Polar depth for Potentially Heavy-tailed data
https://arxiv.org/abs/2606.00343

work in progress

Stephan Clémençon [1] , Carlos Fernándes [1], Pavlo Mozharovskyi [1],
Anne Sabourin [2]

[1] LTCI, Télécom Paris, Institut polytechnique de Paris.
stephan.clemencon@telecom-paris.fr,carlos_ds730@yahoo.es,pavlo.mozharovskyi@telecom-paris.fr

[2] Université Paris Cité, Université Paris Saclay, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-75006, Paris, France anne.sabourin@u-paris.fr

JDS, Clermont Ferrand, June 2026

Outline

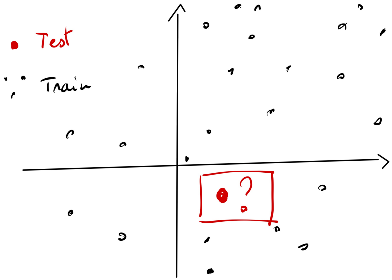
Data Depth

- Dataset $X_i \overset{i.i.d.}{\sim} P, i \leq n$, in $\mathcal{X} = \mathbb{R}^d$.

- **Goal**: quantify abnormality of point x with respect to other points.
  $\longrightarrow$ Anomaly Detection, Multivariate "quantile" contour estimation

- **Statistical depth function**: A function $D(\cdot \mid P) \to \mathbb{R}^+$: the lower, the more shallow is the point and the more abnormal.

- **Empirical (non-parametric) estimator** $D(\cdot \mid P_n)$

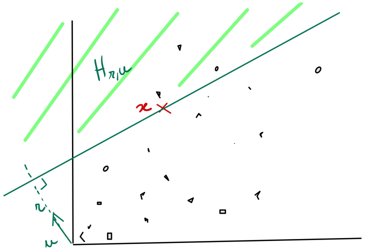


Tukey depth

- Seminal work Tukey (1975), "half-space depth"

$$D(x \mid P) = \inf\{P(H_{r,u}) : x \in H_{r,u}, r \geq 0, u \in \mathbb{S}(\text{unit sphere}).\}$$

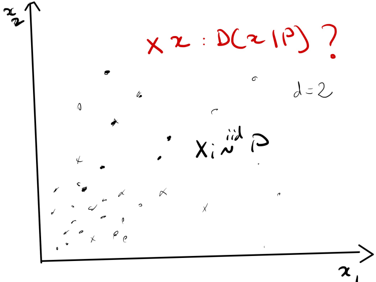
where $H_{r,u} = \{x \in \mathbb{R}^d : \langle x, u \rangle \geq r\}$.



- Applications - computation strategies - alternative depths, *e.g.* Li et al. (2012); Mozharovskyi et al. (2015) (classification), or Mozharovskyi and Valla (2025); Staerman et al. (2023) (Anomaly detection), simplicial depth (Liu, 1990), projection depth (Liu, 1992), majority Monge-Kantorovich depth (Chernozhukov et al., 2017), **Angular depth (Nagy and Laketa, 2025) for directional data** . . .

Extrapolation for Tukey Depth

$D(x|P)$ outside the point cloud, when $D(\cdot\ |P_n) = 0$?



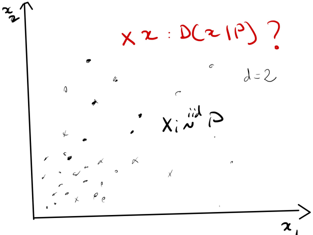
- Einmahl et al. (2015); He and Einmahl (2017): Regular Variation assumptions $b(t)\mathbb{P}(X \in tA) \to \nu(A)$ with $b(tx)/b(t) \to x^{-\alpha}$, even $b(t) = t^{\alpha}$ for the theory.

- **Key result/tool** $D(x|P) \approx D(x|\nu)$ up to rescaling, for large $\|x\|$.

  $\longrightarrow$ nonparametric estimation of ν

Why a new notion of Depth?

$D(x|P)$ outside the point cloud, when $D(\cdot\ |P_n) = 0$?



- **Half-spaces not invariant under rescaling : incompatible with structure of multivariate extremes**

- What if some coordinates are non-negative (typical for physical quantities: precipitations, river discharge, snow depth . . .)?
  $\longrightarrow$ **0 should not be treated as “shallow” even though always on boudnary of point cloud**

- Coordinates with different tail indices: **How to incorporate marginal standardization?**

Outline

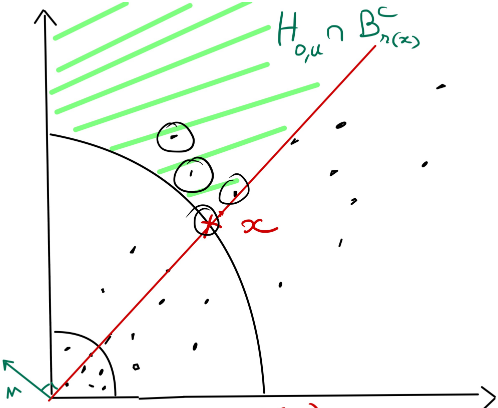
Polar Depth

$\|\cdot\|$ the Euclidean norm; $r(x) = \|x\|$, $\theta(x) = r(x)^{-1}x$.

Definition: polar depth

For π a positive measure on $\mathbb{R}^d$, $x \in \mathbb{R}^d$

$$pD(x \mid P) = \inf_{\|u\|=1,\, x \in H_{0,u}} \pi(B^c_{r(x)} \cap H_{0,u})$$



Key properties of Polar Depth

Lemma: connexion with angular depth "aD" (Dyckerhoff and Nagy, 2023; Nagy and Laketa, 2025)

Let $P_{\theta|t}$ denote conditional distribution of $\theta(X)$ given $r(X) \geq t$; and $P_r(\cdot) = \mathbb{P}(r(X) \in \cdot)$. Then

$$pD(x \mid P) = P_r(r(x), \infty)\ aD(\theta(x) \mid P_{\theta \mid r(x)})$$

Depth Axiomatic properties (see Zuo and Serfling (2000) for desirable axiomatic properties of depth)

- !! No affine invariance: $pD(Ax + B \mid (A \cdot + B)_{\#}P) \neq pD(x \mid P)$.
- Rotational invariance, monotonicity along rays passing through 0, vanishing at infinity, upper semicontinuity w.r.t x, continuity in measure.

Outline

Empirical Polar Depth

1. Empirical radial depth (survival function)

$$\widehat{P}_r([r(x),\infty)) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{r(X_i) \geq r(x)\}.$$

2. Angular outwards dataset $\mathcal{D}_\theta = \{\theta(X_i) : \ r(X_i) \geq r(x) \text{ for } 1 \leq i \leq n\}$
   $\rightarrow$ Empirical measure $\widehat{P}_{\theta|r(x)}$ on $\mathcal{D}_\theta \subset \mathbb{S}$.

3. Angular half-space depth of $\theta(x)$ relative to $\widehat{P}_{\theta|r(x)}$:

$$\widehat{aD}(\theta(x) \mid P_{\theta|r(x)}) := aD(\theta(x)|\widehat{P}_{\theta|r(x)}).$$

   (Algorithm developed in Dyckerhoff and Nagy (2025) or gnomonic projection in Nagy et al. (2023) combined with half-space depth computation in $\mathbb{R}^{d-1}$ Dyckerhoff and Mozharovskyi (2016))

Polar Depth estimator

$$\widehat{pD}(x \mid P) \ = \ \widehat{P}_r\Big([r(x),\infty)\Big)\ \widehat{aD}(\theta(x) \mid P_{\theta|r(x)}).$$

Finite sample bounds: absolute and relative deviations

Theorem: Non-asymptotic bounds for $\widehat{pD}$

(some conditions on $\delta \leq$ reasonable fixed threshold and any fixed $n \geq 4$)

1. With probabiblity at least $1 - \delta$,

$$\sup_{x \in \mathbb{R}^d \setminus \{0\}} \left| \widehat{pD}(x \mid P) - pD(x \mid P) \right| \leq \sqrt{\frac{d \log(n/\delta)}{n}}.$$

2. For $t > 0$, with $p_t = \mathbb{P}(r(X) \geq t)$,

$$p_t^{-1} \sup_{x:\, r(x) \geq t} |\widehat{pD}(x \mid P) - pD(x \mid P)| \leq 3\sqrt{\frac{d \log((np_t + 1)/\delta)}{np_t}} + \frac{2}{3}\frac{\log(1/\delta)}{np_t}.$$

Main proof idea: Empirical Risk Minimization problem over a VC family of sets of dimension $\approx d$; VC inequality for relative deviation in Goix et al. (2015); Lhaut et al. (2022)

Outline

Regular Variation Assumptions

- Minimal assumption

$$p_t^{-1}\mathbb{P}(X \in tA) \xrightarrow[t\to\infty]{} \nu_1(A), \qquad \text{with} \qquad p_t = \mathbb{P}(r(X) > t), \tag{1}$$

for all Borel sets A such that $0 \notin \partial A$ and $\nu_1(\partial A) = 0$.

ν_1: limit probability measure "of extremes" on $\mathbb{B}^c$

- Stronger assumption (means $t^\alpha p_t \to$ constant)

$$t^\alpha \mathbb{P}(X \in tA) \xrightarrow[t\to\infty]{} \nu(A), \tag{2}$$

(then $\nu_1 = c\,\nu|_{\mathbb{B}^c}(\cdot)$ with $c = \nu(\mathbb{B}^c)^{-1}$.

In both cases, consequences:

- ν_1, ν are homogeneous, $\nu_1(tA) = t^{-\alpha}\nu_1(A)$
- Angular *probability* measure $\Phi(B) = \nu\{tw, w \in B, t \geq 1\}$, $B \subset \mathbb{B}$. Finite on $\mathbb{S}$.

Tail behavior of Polar Depth

Theorem: Tail behavior of *pD* under regular variation

1. Under minimal RV assumption + smoothness ($P(\partial H_{0,u}) = 0$): uniform convergence on compact K such that $0 \notin K$:

$$\sup_{x \in K} |p_t^{-1} pD(tx|P) - pD(x|\nu_1)| \xrightarrow[t \to \infty]{} 0,$$

2. If in addition (2) holds, then, for any $\delta > 0$, uniform convergence far from 0:

$$\sup_{x \in \mathbb{R}^d : r(x) \geq \delta} |p_t^{-1} pD(tx|P) - pD(x|\nu_1)| \xrightarrow[t \to \infty]{} 0.$$

Polar Depth estimation of extremes: $\widehat{XpD}$ I

- Sort observations by decreasing values of the norm: $r(X_{(1)}) \geq \ldots \geq r(X_{(n)})$ and let $R_k = r(X_{(k)})$.

- Choose $1 \ll k \ll n$ [...].

- For x such that $r(x) \geq R_k$, use $\widehat{XpD}$ not $\widehat{pD}$:

  Radial survival function of extremes (standard) : based on tail index estimator $\widehat{\alpha}$, say $\widehat{\alpha} = 1/\mathrm{Hill}$.

  $$\widehat{p}_{r(x)} = \frac{k}{n}(r(x)/R_k)^{-\widehat{\alpha}}$$

  Empirical Angular Measure of extremes

  $$\widehat{\Phi_k} = \frac{1}{k}\sum_{i \leq k} \delta_{\theta(X_{(i)})}$$

Polar Depth estimation of extremes: $\widehat{XpD}$ II

Angular depth estimator(Dyckerhoff and Nagy, 2025)

$$\widehat{aD}(\theta(x)|\Phi) = aD(\theta(x)|\widehat{\Phi_k})$$

Extreme Polar Depth Estimator

$$\widehat{XpD}(x|P) \; = \; \widehat{p}_{r(x)} \; \widehat{aD}(\theta(x)|\Phi)$$

Statistical guarantees for $\widehat{XpD}$

$\Delta_\alpha = |\hat{\alpha} - \alpha|$.

Theorem (simplified): non-asymptotic bounds ⇒ consistency at infinity

Suppose that (2) holds and n, k sufficiently large (...) and Δ_α small with high probability. Then with probability at least $1-\delta$:

$$\sup_{\overline{u}_{n,k}, \underline{u}_{n,k} \le r(x) \le \exp\left(\Delta_\alpha^{-1/2}\right)\underline{u}_{n,k}} p_{r(x)}^{-1} \left| \widehat{XpD}(x|P) - pD(x|P) \right| \le C\sqrt{\frac{d \log(k/\delta)}{k}} + \mathrm{Bias}(k,n)$$

where C is explicit, not ridiculously high, $\overline{u}_{n,k} \sim 1 - k/n$ quantile of $r(X)$ and the bias term $\mathrm{Bias}(k,n)$ converges to 0 as $n/k \to +\infty$.

Standardization

- Marginal standardization: customary in EVT to obtain non-degenerate limit measure when the X^j's have different tail behavior

- Multivariate marginal standardization $v(x) = v_1(x^1), \dots, v_d(x^d))$ with $v_j(t) = \frac{1}{1-F_j(t)}$, then work with vectors $V_i = v(X_i)$, $i \leq n$.

- Empirical counterpart (here): empirical $\hat{F}_j$ below threshold, parametric with $\hat{\alpha}_j$ above threshold $(\dots) \to \hat{v}$, $\widehat{V}_i$'s.

- **Key result in MEVT**: If X is in a multivariate domain of attraction, then

$$t\mathbb{P}\left(V \in t(\cdot)\right) \to \mu(\cdot)$$

  where μ is a (standardized) limit measure, like ν but with $\alpha = 1$.
- Standardized angulra measure

$$\Phi_s(B) = \mu(tB, t \geq 1) \qquad B \subset \mathbb{S}.$$

Standardized Polar Depth pD_s and asymptotics

Polar Depth of $v(x)$ with respect to $v_\# P$, the distribution of the V_i's

$$pD_s(x|P) := pD(v(x)|v_\# P).$$

Under minimal standardized RV assumption + some uniform convergence on $\mathbb{B}_t^c$:

Proposition: tail equivalent and role of Φ_s

$$\underbrace{\sup_{y:\|y\|\geq t} \Big| \, \|y\| pD(y|v_\# P) - aD(\theta(y)|\Phi_s) \, \Big|}_{:=B_1(t)} \to 0.$$

Angular Depth w.r.t. Φ_s is key.

Ingredients: Empirical empirical angular measure + assumptions

- **Empirical (standardized) angular measure**: non-asymptotic control in $O(\sqrt{d \log k/k})$ in Clémençon et al. (2023).

$$\widehat{\Phi_s}(\cdot) = \frac{n}{k} \sum_{i \leq n} \mathbb{1}\{\theta(\hat{V}_i) \in \cdot, \|\hat{V}_i\| \geq n/k\},$$

- For their bounds to apply: Assume additionally that Φ_s has a density w.r.t. surface measure σ_{d-1} on $\mathbb{S}$ bounded by $\|\phi\|_\infty < \infty$ + no mass near the axes.

- Choose $\widehat{p_j}(x)$ in $\widehat{XpD}$ as a semi-parametric estimator of $1 - F_j(x)$ (using some $\widehat{\alpha}_j$), using same k as $\widehat{\Phi}_s$.

- $\widehat{v}(x)$: associated empirical transformation to Pareto margins.

Nonasymptotic guarantees for $\widehat{XpD_s}$

Main result (standardized version)

Let $\delta_1 > 0, h < 1/8$, $\Omega \subset \mathbb{R}^d$ s.t. with probability at least $1 - \delta_1$, $\max_{j \leq d} \sup_{x \in \Omega} |\widehat{p}_j(x_j)/p_j(x_j) - 1| < h$. Then for any $t \geq 0$ and $\delta_2 < 1 - \delta_1$ with proba $1 - \delta_1 - \delta_2$,

$$\sup_{x \in \Omega, \|\widehat{v}(x)\| \geq t} \|\widehat{v}(x)\| \left| \widehat{XpD_s}(x) - pD_s(x|P) \right| \leq D_\Phi(k, \delta_2) + B_\Phi(k, n) + \cdots$$

$$\frac{h}{1-h}(\Phi_s(\mathbb{S}) + 2\|\phi\|_\infty) + \cdots$$

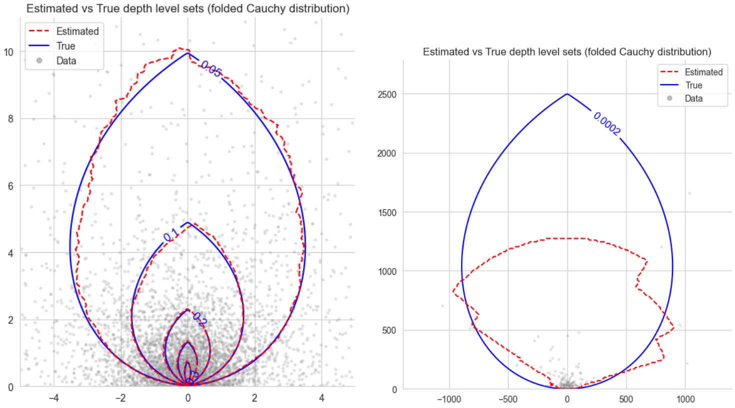
$$\frac{1}{1-h} B_{pD}(t(1-h)),$$

where $D_\Phi(k, \delta), B_\Phi(k, n)$: upper bound on the deviations (resp. Bias) of $\widehat{\Phi_s}$ (Clémençon et al., 2023) ;
B_{pD}: Bias term stemming from the asymptotic approximation of the polar depth, which vanishes at infinity.

Outline

Illustration: Folded Cauchy

- "Folded" Cauchy distribution $X = (X_1, |X_2|)$ where (X_1, X_2) Bivariate Cauchy. $n = 5000$.

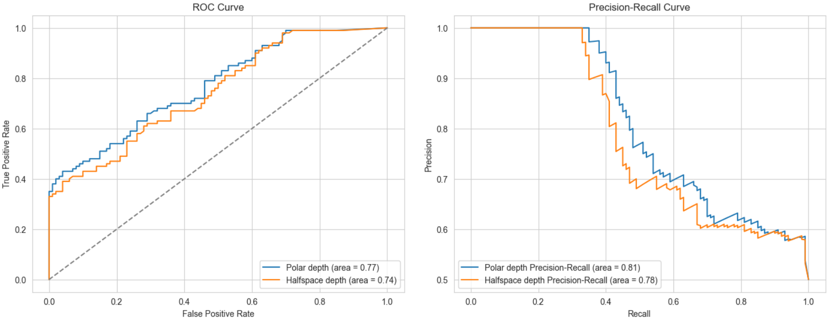


left: Bulk contours, ; right: Tail contours, i.e. $pD = 1/n$.

EXTRAPOLATION IS NECESSARY IN THE TAIL.

Anomaly Detection: Polar versus Tukey Depth (bulk)

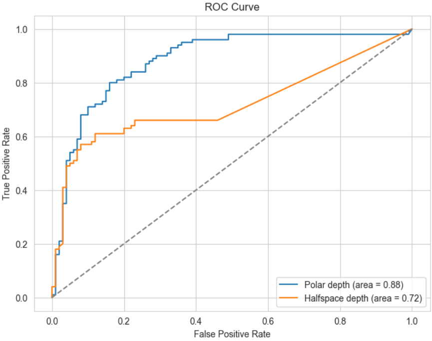
- Anomalies: $r(X)$ super heavy-tailed ($\log R$ is RV(1)) ; $\theta(X) \sim \mathcal{U}[0, \pi]$.
- Anomaly scoring fuction: $s(x) = \widehat{pD}(x)$. The lower, the more abnormal.
- ROC/PR curve: with classifier of the kind $\hat{g}(X) = \mathbb{1}\{s(X) < t\}$, varying t.



similar performance, with pD slightly better.

AD: Comparison Polar versus Tukey Depth (tail)

- Normal data: $r(X) \sim \text{Pareto}(2) \perp\!\!\!\perp \theta(X) \sim \mathcal{U}[0,1]$.
- Anomalies: $r(X) \sim \text{Pareto}(0.1)$ (heavier tail), same angular component.



pD clearly better.

Conclusion

- Novel "Polar Depth" function, suitable for non-extremes and extremes, compatible with marginal standardization.

- **Main scope** Data with one coordinate ≥ 0 such that the larger, the more "extraordinary" the observation.

- **Results:** Nice axiomatic properties and relations with angular depth. Nonasymptotic statistical guarantees scaling as $O(1/\sqrt{n})$ (non-extreme) or $O(1/\sqrt{k})$ (k: effective training sample size) for extremes. Standardization does not break things. Encouraging experiments.

- **Natural applications**: Anomaly detection in bulk or tail or quantile contours in bulk or tail.

References I

Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017). Monge–kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223–256.

Clémençon, S., Jalalzai, H., Lhaut, S., Sabourin, A., and Segers, J. (2023). Concentration bounds for the empirical angular measure with statistical learning applications. *Bernoulli*, 29(4):2797 – 2827.

Dyckerhoff, R. and Mozharovskyi, P. (2016). Exact computation of the halfspace depth. *Computational Statistics & Data Analysis*, 98:19–30.

Dyckerhoff, R. and Nagy, S. (2023). Exact computation of angular halfspace depth.

Dyckerhoff, R. and Nagy, S. (2025). Exact computation of angular halfspace depth. *Statistics and Computing*, 35(6).

Einmahl, J. H., Li, J., and Liu, R. Y. (2015). Bridging centrality and extremity: Refining empirical data depth using extreme value statistics. *Annals of Statistics*, 43(6):2738–2765.

Goix, N., Sabourin, A., Clémen, S., et al. (2015). Learning the dependence structure of rare events: a non-asymptotic study. In *Conference on learning theory*, pages 843–860. PMLR.

References II

He, Y. and Einmahl, J. H. (2017). Estimation of extreme depth-based quantile regions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(2):449–461.

Lhaut, S., Sabourin, A., and Segers, J. (2022). Uniform concentration bounds for frequencies of rare events. *Statistics & Probability Letters*, 189:109610.

Li, J., Cuesta-Albertos, J. A., and Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association*, 107:737–753.

Liu, R. Y. (1990). On a notion of data depth based upon random simplices. *The Annals of Statistics*.

Liu, R. Y. (1992). *Data Depth and Multivariate Rank Tests*, page 279–294. North-Holland, Amsterdam.

Mozharovskyi, P., Mosler, K., and Lange, T. (2015). Classifying real-world data with the $DD\alpha$-procedure. *Advances in Data Analysis and Classification*, 9:287–314.

Mozharovskyi, P. and Valla, R. (2025). Anomaly detection using data depth: multivariate case. *International Journal of Data Science and Analytics*, In print.

Nagy, S., Demni, H., Buttarazzi, D., and Porzio, G. C. (2023). Theory of angular depth for classification of directional data. *Advances in Data Analysis and Classification*.

References III

Nagy, S. and Laketa, P. (2025). Theoretical properties of angular halfspace depth. *Bernoulli*, 31(2):1007–1031.

Staerman, G., Adjakossa, E., Mozharovskyi, P., Hofer, V., Gupta, J. S., and Clémençon, S. (2023). Functional anomaly detection: a benchmark study. *International Journal of Data Science and Analytics*, 16(16):101–117.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, volume 2, pages 523–531. Canadian Mathematical Congress.

Zuo, Y. and Serfling, R. (2000). Structural properties and convergence results for contours of sample statistical depth functions. *The Annals of Statistics*, 28(2):483–499.