PRIMUS GeMS: SOME OPEN PROBLEMS

STANISLAV NAGY

Basic definitions and notation. Let $\mathcal{P}\left(\mathbb{R}^d\right)$ be the set of all probability measures on $\mathbb{R}^d$. For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ and $x \in \mathbb{R}^d$, the halfspace (Tukey) depth [6] of x with respect to (w.r.t.) P is defined as

$$hD(x;P) = \inf_{\mathfrak{H} \in \mathcal{H}(x)} P(\mathfrak{H}),$$

where $\mathcal{H}(x)$ is the set of all halfspaces whose boundary hyperplane passes through x. The halfspace depth quantifies the "centrality" of a point x w.r.t. the distribution P. For $\alpha \geq 0$, consider the upper level set of the depth

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

This collection of the so-called depth regions hD_α, for $\alpha \in [0,1]$, constitutes a generalisation of quantiles to multivariate probability measures. The point (or a set of points) that maximizes the depth w.r.t. P is the generalised (halfspace) median of P in $\mathbb{R}^d$. For α high (near $1/2$), region hD_α forms the locus of points in the "centre" of the distribution P. For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ uniform on a convex body $K \subset \mathbb{R}^d$, the depth regions hD_α coincide with the so-called floating bodies of K studied in geometry [30, 23, 34].

An array of open problems regards generalisations of results known for floating bodies (uniform distributions on those bodies) in $\mathbb{R}^d$ to "reasonable" classes of measures in $\mathcal{P}\left(\mathbb{R}^d\right)$. Often, it is easy to show that a given property does not hold true for all measures in $P \in \mathcal{P}\left(\mathbb{R}^d\right)$. In that case, it is always interesting to see whether the property can hold true at least for P that:

- has finite moments,
- has a density f (w.r.t. the Lebesgue measure),
- its density is bounded, continuous, or smooth,
- is (in some sense) symmetric,
- its density f is unimodal, log-concave, or quasi-concave, etc.

Most of the problems presented in this text is examined in greater detail in the survey paper [30].

1. **Characterisation by depth.** Let $P, Q \in \mathcal{P}\left(\mathbb{R}^d\right)$ be two different probability measures. Does there exist a point $x \in \mathbb{R}^d$ such that $hD(x;P) \neq hD(x;Q)$? We know that this characterisation result holds true under additional assumptions:

- if P is has a finite number of atoms [22];
- if the boundaries of the depth regions hD_α are smooth for both P and Q for all $\alpha \in [0,1/2)$ [21].

We know that the general conjecture is not valid [28]. Though, all the available examples of different distributions with the same depth that are distributions without a finite first moment (expectation). Is the existence of the expectation sufficient for the depth to characterize the distribution? Under which conditions are absolutely continuous probability distributions

E-mail address: `nagy@karlin.mff.cuni.cz`.
Date: February 5, 2019.

characterised by their depth? Is the uniform distribution on a triangle characterized by its depth?

Can we, for a given halfspace $\mathfrak{H} \subset \mathbb{R}^d$, determine $P(\mathfrak{H})$ only from the depth of P at all points in $\mathbb{R}^d$?

2. **Centroids of the cuts of a measure.** For a point $x \in \mathbb{R}^d$ and a measure P we say that a hyperplane $H \subset \mathbb{R}^d$ minimizes the depth at point $x \in \mathbb{R}^d$, if $x \in H$ and $P(\mathfrak{H}) = hD(x;P)$ for one of the halfspaces $\mathfrak{H}$ whose boundary is H. Halfspace $\mathfrak{H}$ is then called a minimizing halfspace of x. For a random vector $X \sim P$ uniformly distributed on a convex body K it holds that at each $x \in K$ there exists a minimizing hyperplane, and for each hyperplane H that minimizes the depth at $x \in K$ we have that x is the centroid of a cut of the body K by the hyperplane H [18]. For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ with continuous density positive on a convex set, a variant of this result is proved in [19, Theorem 3.1]. Does a version of this theorem hold true also for general measures $P \in \mathcal{P}\left(\mathbb{R}^d\right)$?

3. **Cuts of a measure through its centroid.** For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ with a density, we know that there exists a collection minimizing halfspaces of x_P whose union is $\mathbb{R}^d$ [31, Propositions 8 and 12]. By Problem 2, for reasonable measures P, for each bounding hyperplane H of such a halfspace, the centroid of the cut of P by H is x_P.

Does there always exist a collection of $d+1$ hyperplanes $H_i \ni x_P$ such that x_P is the centroid of $H_i \cap K$ for each $i = 1, \dots, d+1$? Is there always a collection H_i, $i = 1, \dots, d+1$ of hyperplanes passing through the expectation (the centroid) $\mathrm{E}\, X$ of a reasonable measure $P \in \mathcal{P}\left(\mathbb{R}^d\right)$, $X \sim P$, such that $\mathrm{E}\, X = \mathrm{E}\left(X \mid H_i\right)$ for each H_i, $i = 1, \dots, d+1$? In the special case of P uniform on a convex body K, this is the open problem A8 from the book [5].

4. **Grünbaum's inequality.** For $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$ uniform on a convex body $K \subset \mathbb{R}^d$ and $x = \mathrm{E}\, X$ the centroid of K it holds that [17]

$$hD(x;P) \geq \left(\frac{d}{d+1}\right)^d \geq \exp(-1) > 0.36. \tag{1}$$

This inequality is interesting, because the lower bound does not depend on the dimension d. Under what conditions does such a result hold true also for measures? For log-concave measures, analogous inequalities are derived in [3, Section 5.2] — can these results be extended to other classes of reasonable measures?

Does an inequality of type (1) hold true also for measures without the condition $x = \mathrm{E}\, X$? That is, find the broadest class of measures $\mathcal{Q}$ such that

$$\inf_{P \in \mathcal{Q}} \sup_{x \in \mathbb{R}^d} hD(x;P) \geq c > 0$$

for some $c > 0$.

5. **Funk's characterisation of symmetry.** We say that a measure $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is halfspace symmetric around $x \in \mathbb{R}^d$ if $hD(x;P) \geq 1/2$.

For a convex body $K \subset \mathbb{R}^d$, K is symmetric around $x \in K$ (in the sense $K - x = -(K-x)$) if and only if the uniform measure P on K is halfspace symmetric around x [14, 33]. This result is known as the Funk theorem.

Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be halfspace symmetric around $x \in \mathbb{R}^d$ such that $P(\{x\}) = 0$. Then, for $X \sim P$, the distributions of the random vectors $(X-x)/\left\|X-x\right\|$ and $-(X-x)/\left\|X-x\right\|$ are identical [32, Theorem 2]. This implies a general version of the Funk theorem for measures. As far as we can tell, this theorem is not known in geometry.

Under what conditions imposed on the measure P (i.e. the existence and continuity of the density f, its quasi-concavity, or log-concavity) does it hold true, that the halfspace symmetry around $x \in \mathbb{R}^d$ is equivalent with the symmetry of its density f around x (i.e. $f(\cdot - x) \equiv f(x - \cdot)$)?

6. **Smoothness of depth contours.** Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution P with a density. Denote by $P_n \in \mathcal{P}\left(\mathbb{R}^d\right)$ the empirical measure of this random sample[1]. We know that for a given $x \in \mathbb{R}^d$, the distribution of the random variable $\sqrt{n}\left(hD(x; P_n) - hD(x; P)\right)$ is asymptotically normal if and only if there is a unique hyperplane minimizing the depth (w.r.t. P) at x [24]. The last condition is equivalent with the smoothness of the boundary of the depth region hD_α for $\alpha = hD(x; P)$ at x. Under what conditions on P and x can we guarantee that this is fulfilled?

Under what conditions on P can we guarantee that for all points $x \in \mathbb{R}^d$ (except for the halfspace median of P) the asymptotic distribution of $\sqrt{n}\left(hD(x; P_n) - hD(x; P)\right)$ if normal? The only known example of such distributions in statistics is the class of elliptically symmetric distributions (e.g. multivariate normal distributions), for which the depth regions hD_α are concentric ellipsoids, and, a bit more generally, a class of very special distributions called α-symmetric distributions [11, Chapter 7]. Another example of such distributions from geometry are the uniform distributions on symmetric, strictly convex bodies with smooth boundaries [25]. For which other measures does this condition hold true?

7. **Detection of rough points.** In practice the measure P is unknown, and we are given only the empirical measure P_n of the random sample $X_1, X_2, \ldots, X_n$ from P. Is it possible, using only P_n, to detect (test) whether for a given point $x \in \mathbb{R}^d$ the contour of the depth $hD(\cdot; P)$ is smooth at x as in Problem 6?

8. **Shape and orientation of depth contours.** As shown by Milman and Pajor [27, page 104], for P uniform on a (symmetric) convex body $K \subset \mathbb{R}^d$, each set hD_α is homothetic to a fixed ellipsoid whose orientation depends only on the variance matrix of P. For log-concave measures P, a generalization of this theorem is mentioned in [12]. Under what conditions is it possible to extend this theorem to general measures P?

Let P be an isotropic measure[2] with a positive density on $\mathbb{R}^d$. Does it hold true that for each $r > 0$ there exists $\alpha > 0$ such that $B(r) \subset hD_\alpha$, where $B(r) = \left\{x \in \mathbb{R}^d \colon \|x\| < r\right\}$?

9. **Gnedenko's law of large numbers.** For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ log-concave and $X_1, X_2, \ldots, X_n$ a random sample from P, Fresen [13] shows that the convex hull $\operatorname{co}\left(X_1, X_2, \ldots, X_n\right)$ of the observations does, for $n \to \infty$, with large probability "behave like" the set hD_α with $\alpha = 1/n$. Does this result hold true for broader classes of measures? Is it possible to sharpen it[3]? Does this result have applications in statistics, e.g. in the analysis of multivariate extremes?

10. **Probabilistic volumes of the depth regions.** Is it possible, from the depth of all points in $\mathbb{R}^d$, to determine $P(hD_\alpha)$? Can we (for reasonable measures) at least well estimate this probability?

What is the relation between the index α and the characteristics of the set hD_α (probability $P(hD_\alpha)$, volume $\operatorname{vol}\left(hD_\alpha\right)$, the diameter of hD_α)?

[1] The uniform distribution concentrated in the sample points.

[2] $\mathrm{E}\, X = 0$, and the variance matrix $X \sim P$ is a multiple of an identity.

[3] The estimates of the distances between sets in [13] are considered only w.r.t. a very special metric in the space of convex bodies.

11. **Affine surface area for measures.** For P uniform on a convex body K the limit

$$\lim_{\alpha\to 0+} \frac{1-P(hD_\alpha)}{\alpha^{2/(d-1)}}$$

is proportional to the affine surface area of the body K [34]. Is it possible to extend this result to measures? What are the properties, and the interpretation of this characteristic in $\mathcal{P}\left(\mathbb{R}^d\right)$?

12. **Affine invariant points.** A continuous mapping p from the space of convex bodies in $\mathbb{R}^d$ (equipped with the Hausdorff metric) to $\mathbb{R}^d$ that is equivariant w.r.t. non-singular affine transformations of $\mathbb{R}^d$ is called an affine invariant point [18, 26]. Examples of affine invariant points are the centroid, or the centre of the John ellipsoid of K (the ellipsoid of maximal volume that is contained in K). Is the halfspace median of a convex body an affine invariant point? Is it possible to consider affine invariant points also w.r.t. probability measures?

13. **Computation of depth.** Finding the depth of a point $x \in \mathbb{R}^d$ w.r.t. the empirical measure P_n of a random sample for larger values of d and n can be computationally very expensive. The best known exact algorithms have complexity $\mathcal{O}(n^{d-1}\log(n))$ [10]. For a recent survey on related results see Chapter 58 in the book [16]. Usually, the exact computations of the halfspace medians, and the depth regions, are even more involved. Is it possible to speed up these algorithms? Can we, in a fast way, compute the depth w.r.t. a measure given by a density? A trivial approach to the last problem is described in [20].

14. **Approximation of depth.** In practice, for larger d we often resort to the approximation of the depth $hD(x;P)$ (or $hD(x;P_n)$) using the function

$$hD_N(x;P) = \min_{i=1,\dots,N} P\left(\left\{y \in \mathbb{R}^d\colon\ \langle x, U_i\rangle \leq \langle y, U_i\rangle\right\}\right), \tag{2}$$

where $U_1,\dots,U_N$ is a random sample from (the uniform) distribution on the unit sphere in $\mathbb{R}^d$. As $N\to\infty$ we know that $hD_N(x;P)\to hD(x;P)$ almost surely [9, Section 6]. Does such an approximation hold true also uniformly over all $x\in\mathbb{R}^d$? How large N do we need to take in order to achieve a sufficiently good approximation of the true depth? First results in this direction can be found in [29].

15. **Depth in non-Euclidean spaces.** For a definition of the halfspace depth in a general space M it suffices to introduce the concept of halfspaces, as a system of some measurable subsets of M. In the literature, the depth has been introduced in this way on, e.g. the unit sphere [35], or in general metric spaces [4]. Similarly, in geometry, the study of floating bodies on manifolds and other general structures is already under way [1, 2]. Which properties of the depth in $\mathbb{R}^d$ hold true also without the assumption of linearity of the space M?

16. **Depth in infinite-dimensional spaces.** Let B be a general Banach space, which can be equipped with halfspaces of the form

$$\mathfrak{H} = \mathfrak{H}(y,\varphi) = \{x\in B\colon \varphi(x)\leq\varphi(y)\},$$

where $y\in B$ and φ is a bounded linear functional from the dual space of B. With such halfspaces it is possible to formally define the halfspace depth also for measures $P\in\mathcal{P}(B)$. Though, it appears that in that case hD can degenerate, i.e. assign $hD(x;P)=0$ to (almost) all $x\in B$ also for very reasonable measures P [8, 15]. On the other hand, inequalities of type (1) suggest that for special measures, some points can still have positive depth, also in B of infinite dimension.

Does it make sense to consider the halfspace depth also in infinite-dimensional spaces? Can we describe the locus of points for which $hD(x;P)>0$ in in a general space B? Is it possible to resolve the problem of the degeneration of the depth?

17. **Existence of Dupin's floating bodies.** For a convex body $K \subset \mathbb{R}^d$ of unit volume and $\alpha > 0$ small, we say that the Dupin's floating body of K is the convex set $K_{[\alpha]}$ such that each supporting hyperplane of $K_{[\alpha]}$ cuts off a set of volume α from K [7]. Dupin's floating body may not exist. For instance, $K_{[\alpha]}$ does not exist for K the triangle in $\mathbb{R}^2$ for any $\alpha > 0$. Though, if $K_{[\alpha]}$ exists, then $K_{[\alpha]} = hD_\alpha$ for P the uniform distribution on K [34]. In other words, if the Dupin's floating body of K exists, then it coincides with the floating body of K. For a measure $P \in \mathcal{P}\left(\mathbb{R}^d\right)$, one can define Dupin's floating bodies analogously. In connection with Problem 1, it appears that [30, Theorem 34] if all Dupin's floating bodies of P exist for all $\alpha \in (0, 1/2)$, then the measure P is characterised by its depth, and $P(\mathfrak{H})$ is given either by $c = \sup_{x \in \partial \mathfrak{H}} hD(x; P)$, or by $1-c$ for any halfspace $\mathfrak{H} \subset \mathbb{R}^d$. In connection with Problem 6, if all depth contours of P are smooth, then all Dupin's floating bodies of P exist for all $\alpha \in (0, 1/2)$.

In Bobkov [3, Theorem 6.1] it is shown that for symmetric, full-dimensional, κ-concave[4] probability measures P with $\kappa > -1$, all Dupin's floating bodies of P exist. Can the condition $\kappa > -1$ be dropped? Does there exist other probability measures P whose Dupin's floating bodies exist?

18. **Strict convexity of hD_α.** For P a uniform distribution on a convex body, hD_α is always a strictly convex set. This follows from the partial positive result stated in Problem 2. If P is such that $X \sim P$ does not have an expectation, hD_α may fail to be strictly convex. Also, if P is distributed uniformly on a non-convex set in $\mathbb{R}^d$, there may exist hD_α whose boundary contains a line segment. Under which conditions are all depth regions of $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ strictly convex? Must all Dupin's floating bodies of a measure be strictly convex?

19. **Homothety conjecture.** For K an ellipsoid in $\mathbb{R}^d$, all (Dupin's) floating bodies of K are concentric ellipsoids with the same centre and the same orientation as K. Is it true that if, for some convex body $K \subset \mathbb{R}^d$, some floating body K_δ is an ellipsoid, then K must be an ellipsoid as well? This problem is called the homothety conjecture [36, 37].

References

[1] Besau, F. and Werner, E. M. (2016). The spherical convex floating body. *Adv. Math.*, 301:867–901.

[2] Besau, F. and Werner, E. M. (2018). The floating body in real space forms. *J. Differential Geom.*, 110(2):187–220.

[3] Bobkov, S. G. (2010). Convex bodies and norms associated to convex measures. *Probab. Theory Related Fields*, 147(1-2):303–332.

[4] Carrizosa, E. (1996). A characterization of halfspace depth. *J. Multivariate Anal.*, 58(1):21–26.

[5] Croft, H. T., Falconer, K. J., and Guy, R. K. (1994). *Unsolved problems in geometry.* Problem Books in Mathematics. Springer-Verlag, New York.

[6] Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20(4):1803–1827.

[7] Dupin, C. (1822). *Applications de Géométrie et de Méchanique.* Bachelier, Paris.

[8] Dutta, S., Ghosh, A. K., and Chaudhuri, P. (2011). Some intriguing properties of Tukey's half-space depth. *Bernoulli*, 17(4):1420–1434.

[9] Dyckerhoff, R. (2004). Data depths satisfying the projection property. *Allg. Stat. Arch.*, 88(2):163–190.

[4] Measure $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is called κ-concave for $-\infty \le \kappa \le \infty$ if $P\left(tA + sB\right) \ge \left(tP(A)^\kappa + sP(B)^\kappa\right)^{1/\kappa}$ for all $A, B \subset \mathbb{R}^d$ Borel, $s, t > 0$, $s + t = 1$. Condition $\kappa > -1$ implies the existence of $\mathrm{E}\, X$ for $X \sim P$. For details see [3, Section 2].

[10] Dyckerhoff, R. and Mozharovskyi, P. (2016). Exact computation of the halfspace depth. *Comput. Statist. Data Anal.*, 98:19–30.

[11] Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London.

[12] Fresen, D. (2012). The floating body and the hyperplane conjecture. *Arch. Math. (Basel)*, 98(4):389–397.

[13] Fresen, D. (2013). A multivariate Gnedenko law of large numbers. *Ann. Probab.*, 41(5):3051–3080.

[14] Funk, P. (1915). Über eine geometrische Anwendung der Abelschen Integralgleichung. *Math. Ann.*, 77(1):129–135.

[15] Gijbels, I. and Nagy, S. (2017). On a general definition of depth for functional data. *Statist. Sci.*, 32(4):630–639.

[16] Goodman, J. E., O'Rourke, J., and Tóth, C. D., editors (2018). *Handbook of discrete and computational geometry*. Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL.

[17] Grünbaum, B. (1960). Partitions of mass-distributions and of convex bodies by hyperplanes. *Pacific J. Math.*, 10:1257–1261.

[18] Grünbaum, B. (1963). Measures of symmetry for convex sets. In *Proc. Sympos. Pure Math., Vol. VII*, pages 233–270. Amer. Math. Soc., Providence, R.I.

[19] Hassairi, A. and Regaieg, O. (2008). On the Tukey depth of a continuous probability distribution. *Statist. Probab. Lett.*, 78(15):2308–2313.

[20] Kaiser, M. J. (2001). The convex floating body. *Appl. Math. Lett.*, 14(4):483–486.

[21] Kong, L. and Zuo, Y. (2010). Smooth depth contours characterize the underlying distribution. *J. Multivariate Anal.*, 101(9):2222–2226.

[22] Koshevoy, G. A. (2002). The Tukey depth characterizes the atomic measure. *J. Multivariate Anal.*, 83(2):360–364.

[23] Leichtweiß, K. (1986). Zur Affinoberfläche konvexer Körper. *Manuscripta Math.*, 56(4):429–464.

[24] Massé, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 10(3):397–419.

[25] Meyer, M. and Reisner, S. (1991). A geometric property of the boundary of symmetric convex bodies and convexity of flotation surfaces. *Geom. Dedicata*, 37(3):327–337.

[26] Meyer, M., Schütt, C., and Werner, E. M. (2015). Affine invariant points. *Israel J. Math.*, 208(1):163–192.

[27] Milman, V. D. and Pajor, A. (1989). Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 64–104. Springer, Berlin.

[28] Nagy, S. (2018a). Halfspace depth does not characterize probability distributions. arXiv:1810.09207.

[29] Nagy, S. (2018b). Uniform rates of convergence for the approximations of the halfspace depth and the projection depth. In preparation.

[30] Nagy, S., Schütt, C., and Werner, E. M. (2018). Data depth and floating body. arXiv:1809.10925.

[31] Rousseeuw, P. J. and Ruts, I. (1999). The depth function of a population distribution. *Metrika*, 49(3):213–244.

[32] Rousseeuw, P. J. and Struyf, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Stat. Plan. Inference*, 122(1-2):161–173.

[33] Schneider, R. (1970). Functional equations connected with rotations and their geometric applications. *Enseignement Math. (2)*, 16:297–305 (1971).

[34] Schütt, C. and Werner, E. M. (1990). The convex floating body. *Math. Scand.*, 66(2):275–290.
[35] Small, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.*, 15(1):31–39.
[36] Stancu, A. (2006). The floating body problem. *Bull. London Math. Soc.*, 38(5):839–846.
[37] Werner, E. M. and Ye, D. (2011). On the homothety conjecture. *Indiana Univ. Math. J.*, 60(1):1–20.