ILLUMINATION DEPTH: APPENDIX

STANISLAV NAGY AND JIŘÍ DVOŘÁK

Abstract. This document complements the article *S. Nagy and J. Dvořák: Illumination depth*. In Appendix A the proofs of all the theoretical results from the main paper are given. Appendix B contains additional simulation studies; in Appendix C the `R` source code of the illumination procedure can be found.

Appendix A. Proofs of the theoretical results

A.1. **Proof of Lemma 1.** For $d = 1$, $\Sigma = \sigma^2 > 0$ and $|x - \mu| > \sigma$, the formula reduces to $\mathcal{I}(x; \mathcal{E}_{\mu,\Sigma}) = |x - \mu| + \sigma$, which is the illumination of x outside $\mathcal{E}_{\mu,\Sigma} = \sigma B^1$ on that ball. For $d > 1$, let us first compute the illumination of a unit ball B^d in $\mathbb{R}^d$. Take $x \notin B^d$. The set difference of the convex hull of x and B^d minus B^d is a cone with height $\|x\| - 1/\|x\|$ and base a $(d-1)$-dimensional ball with radius $\sqrt{1 - 1/\|x\|^2}$, without a spherical cap of B^d of height $1 - 1/\|x\|$. Because $\operatorname{vol}_d(B^d) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}$, the volume of the cone is

$$\frac{1}{d} \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2} + 1\right)} \left(1 - 1/\|x\|^2\right)^{\frac{d-1}{2}} \left(\|x\| - \frac{1}{\|x\|}\right), \tag{A.1}$$

and the volume of the cap is

$$\frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^{\arccos(1/\|x\|)} \sin^d(t) \, \mathrm{d}\, t. \tag{A.2}$$

Altogether, (A.1) and (A.2) give that

$$\mathcal{I}\left(x; B^d\right) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} \left(\frac{\|x\|}{d} \left(1 - 1/\|x\|^2\right)^{\frac{d+1}{2}} - \int_0^{\arccos(1/\|x\|)} \sin^d(t) \, \mathrm{d}\, t \right) + \operatorname{vol}_d\left(B^d\right).$$

It is not difficult to see that $\mathcal{E}_{\mu,\Sigma} = \Sigma^{1/2} B^d + \mu = \bigcup_{x \in B^d} \left\{\Sigma^{1/2} x + \mu\right\}$. Thus, by the affine equivariance of the illumination bodies [9, Proposition 2] we have $\mathcal{I}(x; \mathcal{E}_{\mu,\Sigma}) = \mathcal{I}\left(\Sigma^{-1/2}(x - \mu); B^d\right) \sqrt{|\Sigma|}$. The general assertion then follows from $\left\|\Sigma^{-1/2}(x - \mu)\right\| = \sqrt{(x - \mu)^{\mathsf{T}} \Sigma^{-1} (x - \mu)}$.

A.2. **Lemma A.1.** The next lemma summarizes some analytical properties of the function g_d defined in Section 2.

Lemma A.1. *For all* $d \geq 1$

Charles University, Faculty of Mathematics and Physics, Department of Probability and Math. Statistics, Prague, Czech Republic

E-mail addresses: `nagy@karlin.mff.cuni.cz`.

Date: May 28, 2020.

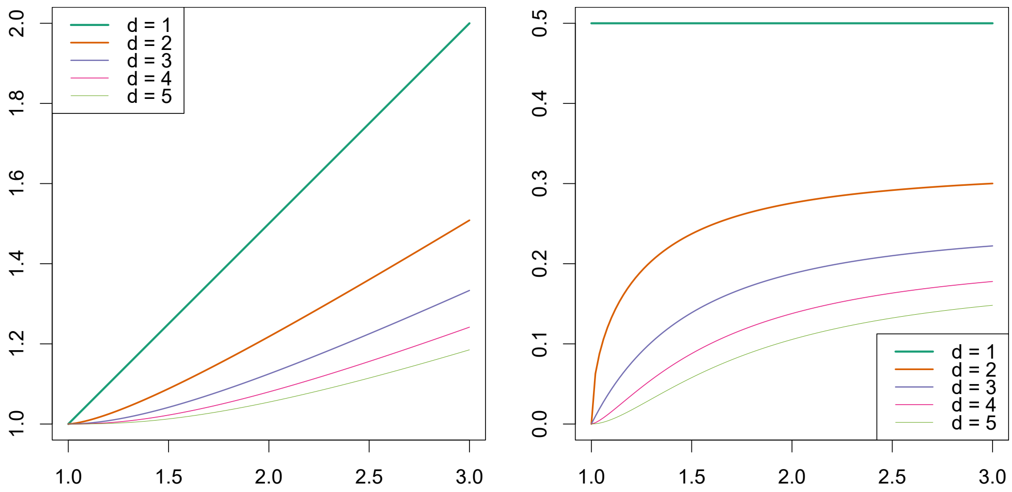

FIGURE A.1. *First five functions g_d, $d = 1, \dots, 5$ (left panel) and their first derivatives (right panel).*

(i) function $g_d\colon [1,\infty) \to [1,\infty)$ is uniformly continuous, strictly increasing, and convex;
(ii) $g_d(1) = 1$, $\lim_{t\to\infty} g_d(t) = \infty$;
(iii) g_d is differentiable on $(1,\infty)$ and

$$g_d'(t) = \frac{\Gamma\left(\frac{d}{2}+1\right)}{\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)}\frac{1}{d}\left(1-\frac{1}{t^2}\right)^{(d-1)/2} \qquad \text{for } t \in (1,\infty); \tag{A.3}$$

(iv) $g_d(t) - 1 = \mathcal{O}\left((t-1)^{(d+1)/2}\right)$ as $t \to 1$ from the right;
(v) the minimal modulus of continuity of the inverse function g_d^{-1} takes the form

$$w_{g_d^{-1}}(h) = \sup_{|s-t|<h} \left|g_d^{-1}(s) - g_d^{-1}(t)\right| = g_d^{-1}(1+h) - 1 \quad \text{for } h \ge 0;$$

(vi) as $h \to 0$ from the right, $w_{g_d^{-1}}(h) = \mathcal{O}\left(h^{2/(d+1)}\right)$;
(vii) as $t \to \infty$, $g_d^{-1}(t) = \mathcal{O}(t)$.

Proof. Using the Leibniz integral formula it is easy to see that the derivative of g_d is (A.3). That function is positive, increasing, and bounded from above. Hence, g_d is strictly increasing, convex, and Lipschitz continuous. Part (iv) follows by an application of l'Hôpital's rule

$$\begin{aligned} \lim_{t\to 1}\frac{g_d(t)-1}{(t-1)^{(d+1)/2}} &= \lim_{t\to 1}\frac{\Gamma\left(\frac{d}{2}+1\right)}{\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)}\frac{2}{d(d+1)}\frac{\left((t-1)(t+1)\right)^{(d-1)/2}}{t^{d-1}(t-1)^{(d-1)/2}} \\ &= \frac{\Gamma\left(\frac{d}{2}+1\right)}{\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)}\frac{2^{(d+1)/2}}{d(d+1)}. \end{aligned} \tag{A.4}$$

For Part (v) first note that because g_d is smooth, strictly increasing and convex, its inverse g_d^{-1} must be smooth, strictly increasing and concave. For such a function the mean value theorem asserts that the greatest difference $g_d^{-1}(s) - g_d^{-1}(t)$ subject to $1 \le t \le s < t+h$ must be attained at the left endpoint of its domain, i.e. for $t = 1$ and $s = 1+h$. To obtain the rate of the modulus of continuity, note that by (A.4) there exists $c > 0$ such

that
$$g_d(t) - 1 \geq c\,(t-1)^{(d+1)/2} \quad \text{for all } t > 1 \text{ close enough to } 1.$$
Apply g_d^{-1} to both sides of this inequality and substitute $h = c\,(t-1)^{(d+1)/2}$ to get
$$g_d^{-1}(1+h) - 1 \leq \left(\frac{h}{c}\right)^{2/(d+1)} \quad \text{for all } h > 0 \text{ small enough,}$$
and the conclusion follows. Finally, using substitution $t = g_d(s)$ and l'Hôpital's rule again,
$$\lim_{t\to\infty} \frac{g_d^{-1}(t)}{t} = \lim_{s\to\infty} \frac{g_d^{-1}(g_d(s))}{g_d(s)} = \lim_{s\to\infty} \frac{s}{g_d(s)} = \lim_{s\to\infty} \frac{1}{g_d'(s)} = \frac{d\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)}.$$
Hence, $g_d^{-1}(t) = \mathcal{O}(t)$ as $t \to \infty$. □

A.3. **Proof of Theorem 2.** We only prove the first part of the theorem. The remaining parts are straightforward, and follow directly from the essential properties of the halfspace depth [5], and the properties of the illumination [9].

By the affine invariance of the halfspace depth [3, Lemma 2.1] we know that $(P_{AX+b})_\alpha = A(P_X)_\alpha + b$. For the illumination, it follows that
$$\begin{aligned}\frac{\mathcal{I}\left(Ax+b;(P_{AX+b})_\alpha\right)}{\mathrm{vol}_d\left((P_{AX+b})_\alpha\right)} &= \frac{\mathrm{vol}_d\left(\mathrm{co}\left((A(P_X)_\alpha + b) \cup \{Ax+b\}\right)\right)}{\mathrm{vol}_d\left(A(P_X)_\alpha + b\right)} \\ &= \frac{\mathrm{vol}_d\left(A\left((P_X)_\alpha \cup \{x\}\right) + b\right)}{\mathrm{vol}_d\left(A(P_X)_\alpha + b\right)} = \frac{\mathcal{I}\left(x;(P_X)_\alpha\right)}{\mathrm{vol}_d\left((P_X)_\alpha\right)}.\end{aligned}$$

A.4. **Proof of Theorem 3.** We start with the illumination. From [4, Theorem 4.2] we know that under the assumptions of the theorem, the central regions P_α are consistent for P in the Hausdorff distance, i.e.
$$\mathsf{d}_H\left(P_{n,\alpha}, P_\alpha\right) \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{A.5}$$
For any $x \in K_n$ we know that almost surely for n large
$$\begin{aligned}\left|\mathcal{I}\left(x;P_{n,\alpha}\right) - \mathcal{I}\left(x;P_\alpha\right)\right| &= \left|\mathrm{vol}_d\left(\mathrm{co}\left(P_{n,\alpha}\cup\{x\}\right)\right) - \mathrm{vol}_d\left(\mathrm{co}\left(P_\alpha \cup \{x\}\right)\right)\right| \\ &\leq c_d \sum_{j=0}^{d-1} \mathsf{d}_H\left(\mathrm{co}\left(P_\alpha\cup\{x\}\right), \mathrm{co}\left(P_{n,\alpha}\cup\{x\}\right)\right)^{d-j} R_n^j \\ &\leq c_d \sum_{j=0}^{d-1} \mathsf{d}_H\left(P_\alpha\cup\{x\}, P_{n,\alpha}\cup\{x\}\right)^{d-j} R_n^j \\ &\leq c_d \sum_{j=0}^{d-1} \mathsf{d}_H\left(P_\alpha, P_{n,\alpha}\right)^{d-j} R_n^j \\ &\leq d\,c_d\,\mathsf{d}_H\left(P_\alpha, P_{n,\alpha}\right) \max\left\{1, R_n^{d-1}\right\}.\end{aligned} \tag{A.6}$$
In the inequalities we used Lemma A.2 stated below for $\mathsf{d}_H\left(P_{n,\alpha}, P_\alpha\right) < 1$, and the properties of the Hausdorff distance [7, p. 64]. Since for a fixed compact set $K = K_n$ for all n the term R_n is constant, the first part of the theorem is verified in view of (A.5).

To derive the rates of convergence, by [1, Theorem 2] we have that $\mathsf{d}_H\left(P_{n,\alpha}, P_\alpha\right) = \mathcal{O}_{\mathsf{P}}\left(n^{-1/2}\right)$, and the last inequality in (A.6) is enough to conclude.

For the affine invariant version of the illumination, write

$$\begin{aligned}\left|\frac{\mathcal{I}\left(x;P_{n,\alpha}\right)}{\mathrm{vol}_d\left(P_{n,\alpha}\right)}-\frac{\mathcal{I}\left(x;P_{\alpha}\right)}{\mathrm{vol}_d\left(P_{\alpha}\right)}\right| &\leq \frac{\left|\mathcal{I}(x;P_{n,\alpha})-\mathcal{I}(x;P_{\alpha})\right|}{\mathrm{vol}_d\left(P_{n,\alpha}\right)} \\ &\quad+\left|\mathcal{I}(x;P_{\alpha})\right|\left|\frac{1}{\mathrm{vol}_d\left(P_{n,\alpha}\right)}-\frac{1}{\mathrm{vol}_d\left(P_{\alpha}\right)}\right|.\end{aligned} \tag{A.7}$$

By the assumptions of the theorem we know that $\mathrm{vol}_d\left(P_{\alpha}\right)>0$. From (A.5) and Lemma A.2 it thus follows that for n large enough $\mathrm{vol}_d\left(P_{n,\alpha}\right)\geq\mathrm{vol}_d\left(P_{\alpha}\right)/2$ almost surely, and that for such n it also holds true that

$$\left|\frac{1}{\mathrm{vol}_d\left(P_{n,\alpha}\right)}-\frac{1}{\mathrm{vol}_d\left(P_{\alpha}\right)}\right|\leq\frac{2\left|\mathrm{vol}_d\left(P_{n,\alpha}\right)-\mathrm{vol}_d\left(P_{\alpha}\right)\right|}{\mathrm{vol}_d\left(P_{\alpha}\right)^2}\leq\frac{2\,d\,c_d\max\left\{1,R_1^{d-1}\right\}\mathsf{d}_H\left(P_{\alpha},P_{n,\alpha}\right)}{\mathrm{vol}_d\left(P_{\alpha}\right)^2}$$

almost surely, for $c_d>0$ the constant from Lemma A.2. By [1, Theorem 2] the last formula can be written also as

$$\left|\frac{1}{\mathrm{vol}_d\left(P_{n,\alpha}\right)}-\frac{1}{\mathrm{vol}_d\left(P_{\alpha}\right)}\right|=\mathcal{O}_{\mathsf{P}}\left(n^{-1/2}\right).$$

Finally, because P_{α} is a fixed bounded set, a trivial upper bound for $\sup_{x\in K_n}\left|\mathcal{I}(x;P_{\alpha})\right|$ is the maximum illumination of $x\in K_n$ w.r.t. the smallest enclosing ball of P_{α}. By Lemmas 1 and A.1 this is of order $\mathcal{O}\left(R_n\right)$. Altogether, all the above bounds and the consistency result for $\mathcal{I}$ can be plugged into (A.7) to obtain the desired rate of convergence

$$\begin{aligned}\sup_{x\in K_n}\left|\frac{\mathcal{I}\left(x;P_{n,\alpha}\right)}{\mathrm{vol}_d\left(P_{n,\alpha}\right)}-\frac{\mathcal{I}\left(x;P_{\alpha}\right)}{\mathrm{vol}_d\left(P_{\alpha}\right)}\right| &=\mathcal{O}_{\mathsf{P}}\left(\frac{\max\{1,R_n^{d-1}\}}{\sqrt{n}}\right)+\mathcal{O}\left(R_n\right)\mathcal{O}_{\mathsf{P}}\left(\frac{1}{\sqrt{n}}\right) \\ &=\mathcal{O}_{\mathsf{P}}\left(\frac{\max\{1,R_n^{d-1}\}}{\sqrt{n}}\right).\end{aligned}$$

The rates of convergence for the illumination depth follow by a combination of the results above, the Lipschitz continuity of φ, and $\sup_{x\in\mathbb{R}^d}\left|hD(x;P_n)-hD(x;P)\right|=\mathcal{O}_{\mathsf{P}}\left(n^{-1/2}\right)$ that follows from [3].

Lemma A.2. *Let $R>0$. There exists a constant $c_d>0$ such that for all convex bodies $K,L\subset\mathbb{R}^d$ with $K\subset B^d\left(x,R\right)$ for some $x\in\mathbb{R}^d$*

$$\left|\mathrm{vol}_d\left(K\right)-\mathrm{vol}_d\left(L\right)\right|\leq c_d\sum_{j=0}^{d-1}\mathsf{d}_H\left(K,L\right)^{d-j}R^j.$$

Proof. Write $\delta=\mathsf{d}_H\left(K,L\right)$. From the definition of the Hausdorff distance (6) we have that

$$K\subset L+\delta B^d\text{ and }L\subset K+\delta B^d. \tag{A.8}$$

If $\mathrm{vol}_d\left(K\right)\leq\mathrm{vol}_d\left(L\right)$, this gives $\mathrm{vol}_d\left(K\right)\leq\mathrm{vol}_d\left(L\right)\leq\mathrm{vol}_d\left(K+\delta B^d\right)$; in the other case $\mathrm{vol}_d\left(L\right)<\mathrm{vol}_d\left(K\right)$ we get $\mathrm{vol}_d\left(L\right)<\mathrm{vol}_d\left(K\right)\leq\mathrm{vol}_d\left(L+\delta B^d\right)$. This results in

$$\left|\mathrm{vol}_d\left(K\right)-\mathrm{vol}_d\left(L\right)\right|\leq\max\left\{\mathrm{vol}_d\left(K+\delta B^d\right)-\mathrm{vol}_d\left(K\right),\mathrm{vol}_d\left(L+\delta B^d\right)-\mathrm{vol}_d\left(L\right)\right\},$$

and it is enough to bound the excess volume of the outer parallel body $K+\delta B^d$ of a convex body K, and analogously for L. For this, use the Steiner formula [7, Formula (4.1)]

$$\mathrm{vol}_d\left(K+\delta B^d\right)=\sum_{j=0}^{d}\delta^{d-j}\,\mathrm{vol}_{d-j}\left(B^{d-j}\right)V_j(K),$$

where $V_j(K)$ stands for the intrinsic volume of the convex body K [7, Chapter 4]. In particular, it holds true that $V_d(K) = \mathrm{vol}_d(K)$, $V_{d-1}(K)$ is proportional to the surface area measure of K, $V_1(K)$ is the so-called intrinsic width of K, and $V_0(K) = 1$.

From the monotonicity of the intrinsic volumes that follows from formulas (5.25) and (5.31) in [7], and $K \subset B^d(x, R)$, we can use the expression for the intrinsic volumes of a ball (4.64) from [7] and bound

$$\begin{aligned}\mathrm{vol}_d\left(K + \delta B^d\right) - \mathrm{vol}_d(K) &\leq \sum_{j=0}^{d-1} \delta^{d-j}\, \mathrm{vol}_{d-j}\left(B^{d-j}\right) V_j(B^d) \\ &= \mathrm{vol}_d\left(B^d\right) \sum_{j=0}^{d-1} \delta^{d-j} \binom{d}{j} R^j.\end{aligned} \tag{A.9}$$

For a bound on the excess volume of $L + \delta B^d$, first note that from (A.8) we have

$$L \subset K + \delta B^d \subset B^d(x, R) + B^d(0, \delta) = B^d(x, R + \delta).$$

Similarly as in (A.9) we can thus write

$$\begin{aligned}\mathrm{vol}_d\left(L + \delta B^d\right) - \mathrm{vol}_d(L) &\leq \mathrm{vol}_d\left(B^d\right) \sum_{j=0}^{d-1} \delta^{d-j} \binom{d}{j} (R+\delta)^j \\ &= \mathrm{vol}_d\left(B^d\right) \sum_{j=0}^{d-1} \delta^{d-j} \binom{d}{j} \sum_{k=0}^{j} \binom{j}{k} R^k \delta^{j-k} \\ &= \mathrm{vol}_d\left(B^d\right) \sum_{k=0}^{d-1} \delta^{d-k} R^k \sum_{j=k}^{d-1} \binom{d}{j} \binom{j}{k}.\end{aligned}$$

From (A.9) and the last inequality we see that our claim holds true for

$$c_d = \mathrm{vol}_d\left(B^d\right) \max_{k=0,\dots,d-1} \sum_{j=k}^{d-1} \binom{d}{j} \binom{j}{k},$$

the maximum of all the terms that are constant in R and δ in the sums on the right-hand sides of the two excess volume bounds. □

A.5. **Consistency of the illumination on unbounded sets.** Over unbounded subsets of $\mathbb{R}^d$ with $d > 1$, the illumination is not uniformly consistent. So see this take a convex body K in $\mathbb{R}^d$, y in the distance of $\varepsilon > 0$ from K, and let $K_y = \mathrm{co}(K \cup \{y\})$. Surely, $\mathsf{d}_H(K_y, K) = \varepsilon$. By the Hahn-Banach separation theorem [7, Theorem 1.3.7], y and K can be strongly separated by two parallel hyperplanes H_1, H_2 whose distance is at least $\varepsilon/2$ and $y \in H_1$. Take $x \in H_2$ far enough from y. The illumination $\mathcal{I}(x; K_y)$ and $\mathcal{I}(x; K)$ then differs by, at least, the illumination of x onto the cone $K_y \cap H_2^+$ for H_2^+ the halfspace whose boundary is H_2 and $y \in H_2^+$. This illumination can be bounded from both below and above by the illumination of x on any two balls B_1 and B_2 centred at some $z \in K_y \cap H_2^+$ such that $B_1 \subset K_y \cap H_2^+ \subset B_2$, respectively. By Lemmas 1 and A.1, the latter two illuminations both grow with increasing $R = \|z - x\|$ at a rate $\mathcal{O}(R)$, i.e. $\mathcal{I}(x; K_y) - \mathcal{I}(x; K) = \mathcal{O}(\|z - x\|)$ with $H_2 \ni x \to \infty$. In other words, for any $\varepsilon > 0$ one can find x far enough so that $\mathcal{I}(x; K_y) - \mathcal{I}(x; K) \geq 1$. Consequently, even if the distance $\mathsf{d}_H(K_n, K)$ converges to zero (almost surely), the illumination differences

$|\mathcal{I}(x; K_n) - \mathcal{I}(x; K)|$ and $|\mathcal{I}(x; K_n) / \operatorname{vol}_d(K_n) - \mathcal{I}(x; K) / \operatorname{vol}_d(K)|$ cannot, in general, vanish uniformly over unbounded sets.

A.6. **Proof of Theorem 4.** For x fixed, the illumination of x tends to infinity if and only if the halfspace depth central region functional $P_n \mapsto P_{n,\alpha}$ breaks down. Hence, it is enough to evaluate the breakdown point of $P_{n,\alpha}$ with respect to the Hausdorff distance. We follow the derivations in the proofs of [2, Proposition 2.2] and [3, Proposition 3.2]. Let $x_M \in \mathbb{R}^d$ be (any) halfspace median of P_n, that is let $hD(x_M; P_n) = \Pi(P_n)$. By the argument used in the proof of [3, Lemma 3.1] to upset the set $P_{n,\alpha} = \left\{y \in \mathbb{R}^d \colon hD(y; P_n)\, n \geq \lceil \alpha n \rceil\right\}$ entirely, the smallest number of additional points that need to be added to the data is m, the smallest integer that satisfies $m \geq \lceil \alpha (m + n) \rceil$ (compare with formula (6.19) in [3]). This inequality is solved by $m = \lceil (\alpha/(1-\alpha)) n \rceil$. The additional condition $\alpha \leq \Pi(P_n)/(1+\Pi(P_n))$ ensures that $m \leq \lceil \Pi(P_n)\, n \rceil = \Pi(P_n)\, n$. From this it follows that the depth of x_M with respect to the contaminated dataset must be at least $\Pi(P_n)\, n/(n+m) \geq \Pi(P_n)/(1+\Pi(P_n)) \geq \alpha$. Hence, after the contamination procedure, the central region of points whose depth is at least α must be non-empty.

In the situation when $\alpha > \Pi(P_n)/(1+\Pi(P_n))$, due to the nestedness of the central regions $P_{n,\alpha}$, by the previous part of the proof at least

$$m = \left\lceil \frac{\Pi(P_n)/(1+\Pi(P_n))}{1 - \Pi(P_n)/(1+\Pi(P_n))} n \right\rceil = \lceil \Pi(P_n)\, n \rceil = \Pi(P_n)\, n$$

contaminating points are needed.

The corollary with the asymptotic value of the breakdown point follows the same argument as in the proof of [3, Propositions 3.2 and 3.3].

A.7. **Proof of Theorem 5.** The proofs of parts (i), (ii) and (iii) are straightforward and analogous to the proof of Theorem 2. For part (iv) it is sufficient to realise that according to the non-degeneracy of P, and symmetry conditions imposed on the estimator F_n, the lower level set of $M_\alpha(\cdot; P_n)$ is large if and only if either (i) the central region $P_{n,\alpha}$ is extremely large; or (ii) $F_n^{-1}(1-\alpha)$ is extremely small. By Theorem 4, for the former case, asymptotically at least $m \approx n \min\{\alpha, 1/3\}/(1 - \min\{\alpha, 1/3\})$ contaminating points have to be added to the random sample to disrupt the central region entirely. In the latter case, unless there exists a configuration of m points that make $F_n^{-1}(1-\alpha)$ arbitrarily small, the set $P_{n,\alpha}$ cannot be made arbitrarily large. By extension, no fixed lower level set (12) can then be made too big. By the assumption on the breakdown point of $F_n^{-1}(1-\alpha)$, in the second scenario it is even more difficult to break down the estimator (12) than in the first one. Another option when the lower level set (12) breaks down is when it is an empty set. But, that can happen only if for some $\delta > 0$ small enough, $F_n^{-1}(1 - \Pi(P_n)) > \delta$. This is ruled out by the additional condition imposed on δ. Thus, the resulting limiting breakdown point of the level set is the same as that of $P_{n,\alpha}$.

A.8. **Proof of Theorem 6.** By (9) and (10), the Mahalanobis distance $\mathsf{d}_\Sigma(x, \mu)$ can be written either as $F^{-1}(1 - hD(x; P))$ for any $x \in \mathbb{R}^d$, or, in case when $x \notin P_\alpha$, also as

$F^{-1}(1-\alpha)\, g_d^{-1}(\mathcal{I}(x;P_\alpha)/\operatorname{vol}_d(P_\alpha))$. It is thus sufficient to bound

$$\begin{aligned}
&\sup_{x\in K_n}|M_\alpha(x;P_n)-\mathsf{d}_\Sigma(x,\mu)| \\
&\quad\le \sup_{x\in K_n\cap P_{n,\alpha}}|M_\alpha(x;P_n)-\mathsf{d}_\Sigma(x,\mu)| + \sup_{x\in K_n\setminus P_{n,\alpha}}|M_\alpha(x;P_n)-\mathsf{d}_\Sigma(x,\mu)| \\
&\quad\le \sup_{x\in K_n\cap P_{n,\alpha}}\left|F_n^{-1}(1-hD(x;P_n))-F^{-1}(1-hD(x;P))\right| \\
&\qquad+\sup_{x\in K_n\setminus(P_{n,\alpha}\cup P_\alpha)}\left|F_n^{-1}(1-\alpha)\,g_d^{-1}\left(\frac{\mathcal{I}(x;P_{n,\alpha})}{\operatorname{vol}_d(P_{n,\alpha})}\right)-F^{-1}(1-\alpha)\,g_d^{-1}\left(\frac{\mathcal{I}(x;P_\alpha)}{\operatorname{vol}_d(P_\alpha)}\right)\right| \\
&\qquad+\sup_{x\in K_n\cap(P_\alpha\setminus P_{n,\alpha})}|M_\alpha(x;P_n)-\mathsf{d}_\Sigma(x,\mu)|\,.
\end{aligned}$$

The three suprema on the right hand side will be treated separately. Denote them by **I**, **II**, and **III**, respectively.

A.8.1. *Supremum **I**.* The sample halfspace depth $hD(\cdot;P_n)$ is known [3, formula (6.6)] to be a uniformly consistent estimator of its population version

$$\sup_{x\in\mathbb{R}^d}|hD(x;P_n)-hD(x;P)| \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{A.10}$$

Because P is halfspace symmetric, yet its centre of symmetry has zero probability mass, for any $x\in K_n\cap P_{n,\alpha}$ we have $\alpha\le hD(x;P_n)\le 1/2$, with the second inequality almost surely for all n large enough due to (A.10). We may use the consistency (A.10) again to get that for any $\varepsilon>0$ small, $\alpha-\varepsilon\le hD(x;P)\le 1/2$ for all $x\in K_n\cap P_{n,\alpha}$ and n large enough.

Function F is strictly increasing in a neighbourhood of $[0,F^{-1}(1-\alpha)]$. Thus, F^{-1} is (uniformly) continuous on $I=[1/2,1-\alpha]$. Its approximating sequence $\{F_n^{-1}\}_{n=1}^\infty$ is a sequence of functions that are non-decreasing, and converge to F^{-1} at each $t\in I$ by the uniform consistency of F_n from (13), and [8, Lemma 21.2]. A lemma of Pólya [6, Problem 127, part II] gives that this convergence is uniform on I. We can thus write for n large enough

$$\begin{aligned}
\mathbf{I}&\le \sup_{x\in K_n\cap P_{n,\alpha}}\left|F_n^{-1}(1-hD(x;P_n))-F^{-1}(1-hD(x;P_n))\right| \\
&\quad+\sup_{x\in K_n\cap P_{n,\alpha}}\left|F^{-1}(1-hD(x;P_n))-F^{-1}(1-hD(x;P))\right| \\
&\le \sup_{t\in I}\left|F_n^{-1}(t)-F^{-1}(t)\right|+w_{F^{-1}}\left(\sup_{x\in\mathbb{R}^d}|hD(x;P_n)-hD(x;P)|\right),
\end{aligned}$$

where $w_{F^{-1}}$ is the minimal modulus of continuity of F^{-1} restricted to the interval I. The first supremum on the right hand is small almost surely for n large by the uniform convergence of the quantile functions established above. The second will vanish almost surely because of (A.10) and the uniform continuity of F^{-1} on I.

A.8.2. *Supremum **II**.* Let us first introduce the notation

$$\begin{aligned}
a_x&=-g_d^{-1}\left(\frac{\mathcal{I}(x;P_\alpha)}{\operatorname{vol}_d(P_\alpha)}\right), & b&=F^{-1}(1-\alpha)\,, \\
a_{n,x}&=-g_d^{-1}\left(\frac{\mathcal{I}(x;P_{n,\alpha})}{\operatorname{vol}_d(P_{n,\alpha})}\right), & b_n&=F_n^{-1}(1-\alpha)\,.
\end{aligned} \tag{A.11}$$

In supremum **II** we bound for $L_n^{II} = K_n \setminus (P_{n,\alpha} \cup P_\alpha)$

$$\sup_{x\in L_n^{II}} |a_{n,x} b_n - a_x b| \leq |b_n - b| \sup_{x\in L_n^{II}} |a_x| + |b_n| \sup_{x\in L_n^{II}} |a_{n,x} - a_x| . \tag{A.12}$$

For the supremum in the first summand in (A.12) we know from (10) that

$$\begin{aligned} \sup_{x\in L_n^{II}} |a_x| &\leq \sup_{x\in B^d(\mu,R_n)} \frac{\mathsf{d}_\Sigma(x,\mu)}{F^{-1}(1-\alpha)} = \frac{R_n}{F^{-1}(1-\alpha)} \sup_{x\in B^d(\mu,1)} \mathsf{d}_\Sigma(x,\mu) \\ &= \frac{R_n}{F^{-1}(1-\alpha)} \sqrt{\sup_{\|x\|=1} x^\mathsf{T}\Sigma^{-1}x} = \frac{R_n}{F^{-1}(1-\alpha)} \sqrt{1/\lambda} = \mathcal{O}(R_n) , \end{aligned} \tag{A.13}$$

where $\lambda > 0$ is the smallest eigenvalue of Σ. Using the assumption (14) we see that the first summand on the right hand side of (A.12) vanishes in probability as $n \to \infty$. Furthermore, by (14) we also have that $|b_n| = \mathcal{O}_\mathsf{P}(1)$, and by Lemma A.1 together with Theorem 3

(A.14)

$$\begin{aligned} \sup_{x\in L_n^{II}} |a_{n,x} - a_x| &\leq w_{g_d^{-1}} \left(\sup_{x\in L_n^{II}} \left| \frac{\mathcal{I}(x;P_{n,\alpha})}{\mathrm{vol}_d(P_{n,\alpha})} - \frac{\mathcal{I}(x;P_\alpha)}{\mathrm{vol}_d(P_\alpha)} \right| \right) = w_{g_d^{-1}} \left(\mathcal{O}_\mathsf{P} \left(\frac{\max\left\{1, R_n^{d-1}\right\}}{\sqrt{n}} \right) \right) \\ &= \mathcal{O}_\mathsf{P} \left(\left(\frac{\max\left\{1, R_n^{d-1}\right\}}{\sqrt{n}} \right)^{2/(d+1)} \right) = o_\mathsf{P}(1), \end{aligned}$$

where $w_{g_d^{-1}}$ is the minimal modulus of continuity of g_d^{-1} from Lemma A.1. Together, we have verified that

(A.15)

$$\sup_{x\in L_n^{II}} |a_{n,x} b_n - a_x b| = \mathcal{O}(R_n)\, o_\mathsf{P}(1/R_n) + \mathcal{O}_\mathsf{P}(1)\mathcal{O}_\mathsf{P} \left(\left(\frac{\max\left\{1, R_n^{d-1}\right\}}{\sqrt{n}} \right)^{2/(d+1)} \right) = o_\mathsf{P}(1) .$$

A.8.3. *Supremum **III**.* Here it will be crucial that under the conditions of the theorem, the set $P_\alpha \setminus P_{n,\alpha}$ is negligible as $n \to \infty$ by the consistency of the halfspace depth contours (A.5). First, without loss of generality, suppose that both P_α and $P_{n,\alpha}$ are contained in K_n. This is possible, because P_α is a fixed set, and the sequence $P_{n,\alpha}$ is convergent almost surely by (A.5). Thus, possible enlargement of K_n by a fixed set does not affect any results in this proof. Take $x \in L_n^{III} = K_n \cap (P_\alpha \setminus P_{n,\alpha})$. As $x \notin P_{n,\alpha}$,

$$M_\alpha(x;P_n) = F_n^{-1}(1-\alpha)\, g_d^{-1} \left(\frac{\mathcal{I}(x;P_{n,\alpha})}{\mathrm{vol}_d(P_{n,\alpha})} \right) .$$

In terms of x, this expression varies monotonically with $\mathcal{I}(x;P_{n,\alpha})$. Note that for K a convex body, the illumination $\mathcal{I}(\cdot;K)$ strictly increases on any straight halfline L that starts from $x \in \partial K$ (the boundary of K) and does not intersect K elsewhere, i.e. $K \cap L = \{x\}$. Thus, in our situation, if one considers any halfline that starts at a boundary point of $P_{n,\alpha}$ and passes through x,

$$\inf_{y\in\partial P_{n,\alpha}} M_\alpha(y;P_n) \leq M_\alpha(x;P_n) \leq \sup_{y\in\partial P_\alpha} M_\alpha(y;P_n) .$$

On the boundary of $P_{n,\alpha}$ we are in the situation dealt with in supremum **I**, and by that part of the proof we know that for $\varepsilon > 0$ given, almost surely for any n large enough,

$$\inf_{y\in\partial P_{n,\alpha}} \mathsf{d}_\Sigma(y,\mu) - \varepsilon \leq \inf_{y\in\partial P_{n,\alpha}} M_\alpha(y;P_n) .$$

Likewise, for the upper bound, by part **II** of this proof, and the continuity of $\mathsf{d}_\Sigma(x,\mu)$, we have an analogous restriction, and with high probability, for n large enough,

$$\sup_{y\in\partial P_\alpha} M_\alpha(y;P_n) \leq \sup_{y\in\partial P_\alpha} \mathsf{d}_\Sigma(y,\mu)+\varepsilon = F^{-1}(1-\alpha)+\varepsilon.$$

Finally, we use (A.5) and the fact that the Hausdorff distances of convex bodies, and of their boundaries, are the same [7, Lemma 1.8.1]. This gets that almost surely, for any $\delta > 0$, for all n large enough, and any $y \in \partial P_{n,\alpha}$, there exists $z \in \partial P_\alpha$ such that $\|y-z\| < \delta$. Now, because $\mathsf{d}_\Sigma(x,\mu)$ is in x (uniformly) continuous in a uniform neighbourhood of P_α, this means that almost surely, for n large enough,

$$\inf_{y\in\partial P_{n,\alpha}} \mathsf{d}_\Sigma(y,\mu) \geq \inf_{y\in\partial P_\alpha} \mathsf{d}_\Sigma(y,\mu)-\varepsilon = F^{-1}(1-\alpha)-\varepsilon,$$

and for any $x \in L_n^{III}$

$$F^{-1}(1-\alpha)-\varepsilon \leq \mathsf{d}_\Sigma(x,\mu) \leq F^{-1}(1-\alpha).$$

Altogether, collect all the bounds in this part of the proof to get that for any $\varepsilon > 0$, with high probability, for n large enough,

$$\sup_{x\in L_n^{III}} \left|M_\alpha(x;P_n)-\mathsf{d}_\Sigma(x,\mu)\right| \leq 2\varepsilon,$$

which finishes the proof.

A.9. **Proof of Theorem 7.** In view of the uniform consistency of the halfspace depth (A.10) it suffices to show that

$$\sup_{x\in K_n\setminus P_{n,\alpha}} \left|F_n\left(-g_d^{-1}\left(\frac{\mathcal{I}(x;P_{n,\alpha})}{\mathrm{vol}_d(P_{n,\alpha})}\right)F_n^{-1}(1-\alpha)\right)-hD(x;P)\right| \xrightarrow[n\to\infty]{\mathsf{P}} 0.$$

Proceed analogously as in the proof of Theorem 6, and consider two situations — the supremum above over $x \in L_n^{II} = K_n \setminus (P_{n,\alpha}\cup P_\alpha)$, and the the supremum over $x \in L_n^{III} = K_n \cap (P_\alpha \setminus P_{n,\alpha})$.

Suppose first that $x \in L_n^{II}$. By (9) and (10), in the notation from (A.11) we have that

$$\begin{aligned}
|F_n(a_{n,x}b_n)-hD(x;P)| &= |F_n(a_{n,x}b_n)-F(-\mathsf{d}_\Sigma(x,\mu))| \\
&= |F_n(a_{n,x}b_n)-F(a_xb)| \\
&\leq |F_n(a_{n,x}b_n)-F(a_{n,x}b_n)|+|F(a_{n,x}b_n)-F(a_xb)|.
\end{aligned}$$

Therefore,

$$\sup_{x\in L_n^{II}} |F_n(a_{n,x}b_n)-hD(x;P)| \leq \sup_{t\in\mathbb{R}}|F_n(t)-F(t)|+\sup_{x\in L_n^{II}}|F(a_{n,x}b_n)-F(a_xb)|.$$

The first summand on the right hand side above vanishes almost surely as $n\to\infty$ by (13). For the second summand we already have a bound from (A.15) from the proof of Theorem 6. Since F has a density, it must be uniformly continuous on $\mathbb{R}$. Denote by $w_F\colon(0,\infty)\to\mathbb{R}$ its minimal modulus of continuity. We obtain

$$\sup_{x\in L_n^{II}}|F(a_{n,x}b_n)-F(a_xb)| \leq w_F\left(\sup_{x\in L_n^{II}}|a_{n,x}b_n-a_xb|\right) = w_F(o_{\mathsf{P}}(1)) = o_{\mathsf{P}}(1).$$

This completes the part of the proof with L_n^{II}.

For the second part, consider $x \in L_n^{III}$. Note that thanks to (9) and the continuity of F in a neighbourhood of $F^{-1}(1-\alpha)$, the halfspace depth $hD(\cdot;P)$ must be (uniformly) continuous in a uniform neighbourhood of P_α. Furthermore, for $x \notin P_{n,\alpha}$, $RhD_\alpha(x; P_n)$ varies monotonically with $\mathcal{I}(x; P_{n,\alpha})$. Thus, derivation analogous to that from part **III** in the proof of Theorem 6 gives that the convergence of the halfspace depth contours (A.5) implies that with $n \to \infty$

$$\sup_{x \in L_n^{III}} |RhD_\alpha(x; P_n) - hD(x; P)| = o_{\mathsf{P}}(1),$$

and the proof is finished.

A.10. **Proof of Theorem 8.** By the uniform consistency of the halfspace depth (A.10) we can bound for n large enough

$$\begin{aligned}\sup_{x \in \mathbb{R}^d \colon hD(x;P_n) \geq \alpha} \left| \frac{RhD_\alpha(x; P_n)}{hD(x; P)} - 1 \right| &\leq \sup_{x \in \mathbb{R}^d \colon hD(x;P_n) \geq \alpha} 2 \left| \frac{RhD_\alpha(x; P_n) - hD(x; P)}{hD(x; P_n)} \right| \\ &\leq \frac{2}{\alpha} \sup_{x \in \mathbb{R}^d} |hD(x; P_n) - hD(x; P)|,\end{aligned}$$

where the last term vanishes almost surely as $n \to \infty$. Thus, in the notation established in (A.11) in the proof of Theorem 6, it suffices to show that also the right hand size of

$$\sup_{x \in K_n \setminus P_{n,\alpha}} \left| \frac{F_n(a_{n,x} b_n)}{hD(x; P)} - 1 \right| \leq \sup_{x \in L_n^{II}} \left| \frac{F_n(a_{n,x} b_n)}{F(a_x b)} - 1 \right| + \sup_{x \in L_n^{III}} \left| \frac{F_n(a_{n,x} b_n)}{hD(x; P)} - 1 \right|$$

is asymptotically negligible, where $L_n^{II} = K_n \setminus (P_{n,\alpha} \cup P_\alpha)$ and $L_n^{III} = K_n \cap (P_\alpha \setminus P_{n,\alpha})$. We used (9) and (10) to obtain the expression on the right hand side. We already have everything prepared to bound the second summand above. Indeed, by Theorem 7

$$\begin{aligned}\sup_{x \in L_n^{III}} \left| \frac{F_n(a_{n,x} b_n)}{hD(x; P)} - 1 \right| &\leq \sup_{x \in L_n^{III}} \frac{|RhD_\alpha(x; P_n) - hD(x; P)|}{\alpha} \\ &\leq \frac{1}{\alpha} \sup_{x \in K_n} |RhD_\alpha(x; P_n) - hD(x; P)| = o_{\mathsf{P}}(1).\end{aligned}$$

Let us now focus on the supremum over L_n^{II}. For $x \in L_n^{II}$ we can write

$$\begin{aligned}\left| \frac{F_n(a_{n,x} b_n)}{F(a_x b)} - 1 \right| &= \left| \frac{F(a_{n,x} b_n)}{F(a_x b)} \right| \left| \frac{F_n(a_{n,x} b_n)}{F(a_{n,x} b_n)} - \frac{F(a_x b)}{F(a_{n,x} b_n)} \right| \\ &\leq \left| \frac{F(a_{n,x} b_n)}{F(a_x b)} \right| \left(\left| \frac{F_n(a_{n,x} b_n)}{F(a_{n,x} b_n)} - 1 \right| + \left| \frac{F(a_x b)}{F(a_{n,x} b_n)} - 1 \right| \right).\end{aligned} \tag{A.16}$$

In the same way as in (A.12), (A.13), (A.14) and (A.15) in the proof of Theorem 6 we have, using (16), that

$$\begin{aligned}&\sup_{x \in L_n^{II}} |a_{n,x} - a_x| = \mathcal{O}_{\mathsf{P}}(\omega_n), \\ &\sup_{x \in L_n^{II}} |a_{n,x} b_n - a_x b| = \mathcal{O}(R_n)\, \mathcal{O}_{\mathsf{P}}(\xi_n) + \mathcal{O}_{\mathsf{P}}(\omega_n) = \mathcal{O}_{\mathsf{P}}(\max\{R_n \xi_n, \omega_n\}), \\ &b \sup_{x \in L_n^{II}} |a_x| = R_n / \sqrt{\lambda} = \mathcal{O}(R_n).\end{aligned} \tag{A.17}$$

By the definition of the refined depth (15) we also see that $a_{n,x} < -1$ for any $x \in L_n^{II}$. Combine this with (A.17) to obtain that there exists $c > 0$ such that for all $n \geq 1$ and $x \in L_n^{II}$ we can write $(1 - c\,\omega_n)\, b < |a_x b|$, which means that for n large enough

$b/2 < |a_x b| < R_n/\sqrt{\lambda}$ for all $x \in L_n^{II}$. Formulas (A.17) therefore allow us to write for some $c > 0$ large enough for all $\varepsilon > 0$ and n large

$$\mathsf{P}\left(\sup_{x\in L_n^{II}}\left|\frac{F(a_x b)}{F(a_{n,x}b_n)}-1\right|>\varepsilon\right)\leq \mathsf{P}\left(\sup_{x\in L_n^{II}}|a_{n,x}b_n-a_x b|>c\max\left\{R_n\xi_n,\omega_n\right\}\right)$$
$$+\,\mathsf{P}\left(\sup_{x\in L_n^{II}}|a_{n,x}-a_x|>c\,\omega_n\right)+\mathsf{P}\left(\sup_{\substack{b/2<|s|<R_n/\sqrt{\lambda}\\ |s-t|<c\,\max\{R_n\xi_n,\omega_n\}}}\left|\frac{F(s)}{F(t)}-1\right|>\varepsilon\right).$$

The first two summands on the right hand side vanish with $n \to \infty$ because of the first two formulas in (A.17). The argument in the last summand is non-random, and the probability is therefore equal to zero for n large by (18). Thus,

$$\sup_{x\in L_n^{II}}\left|\frac{F(a_x b)}{F(a_{n,x}b_n)}-1\right|\xrightarrow[n\to\infty]{\mathsf{P}}0. \tag{A.18}$$

Using similar argumentation we have that there is $c > 0$ with the property that for all $\varepsilon > 0$ and n large enough

$$\mathsf{P}\left(\sup_{x\in L_n^{II}}\left|\frac{F_n\left(a_{n,x}b_n\right)}{F\left(a_{n,x}b_n\right)}-1\right|>\varepsilon\right)$$
$$\leq \mathsf{P}\left(\sup_{x\in L_n^{II}}|a_{n,x}b_n-a_x b|>c\,\omega_n\right)+\mathsf{P}\left(\sup_{|a_{n,x}b_n|<R_n/\sqrt{\lambda}+c\,\omega_n}\left|\frac{F_n\left(a_{n,x}b_n\right)}{F\left(a_{n,x}b_n\right)}-1\right|>\varepsilon\right)$$
$$\leq \mathsf{P}\left(\sup_{x\in L_n^{II}}|a_{n,x}b_n-a_x b|>c\,\omega_n\right)+\mathsf{P}\left(\sup_{|t|<2R_n/\sqrt{\lambda}}\left|\frac{F_n\left(t\right)}{F\left(t\right)}-1\right|>\varepsilon\right),$$

and the last expression tends to zero in as $n \to \infty$ thanks to the second rate in (A.17) and (17). Thus,

$$\sup_{x\in L_n^{II}}\left|\frac{F_n\left(a_{n,x}b_n\right)}{F\left(a_{n,x}b\right)}-1\right|\xrightarrow[n\to\infty]{\mathsf{P}}0. \tag{A.19}$$

Altogether, we can start from (A.16) and bound

$$\sup_{x\in L_n^{II}}\left|\frac{F_n\left(a_{n,x}b_n\right)}{F\left(a_x b\right)}-1\right|\leq \sup_{x\in L_n^{II}}\left|\frac{F(a_{n,x}b_n)}{F(a_x b)}\right|\left(\sup_{x\in L_n^{II}}\left|\frac{F_n\left(a_{n,x}b_n\right)}{F\left(a_{n,x}b_n\right)}-1\right|+\sup_{x\in L_n^{II}}\left|\frac{F(a_x b)}{F(a_{n,x}b_n)}-1\right|\right).$$

The first term on the right hand side is bounded in probability due to (A.18). The summands vanish in probability thanks to (A.19) and (A.18), respectively. The theorem is proved.

A.11. **Proof of Remark (R8).** Suppose first that $F(t) \geq c\,|t|^{\gamma}$ for some $c > 0$, t small enough and $\gamma < 0$. Consider $R_n = \mathcal{O}(n^{\alpha})$ for $\alpha > 0$. We have

$$\frac{\omega_n}{F\left(-R_n/\sqrt{\lambda}\right)}\leq\frac{\left(R_n^{d-1}/\sqrt{n}\right)^{2/(d+1)}}{c\left|R_n/\sqrt{\lambda}\right|^{\gamma}}=\mathcal{O}\left(n^{2\alpha(d-1)/(d+1)-1/(d+1)-\alpha\gamma}\right).$$

For the right hand side to be $o(1)$, it is sufficient that $\alpha < (2(d-1) - \gamma(d+1))^{-1}$.

If $F(t) \geq c\, e^{-|t|^\gamma}$ for some $c > 0$, $\gamma > 0$ and all t small enough, we get for $R_n \leq \left(\left(\frac{1}{d+1} - \varepsilon\right)\log(n)\right)^{1/\gamma}\sqrt{\lambda}$ and for $\varepsilon > 0$

$$\frac{\omega_n}{F\left(-R_n/\sqrt{\lambda}\right)} \leq \frac{\left(R_n^{d-1}/\sqrt{n}\right)^{2/(d+1)}}{c\, e^{-\left|R_n/\sqrt{\lambda}\right|^\gamma}} \leq \frac{R_n^{2(d-1)/(d+1)}}{c\, n^\varepsilon} = o(1).$$

For $F(t) \geq c \exp\left(-e^{|t|^\gamma}\right)$ for some $c > 0$, $\gamma > 0$ and all t small enough, $R_n \leq \left(\log\left(\frac{1}{d+1} - \varepsilon\right) + \log\log n\right)^{1/\gamma}\sqrt{\lambda}$ with $\varepsilon > 0$ gives

$$\frac{\omega_n}{F\left(-R_n/\sqrt{\lambda}\right)} \leq \frac{\left(R_n^{d-1}/\sqrt{n}\right)^{2/(d+1)}}{c \exp\left(-e^{\left|R_n/\sqrt{\lambda}\right|^\gamma}\right)} \leq \frac{R_n^{2(d-1)/(d+1)}}{c\, n^\varepsilon} = o(1).$$

A.12. **Proof of Theorem 9.** The logarithm of the density of $P^{(j)}$ at $x \in \mathbb{R}^d$ can be written as

$$\log\left(f_j(x)\right) = -\log\left(\sqrt{(2\pi)^d}\right) - \log\left(\sqrt{|\Sigma_j|}\right) - \frac{1}{2}\mathsf{d}_{\Sigma_j}\left(x, \mu_j\right)^2.$$

By (9) we know that for any $0 < \delta < 1/2$

$$\begin{aligned}\operatorname{vol}_d\left(P_\delta^{(j)}\right) &= \operatorname{vol}_d\left(\left\{y \in \mathbb{R}^d \colon \Phi\left(-\mathsf{d}_{\Sigma_j}\left(y, \mu_j\right)\right) \geq \delta\right\}\right) = \operatorname{vol}_d\left(\left\{y \in \mathbb{R}^d \colon \mathsf{d}_{\widetilde{\Sigma}_j}\left(y, \mu_j\right) \leq 1\right\}\right)\\ &= \operatorname{vol}_d\left(\widetilde{\Sigma}_j^{1/2} B^d + \mu_j\right) = \left|\widetilde{\Sigma}_j^{1/2}\right| \operatorname{vol}_d\left(B^d\right) = \Phi^{-1}\left(1-\delta\right)^d \sqrt{|\Sigma_j|}\operatorname{vol}_d\left(B^d\right),\end{aligned}$$

where $\widetilde{\Sigma}_j = \Phi^{-1}\left(1-\delta\right)^2 \Sigma_j$. Thus,

$$2\log\left(\pi_j f_j(x)\right) = 2\log\left(\frac{\Phi^{-1}\left(1-\delta\right)^d \operatorname{vol}_d\left(B^d\right)}{\sqrt{(2\pi)^d}}\right) + 2\log\left(\frac{\pi_j}{\operatorname{vol}_d\left(P_\delta^{(j)}\right)}\right) - \mathsf{d}_{\Sigma_j}\left(x, \mu_j\right)^2,$$

and $\pi_1 f_1(x) > \pi_2 f_2(x)$ if and only if (19) is true.

The uniform consistency follows from Theorem 6, formula (A.5), and Lemma A.2.

Appendix B. Additional simulations and results

B.1. **Robust classification.**

B.1.1. *Bivariate normal distribution, location difference.* We repeat the same classification experiment as in Section 5.3.1, with $P^{(1)} = P_X, P^{(2)} = P_{X+(2,2)^\intercal}, P^{(3)} = P_{X+(20,20)^T}$. This accounts for classification in presence of only location difference. The results are summarized in Figure B.1 and Table B.1. We observe similar results as in Section 5.3.1: the optimal (Bayes) error rate is nearly achieved by the illumination-based approach and the classical QDA. The approach based on the refined depth performs worse, especially in the extremes, and it is very sensitive to possible contamination.

B.1.2. *Bivariate elliptical distribution, location difference.* Finally, consider the experiment from Section 5.3.2 with $P^{(1)} = P_Y, P^{(2)} = P_{Y+(2,2)^\intercal}, P^{(3)} = P_{X+(20,20)^\intercal}$. Our results are summarized in Figure B.2 and Table B.2. We observe that in the case with no contamination, the illumination-based approach and the classical QDA, used only as a reference method here, perform slightly better than the method based on the refined depth. If some contamination is present, the robust QDA appears to outperform both competitors.

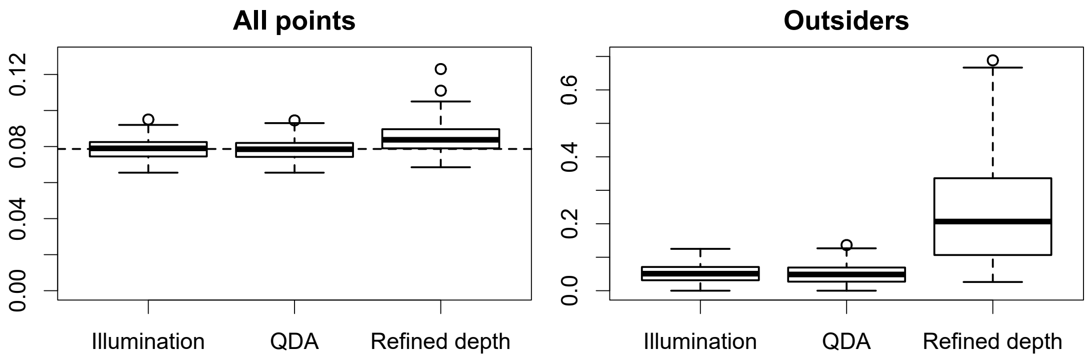

Figure B.1. *Misclassification rates, based on 100 replications of the experiment with two bivariate normal distributions with different location and same scale. Based on all testing points (left panel) or the outsiders (right panel). Dashed horizontal line in the left panel corresponds to the theoretical Bayes error rate.*

	All points			Outsiders		
	Illumination	QDA	Ref. depth	Illumination	QDA	Ref. depth
0 %	0.079 (0.006)	0.079 (0.006)	0.085 (0.009)	0.054 (0.030)	0.051 (0.031)	0.236 (0.165)
1 %	0.079 (0.006)	0.089 (0.007)	0.101 (0.010)	0.039 (0.032)	0.059 (0.056)	0.236 (0.235)
5 %	0.081 (0.006)	0.118 (0.010)	0.113 (0.011)	0.047 (0.038)	0.065 (0.052)	0.216 (0.109)
10 %	0.087 (0.006)	0.135 (0.012)	0.120 (0.011)	0.066 (0.045)	0.069 (0.054)	0.210 (0.117)

Table B.1. *Average misclassification rates and their standard deviations (in brackets), bivariate normal distributions with different location and same scale, level of contamination in one of the training samples ranging from 0 to 10 %. Based on 100 replications of the experiment and all testing points (left part) and outsiders (right part), respectively.*

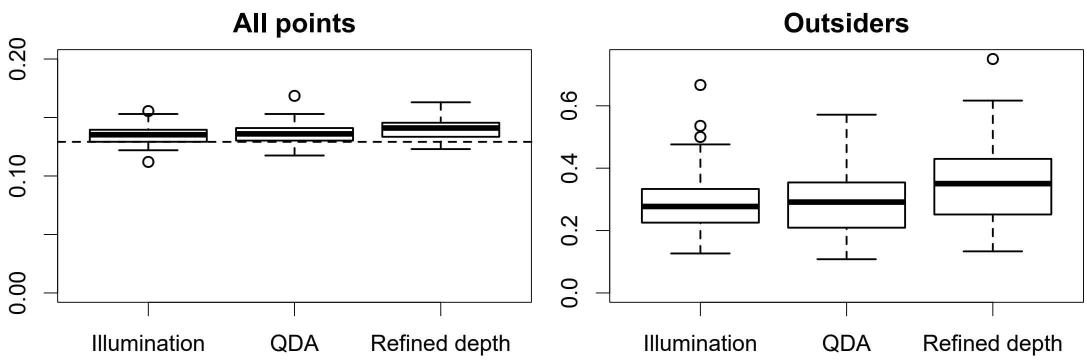

Figure B.2. *Misclassification rates, based on 100 replications of the experiment with two bivariate elliptical distributions with different location and same scale. Based on all testing points (left panel) or the outsiders (right panel). Dashed horizontal line in the left panel corresponds to the theoretical Bayes error rate.*

Appendix C. R source code

```
library(TukeyRegion)
library(geometry)
```

	All points			Outsiders		
	Illumination	QDA	Ref. depth	Illumination	QDA	Ref. depth
0 %	0.135 (0.007)	0.136 (0.008)	0.140 (0.009)	0.286 (0.089)	0.296 (0.109)	0.353 (0.127)
1 %	0.134 (0.008)	0.168 (0.015)	0.142 (0.010)	0.317 (0.111)	0.374 (0.119)	0.382 (0.136)
5 %	0.140 (0.007)	0.242 (0.017)	0.162 (0.013)	0.285 (0.092)	0.354 (0.116)	0.384 (0.107)
10 %	0.165 (0.014)	0.265 (0.020)	0.175 (0.014)	0.311 (0.110)	0.336 (0.114)	0.365 (0.120)

Table B.2. *Average misclassification rates and their standard deviations (in brackets), bivariate elliptical distributions with different location and same scale, level of contamination in one of the training samples ranging from 0 to 10 %. Based on 100 replications of the experiment and all testing points (left part) and the outsiders (right part), respectively.*

```
Illumination = function(X,x,alpha){
  # X: n-times-d matrix of the sample points (n points in d dimensions)
  # x: vector of length d whose illumination is computed
  # alpha: cut-off value for the illumination
  # returns
  # I: the illumination of x onto the depth central region
  # (volume of the convex hull of points with hD at least alpha, and x)
  # volPa: volume of the depth central region
  # (volume of the region of points whose hD is at least alpha)
  Pa = TukeyRegion(X,depth=alpha*nrow(X),retVertices=TRUE,retVolume=TRUE)
  volPax = convhulln(rbind(Pa$vertices,x),options="FA")$vol
  return(list(I=volPax,volPa=Pa$volume))
}
```

References

[1] Brunel, V.-E. (2019). Concentration of the empirical level sets of Tukey's halfspace depth. *Probab. Theory Related Fields*, 173(3–4):1165–1196.

[2] Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Qualifying paper, Harvard University.

[3] 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.

[4] Dyckerhoff, R. (2017). Convergence of depths and depth-trimmed regions. *arXiv preprint arXiv:1611.08721.*

[5] Nagy, S., Schütt, C., and Werner, E. M. (2019). Halfspace depth and floating body. *Stat. Surv.*, 13:52–118.

[6] Pólya, G. and Szegő, G. (1998). *Problems and theorems in analysis. I.* Classics in Mathematics. Springer-Verlag, Berlin. Series, integral calculus, theory of functions, Translated from the German by Dorothee Aeppli, Reprint of the 1978 English translation.

[7] Schneider, R. (2014). *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, expanded edition.

[8] van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge.

[9] Werner, E. M. (2006). Floating bodies and illumination bodies. In *Integral geometry and convexity*, pages 129–140. World Sci. Publ., Hackensack, NJ.