. Let us now develop this key structural result.

Formally, the properties to be invoked for \( P \) are:

(ZHP) Zero Hyperplane Probability: \( P(H[x, u]) = 0 \), \( u \in \mathcal{X}^*, x \in \mathcal{X} \).

(PHP) Positive Halfspace Probability: \( P(H[x, u]) > 0 \), \( u \in \mathcal{X}^*, x \in (\text{supp} \ P)^\circ \).

Proposition 4.1. Let \( \mathcal{X} \) be a separable normed space or a reflexive Banach space, and let \( P \) be a distribution on \( \mathcal{X} \).

(a) If \( P \) satisfies (ZHP), then for each \( x \in \mathcal{X} \) there exists \( u_x \in \mathcal{B}^* \) such that

\[ D_T(x, P) = P(H[x, u_x]). \tag{2} \]

(b) If, further, \( P \) satisfies (PHP), then \( D_T(x, P) > 0 \) for \( x \in (\text{supp} \ P)^\circ \).

The preceding result encompasses, for example, the spaces \( C(V) \) and, for \( p \neq \infty \), \( \ell_p, L_p, \) and \( \mathbb{W}^{k,p} \). By Lemma 4.2, the requirement \( x \in (\text{supp} \ P)^\circ \) in statement (b) above is equivalent to requiring \( u_x(x) \in (\text{supp} \ P_{u_x(X)})^\circ \). In the univariate case with \( \mathcal{X} = \mathbb{R} \), the above proposition reduces to Lemma 4.1.

The following proposition provides conditions under which the key properties desired for depth functions hold for the Tukey depth.

Proposition 4.2. Let \( \mathcal{X} \) be a topological vector space and let \( P \in \mathcal{P}_0(\mathcal{X}) \).

(a) The Tukey depth function fulfills Property (M).

(b) The Tukey depth function fulfills Property (I) with \( \mathcal{G}(\mathcal{X}) \) including \( \mathcal{G}_1 - \mathcal{G}_3 \), or including \( \mathcal{G}^{(0)} \) if \( \mathcal{X} \) is normed.

(c) For \( \mathcal{X} \) normed, the Tukey depth function fulfills Property (Z).

(d) For \( \mathcal{X} \) a separable normed or reflexive Banach space, and for continuous \( P \) satisfying (ZHP) and (PHP), the Tukey depth function fulfills Property (N).

We emphasize that Property (I) is merely a minimal invariance requirement. For particular spaces \( \mathcal{X} \), broader invariance is typically required. As an example, full affine invariance is desired when \( \mathcal{X} = \mathbb{R}^d \), and the Tukey depth indeed meets this more stringent invariance criterion.
It is also seen that for $X = \mathbb{R}^d$ the Tukey depth satisfies Property (N). However, in the infinite-dimensional setting, the great wealth of hyperplanes provided by the richness of $X^*$ permits the essential condition (PHP) to fail to hold for certain $P$ on certain $X$. Example 4.1 below illustrates this for a wide class of distributions $P$ on the sequence space $X = \ell_2$.

**Example 4.1.** Let $X = \ell_2$ with the usual norm and associated Borel $\sigma$-field. Let $P$ be the distribution of a random element $X = (X_1, X_2, \ldots)$ of $\ell_2$ with the $X_i$ independent, $EX_i \equiv 0$, $EX_i^2 = \sigma_i^2$, and $\sum_{i \geq 1} \sigma_i^2 < \infty$. Then, for $x = (x_1, x_2, \ldots) \neq (0, 0, \ldots)$, we have $D_T(x, P) \leq \left( \sum_{i \geq 1} x_i^2 / \sigma_i^2 \right)^{-1}$. If, further, either

(a) $\inf_{i \geq 1} i P(|X_i| > \varepsilon \sigma_i) > 0$ for some $\varepsilon > 0$, or

(b) $\sum_i E(X_i^4)/i^2 \sigma_i^4 < \infty$,

then $\sum_{i \geq 1} X_i^2 / \sigma_i^2 = \infty$ with $P$-probability 1, whence $P(D_T(X, P) = 0) = 1$ and so Property (N) fails completely for the Tukey depth.

Example 4.1 with assumption (b) is given by Dutta, Ghosh, and Chaudhuri (2011, Thm. 3), and here we add the alternative assumption (a). Similar examples can be exhibited for any separable Hilbert space, since such spaces are isometrically isomorphic to $\ell_2$. In particular, Chakraborty and Chaudhuri (2014b, Thms. 1 and 2) show failure of (N) for the Tukey depth in a further example with $P$ on $\ell_2$ and in an example with $P$ Gaussian on the separable space $C[0, 1]$.

**Remark 4.1.** Some analysis of Example 4.1. (a) First we note that, as shown in Serfling (2019), the distribution $P$ belongs to the class $\mathcal{P}_2(\ell_2)$ defined in Sect. 1.

(b) Consequently, on the basis of Guidelines B and C of Sect. 1, Example 4.1 is of technical interest but not necessarily applicable to data settings of practical interest for the depth approach. Rather, this model is for a single data point, a sequence in $\mathbb{R}^\infty$, instead of a data cloud.

(b) For independent Gaussian $X_i$ with $EX_i \equiv 0$ and $EX_i^2 = \sigma_i^2$, both assumptions (a) or (b) of Example 4.1 are satisfied, so that $\sum_{i \geq 1} X_i^2 / \sigma_i^2 = \infty$ with $P$-probability 1. It follows that the density of $(X_1, \ldots, X_m)$ is $\propto \exp\{- (1/2) \sum_i x_i^2 / \sigma_i^2 \}$, with limit 0 on a set of $P$-probability 1 as $m \to \infty$, which is consistent with the absence of a density for this $P$ and which shows that if there were a density it too would be zero with $P$-probability 1, compatibly with the Tukey depth being zero with $P$-probability 1.

(c) As seen from the proof of this example, the failure of Property (N) is because for each $x$ in a set of $P$-probability 1 a hyperplane through $x$ may be chosen with arbitrarily small probability content for one of the corresponding halfspaces. Thus the infimum defining the Tukey depth is taken over too great a wealth of possible hyperplanes, reflecting the richness of the dual space $\ell_2$ as well as the vulnerability of defining a depth as an infimum.

Thus issues regarding a depth on infinite-dimensional spaces do not establish the infeasibility of sample versions computed on a finite-dimensional surrogate. Failure
of Property (N) for $D_T(x, P)$ on certain infinite-dimensional $X$ is important as an insight but not a criterion for judging the sample version $D_T(x, X_n)$, which is evaluated on the basis of performance considerations. Of course, even the finite-dimensional case poses dimensionality issues, as in the requirement $d < n$ for the sample Tukey depth, which in some cases is problematic.

As seen in Sect. 5, Prop. 5.4, when $P$ satisfies a certain very broad notion of symmetry about some point $\theta_0$, as indeed holds for the Gaussian case of Example 4.1 with $\theta_0$ = the origin, then $D_T(\cdot, P)$ is maximal and positive at $\theta_0$. This is without restriction on $X$ and regardless of whether Property (N) is satisfied. Under such symmetry of $P$, the “Tukey median” and the center-locating feature of the Tukey depth thus have general validity.

4.2 The projection depth

Another depth well-established for $X = \mathbb{R}^d$ is the projection depth. In the general setting, relative to a distribution $P$ on $X$, the “projection depth” of $x \in X$ is defined, like the Tukey depth, in terms of the “projections” $u(x)$ in $\mathbb{R}$ for $u \in X^*$. However, instead of probabilities of halfspaces, one examines $u(x)$ relative to the univariate distribution $F_{u(x)}$ of a random projection $u(X)$, for $X$ having distribution $P$.

**Definition 4.2 (Projection Depth and Outlyingness).** Let $X$ be any vector space. For a distribution $P$ on $X$, and for given location and spread measures $\mu(F)$ and $\sigma(F)$ defined on univariate distributions $F$, the associated projection outlyingness function is defined as $O_{P}(x, P) = \sup_{u \in S^*} |u(x) - \mu(F_{u(x)})|/\sigma(F_{u(x)}), x \in X$, and the associated projection depth function as $D_P(x, P) = (1 + O_{P}(x, P))^{-1}, x \in X$, where “$S^*$” may be $X^*$ or its closed unit ball $B^*$ or simply its closed unit sphere $\{u : u \in X^*, \|u\| = 1\}$, chosen as convenient in any particular context.

For example, $\mu(F)$ and $\sigma(F)$ might be the mean and standard deviation of $F$, or, for more robust sample versions, the Median and MAD of $F$. For $X = \mathbb{R}$, the projection outlyingness function is $O_{P}^{(1)}(x, F) = |x - \mu(F)|/\sigma(F), x \in \mathbb{R}$, the classical scaled-deviation outlyingness function, yielding the representation

$$O_{P}(x, P) = \sup_{u \in S^*} O_{P}^{(1)}(u(x), F_{u(x)}), x \in X.$$  

**Proposition 4.3.** Let $X$ be a separable normed space or a reflexive Banach space, and let $P$ be a distribution on $X$.

(a) For each $x \in X$, there exists $u_x \in B^*$ such that

$$O_P(x, P) = \frac{|u_x(x) - \mu(F_{u_x(x)})|}{\sigma(F_{u_x(x)})}. \quad (3)$$

(b) If all induced univariate distributions $F_{u(x)}, u \in B^*$, are nondegenerate, then $D_P(x, P) > 0$, for $x \in X$. 

14
Proposition 4.4. Let $\mathcal{X}$ be any vector space and let $P \in \mathcal{P}_0(\mathcal{X})$.

(a) The projection outlyingness $O_P(x, P)$ is convex in the argument $x$.
(b) The projection depth function fulfills Property (M).
(c) The projection depth function fulfills Property (I) with $\mathcal{G}(\mathcal{X})$ including $\mathcal{G}_1-\mathcal{G}_3$, or including $\mathcal{G}^{(0)}$ if $\mathcal{X}$ is normed.
(d) For $\mathcal{X}$ normed, the projection depth function fulfills Property (Z).

Like the Tukey depth, the projection depth in some infinite-dimensional cases fails to satisfy Property (N).

Example 4.2 (Example 4.1 revisited). For $P$ as given in Example 4.1, the projection depth with mean and standard deviation as univariate location and spread measures does not satisfy Property (N).

Remarks similar to those in Remark 4.1 regarding Example 4.1 for the Tukey depth apply in kind for the projection depth. Also, incorporating the median as the univariate location measure, the projection depth deepest point is well-behaved for general $\mathcal{X}$. Under a broad symmetry assumption on $P$, as seen in Sect. 5, Prop. 5.4, $D_P(\cdot, P)$ satisfies a “maximality at center” property with maximal depth $= 1$, and the “projection median” is valid without restriction on $\mathcal{X}$ (at least with the median as univariate location measure).

4.3 The spatial depth

The spatial depth, well-established in Euclidean spaces as well as in functional data spaces, is easily formulated in the general setting of a normed vector space $\mathcal{X}$ and fulfills properties I, Z, and N without issues. Further, a “maximality at center” property is straightforward to show. However, instead of fulfilling property (M), the spatial depth satisfies an alternative notion of monotonicity along rays from the center. We study its relationship with (M), augmenting useful technical treatments of the spatial depth in $\mathbb{R}^d$ by Koltchinskii (1997) and in the general setting provided by Chakraborty and Chaudhuri (2014a,b) and Nagy (2017).

In developing our treatment, it is very convenient to first define the spatial rank function. This anticipates the general treatment of spatial-depth-related rank functions as popularized in Oja (2010), for example.

Definition 4.3 (Spatial Rank, Depth, and Outlyingness Functions). Let $\mathcal{X}$ be a normed vector space. For a distribution $P$ on $\mathcal{X}$ and $X$ a random element with distribution $P$, and with $S(\cdot)$ the sign function, the associated spatial rank function is defined as $R_S(x, P) = ES(x - X), x \in \mathcal{X}$, a “directional rank” giving for each point $x$ the expected direction from $X$, taking values in the unit ball $\mathfrak{B}(\mathcal{X})$. Its magnitude is the associated spatial outlyingness function, $O_S(x, P) = \|R_S(x, P)\|$, $x \in \mathcal{X}$, taking values in $[0, 1]$, and the associated spatial depth function is $D_S(x, P) = 1 - O_S(x, P) = 1 - \|R_S(x, P)\|$, $x \in \mathcal{X}$, likewise taking values in $[0, 1]$. \[\Box\]
For univariate $X$ having continuous distribution $F$, the spatial $R$, $O$, and $D$ functions are given by $2F(x) - 1$ (the classical “centered rank function”), $|2F(x) - 1|$, and $1 - |2F(x) - 1| = 2 \min \{ F(x), 1 - F(x) \}$, respectively.

The spatial deepest point is the solution $x = \theta_S$ of the equation $R_S(x, P) = 0$. It is evident, intuitively, that for $x$ deep within the high-probability region of $P$, the vectors $S(x - X)$ toward $x$ from random $X$ tend to balance each other, yielding $\| R_S(x, P) \| \approx 0$. On the other hand, for $x$ far outside the high-probability region, the directional vectors $S(x - X)$ tend to approximate $S(x - \tilde{\theta}_S)$ yielding $\| R_S(x, P) \| \approx \| S(x - \tilde{\theta}_S) \| = 1$. For $x$ between these extremes, one expects $\| R_S(x, P) \|$ to range from 0 to 1. Thus Properties (Z) and (N) should hold. Indeed, these along with Property (I) with suitable $G(\mathcal{X})$ are established in Prop. 4.5.

However, for the spatial depth, monotonicity along rays from the center holds not according to (M), which in this case would entail monotonicity of the magnitude of the rank function, but rather according to a monotonicity property of the rank function itself. The following definition is relevant.

**Definition 4.4** (Monotone operator on a vector space). Let $\mathcal{X}$ be a vector space. An operator $T : \mathcal{X} \rightarrow \mathcal{X}^*$ is called monotone if $(T(x) - T(y))(x - y) \geq 0$, for all $x, y \in \mathcal{X}$. \hfill $\Box$

For $\mathcal{X}$ an inner product space, with $u(x)$ given by $\langle u, x \rangle$ for $u \in \mathcal{X}^*$ and $x \in \mathcal{X}$, this monotonicity may be expressed as $$\langle T(x) - T(y), x - y \rangle = \| T(x) - T(y) \| \| x - y \| \cos \angle(T(x) - T(y), x - y) \geq 0,$$

or equivalently as $\cos \angle(T(x) - T(y), x - y) \geq 0$. This means that the vector $T(x) - T(y)$ points into the same halfspace as does $x - y$, relative to the hyperplane to which $x - y$ is normal.

For $\mathcal{X}$ a smooth normed space, consider the Gâteaux derivative of the norm, $\tilde{S}(x)(h) \in \mathcal{X}^*$ defined by Eqn. (B.1) of Appendix B. For distribution $P$ on $\mathcal{X}$ and $x \in \mathcal{X}$, define the related functional $$\tilde{R}(x, P)(h) = E\tilde{S}(x - X)(h), \ h \in \mathcal{X}.$$

For $\mathcal{X}$ a smooth inner product space, $\tilde{S}(x)$ and $\tilde{R}(x, P)$ are given by the usual sign function $S(x)$ and the spatial rank function $R_S(x, P)$, respectively.

**Lemma 4.3.** Let $\mathcal{X}$ be a smooth normed space.

(a) For $x \in \mathcal{X}$, $\tilde{S}(x)$ is a monotone operator.

(b) For $x \in \mathcal{X}$ and distribution $P$ on $\mathcal{X}$, $\tilde{R}(x, P)$ is a monotone operator.

(c) If also $\mathcal{X}$ is strictly convex and $P$ is nonatomic and not concentrated on a line, then $\tilde{R}(x, P)$ is strictly monotone, i.e., the inequality in Def. 4.4 is strict.

**Proposition 4.5.** Let $\mathcal{X}$ be an inner product space, and assume that it is smooth, strictly convex, and reflexive (which includes Hilbert spaces). Let $P$ be a distribution on $\mathcal{X}$ that is nonatomic and not concentrated on a line.
(a) There exists a unique maximal (minimal) spatial depth (outlyingness) point \( \tilde{\theta}_S \) as “center”.

(b) Putting \( x_\alpha = (1 - \alpha)\tilde{\theta}_S + \alpha x, \alpha \in (0, 1), \) and \( \Delta(x, P) = \angle(R(x, P), x - \tilde{\theta}_S), \) for \( x \in \mathcal{X} \), two monotonicities hold along rays outward from \( \tilde{\theta}_S \):

\[
\text{(SAM) Spatial Angle Monotonicity:} \\
\cos \Delta(x, P) \geq \cos \Delta(x_\alpha, P)
\]

\[
\text{(SRM) Spatial Rank Monotonicity:} \\
\langle R(x, P) - R(x_\alpha, P), x - \tilde{\theta}_S \rangle > 0.
\]

(c) The spatial depth and outlyingness functions fulfill Property (I) with \( \mathcal{G}(\mathcal{X}) \) including \( \mathcal{G}^{(0)} \) and without the proviso of \( E_0 \)-equivalence.

(d) The range of the rank function \( R(x, P) \) is the entire open unit ball in \( \mathcal{X}^* \), and Properties (N) and (Z) trivially follow.

It follows from Prop. 4.5 that the spatial depth is not degenerate in the setting of Example 4.1.

\textbf{Remark 4.2.} Let us interpret and clarify the two monotonicity properties stated in Prop. 4.5 for the spatial rank function \( R(\cdot, P) \). (a) Note from the proof that the two monotonicities actually hold for rays outward from any point \( \theta_0 \), not just the deepest point \( \tilde{\theta}_S \), which, however, is the case of primary interest.

(b) The spatial angle monotonicity (SAM) is simple and intuitive. As \( x \) moves outward along any ray from \( \tilde{\theta}_S \), the angle between the directional rank \( R(x, P) \) of \( x \) and the direction of the ray monotonically decreases to 0, bringing the directional rank of \( x \) into alignment with the direction of the ray and into coincidence with the vector \( S(x - \tilde{\theta}_S) \) having magnitude 1. Thus also follows that, along any ray, the outlyingness of \( x \) goes to 1 in the limit, and its depth to 0, although not necessarily monotonically as property (M) would require.

(c) The spatial rank monotonicity (SRM) characterizes the rank function \( R(x, P) \) as a “monotone operator”, in an equivalent form directly involving the ray from \( \tilde{\theta}_S \) to \( x \). This type of monotonicity is a well-known technical notion. For the case \( \mathcal{X} = \mathbb{R}^d \), a proof of (SRM) is given by Koltchinskii (1997) and detailed discussion and illustration is provided by Nagy (2017).

(d) Using a standard expression for the inner product, we readily obtain the following equivalent expressions for (SRM):

\[
|R(x, P)||x - \tilde{\theta}_S| \cos \Delta(x, P) \geq |R(x_\alpha, P)||x - \tilde{\theta}_S| \cos \Delta(x_\alpha, P), \\
|R(x, P)| \cos \Delta(x, P) \geq |R(x_\alpha, P)| \cos \Delta(x_\alpha, P), \\
O(x, P) \cos \Delta(x, P) \geq O(x_\alpha, P) \cos \Delta(x_\alpha, P).
\]

(e) Let us now compare (M) as given by \( O(x, P) \geq O(x_\alpha, P) \), (SAM) as given by Eqn. (4), and (SRM) as given by Eqn. (5). Although none of these imply
either of the others, they are closely interrelated: together, (M) and (SAM) imply (SRM). Indeed, the left and right hand sides of (SRM) are the products of the left and right hand sides of (M) and (SAM). No other useful implications hold. See also Nagy (2017) for useful comparison of (M) and (SRM) in the case $X = \mathbb{R}^d$.

(f) Thus together the monotonicities (SAM) and (SRM) offer an acceptable alternative to (M), which, however, is the more appealing when it holds. Also, for many depths, not only does (M) hold, but also the rank vector is defined so as to lie on the ray from the center, in which case (SAM) trivially holds, yielding (SRM) as well. Thus the monotonicity properties satisfied by the spatial depth are slightly weaker than those satisfied by typical other depths.

\[ \text{(7)} \]

### 4.4 Contour-generated depth functions

Does a given family of nested contours generate a depth function in a valid sense? Suppose that a family of nested and bounded contours in a vector space $X$ is given. Under some conditions on the family, an associated depth function can be generated. Let us formulate this precisely and then evaluate its appeal.

Let \( C = \{C_\beta, \; 0 \leq \beta < 1\} \) be a family of contours in \( X \) that

\begin{align*}
C_1 & \text{ are nested,} \\
C_2 & \text{ partition } X, \\
C_3 & \text{ are star-shaped about a point } x_0, \text{ with } C_0 = \{x_0\}, \text{ and} \\
C_4 & \text{ have enclosed regions increasing with } \beta: T_\beta \subseteq T_{\beta'}, \text{ for } \beta < \beta',
\end{align*}

where \( T_\beta \) denotes the closure of the region enclosed by \( C_\beta \). Relative to any \( P \in \mathcal{P}_0(X) \), the probability weight of \( T_\beta \) represents a measure of the common degree of outlyingness of the points in the contour \( C_\beta \). Therefore, associated with \( C \), a prospective “outlyingness” function is defined by \( O_C(x, P) = P(T_\beta) \) for \( x \in C_\beta, \; 0 \leq \beta < 1 \), and a corresponding prospective “depth” function by

\[
D_C(x, P) = 1 - O_C(x, P) = 1 - P(T_\beta) \text{ for } x \in C_\beta, \; 0 \leq \beta < 1.
\]

Note that \( D_C(\cdot, P) \) provides a meaningful labeling of the contours in \( C \), whether or not they already possess some labeling other than the index \( \beta \).

Let us now investigate how well \( D_C(\cdot, P) \) qualifies as a depth function. Properties \((\text{M}), (\text{N}), \text{ and } (\text{Z})\) of Def. 3.1 are immediately fulfilled. However, for Property \((\text{I})\) to hold for a given class of coordinate-change transformations \( \mathcal{G}(X) \), certain conditions on \( \mathcal{G}(X) \) are required. For \( g \in \mathcal{G}(X) \), define \( C^{(g)} = \{C^{(g)}_\beta, \; 0 \leq \beta < 1\} \), with \( C^{(g)}_\beta = \{gx : x \in C_\beta\} \), the contour in \( gX \) produced by applying \( g \) to \( C_\beta \). Let \( T^{(g)}_\beta \) denote the closed region enclosed by \( C^{(g)}_\beta \).

**Proposition 4.6.** Let a family of contours \( C \in X \) satisfy \( C_1-C_4 \). Then \( D_C(\cdot, P) \) defined by Eqn. (7) is a \( (\mathcal{G}(X), E_0) \) depth function in the sense of Def. 3.1 if, for \( g \in \mathcal{G}(X) \), the family of contours \( C^{(g)} \) satisfies \( C_1-C_4 \) with \( x_0 \) replaced by \( gx_0 \) and \( T_\beta \) by \( T^{(g)}_\beta \). Moreover, the \( \mathcal{G}(X) \)-invariance holds without invoking \( E_0 \)-equivalence.
It is easily checked that the requirements on $G(X)$ in Prop. 4.6 are readily met, in particular, by the transformation classes $G_1$-$G_4$. It is clearly appropriate to consider $D_C(\cdot, P)$ defined by Eqn. (7) a proper depth function, a natural construction given a family of contours. Let us call it a “contour depth”.

Nevertheless, the appeal of a “contour depth” is primarily technical. In practice, one starts with a particular notion of centrality and develops a depth function whose contours are of practical interest. Or, one derives a quantile function according to some notion of centrality, and this in turn generates an associated depth function (see Serfling, 2006, 2010). Given such contours, the probability weights and volumes of the enclosed “central regions” represent an additional artifacts having special roles (see Liu, Parelis, and Singh, 1999, for detailed treatment), but these roles are carried out without formally constructing the “contour depth”.

One might consider the contour depth generated by the family of contours of a unimodal density function. However, as pointed out in Serfling (2019), Sect. 2.7.2, the associated contour depth fails to be a nonparametric functional defined over all the $P$ allowed under Guideline C and thus is inadmissible as a depth function. In sum, “contour depths” carry limited practical appeal as depth functions, although they are implicitly in the background in the practical use of a depth function.

5 On maximality of a depth at a given center of symmetry

If a distribution $P$ is symmetric about a point $\theta \in X$ in some sense, it is desirable that any depth function $D(\cdot, P)$ be maximal at $\theta$ and thus orient to $\theta$ as “center”. It is also desirable that the sample deepest point have distribution symmetric about $\theta$ in the same sense. For breadth of application, one would like these properties to hold under as broad a possible notion of symmetry and with as minimal restrictions as possible imposed on $X$ or $D$. Let us examine this relative to three increasingly broad well-known nonparametric notions of symmetry: central symmetry\footnote{This includes elliptical symmetry for $X = \mathbb{R}^d$.} (for any vector space), angular symmetry (for any normed vector space), and halfspace symmetry (for any topological vector space, including the normed case).

**Definition 5.1 (Central Symmetry of $P$).** The distribution $P$ is centrally symmetric (C-symmetric) about $\theta \in X$ if $P(S) = P(S^{(\theta)})$ for any set $S \in \mathcal{A}$ and its reflection $S^{(\theta)}$ about $\theta$, or, equivalently, if $X - \theta$ and $\theta - X$ are identically distributed for $X$ having distribution $P$. \qed

For $P$ C-symmetric about a point $\theta_0$, most typical depth functions are maximal at $\theta_0$ deepest points have distributions C-symmetric about $\theta_0$. In particular, this holds for the depth functions $D_T(\cdot, P)$, $D_P(\cdot, P)$, and $D_S(\cdot, P)$.

**Proposition 5.1.** Let $P$ be C-symmetric about $\theta_0$ in a vector space $X$, and let $D(\cdot, P)$ be a depth function fulfilling Property (I) of Def. 3.1 with $G(X)$ containing...
Also, suppose that \( D(\cdot, P) \) has a unique deepest point \( \tilde{\theta}(P) \). Then the following hold.

(a) \( D(\cdot, P) \) is maximal at \( \theta_0; \tilde{\theta}(P) = \theta_0 \).

(b) For a sample \( X \), the distribution of the sample deepest point \( \tilde{\theta}(X) \) is \( C \)-symmetric about \( \theta_0 \).

Statement (b) of Prop. 5.1 reflects the well-known result in the univariate setting that, for a sample of observations with distribution symmetric about \( \theta \in \mathbb{R} \), the distribution of any odd translation statistic is symmetric about \( \theta \) (see Randles and Wolfe, 1979, Cor. 1.3.19).

**Definition 5.2 (Angular Symmetry of \( P \)).** A distribution \( P \) is angularly symmetric (A-symmetric) about \( \theta \) in a normed \( X \) if \( (X - \theta)/|X - \theta| \) and \( -(X - \theta)/|X - \theta| \) are identically distributed, for \( X \) having distribution \( P \).

The following result gives an important characterization of A-symmetry. Its proof follows lines of development in Zuo and Serfling (2000c) for \( X = \mathbb{R}^d \).

**Proposition 5.2.** Let \( P \) be a distribution on a normed vector space \( X \).

(a) If \( P \) is A-symmetric about \( \theta_0 \), then any hyperplane passing through \( \theta_0 \) divides \( X \) into two closed halfspaces \( H_1 \) and \( H_2 \) such that

\[
P(X \in H_1) = P(X \in H_2) \geq 1/2.
\]

(b) The converse holds.

(c) The probabilities in Eqn. (8) equal \( 1/2 \) if \( P \) is continuous.

Statement (a) of Prop. 5.2 quickly yields that the Tukey depth equals \( 1/2 \) at the center of A-symmetry. However, a more general result is given below in Prop. 5.4(a).

For the spatial depth, the maximality result in Prop. 5.1 holds more broadly, under A-symmetry, and, moreover, the uniqueness requirement for a deepest point is not needed.

**Proposition 5.3.** Let \( P \) be a distribution on a normed vector space \( X \). If \( P \) is A-symmetric about a point \( \theta_0 \) in \( X \), then \( D_\mathcal{S}(\cdot, P) \) is maximal at \( \theta_0 \).

Let us now examine a still more general notion of symmetry of \( P \). The following generalizes the definition given for \( X = \mathbb{R}^d \) by Zuo and Serfling (2000c).

**Definition 5.3 (Halfspace Symmetry of \( P \)).** Let \( X \) be a topological vector space. A distribution \( P \) on \( X \) is halfspace symmetric (H-symmetric) about \( \theta \in X \) if \( P(X \in H) \geq 1/2 \) for each closed halfspace \( H \) with \( \theta \) on the boundary.

Equivalently, \( P(X \in H) \geq 1/2 \) for any halfspace \( H \) containing \( \theta \). Equivalently, any hyperplane passing through \( \theta \) must divide \( X \) into two closed halfspaces each of which has probability at least \( 1/2 \). It then is immediate from Prop. 5.2 that A-symmetry implies H-symmetry. Although the converse is not true, it turns out that H-symmetry about a unique point \( \theta \) reduces to A-symmetry about \( \theta \) if \( P \) is
continuous or if \( P \) is discrete with null mass at \( \theta \). This is proved (details omitted) using similar lines of argument as in Zuo and Serfling (2000c, Thm. 2.6) for the case \( \mathcal{X} = \mathbb{R}^d \).

The breadth of H-symmetry makes it challenging to establish maximality at a unique point of H-symmetry under simple invariance conditions on the depth \( D \), as under C-symmetry in Prop. 5.1. However, for the Tukey depth, maximality at a unique point of H-symmetry can be proved readily by taking advantage of the fact that both \( D_T(\cdot, P) \) and H-symmetry are defined in terms of probabilities on halfspaces. This maximality is also straightforward to show for the projection depth. The following proposition expresses these results.

**Proposition 5.4.** Let \( P \) be a distribution on a topological vector space \( \mathcal{X} \). If \( P \) is H-symmetric about a unique point \( \theta_0 \) in \( \mathcal{X} \), then the following hold.

(a) \( D_T(\cdot, P) \) is maximal at \( \theta_0 \): \( \widehat{\theta}_T(P) = \theta_0 \), with \( D_T(\theta_0, P) = 1/2 \).

(b) With univariate location measure \( \mu(\cdot) \) given by the median "Med", \( D_P(\cdot, P) \) is maximal at \( \theta_0 \): \( \widehat{\theta}_P(P) = \theta_0 \), with \( D_P(\theta_0, P) = 1 \).

It is of interest to investigate extensions of Prop. 5.1(b) to A-symmetry or H-symmetry.

Note for the contour depth that, if the defining contours follow a given notion of symmetry about \( \theta \), then immediately the contour depth is maximal at \( \theta \), without imposing conditions. However, this is only of technical interest.

**Further comments on the Tukey, projection, and spatial depths.** As seen above, all three succeed quite well at the very basic task of center-locating in the case of a symmetric \( P \). This is valid for general \( \mathcal{X} \). (Let us note that not all popular depth functions succeed in this so well.)

However, regarding the defining properties that concern the broader role of a depth function, we have seen in Sect. 4 issues with (N) for the Tukey and projection depths for certain cases of infinite-dimensional \( \mathcal{X} \) and with (M) for the spatial depth for any \( \mathcal{X} \). Thus none of these dominates the others overall for general \( \mathcal{X} \).

Confining to finite-dimensional \( \mathcal{X} \), however, essentially to \( \mathbb{R}^d \), the Tukey and projection depths perform quite well regarding all properties (M), (I), (N), and (Z). Indeed, each satisfies not only (I) but also the full affine invariance appropriate for this choice of \( \mathcal{X} \). However, for \( \mathbb{R}^d \), the spatial depth fails to satisfy the desired full affine invariance. This is corrected by standardizing the argument within the depth function, as discussed in Serfling (2010), but then one obtains a different version that entails additional computational and robustness issues as a result of the standardization. Thus, for finite-dimensional \( \mathcal{X} \), the Tukey and projection depths dominate the spatial depth with respect to the defining and related properties. Of course, this is achieved at a greater computational cost.

For infinite-dimensional \( \mathcal{X} \) as with functional data, full affine invariance is not desired, even when a discretization to \( \mathbb{R}^d \) is applied (see Serfling and Wijesuriya, 2017, for discussion). Here the spatial depth not only meets the desired invariance requirements but also carries the added appeal of simplicity and computational ease. The full affine invariance of the Tukey and projection depths does not endow
them with added status in this case. Indeed, in the functional data setting, the straightforward formulations of the Tukey and projection depths do not succeed. This is not because of their anomalous behavior with respect to property (N), but rather due to lack of sensitivity to shape outliers. On the other hand, the spatial depth performs well in this setting (see Serfling and Wijesuriya, 2017). Thus, for infinite-dimensional $\mathcal{X}$, the spatial depth dominates the Tukey and projection depths with respect to the defining and related properties.

Of course, in practice other criteria besides the defining properties must be considered. We have already mentioned computational burden. Robustness with respect to outlier detection is also especially important. In this regard, the Tukey, projection, and spatial depths on $\mathbb{R}^d$ perform reasonably well in typical scenarios, but they differ in how they balance across masking and swamping robustness, none absolutely dominating the others (see Wang and Serfling, 2018).

6 Concluding remarks

Together, the general perspectives on the depth approach in Serfling (2019) and the results on formulation and maximality of depth functions in the present paper are intended to be instrumental to the continuing development of the depth approach. Let us summarize key points of interest. (i) The interplay between population and sample depth functions is clarified. (ii) Aspects of dimensionality are clarified. (iii) The interplay between sample size and dimension is clarified. (iv) Distributions that may be ignored for evaluation of population depths are characterized. (v) Inference aspects of the depth approach are elucidated. (vi) Axiomatic arguments for the defining properties of depth functions are provided. (vii) The notions of “density depth”, “local depth”, and “contour depth” are clarified. (viii) Further insights into the Tukey, projection, and spatial depths are obtained. Since these three now classical depths have straightforward formulations in quite general data settings, as well as incarnations in the setting of depth functions on parameter spaces (see Rousseeuw and Hubert, 1999, and Mizera, 2002, for use of the halfspace notion), their potential roles are broad and it is productive to deepen our understanding of them. (ix) It is seen that the central, angular, and halfspace nonparametric notions of symmetry, well-established in $\mathbb{R}^d$, have straightforward formulations in the case of general $\mathcal{X}$. (x) It is seen that the center-locating property for symmetric $P$ and the median-defining feature in general can hold for depth functions even when there are $\mathcal{X}$-related issues affecting the overall performance of the depth. This aspect of the depth approach has a status all its own.

Acknowledgements

The author gratefully acknowledges useful input from G. L. Thompson.
A Proofs

Technical background is provided in Appendix B.

Proof of Proposition 4.1. (a) Fix \( x \in \mathcal{X} \). Taking \( D_T(x, P) \) as the infimum over \( u \in \mathcal{B}^* \) of the probabilities \( P(H[x, u]) \), there exists some sequence \( \{\tilde{u}_n\} \in \mathcal{B}^* \) such that \( P(H[x, \tilde{u}_n]) \to D_T(x, P) \). Applying Lemma B.1(b,c), \( \mathcal{B}^* \) is weak-* sequentially compact under our assumptions on \( \mathcal{X} \), whereby the sequence \( \{\tilde{u}_n\} \) has a weak-* convergent subsequence \( \{u_n\} \) with limit \( u_x \in \mathcal{B}^* \). The corresponding halfspace \( H[x, u_x] \) satisfies Eqn. (2), as will now be shown. We have \( |P(H[x, u_n]) - P(H[x, u_x])| \leq P(\Delta_n) \), with \( \Delta_n = H[x, u_x] \Delta H[x, u_n] \), where \( A \Delta B \) denotes the symmetric difference of \( A \) and \( B \). It now suffices to show that \( P(\Delta_n) \to 0 \) as \( n \to \infty \). Write \( P(\Delta_n) = EI_{\Delta_n}(X) \). Check that \( I_{\Delta_n}(y) \equiv 0 \) for \( y \in H[x, u_x] \cap H[x, u_n] \) and also for \( y \in H[x, u_x]^c \cap H[x, u_n]^c \). Now \( \Delta_n = \Delta_n^{(1)} + \Delta_n^{(2)} \), where \( \Delta_n^{(1)} = H[x, u_x] \setminus H[x, u_n] \) and \( \Delta_n^{(2)} = H[x, u_n] \setminus H[x, u_x] \). For \( y \in \Delta_n^{(1)} \) we must have \( u_x(y) > u_x(x) \) and \( u_n(y) < u_n(x) \). However, the weak-* convergences \( u_n(y) \to u(y) \) and \( u_n(x) \to u(x) \) yield that \( u_n(y) > u_n(x) \) for all large \( n \), i.e., that \( y \in H[x, u_n] \) for all large \( n \), and thus \( I_{\Delta_n}(y) = 0 \) for all large \( n \). Likewise, for \( y \in \Delta_n^{(2)} \), a similar argument yields that again \( I_{\Delta_n}(y) = 0 \) for all large \( n \). It follows that for \( y \in \mathcal{X} \setminus H[x, u_x] \), a set of \( P \)-measure 1 under assumption (ZHP), we have \( I_{\Delta_n}(y) \to 0 \), \( n \to \infty \). Then the Lebesgue dominated convergence theorem yields \( P(\Delta_n) = EI_{\Delta_n}(X) \to 0 \), \( n \to \infty \), in turn yielding \( P(H[x, u_n]) \to P(H[x, u_x]) \), \( n \to \infty \), establishing Eqn. (2).

(b) This follows immediately from the assumed (PHP) property. \( \square \)

The above proof is inspired by the treatment of Massé (2004) for \( \mathbb{R}^d \).

Proof of Proposition 4.2. (a) For \( x \in \mathcal{X} \) and \( \alpha \in (0,1) \), we compare \( D_T(x, P) \) and \( D_T(\theta + \alpha(x - \theta), P) \), where for convenience \( \theta \) denotes the deepest point \( \theta_T(P) \).

Let \( H_x \) be a closed halfspace with \( x \) on the boundary and \( \theta + \alpha(x - \theta) \notin H_x^\circ \). Then, by the separating hyperplane theorem (Lemma B.2), there exists a closed hyperplane \( h_{\theta + \alpha(x - \theta)} \) with \( \theta + \alpha(x - \theta) \) on the boundary and with \( h_{\theta + \alpha(x - \theta)} \cap H_x^\circ = \emptyset \). Let \( H_{\theta + \alpha(x - \theta)} \) be the halfspace with \( h_{\theta + \alpha(x - \theta)} \) as boundary and containing \( x \). Then \( H_x \subset H_{\theta + \alpha(x - \theta)} \) and hence \( P(H_x) \leq P(H_{\theta + \alpha(x - \theta)}) \). It follows that \( D_T(x, P) \leq D_T(\theta + \alpha(x - \theta), P) \), \( \forall \alpha \in (0,1) \), establishing (M). (For \( \mathcal{X} = \mathbb{R}^d \), a similar proof is given in Zuo and Serfling, 2000a, Thm. 2.1.)

(b) Straightforward.

(c) Let \( X \) have distribution \( P \). For each \( x \) and \( X \), there exists a halfspace \( H_x \subset \{||X|| \geq ||x||\} \) with \( x \) on the boundary. Hence \( D_T(x, P) \leq P(||X|| \geq ||x||) \to 0 \) as \( ||x|| \to \infty \), establishing (Z).

(d) Immediate, for \( X \in (\text{supp } P)^{\circ} \) with \( P \)-probability 1 under the conditions. \( \square \)

Proof of Example 4.1. For \( \mathcal{X} = \ell_2 \), an inner product space with dual \( \mathcal{X}^* \) also = \( \ell_2 \), the Tukey depth over \( x \in \ell_2 \) may be represented as \( D_T(x, P) = \inf_{u \in \ell_2} P(\langle u, X \rangle \geq \langle u, x \rangle) \). Then for \( X \) having distribution \( P \). For \( x = (x_1, x_2, \ldots) \neq 0 = (0, 0, \ldots) \in \ell_2 \) and \( u = (u_1, u_2, \ldots) \in \ell_2 \) such that \( \langle u, x \rangle > 0 \),
Markov’s inequality yields
\[ P(\langle u, X \rangle \geq \langle u, x \rangle) \leq \frac{E(\langle u, X \rangle)^2}{\langle u, x \rangle^2} = \frac{\sum_{i \geq 1} u_i^2 \sigma_i^2}{\left( \sum_{i \geq 1} u_i x_i \right)^2}. \tag{A.1} \]

Define \( u_x(m) = (x_1/\sigma_1^2, \ldots, x_m/\sigma_m^2, 0, 0, \ldots) \in \ell_2 \), for each \( m \geq 1 \). Inserting \( u = u_x(m) \) into Eqn. (A.1) for \( m \) sufficiently large that \( x_m \neq 0 \) and hence \( \langle u_x(m), x \rangle > 0 \), we readily obtain \( P(\langle u_x(m), X \rangle \geq \langle u_x(m), x \rangle) \leq \left( \sum_{i=1}^m x_i^2/\sigma_i^2 \right)^{-1} \). Then \( D_T(x, P) \leq \lim_{m \to \infty} \left( \sum_{i=1}^m x_i^2/\sigma_i^2 \right)^{-1} = \left( \sum_{i=1}^\infty x_i^2/\sigma_i^2 \right)^{-1}. \) The remaining assertions of this example follow immediately through application of Lemma C.1 of Serfling (2019) with \( \{X_i/\sigma_i\} \) substituted for \( \{X_i\} \).

The above proof is a variation on that given for part (b) in [3, Thm. 3].

**Proof of Proposition 4.3.** Fix \( x \in \mathcal{X} \). Since \( O^*_T(x, P) \) is the supremum over \( u \in B^* \) of \( |\langle u(x) - \mu(F_u(X)) \rangle|/\sigma(F_u(x)) \), there exists a sequence \( \{\overline{u}_n\} \in B^* \) for which \( |\overline{u}_n(x) - \mu(F_{\overline{u}_n(X)})|/\sigma(F_{\overline{u}_n(x)}) \to O^*_P(x, P) \). Under the assumptions on \( \mathcal{X} \), it follows by Lemma B.1(b,c) that \( B^* \) is weak*-sequentially compact, whereby \( \{\overline{u}_n\} \) has a weak*-convergent subsequence \( \{u_n\} \) with limit \( u_x \in B^* \). This convergence also implies weak convergence of \( \{F_{\overline{u}_n(X)}\} \) and thus convergence of the moments of the bounded functions \( u_x(X) \). Thus follows Eqn. (3), whose right-hand side is finite when the assumption on \( P \) in (b) holds.

**Proof of Proposition 4.4.** (a) Note that, for any \( \theta \in \mathcal{X} \), \( x \in \mathcal{X} \), and \( \alpha \in (0, 1) \),
\[ |u(\theta + \alpha(x - \theta)) - \mu(F_u(X))| = |u(1 - \alpha \theta + \alpha x) - \mu(F_u(X))| = |(1 - \alpha)(u(\theta) - \mu(F_u(X))) + \alpha(u(x) - \mu(F_u(X)))| \leq (1 - \alpha)|u(\theta) - \mu(F_u(X))| + \alpha|u(x) - \mu(F_u(X))|. \]
This immediately yields the desired convexity, \( O_P(\theta + \alpha(x - \theta), P) \leq (1 - \alpha)O_P(\theta, P) + \alpha O_P(x, P) \).

(b) Now let \( \theta \) denote the deepest point \( \theta_P(P) \), at which \( O(, P) \) is minimal. For \( x \in \mathcal{X} \) and \( \alpha \in (0, 1) \), we compare \( O_P(x, P) \) and \( O_P(\theta + \alpha(x - \theta), P) \). An immediate consequence of (a) is that \( O_P(\theta + \alpha(x - \theta), P) \leq O_P(x, P) \). Thus \( O_P(, P) \) satisfies (M) and then so does \( D_P(x, P) \).

(c) Straightforward.

(d) Let \( x_0 \in \mathcal{X} \) and take \( x_\alpha = \alpha x_0 \), \( \alpha > 0 \). Choose any \( u \in \mathcal{X}^* \). Then \( \|x_\alpha\| = \|\alpha\|\|x_0\| \) and \( |\langle u(x_\alpha) \rangle| = \|\alpha\||\langle u(x_0) \rangle| \) both \( \to \infty \) as \( \alpha \to \infty \). It readily follows that \( \sup_{x \in \mathcal{X}} O_P(x, P) = \sup_{\alpha > 0} O_P(x_\alpha, P) = \infty \), yielding (Z) for \( D_P(, P) \).

**Proof of Example 4.2.** For \( x \in \mathcal{X} \), define \( u_x(m) \), \( m \geq 1 \), as in the proof of Example 5.1. Then it is readily checked that \( O_P(x, P) \geq \sup_{m \geq 1} |\langle u_x(m), x \rangle - \mu(F_{u_x(m), X})|/\sigma(F_{u_x(m), X}) = \sup_{m \geq 1} \left( \sum_{i=1}^m x_i^2/\sigma_i^2 \right)^{1/2} = \left( \sum_{i=1}^\infty x_i^2/\sigma_i^2 \right)^{1/2} \), which, as shown previously, equals \( \infty \) except for \( x \) in a set of \( P \)-measure 0. Thus (N) fails for \( D_P(, P) \) in this example.

**Proof of Lemma 4.3.** (a) Monotonicity of \( \tilde{S}(x) \) follows from the well-known monotonicity of the Gâteaux derivative of a convex function. Nevertheless, for convenience and instructiveness, we provide the proof for the case at hand, the convex function being the norm. We have \( \tilde{S}(x)(y - x) = \lim_{t \to 0} t^{-1}(|x + t(x - y)| - |x|) \leq \lim_{t \to 0} t^{-1}((1 - t)|x| + t|y| - |x|) = |y| - |x| \), from which follows \( \tilde{S}(x)(y - x) \geq |x| - |y| \).
Similarly, \( \tilde{S}(y)(x-y) \geq |y| - |x| \). Adding these relations, monotonicity of \( \tilde{S}(x) \) follows:

\[
(\tilde{S}(x) - \tilde{S}(y))(x-y) \geq 0.
\] (A.2)

(b) Now let \( X \) have distribution \( P \) and use Eqn. (A.2) with \( x \) and \( y \) replaced by \( x - X \) and \( y - X \), respectively. Then \( (\tilde{S}(x-X) - \tilde{S}(y-Y))(x-y) \geq 0 \). Taking expectations, we obtain \( (\bar{R}(x,P) - \bar{R}(y,P))(x-y) \geq 0 \), monotonicity of \( \bar{R}(x,P) \).

(c) Follows from Chakraborty and Chaudhuri (2014a, Thm. 3.1).

**Proof of Proposition 4.5.** (a) Using the strict convexity and reflexiveness of \( \mathcal{X} \) and that \( P \) is not concentrated on a line, this is proved in Kemperman (1987).

(b) \( (\text{SAM}) \). Let \( X \) have distribution \( P \). By straightforward geometry, \( \angle(S(x-X), x-\theta_S) \leq \angle(S(x_\alpha-X), x_\alpha-\theta_S) \). Taking expectations, we obtain \( \angle(R(x,P), x-\theta_S) \leq \angle(R(x_\alpha,P), x_\alpha-\theta_S) \), or \( \Delta(x,P) \leq \Delta(x_\alpha,P) \), or equivalently Eqn. (4). (This derivation uses only that \( \mathcal{X} \) is normed and not the restrictions on \( P \).)

(c) \( (\text{SRM}) \). By Lemma 4.3(c), applying the strict convexity and the condition on \( P \), we have \( (R(x,P) - R(x_\alpha,P), x-x_\alpha) > 0 \). Equivalently, via the relation \( x-x_\alpha = (1-\alpha)(x-\theta_S) \), we have Eqn. (5). (Using only that \( \mathcal{X} \) is normed and without restrictions on \( P \), Lemma 4.3(b) yields the same monotonicity, except not strictly.)

(d) Proved in Chakraborty and Chaudhuri (2014b, Thm. 3.1).

**Proof of Proposition 4.6.** The properties \( (\text{M}) \), \( (\text{N}) \), and \( (\text{Z}) \) follow immediately from the definition of \( D_C(\cdot,P) \). The formulation of \( D_C(\cdot,\cdot) \) on \( g\mathcal{X} \) is \( D_{C(g)}(gx, Pg^{-1}) \) defined by Eqn. (7) in terms of the probability weights on the regions \( T_\beta(g) \). Since the probability weights are the same for \( T_\beta \) and \( T_\beta^{(g)} \), we have \( D_{C(g)}(gx, Pg^{-1}) = D_C(x,P), x \in \mathcal{X} \), fulfilling Property \( (\text{I}) \) without needing to invoke \( E_0 \)-equivalence.

**Proof of Proposition 5.1.** (a) Consider the reflection transformation \( g_0 : x \mapsto 2\theta_0 - x \) in \( G_\beta \). By the C-symmetry of \( P \) about \( \theta_0 \), the distributions \( P \) and \( Pg_0^{-1} \) agree, in which case the depth functions \( D(\cdot,P) \) and \( D(\cdot, Pg_0^{-1}) \) are identical and hence the respective deepest points also are identical: \( \bar{\theta}(P) = \bar{\theta}(Pg_0^{-1}) \). By the equivariance of deepest points under \( G_\beta \) (Remark 4.3), \( \bar{\theta}(Pg_0^{-1}) = 2\theta_0 - \bar{\theta}(P) \). Thus follows \( \bar{\theta}(P) = \theta_0 \).

(b) Again applying the C-symmetry of \( P \) about \( \theta_0 \), along with the independence of the sample observations, \( X \) and \( 2\theta_0 - X \) are identically distributed and thus likewise \( \bar{\theta}(X) \) and \( \bar{\theta}(2\theta_0 - X) \). Therefore, by (a) and the reflection equivariance, \( \bar{\theta}(X) \) and \( 2\theta_0 - \bar{\theta}(X) \) are identically distributed and thus so are \( \bar{\theta}(X) - \theta_0 \) and \( \theta_0 - \bar{\theta}(X) \).

**Proof of Proposition 5.2 (a)** Let \( P \) be \( A \)-symmetric about \( \theta \). Then \( Y = (X - \theta)/(X - \theta) \) is \( C \)-symmetric about the origin 0 in \( \mathcal{X} \). Consider any closed halfspace \( H \) with 0 on the boundary. Then, immediately, \( P(Y \in H) = P(Y \in -H) \). Since the support of \( Y \) is the unit hypersphere \( S(\mathcal{X}) \), this is equivalent to \( P(Y \in H \cap S(\mathcal{X})) \).
= P(Y \in -H \cap S(\mathcal{X}))), which yields P(X - \theta \in H) = P(X - \theta \in -H) and in turn P(X - \theta \in H) = P(X - \theta \in -H) \geq 1/2.

(b) The implications in these steps hold in reverse. The details are omitted.

\text{Proof of Proposition 5.3} Let X have distribution P. Under A-symmetry of P about \theta_0, S(\theta_0 - X) and S(X - \theta_0) are equal in distribution. But S(X - \theta_0) = -S(\theta_0 - X). Hence, immediately, O_S(\theta_0, P) = -O_S(\theta_0, P), whence O_S(\theta_0, P) = 0 (minimality of O, maximality of D).

\text{Proof of Proposition 5.4} (a) By the definition of H-symmetry about \theta_0, P(H) \geq 1/2 for any closed halfspace H with \theta_0 on the boundary. Hence D_T(\theta_0, P) \geq 1/2.

Now suppose that D_T(x_0, P) > 1/2 for some point x_0 \neq \theta_0 \in \mathcal{X}. Then P(H) > 1/2 for any closed halfspace H with x_0 on the boundary, in which case P is also H-symmetric about x_0, contradicting the uniqueness assumption. Therefore, D_T(\theta_0, P) = \sup_x D_T(x, P).

(b) The H-symmetry of P about \theta_0 yields that the halfspaces \{u(X - \theta_0) \geq 0\} and \{u(X - \theta_0) \leq 0\} each have probability \geq 1/2, which implies Med(u(X)) = u(\theta_0), each u \in \mathcal{X}^*. Then O_P(\cdot, P) is given by O_P(x, P) = \sup_{u \in \mathcal{X}^*} |u(x) - u(\theta_0)|/\sigma(F_u(X)), x \in \mathcal{X}, which is minimal at x = \theta_0, with O_P(\theta_0, P) = 0 and thus D_P(\theta_0, P) = 1.

B Some basic vector space background

Let \mathcal{X} be a vector space. If equipped with a topology, \mathcal{X} is a \textit{topological} vector space.

B.1 Some notation, definitions, and key facts

\begin{itemize}
    \item For x, y \in \mathcal{X}, denote by \overrightarrow{xy} the line joining x and y, and by \angle(x, y) the angle between x and y.
    \item Denote by \mathcal{X}^* the \textit{dual space} of linear functions u : \mathcal{X} \mapsto \mathbb{R}. For \mathcal{X} \textit{normed} with norm \| \cdot \|, \mathcal{X}^* is defined to be the space of \textit{continuous} linear functions, and these are equivalently \textit{bounded}, and the usual norm on \mathcal{X}^* is given by

    \[ \|u\| = \sup_{\|x\| \leq 1} |u(x)| = \sup_{\|x\| = 1} |u(x)| = \sup_{x \neq 0} \frac{|u(x)|}{\|x\|}, \text{ for } u \in \mathcal{X}^*. \]

    For a \textit{topological} vector space, the dual is also defined to be the space of \textit{continuous} linear functions and is a Banach space (defined below). For \mathcal{X} \textit{finite-dimensional}, \mathcal{X}^* is \textit{linearly isomorphic} to \mathcal{X}.

    \item Some common spaces:
    \begin{itemize}
        \item \mathbb{R}^d \text{ Euclidean space of dimension } d \geq 1: \text{ vectors } \vec{x} = (x_1, \ldots, x_d), \text{ with norm } \|\vec{x}\| = \left( \sum_{i=1}^{d} x_i^2 \right)^{1/2}. \text{ The dual is } \mathbb{R}^d \text{ itself.}
        \item L_p(V, \mu) \text{ Space of } p\text{th power } \mu\text{-integrable functions on compact } V \subset \mathbb{R}^d, \text{ with norm } \|f\|_p = \left( \int |f(x)|^p d\mu(x) \right)^{1/p}, \text{ for } 1 \leq p < \infty. \text{ The dual is } L_q(V, \mu), \text{ where } 1/p + 1/q = 1.
    \end{itemize}
\end{itemize}
$L_{\infty}(V, \mu)$ Space of essentially bounded functions, with $\|f\|_{\infty} = \text{ess sup}_x |f(x)|$ with respect to measure $\mu$. However, the dual of $L_{\infty}(V, \mu)$ is not $L_1(V, \mu)$.

$\ell_p(\mathbb{R})$ Space of sequences $x = (x_1, x_2, \ldots)$ in $\mathbb{R}$, with $\|x\|_p = (\sum |x_i|^p)^{1/p} < \infty$ for $1 \leq p < \infty$, and sequences with $\|x\|_\infty = \text{sup} |x_i| < \infty$ for $p = \infty$. For $1 \leq p < \infty$ the dual is $\ell_q$, where $1/p + 1/q = 1$. (However, for $p = \infty$ it is not $\ell_1$.)

c₀ Space of sequences $x = (x_1, x_2, \ldots)$ in $\mathbb{R}$ with $\lim_{n\to\infty} x_n = 0$ and $\|x\| = \text{sup} |x_i|$. The dual is $\ell_1$.

$C(V)$ Space of continuous functions on compact $V \subset \mathbb{R}$, with norm $\|f(x)\| = \text{sup} |f(x)|$ or with inner product $\langle f, g \rangle = \int fg$. With the supremum norm, the dual is the space of finite signed Borel measures on $V$ equipped with its total variation norm. For the inner product norm, the dual may be identified with the space of Radon¹³ measures on $V$.

$C_b(V)$ Space of bounded and continuous functions on $V \subset \mathbb{R}$.

$C^\infty(V)$ Space of functions on $V \subset \mathbb{R}$ with continuous derivatives of all orders.

$W^{k,p}(V, \mu)$ Sobolev space of functions whose weak derivatives $f^{(j)}$ of order $j \leq k$ belong to $L_p(V, \mu)$, with $\|f\|_{k,p} = \left(\sum_{j=0}^{k} \int f^{(j)}(x)d\mu(x)\right)^{1/p}$. The dual is $W^{-k,d}(V, \mu)$, where $1/p + 1/q = 1$.

★ Some Important Types of Space:

**Metric.** Vector space equipped with a distance $d(x, y)$.

**Separable Metric.** Metric space with a countable dense subset.

**Banach.** Complete¹⁴ normed vector space, i.e., a complete metric space with the norm-induced metric. Expectations $E(X)$ of random elements are in the Bochner sense¹⁵ if $E\|X\| < \infty$, otherwise in the more broadly defined Pettis sense¹⁶.

**Inner Product.** Normed space with inner product norm $\|x\| = \sqrt{\langle x, x \rangle}$, a norm satisfying the parallelogram law. Also, $u(x)$ is given by $\langle u, x \rangle$ for $u \in X^*$ and $x \in X$. Also, $\cos \angle(x, y) = \langle x, y \rangle/\|x\|\|y\|$.

**Hilbert.** Complete inner product space. Equivalently, Banach space with norm determined by an inner product.

**Compact.** A space closed and bounded and, if infinite-dimensional, also flat: for every $\varepsilon > 0$, contained in the $\varepsilon$-neighborhood of a finite-dimensional space.

¹³A Radon measure on the $\sigma$-algebra of Borel sets of a Hausdorff topological space $X$ is a measure that is finite on all compact sets, outer regular on all Borel sets, and inner regular on open sets.

¹⁴Completeness: Every Cauchy sequence in $X$ converges in the norm to a limit in $X$.

¹⁵Bochner integral: $E(X) = \text{the limit of integrals of simple functions defined in straightforward fashion.}$

¹⁶Pettis integral: $E(X) = t$ such that the integral of $u(X)$ equals $u(t)$, all $u \in X^*$. 
Convex. Normed vector space satisfying: for $x \neq y$ satisfying $\|x\| = \|y\| = 1$, $\|(x + y)/2\| \leq 1$ (strictly convex if $< 1$).

Reflexive. Vector space with $(\mathcal{X}^*)^* = \mathcal{X}$.

Hausdorff. Topological space for which any distinct points can be separated by disjoint neighborhoods.

Locally Convex Space. Hausdorff space in which any neighborhood of the zero element contains a convex neighborhood of the zero element.

* For normed $\mathcal{X}$, define the sign function (or signum function) as $S(x) = x/\|x\|$ for $x \neq 0$ and $0$ for $x = 0$. If the Gâteaux derivative of $\|x\|$ exists, denote it by

$$
\tilde{S}(x)(h) = \lim_{t \to 0} \frac{\|x + th\| - \|x\|}{t}, \quad h \in \mathcal{X},
$$

which is a bounded linear functional of $h$ and thus an element of $\mathcal{X}^*$. For an inner product space, $\tilde{S}(\cdot)$ is given by $S(\cdot)$.

* For $\mathcal{X}$ a topological vector space and $x \in \mathcal{X}$ and $u \in \mathcal{X}^*$, the set $H[x, u] = \{x' \in \mathcal{X} : u(x' - x) \geq 0\}$ is a (closed) "halfspace" in $\mathcal{X}$ with $x$ on the boundary $h[x, u] = \{x' \in \mathcal{X} : u(x' - x) = 0\}$, which is a "hyperplane" in $\mathcal{X}$. Denote the class of closed halfspaces in $\mathcal{X}$ by $\mathcal{H}(\mathcal{X})$.

* For normed $\mathcal{X}$, denote by $\mathcal{B}$ the closed unit ball in $\mathcal{X}$: $\{x \in \mathcal{X} : \|x\| \leq 1\}$. Here "closure" is with respect to convergence in the norm on $\mathcal{X}$ ("strong closure"): for $x_n \in \mathcal{B}$ and $x \in \mathcal{X}$ with $\|x_n - x\| \to 0$, we have $x \in \mathcal{B}$. In comparison, "weak closure" corresponds to weak convergence\(^{17}\) in $\mathcal{X}$. For $\mathcal{X}$ finite-dimensional, these convergences are equivalent.

* For normed $\mathcal{X}$, denote by $\mathcal{B}^*$ the closed unit ball in $\mathcal{X}^*$: $\{u \in \mathcal{X}^* : \|u\| \leq 1\}$. Here "closure" is with respect to weak-$*$ convergence\(^{18}\) ("weak closure"): for weak-$*$ convergence of $u_n \in \mathcal{B}^*$ to $u \in \mathcal{X}^*$, we have $u \in \mathcal{B}^*$. (In comparison, "strong closure" corresponds to convergence in the norm on $\mathcal{X}^*$.)

* Conditions for $\mathcal{B}^*$ to be weak-$*$ compact and weak-$*$ sequentially compact are given in the following lemma.

Lemma B.1. Let $\mathcal{X}$ be a normed vector space.

(a) $\mathcal{B}^*$ is weak-$*$ compact (Alaoglu’s Theorem [14, p. 299]), but not compact with respect to the “strong” topology induced by the norm.

(b) If $\mathcal{X}$ is separable, then $\mathcal{B}^*$ is weak-$*$ sequentially compact\(^{19}\) (Helley’s Theorem [14, p. 283]).

(c) If $\mathcal{X}$ is a reflexive Banach space, then $\mathcal{B}^*$ is weak-$*$ sequentially compact (Corollary of Kakutani’s Theorem [14, p. 301]).

\(^{17}\)Weak convergence of $x_n \in \mathcal{X}$ to $x \in \mathcal{X}$ means: $u(x_n) \to u(x)$, $n \to \infty$, each $u \in \mathcal{X}^*$.

\(^{18}\)Weak-$*$ convergence of $u_n \in \mathcal{X}^*$ to $u \in \mathcal{X}^*$ means: $u_n(x) \to u(x)$, $n \to \infty$, each $x \in \mathcal{X}$.

\(^{19}\)Weak-$*$ sequential compactness: every sequence $\{u_n\}$ in $\mathcal{B}^*$ contains a subsequence $\{u_{n_k}\}$ which is weak-$*$ convergent to a limit $u \in \mathcal{B}^*$. 
The well-known separating hyperplane theorem can be proved in the setting of a general topological vector space and is often called the Geometric Hahn-Banach Theorem or the Hahn-Banach Separating Theorem. See, for example, Trèves (1995).

**Lemma B.2** (Hahn-Banach Separating Theorem). Let \( \mathcal{X} \) be a topological vector space. Let \( M \) be a vector subspace of \( \mathcal{X} \) and \( K \) a nonempty convex open subset of \( \mathcal{X} \) with \( K \cap M = \emptyset \). Then there exists a closed hyperplane \( N \) in \( \mathcal{X} \) containing \( M \) with \( K \cap N = \emptyset \).

**References**


