HALFSPACE DEPTH FOR DIRECTIONAL DATA: THEORY AND COMPUTATION

Stanislav Nagy and Rainer Dyckerhoff

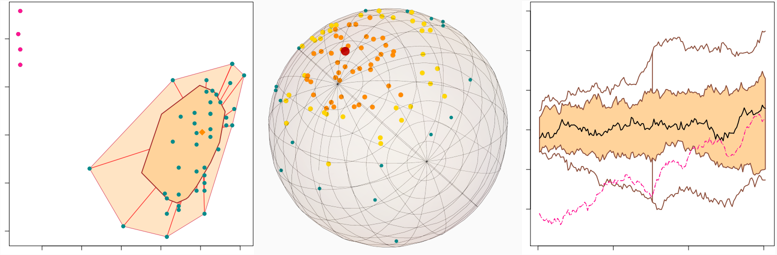
Fairfax 2024

Charles University, Prague, Czech Republic,
University of Cologne, Germany

Multivariate nonparametrics

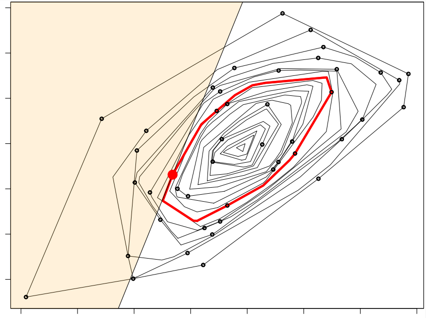
Nonparametric statistics:

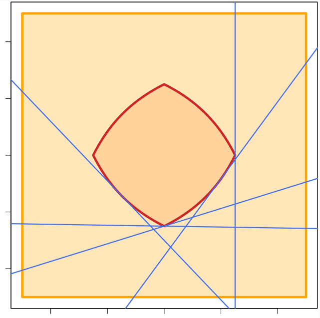
- Inference without assumptions on distributions.
- In $\mathbb{R}$ using the ordering — median, quantiles, ranks...
- What are ranks or quantiles for multivariate (non-Euclidean) data?



Statistical depth

Statistical depth function: Ordering data in multivariate spaces.



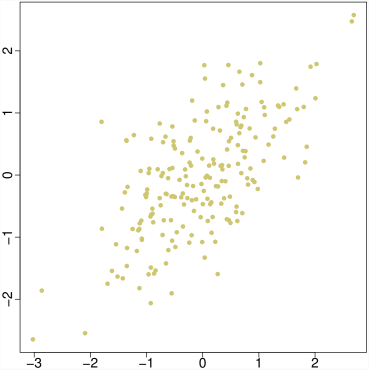


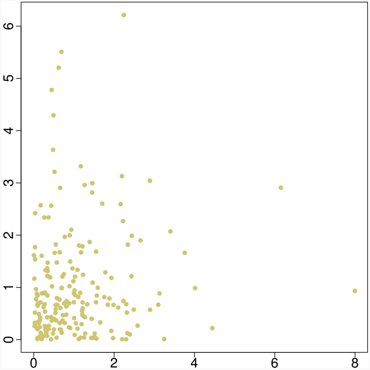
Introduced in 1975 (Tukey); studied intensively since the 1990s.

Statistical depth function

For $\mathcal{P}(\mathcal{S})$ Borel probability measures on $\mathcal{S}$, the depth is

$$D\colon \mathcal{S} \times \mathcal{P}(\mathcal{S}) \to [0,1]\colon (x, P) \mapsto D(x, P).$$

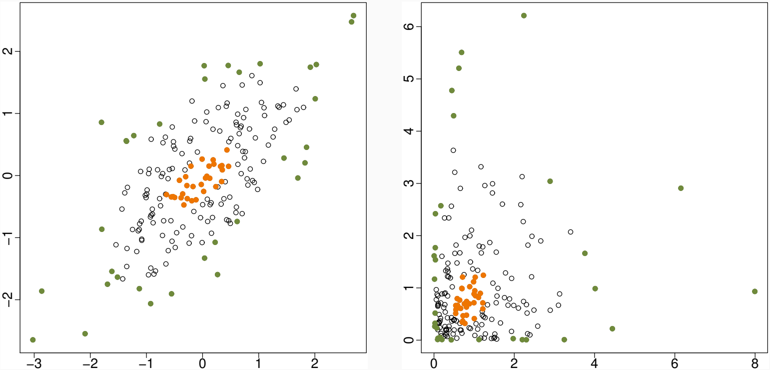




Statistical depth function

For $\mathcal{P}(\mathcal{S})$ Borel probability measures on $\mathcal{S}$, the depth is

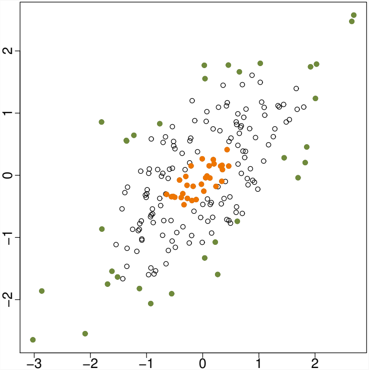
$$D\colon \mathcal{S} \times \mathcal{P}(\mathcal{S}) \to [0,1]\colon (x, P) \mapsto D(x, P).$$

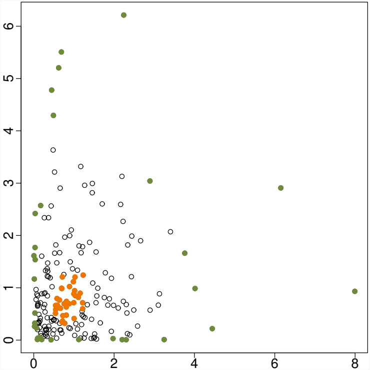


Halfspace depth

Halfspace depth (Tukey, 1975) of a point $x \in \mathbb{R}^d$ w.r.t. $P \in \mathcal{P}\left(\mathbb{R}^d\right)$

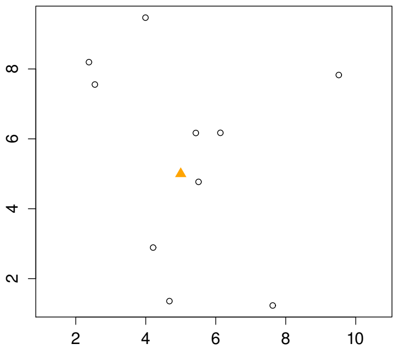
$$HD(x; P) = \inf \left\{P\left(H\right) : H \in \mathcal{H}(x)\right\}.$$





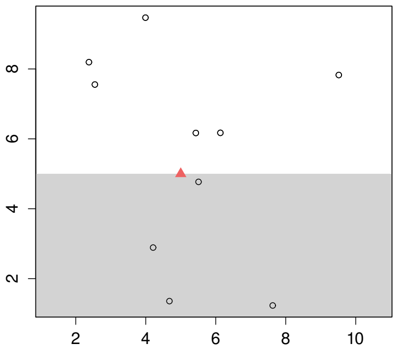
Halfspace depth

$$HD\left(x; P_n\right) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



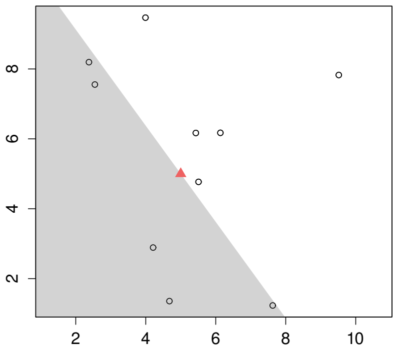
Halfspace depth

$$HD(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



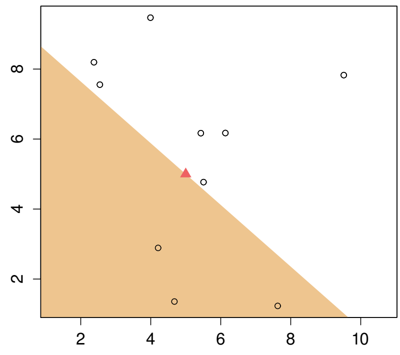
Halfspace depth

$$HD(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$

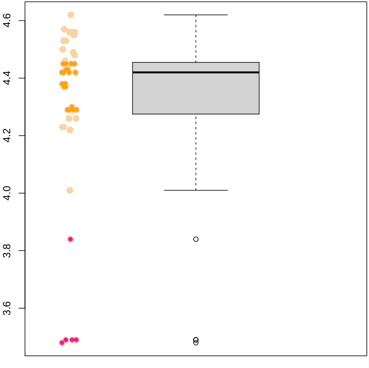


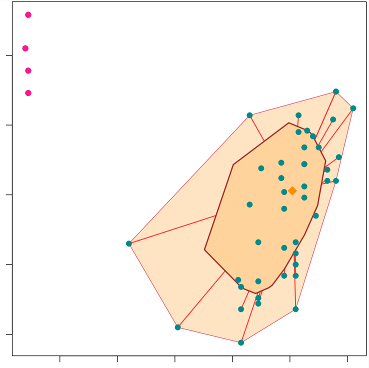
Halfspace depth

$$HD(x; P_n) = \min \frac{\#\text{ of observations in a halfspace that contains } x}{n}$$



Bagplot: A multivariate boxplot (Rousseeuw et al., 1999)





Depth in Euclidean spaces: Properties

Desirable properties of a depth D in $\mathbb{R}^d$:

(Liu, 1990; Zuo and Serfling, 2000; Serfling, 2006)

(L_1) *Affine invariance*

(L_2) *Maximality at center of symmetry for symmetric distributions*

(L_3) *Monotonicity along rays from the maximum point*

(L_4) *Vanishing at infinity*

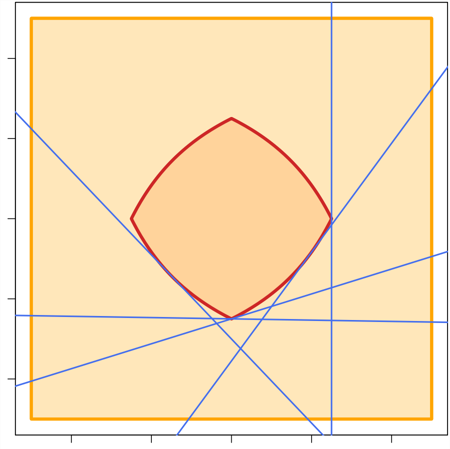
(L_5) *(Semi-)continuity*

(L_6) *Quasi-concavity*

Depth: Quasi-concavity

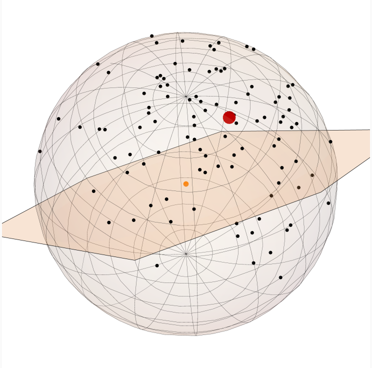
$HD(\cdot; P)$ is always quasi-concave, i.e. for each $\alpha \in [0,1]$

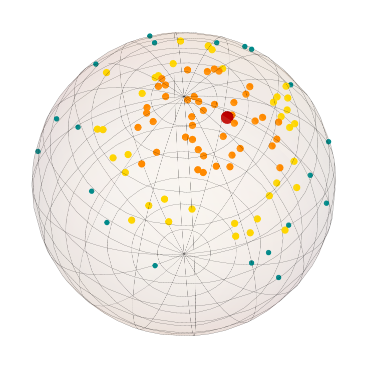
$$HD_\alpha(P) = \left\{ x \in \mathbb{R}^d \colon HD(x; P) \geq \alpha \right\} \text{ is convex}$$



Angular halfspace depth (Small, 1987) of $x \in \mathbb{S}^{d-1}$ w.r.t. $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$

$$aHD(x;P) = \inf\left\{P\left(H\right) : H \in \mathcal{H}(0), x \in H\right\}.$$





Depth for directional data: Affine invariance

Let

$$aD\colon \mathbb{S}^{d-1} \times \mathcal{P}\left(\mathbb{S}^{d-1}\right) \to [0,1]$$

be a general (angular) depth for directional data.

(L_1) *Affine invariance:* For all $x \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ non-singular, and $b \in \mathbb{R}^d$

$$D(x; P) = D(Ax + b; P_{AX+b}).$$

▼

(D_1) *Rotational invariance:* For all $x \in \mathbb{S}^{d-1}$ and any **orthogonal matrix** $O \in \mathbb{R}^{d\times d}$

$$aD(x; P) = aD(Ox; P_{OX}).$$

Maximality and monotonicity

(L_2) *Maximality at center:* If $\mu \in \mathbb{R}^d$ is a *center of symmetry* of P then

$$D(\mu; P) = \sup_{x \in \mathbb{R}^d} D(x; P); \qquad (\star)$$

(L_3) *Monotonicity along rays:* For all $x \in \mathbb{R}^d$ and $\alpha \in [0, 1]$,

$$D(x; P) \leq D(\mu + \alpha(x - \mu); P),$$

where $\mu \in \mathbb{R}^d$ is any point that satisfies ($\star$);

▼

(D_2) *Maximality at center:* For any P with *center of symmetry* at $\mu \in \mathbb{S}^{d-1}$

$$aD(\mu; P) = \sup_{x \in \mathbb{S}^{d-1}} aD(x; P); \qquad (\maltese)$$

(D_3) *Monotonicity along great circles:* For $x \in \mathbb{S}^{d-1} \setminus \{-\mu\}$ and $\alpha \in [0, 1]$

$$aD(x; P) \leq aD\left(\frac{\mu + \alpha(x - \mu)}{\|\mu + \alpha(x - \mu)\|}; P\right),$$

where $\mu \in \mathbb{S}^{d-1}$ is any point that satisfies ($\maltese$);

Maximality and monotonicity

(D_2) *Maximality at center:* For any P with *center of symmetry* at $\mu \in \mathbb{S}^{d-1}$

$$aD(\mu; P) = \sup_{x \in \mathbb{S}^{d-1}} aD(x; P).$$

What is a *center of symmetry* of $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$?

$P \in \mathbb{S}^{d-1}$ is **rotationally symmetric** around μ (Ley and Verdebout, 2017)
if $P \stackrel{d}{=} P_{OX}$ for all $O \in \mathbb{R}^{d \times d}$ orthogonal that fix $\mu \in \mathbb{S}^{d-1}$ (i.e., $O\mu = \mu$).

- μ is never unique ($-\mu$ is also a center of symmetry);
- What about the uniform distribution on $\mathbb{S}^{d-1}$?

Maximality and monotonicity

(D_2) *Maximality at center:* For any P with *center of symmetry* at $\mu \in \mathbb{S}^{d-1}$

$$\max\{aD(\mu; P), aD(-\mu; P)\} = \sup_{x \in \mathbb{S}^{d-1}} aD(x; P).$$

What is a *center of symmetry* of $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$?

$P \in \mathbb{S}^{d-1}$ is **rotationally symmetric** around μ (Ley and Verdebout, 2017)
if $P \overset{d}{=} P_{OX}$ for all $O \in \mathbb{R}^{d \times d}$ orthogonal that fix $\mu \in \mathbb{S}^{d-1}$ (i.e., $O\mu = \mu$).

- μ is never unique ($-\mu$ is also a center of symmetry);
- What about the uniform distribution on $\mathbb{S}^{d-1}$?

Vanishing at infinity and semi-continuity

(L_4) *Vanishing at infinity:* $\lim_{\|x\|\to\infty} D(x;P) = 0$;

(L_5) *Upper semi-continuity:* $D(\cdot;P)\colon \mathbb{R}^d \to [0,\infty)\colon x \mapsto D(x;P)$ is upper semi-continuous.

▼

(D_4) *Minimality at the anti-median:* $aD(-\mu;P) = \inf_{x\in\mathbb{S}^{d-1}} aD(x;P)$, for any $\mu \in \mathbb{S}^{d-1}$ that maximizes aD;

(D_5) *Upper semi-continuity:* $aD(\cdot;P)\colon \mathbb{S}^{d-1} \to [0,\infty)\colon x \mapsto aD(x;P)$ is upper semi-continuous.

Quasi-concavity

(L_6) *Quasi-concavity:* The central regions

$$D_\alpha(P) = \left\{ x \in \mathbb{R}^d \colon D(x; P) \geq \alpha \right\}$$

are convex for all $\alpha \geq 0$.

▼

(D_6) *Spherical quasi-concavity:* The central regions

$$aD_\alpha(P) = \left\{ x \in \mathbb{S}^{d-1} \colon aD(x; P) \geq \alpha \right\}$$

are **spherical convex** for all $\alpha \geq 0$.

Spherical convexity

(D_6) *Spherical quasi-concavity:* The central regions

$$aD_\alpha(P) = \left\{x \in \mathbb{S}^{d-1} \colon aD(x;P) \geq \alpha\right\}$$

are **spherical convex** for all $\alpha \geq 0$.

$A \subseteq \mathbb{S}^{d-1}$ is *spherical convex* (Besau and Werner, 2016) if its radial extension

$$\mathrm{rad}\,(A) = \left\{\lambda\, a \in \mathbb{R}^d \colon a \in A \text{ and } \lambda \geq 0\right\} \subseteq \mathbb{R}^d$$

is convex in $\mathbb{R}^d$.

- Any spherical convex set $A \neq \mathbb{S}^{d-1}$ is contained in a hemisphere.
- A depth satisfying (D_6) must be **constant on a hemisphere**.

Quasi-concavity of aHD

Theorem (Nagy and Laketa, 2024)
For any $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$ *and* $\alpha \geq 0$ *we have*

$$\begin{aligned} aHD_{\alpha}(P) &= \left\{x \in \mathbb{S}^{d-1} \colon aHD(x;P) \geq \alpha\right\} \\ &= \bigcap \left\{H \colon H^{c} \in \mathcal{H}(0) \text{ and } P(H^{c}) < \alpha\right\}. \end{aligned}$$

In particular, aHD satisfies (D_3), (D_5), and (D_6).

➥ *aHD* is **constant on the hemisphere of minimum *P*-mass**.

Depths for directional data: Properties

Reference	Angular Depth	Properties						
		(D_1)	(D_2)	(D_3)	(D_4)	(D_5)	(D_6)	
Ley et al. (2014)	Angular Mahalanobis[1]	✓	✓	✓	✓	✓	✗	
Pandolfo et al. (2018)	Cosine distance	✓	✓	✓	✓	✓	✗	
Pandolfo et al. (2018)	Arc distance	✓	✗	✗	✓	✓	✗	
Pandolfo et al. (2018)	Chord	✓	✗	✗	✗	✓	✗	
Liu and Singh (1992)	Angular simplicial	✓	✗	✗	✗	✓	✗	
Small (1987)	Angular halfspace	✓	✓	✓	✓	✓	✓	

[1] Defined only if the Fréchet median of $X \sim P$ is unique.

Depths for directional data: Properties

Reference	Angular Depth	Properties						
		(D_1)	(D_2)	(D_3)	(D_4)	(D_5)	(D_6)	
Ley et al. (2014)	Angular Mahalanobis[1]	✓	✓	✓	✓	✓	✗	
Pandolfo et al. (2018)	Cosine distance	✓	✓	✓	✓	✓	✗	
Pandolfo et al. (2018)	Arc distance	✓	✗	✗	✓	✓	✗	
Pandolfo et al. (2018)	Chord	✓	✗	✗	✗	✓	✗	
Liu and Singh (1992)	Angular simplicial	✓	✗	✗	✗	✓	✗	
Small (1987)	Angular halfspace	✓	✓	✓	✓	✓	✓	

[1] Defined only if the Fréchet median of $X \sim P$ is unique.

The ***angular Mahalanobis depth*** of $x \in \mathbb{S}^{d-1}$ w.r.t. $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$ is

$$aMD\,(x; P) = \mathrm{P}\,(\langle X, \mu\rangle \leq \langle x, \mu\rangle) = F_{\langle X,\mu\rangle}(\langle x, \mu\rangle),$$

for $\mu \in \mathbb{S}^{d-1}$ the Fréchet median of $X \sim P$.

Depths for directional data: Properties

Reference	Angular Depth	Properties						
		(D$_1$)	(D$_2$)	(D$_3$)	(D$_4$)	(D$_5$)	(D$_6$)	(D$_7$)
Ley et al. (2014)	Angular Mahalanobis[1]	✓	✓	✓	✓	✓	✗	✗
Pandolfo et al. (2018)	Cosine distance	✓	✓	✓	✓	✓	✗	✗
Pandolfo et al. (2018)	Arc distance	✓	✗	✗	✓	✓	✗	✓
Pandolfo et al. (2018)	Chord	✓	✗	✗	✗	✓	✗	✓
Liu and Singh (1992)	Angular simplicial	✓	✗	✗	✗	✓	✗	✓
Small (1987)	Angular halfspace	✓	✓	✓	✓	✓	✓	✓

[1] Defined only if the Fréchet median of $X \sim P$ is unique.

(D$_7$) *Non-rigidity of central regions:* There exists $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$ and $\alpha > 0$ so that $aD_\alpha(P)$ is not a spherical cap.

Angular halfspace depth: almost forgotten

Pandolfo, Paindaveine, Porzio (2018):

> *"The main drawback of the angular halfspace depth is the computational effort it requires, especially for higher dimensions."*

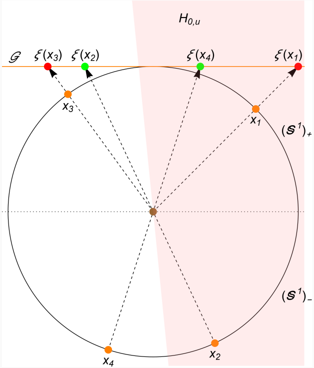
In *R*, function ***sdepth*** in package ***depth*** (Genest et al., 2019)

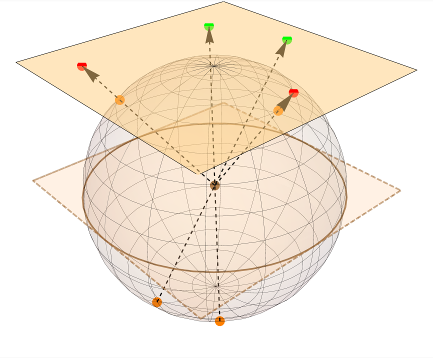
- works only for $d = 2, 3$,
- computing *aHD* for **a single point** $x \in \mathbb{S}^2$:
  - for sample size $n = 200$ takes 7 seconds,
  - for sample size $n = 500$ takes 2 minutes,
  - for sample size $n = 1000$ takes forever.

Computing aHD: Gnomonic projection

The gnomonic projection of $\mathbb{S}^{d-1}$ is the map

$$\xi\colon \left\{y \in \mathbb{S}^{d-1} \colon y_d \neq 0\right\} \to \mathcal{G} = \left\{y \in \mathbb{R}^d \colon y_d = 1\right\} : x \mapsto x/x_d$$





Computing aHD: Gnomonic projection

- ξ maps halfspaces in $\mathcal{H}(0)$ to halfspaces in $\mathcal{G}$.
- Denote $\mathbb{S}^{d-1}_{+} = \left\{x \in \mathbb{S}^{d-1} \colon x_d > 0\right\}$ and $\mathbb{S}^{d-1}_{-} = \left\{x \in \mathbb{S}^{d-1} \colon x_d < 0\right\}$.
- For simplicity, let $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$ satisfy the *smoothness condition*

$$P(\partial H) = 0 \quad \text{for all halfspaces } H \in \mathcal{H}(0). \tag{S}$$

Theorem (Nagy and Laketa, 2024)

For $P \in \mathbb{S}^{d-1}$ satisfying (S)*,*

$$aHD(x; P) = P\left(\mathbb{S}^{d-1}_{-}\right) + HD(\xi(x); P_{\pm}),$$

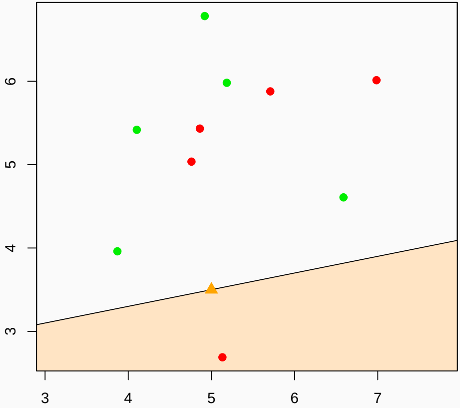
*where $P_{\pm}$ is a **signed measure** in $\mathcal{G}$ (or $\mathbb{R}^{d-1}$) given by*

$$P_{\pm}(B) = P\left(\left\{y \in \mathbb{S}^{d-1}_{+} \colon \xi(y) \in B\right\}\right) - P\left(\left\{y \in \mathbb{S}^{d-1}_{-} \colon \xi(y) \in B\right\}\right).$$

➥ Calculation of HD in $\mathbb{R}^{d-1}$ for signed measures.

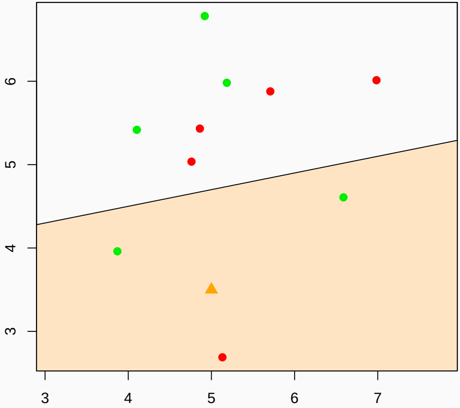
Halfspace depth for signed measures

$$HD(x; P_{\pm}) = \min \frac{(\#\text{ of red - \# of green points) in a halfspace } H \ni x}{n}$$



Halfspace depth for signed measures

$$HD(x; P_{\pm}) = \min \frac{(\text{\# of red - \# of green points) in a halfspace } H \ni x}{n}$$



Exact computation of aHD

➥ Adapt exact algorithms for *HD* (Dyckerhoff and Mozharovskyi, 2016).

Computational complexity for sample size n in $\mathbb{S}^{d-1}$

Algorithm	$aHD(y; P_n)$	$aHD(x_i; P_n), i = 1, \ldots, n$
C1	$\mathcal{O}(n^d)$	$\mathcal{O}(n^d)$
C2(s), **C3**	$\mathcal{O}(n^{d-1} \log n)$	$\mathcal{O}(n^d)$
C2(m)	$\mathcal{O}(n^{d-1} \log n)$	$\mathcal{O}(n^{d-1} \log n)$

Compare with $\mathcal{O}(n^{d-1} \log n)$ for $HD(x; P_n)$ of **a single point** in $\mathbb{R}^d$.

Comparison with sdepth, $d = 3$ (in seconds)

$d = 3$	single Point		$m = 1000$ points		all data points	
n	C2(m)	*sdepth*	C2(m)	*sdepth*	C2(m)	*sdepth*
40	0.00030	0.00030	0.00308	0.250	0.00024	0.0103
80	0.00075	0.00182	0.00697	1.88	0.00089	0.143
160	0.00292	0.0136	0.0165	13.6	0.00318	2.35
320	0.0110	0.111	0.0421	111	0.0120	35.6
640	0.0436	1.08	0.109	995	0.0477	637
1280	0.180	8.90	0.315	8690	0.197	11100
2560	0.786	72.1	1.11		0.797	
5120	3.18	583	3.83		3.47	
10240	12.7		14.1		14.0	
20480	61.3		55.7		67.1	
40960	239		244		278	
81920	1010		1020		1170	

Algorithm *sdepth* was implemented in *C++* to get a fair comparison.

Parallelized algorithm, higher d (in seconds)

d	Task	n =40	80	160	320	640	1280	2560	5120	10240	20480	40960	81920
3	single point	0.00008	0.00023	0.00049	0.00153	0.00510	0.0194	0.0747	0.286	1.14	4.61	19.1	85.6
3	all data points	0.00007	0.00024	0.00064	0.00190	0.00620	0.0238	0.0845	0.326	1.31	5.37	22.3	103
4	single point	0.00044	0.00325	0.0200	0.155	1.18	9.36	77.3					
4	all data points	0.00048	0.00370	0.0252	0.174	1.37	10.8	89.7					
5	single point	0.00486	0.0772	0.970	15.0	244							
5	all data points	0.00572	0.0783	1.11	17.9	284							
6	single point	0.0482	1.47	37.6									
6	all data points	0.0534	1.53	43.2									
7	single point	0.370	24.6										
7	all data points	0.386	23.6										
8	single point	2.07	292										
8	all data points	2.39	305										
9	single point	10.8											
9	all data points	12.1											
10	single point	54.0											
10	all data points	52.7											

Further properties: Qualitative robustness

For simplicity, let $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$ satisfy the smoothness condition

$$P(\partial H) = 0 \quad \text{for all halfspaces } H \in \mathcal{H}(0). \tag{S}$$

Theorem (Nagy and Laketa, 2024)
Under (S), for any $P_n \xrightarrow{w} P$ in $\mathcal{P}\left(\mathbb{S}^{d-1}\right)$

$$\lim_{n\to\infty} \sup_{x\in\mathbb{S}^{d-1}} \left| aHD\left(x; P_n\right) - aHD\left(x; P\right)\right| = 0.$$

The median map $\mathfrak{M}(P) = \left\{x \in \mathbb{S}^{d-1} \colon aHD\left(x; P\right) = \sup_{y\in\mathbb{S}^{d-1}} aHD(y; P)\right\}$

- *is always outer semi-continuous, and*
- *is continuous (in Hausdorff distance) if $\mathfrak{M}(P) = \{x_0\}$ is a singleton.*

➥ For P_n empirical measures from P (i.e., random samples from P), all this holds also without (S).

Further properties: Qualitative robustness

For simplicity, let $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$ satisfy the smoothness condition

$$P(\partial H) = 0 \quad \text{for all halfspaces } H \in \mathcal{H}(0). \tag{S}$$

Theorem (Nagy and Laketa, 2024)
Under (S), for any $P_n \xrightarrow{w} P$ in $\mathcal{P}\left(\mathbb{S}^{d-1}\right)$

$$\lim_{n\to\infty} \sup_{x \in \mathbb{S}^{d-1}} \left| aHD\,(x; P_n) - aHD\,(x; P) \right| = 0.$$

The median map $\mathfrak{M}(P) = \left\{x \in \mathbb{S}^{d-1} \colon aHD\,(x; P) = \sup_{y \in \mathbb{S}^{d-1}} aHD(y; P)\right\}$

- *is always outer semi-continuous, and*
- *is continuous (in Hausdorff distance) if $\mathfrak{M}(P) = \{x_0\}$ is a singleton.*

➥ The sample *aHD* and its medians are always uniformly consistent.

Further properties: Central regions

The central region

$$aHD_\alpha(P) = \left\{x \in \mathbb{S}^{d-1} \colon aHD(x;P) \geq \alpha\right\}.$$

Theorem (Nagy and Laketa, 2024)
Under (S) *and a further mild condition, for any* $P_n \xrightarrow{w} P$ *in* $\mathcal{P}\left(\mathbb{S}^{d-1}\right)$ *and* $A \subset (\min_{x\in\mathbb{S}^{d-1}} aHD(x;P), \max_{x\in\mathbb{S}^{d-1}} aHD(x;P))$ *closed we have*

$$\lim_{n\to\infty} \sup_{\alpha\in A} \delta_H\left(aHD_\alpha(P_n), aHD_\alpha(P)\right) = 0,$$

where δ_H *is the Hausdorff distance.*

➥ The sample *aHD*-central regions are usually uniformly consistent.

Conclusion: Angular halfspace depth

Advantages:

- ranks directional data in $\mathbb{S}^{d-1}$,
- has the nice properties of *HD* from $\mathbb{R}^d$ (as the only one in $\mathbb{S}^{d-1}$),
- is **very fast** to compute in lower dimensions.

➥ Field is open for applications to directional data analysis.

Disadvantages:

- Constancy on a hemisphere of lowest probability.
  - ✪ bagdistance (Hubert, Rousseeuw, Segaert, 2017),
  - ✪ illumination (Nagy and Dvořák, 2021).
- Slow exact computation with higher *d*.
  - ✪ approximate algorithms (Dyckerhoff, Mozharovskyi, Nagy, 2021).

Selected literature

[1] R. Dyckerhoff and S. Nagy. Exact computation of angular halfspace depth. *Under review*, 2024.

[2] Christophe Ley, Camille Sabbah, and Thomas Verdebout. A new concept of quantiles for directional data and the angular Mahalanobis depth. *Electron. J. Stat.*, 8(1):795–816, 2014.

[3] Christophe Ley and Thomas Verdebout. *Modern directional statistics*. Chapman & Hall/CRC Interdisciplinary Statistics Series. CRC Press, Boca Raton, FL, 2017.

[4] Regina Y. Liu and Kesar Singh. Ordering directional data: concepts of data depth on circles and spheres. *Ann. Statist.*, 20(3):1468–1484, 1992.

[5] S. Nagy, H. Demni, D. Buttarazzi, and G. C. Porzio. Theory of angular depth for classification of directional data. *Adv. Data Anal. Classif.*, 2024. To appear.

[6] S. Nagy and P. Laketa. Theoretical properties of angular halfspace depth. *Bernoulli*, 2024. To appear.

[7] Giuseppe Pandolfo, Davy Paindaveine, and Giovanni C. Porzio. Distance-based depths for directional data. *Canad. J. Statist.*, 46(4):593–609, 2018.

[8] Christopher G. Small. Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.*, 15(1):31–39, 1987.

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