# Methods for Estimation of Convex Sets

## Victor-Emmanuel Brunel

*Abstract.* In the framework of shape constrained estimation, we review methods and works done in convex set estimation. These methods mostly build on stochastic and convex geometry, empirical process theory, functional analysis, linear programming, extreme value theory, etc. The statistical problems that we review include density support estimation, estimation of the level sets of densities or depth functions, nonparametric regression, etc. We focus on the estimation of convex sets under the Nikodym and Hausdorff metrics, which require different techniques and, quite surprisingly, lead to very different results, in particular in density support estimation. Finally, we discuss computational issues in high dimensions.

*Key words and phrases:* Convex body, set estimation, Nikodym metric, Hausdorff metric, support function.

## 1. PRELIMINARIES

### 1.1 Introduction

In nonparametric inference, the unknown object of interest cannot be described in terms of a finite number of parameters. Examples include density estimation, nonparametric and high dimensional regression, support estimation, etc. Since the number of observations is only finite, it is necessary to make assumptions on the object of interest in order to make statistical inference significant. Two types of assumptions are most common: smoothness assumptions and shape constraints. A smoothness assumption usually imposes differentiability up to some fixed order, with bounded derivatives (the reader could find an introduction to the estimation of smooth density or regression functions in [88], Chapter 1; [61] imposes smoothness assumptions on the boundary of the support of an unknown density or on the boundary of an unknown set in image reconstruction from random observations). Shape constraints rather impose conditions such as monotonicity, convexity, log-concavity, etc. (e.g., [53] assumes log-concavity of the unknown density; [23] imposes monotonicity or more general shape constraints on the unknown regression function; [55] imposes a monotonicity or a convexity constraint on the boundary of

*Victor-Emmanuel Brunel is Applied Mathematics Instructor, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA (e-mail: vebrunel@mit.edu).*

the support of the unknown density; [59, 16] impose convexity on the support of the unknown distribution).

Smoothness is a quantitative condition, whereas a shape constraint is usually qualitative. Smoothness classes of functions or sets depend on meta parameters, such as the number of existing derivatives or upper bounds on some functional norms. However, in statistical applications, these meta parameters are unlikely to be known to the practitioner. Yet, statistical inference usually requires to choose tuning parameters that depend on these meta parameters. One way to overcome this issue is to randomize the tuning parameters and apply data driven adaptive procedures such as cross validation. However, such procedures are often technical and computationally costly. On the opposite, shape constraints usually do not introduce extra parameters, which makes them particularly attractive.

Many different shape constraints can be imposed on sets. For instance, [85, 56, 55] consider boundary fragments, which are the subgraphs of positive functions defined on a hypercube (or, more generally, on a metric space). Shape constraints on such sets directly translate into shape constraints on their edge functions. For general sets, convexity is probably the most simple shape constraint, even though it leads to a very rich field in geometry. Convexity can be extended to the notion of $r$-convexity, where an $r$-convex set is the complement of the union of open Euclidean balls of radius $r$, $r > 0$ (see, e.g., [63] and [77, 69] for set estimation under $r$-convexity and, more generally, [29, 30] for broader

shape constraints in set estimation). Informally, convexity is the limit of $r$-convexity as $r$ goes to $\infty$. In set estimation, if it is assumed that the unknown set is $r$-convex for some $r > 0$, the meta parameter $r$ may also be unknown to the practitioner and [78] defines a data-driven procedure that adapts to $r$. In the present article, we only focus on convexity, which is a widely treated shape constraint in statistics. On top of convexity, two additional constraints are common in statistics: the rolling ball condition and standardness. A convex set $G$ is said to satisfy the $r$-rolling ball condition ($r > 0$) if, for all $x$ on the boundary of $G$, there is a Euclidean ball $B(a, r)$ such that $x \in B(a, r) \subseteq \bar{G}$, where $\bar{G}$ is the closure of $G$ (see [92, 93] for characterizations of the rolling ball condition, connections with $r$-convexity and statistical applications in set estimation). An equivalent condition is that the complement of $G$ has reach at least $r$. The reach of a set is the supremum of all positive numbers $\varepsilon$ such that any point within a distance $\varepsilon$ of that set has a unique metric projection onto the closure of that set (see [84], Definition 11). A convex set $G$ is called $\nu$-standard ($\nu \in (0, 1)$) if for all $x$ on its boundary, $\mathsf{Vol}(G \cap B(x, \varepsilon)) \geq \nu \mathsf{Vol}(B(x, \varepsilon))$, for all $\varepsilon > 0$ small enough. This roughly means that the set $G$ does not have peaks.

In general, two main types of convex bodies are distinguished in the literature.

- Convex bodies with smooth boundary: The boundary $\partial G$ of a convex body $G$ is *smooth* if for all $x \in \partial G$, $G$ has a unique supporting hyperplane that contains $x$. In that case, let $H_x$ be the unique supporting hyperplane containing $x$ and let $\eta_x$ be the unit vector orthogonal to $H_x$ and pointing towards the inside of $G$. Identify the $(d-1)$-dimensional linear subspace $H_x - x = \{z - x : z \in H_x\}$ with $\mathbb{R}^{d-1}$; Then every $y \in \partial G$ that is in some neighborhood of $x$ can be written uniquely as $y = x + t + f_x(t)\eta_x$, where $t \in H_x - x$ and $f_x$ is a nonnegative convex function defined in a neighborhood of $0$ in $H_x - x$. If the Hessian of $f_x$ at $0$ is positive definite, $\partial G$ is said to have *positive curvature* at $x$. Otherwise, $\partial G$ has *zero curvature* at $x$.

- Convex polytopes: A convex polytope (in short, a polytope) is the convex hull of finitely many points in $\mathbb{R}^d$. By the Minkowski–Weyl theorem, a polytope can also be represented as the intersection of finitely many closed half-spaces. The supporting hyperplane of a polytope $P$ containing $x \in \partial P$ is unique if $x$ is not in a $k$-dimensional face of $P$ for some $k \leq d - 2$, and $P$ has zero curvature at all such boundary points $x$.

We refer the readers who are interested in learning more about convex bodies to [80], and to [96] for a comprehensive study of convex polytopes.

In the field of nonparametric statistics, the problem of set estimation arose essentially with the works [42] (*On a geometric estimation problem*) and [27], which deal with the estimation of the support of a density in a general setup. A simple and natural estimator of the support of an unknown density was introduced in [31], where the estimator is defined as the union of small Euclidean balls centered around the data points. In fact, this estimator is equal to the support of a kernel density estimator for the kernel that is the indicator function of the Euclidean unit ball.

The scope of this survey is the estimation of convex sets. We aim to give an exposition of several methods that build on stochastic and convex geometry, empirical process theory, functional analysis, linear programming, order statistics and extreme value theory, etc. Different models associated with the estimation of convex sets include density support estimation [57, 56, 55, 16, 17, 13], density level set estimation [51, 70, 86], inverse problems in density support estimation [19], estimation of the support of a regression function [58, 85, 14], estimation of the level sets of the Tukey depth function [18], estimation of support functions [40, 48], etc.

Throughout this survey, a *set estimator* is a set-valued statistic, that is, a set which depends on the observed random variables. A precise definition would be necessary in order to rule out measurability issues. However, in order to keep the focus on convex set estimation, we rather choose not to mention these issues, and all probabilities (resp., expectations) should be understood as outer probabilities (resp., expectations). For detailed accounts on set-valued random variables, we refer to [64].

Before going more into the details, let us introduce some notation and definitions.

### 1.2 Notation and Definitions

In the sequel, $d$ is a positive integer, standing for the ambient dimension. For a positive integer $p$, the closed $p$-dimensional Euclidean ball with center $a \in \mathbb{R}^p$ and radius $r \geq 0$ is denoted by $B_p(a, r)$. If $p = d$, we may omit the subscript $p$. The $(p - 1)$-dimensional unit sphere is denoted by $\mathbb{S}^{p-1}$ and the volume of the $p$-dimensional unit Euclidean ball is denoted by $\beta_p$. The Euclidean norm in $\mathbb{R}^d$ is denoted by $\| \cdot \|$, the Euclidean distance is $\rho$ and we write $\langle \cdot, \cdot \rangle$ for the canonical dot product.

A convex body $G \subseteq \mathbb{R}^d$ is a compact and convex set with nonempty interior. We denote by $\mathcal{K}_d$ the collection of all convex bodies and by $\mathcal{K}_d^{(1)}$ the collection of all convex bodies included in $B(0, 1)$. The support function $h_G$ of a convex body $G$ is defined as $h_G(u) = \max\{\langle u, x \rangle : x \in G\}$, for all $u \in \mathbb{S}^{d-1}$: It is the signed distance of tangent hyperplanes to the origin.

The volume of a measurable set $A \subseteq \mathbb{R}^d$ is denoted by $|A|$.

The Nikodym distance between two measurable sets $K, L \subseteq \mathbb{R}^d$ is the volume of their symmetric difference: $\mathsf{d}_\triangle(K, L) = |K \triangle L|$. The Hausdorff distance between any two sets $K, L \subseteq \mathbb{R}^d$ is defined as $\mathsf{d}_\mathsf{H}(K, L) = \inf\{\varepsilon \geq 0 : K \subseteq L + \varepsilon B(0, 1), L \subseteq K + \varepsilon B(0, 1)\}$.

The cardinality of a finite set $A$ is denoted by $\#A$.

When i.i.d. random points $X_1, \ldots, X_n \in \mathbb{R}^d$ have a density $f$ with respect to the Lebesgue measure, we denote by $\mathbb{P}_f$ their joint distribution and by $\mathbb{E}_f$ the corresponding expectation operator, where we omit the dependency on $n$ for simplicity. When $f$ is the uniform density on a compact set $G$, we simply write $\mathbb{P}_G$ and $\mathbb{E}_G$. The convex hull of $X_1, \ldots, X_n$ is denoted by $\hat{K}_n$.

In this article, most, if not all, set-valued estimators are polytopes that depend on a finite random sample. Nonetheless, in order to be consistent with the literature, we reserve the name *random polytope* for $\hat{K}_n$ only.

## 1.3 Outline

In order to assess the quality of a set estimator, the Nikodym and the Hausdorff metrics are most commonly used. Depending on which of these two metrics is to be used, the techniques in estimation of convex sets may differ a lot.

Section 2 is devoted to the estimation of convex sets under Nikodym-type metrics, especially in density support estimation. We first review essential properties of random polytopes and we relate them to the problem of support estimation under the Nikodym metric. We also recall well-known results on the covering numbers of classes of convex bodies and show how these can be used in order to obtain deviation inequalities in convex support estimation. Then we review extensions of these results to the estimation of density level sets under convexity and we discuss other convex set estimation problems under the Nikodym metric.

In Section 3, we switch to the estimation of convex bodies under the Hausdorff metric. An elementary, yet essential result, stated in Lemma 2, shows that the Hausdorff distance between two convex bodies can be computed through their respective support functions.

We review important properties related to the support functions of convex bodies and we show how they apply to the estimation of convex sets under the Hausdorff metric.

Finally, in Section 4, we briefly discuss the computational aspects of convex set estimation in high dimensions. We show, through two examples, how to reduce the computational cost without affecting the rate of convergence of convex set estimators.

## 2. ESTIMATION OF CONVEX SETS UNDER THE NIKODYM METRIC

### 2.1 Random Polytopes and Density Support Estimation

The most common representation of random polytopes consists of taking the convex hull of i.i.d. random points in $\mathbb{R}^d$. Stochastic and convex geometry have provided powerful tools to understand the properties of random polytopes, since the seminal works [74, 75]. In these two papers, $d = 2$ and the random polygon is the convex hull of $n$ i.i.d. random points with the uniform distribution in a planar convex body. The expectation of the missing area and of the number of vertices of the random polygon are computed, up to negligible terms as $n$ goes to infinity. The results substantially depend on the structure of the boundary of the support. Namely, the expected missing area decreases significantly faster when the support is itself a polygon than when its boundary has positive curvature everywhere. The missing area is exactly the Nikodym distance between the random polygon and the support of the random points. Hence, [74, 75] give an approximate value of the risk of the random polygon as an estimator of the convex support. Later, much effort has been devoted to extend these results to higher dimensions, starting with [35], that proves integral formulas for the expected missing volume, surface area, number of vertices, etc. in dimension 3. Among most general results, a ground breaking one is due to [5]. Define the $\varepsilon$-*floating body* of a convex body $G \in \mathcal{K}_d$ as the set of points $x \in G$ such that any closed half-space $H \subseteq \mathbb{R}^d$ containing $x$ has an intersection with $G$ whose volume is at least a fraction $\varepsilon$ of the total volume of $G$, that is, satisfies $|G \cap H| \geq \varepsilon|G|$, where $\varepsilon \in (0, 1)$ (see [34, 9, 82]). The $\varepsilon$-*wet part* of $G$, denoted by $G(\varepsilon)$, is the complement of the $\varepsilon$-floating body of $G$ in $G$. If one thinks of $G$ as an iceberg seen from above, the floating body is the part of $G$ that is above the surface of the water, whereas the wet part is the immersed part of the iceberg.

THEOREM 1 ([5]). *Let $G \in \mathcal{K}_d$ have volume one. Then*

$$c_1 |G(1/n)| \leq \mathbb{E}_G[|G \setminus \hat{K}_n|]$$
$$\leq c_2(d) |G(1/n)| \quad \forall n \geq n_0(d),$$

*where $c_1$ is a universal positive constant, $c_2(d)$ is a positive constant that depends on $d$ only and $n_0(d)$ is a positive integer that depends on $d$ only.*

As a consequence, computing the expected missing volume of $\hat{K}_n$ asymptotically reduces to computing the volume of the $(1/n)$-wet part of $G$, which is no longer a probabilistic question. In addition, it is also known [5] that if $G$ has volume one and $\varepsilon$ goes to zero, $|G(\varepsilon)|$ is of the order at least $\varepsilon \ln(1/\varepsilon)$ and at most $\varepsilon^{2/d+1}$ [5]. The former rate is achieved when $G$ is a polytope whereas the latter rate is achieved when $G$ has a smooth boundary with positive curvature everywhere.

In fact, when $G$ is a smooth convex body with positive curvature everywhere, it is shown in [81] that

$$(1) \quad n^{2/(d+1)} \mathbb{E}_G\left[\frac{|G \setminus \hat{K}_n|}{|G|}\right] \longrightarrow c(d, G), \quad n \to \infty,$$

where $c(d, G)$ is an explicit positive constant that depends on $d$ and $G$ and that is affine invariant in $G$, that is, $c(d, G) = c(d, T(G))$ for all invertible affine transormations $T$. [81] actually shows that this convergence holds for all convex bodies $G$, by noting that all convex bodies have a unique supporting hyperplane at almost all their boundary points (e.g., almost all boundary points of a polytope lie on a $(d-1)$-dimensional face), and where $c(d, G)$ is equal to zero if and only if $\partial G$ has zero curvature almost everywhere (e.g., if $G$ is a polytope).

An interesting result, due to [46], shows that the quantity $\mathbb{E}_G[\frac{|G \setminus \hat{K}_n|}{|G|}]$ is maximum when $G$ is an ellipsoid. In that case, it can be derived from [81] that the constant $c(d, G)$ in (1) is of the order $d^{d+o(d)}$, as $d$ becomes large. As a consequence, when the dimension $d$ becomes too large, the random polytope $\hat{K}_n$ performs poorly as an estimator of $G$ in the worst case, because it suffers the curse of dimensionality, both in the rate $n^{-2/(d+1)}$ and in the constant factor $d^{d+o(d)}$. Yet, it is known that the rate $n^{-2/(d+1)}$ cannot be improved in a minimax sense. The following result is proven in [16] where, for two sequences $a_n$ and $b_n$ of positive numbers, we write $a_n \lesssim_\theta b_n$ if $a_n \leq c(\theta) b_n$, $\forall n \geq 1$, for some positive constant $c(\theta)$ that depends on a parameter $\theta$. In the sequel, we also write $\lesssim$ with no subscript if the involved constant is universal.

THEOREM 2 ([16]). *The following inequalities hold*:

$$n^{-\frac{2}{d+1}} \lesssim_d \inf_{\tilde{G}_n} \sup_{G \in \mathcal{K}^d} \mathbb{E}_G\left[\frac{|G \triangle \tilde{G}_n|}{|G|}\right]$$

$$\leq \sup_{G \in \mathcal{K}^d} \mathbb{E}_G\left[\frac{|G \triangle \hat{K}_n|}{|G|}\right] \lesssim_d n^{-\frac{2}{d+1}},$$

*where the infimum is taken over all estimators $\tilde{G}_n$ based on $n$ i.i.d. observations.*

As a consequence, the random polytope $\hat{K}_n$ is rate optimal over the class $\mathcal{K}_d$ in a minimax sense, with respect to the Nikodym metric. The upper bound in Theorem 2 is a direct consequence of Theorem 1, together with Groemer's result [46]: It suffices to evaluate the volume of the $(1/n)$-wet part of a Euclidean ball of volume one. However, it is not clear that $\hat{K}_n$ is optimal in terms of the constant factors that become exponentially large with the dimension. Note that $\hat{K}_n \subseteq G$ with probability one, hence, $\hat{K}_n$ always underestimates the support $G$. This is why the estimation of $G$ through a dilation of $\hat{K}_n$ could be appealing. It has been considered, for example, [76] in the planar case for Poisson polytopes, and in [65] for $d \leq 2$, but only heuristics are given in the general case, except for the estimation of the volume of $G$ in [1]. [1] poses the question of the performance of a dilated version of $\hat{K}_n$ compared to that of $\hat{K}_n$ itself, but the question remains open.

Note that the lower bound in Theorem 2 is also used in log-concave density estimation. The uniform density on any convex body is log-concave, and for any two convex bodies $G$ and $G'$ of volume 1, the corresponding uniform densities $f_G$ and $f_{G'}$ satisfy $\|f_G - f_{G'}\|_2^2 = |G \triangle G'|$, where $\| \cdot \|$ stands for the $L^2$ norm with respect to the Lebesgue measure in $\mathbb{R}^d$. Hence, some proof techniques for lower bounds on minimax risks in [53] are based on similar arguments as those used to prove the lower bound in Theorem 2.

## 2.2 Adaptation to Polytopal Supports

As we already mentioned earlier, an attractive feature of most shape constraints is that no meta parameters are needed to describe the objects of interest, unlike in smoothness classes. Nonetheless, classes of functions or sets with a shape constraint usually contain parametric subclasses that correspond to simpler structures, which may depend on meta parameters. For instance, classes of monotone (resp. convex) functions contain piecewise constant (resp. affine) functions. A desirable property of an estimator is adaptation to these simpler structures: If the unknown object

belongs to a subparametric class, then the rate of convergence of the estimator should be nearly as good as that of an estimator that would not be agnostic to that simpler structure. In recent years, there have been considerable efforts put in understanding this automatic adaptive features in shape constrained estimators [52, 24, 95, 23, 8, 49, 50].

Turning to the case of convex set estimation, the class $\mathcal{K}_d$ contains subclasses of polytopes with bounded number of vertices, hence, whose support functions are piecewise linear with a bounded number of pieces, each piece corresponding to a vertex.

For the estimation of the convex support of a uniform distribution, the random polytope $\hat{K}_n$ is the maximum likelihood estimator on the class $\mathcal{K}_d$. Indeed, the likelihood function is given by $L_n(C) := |C|^{-n} \prod_{i=1}^{n} \mathbb{1}_{X_i \in C} = |C|^{-n} \mathbb{1}_{\hat{K}_n \subseteq C}$, for all $C \in \mathcal{K}_d$ which is maximized when $C = \hat{K}_n$ (note that $\hat{K}_n \in \mathcal{K}_d$ with probability 1 as long as $n \geq d+1$). Recall that, as a consequence of Theorem 2, in the Nikodym metric, $\hat{K}_n$ estimates $G$ at the speed $n^{-2/(d+1)}$ in the worst case, that is, when $G$ has a smooth boundary. When $G$ is a polytope, Theorem 1 implies that $\hat{K}_n$ estimates $G$ at a much faster speed, namely, $n^{-1}(\ln n)^{d-1}$. A more refined (but not uniform in $G$) result was proven in [4]. For a polytope $P \subseteq \mathbb{R}^d$, let $T(P)$ be the number of flags of $P$, that is, the number of increasing sequences $F_0 \subseteq F_1 \subseteq \cdots \subseteq F_{d-1} \subseteq F_d = P$ of faces of $P$ where $F_k$ is a $k$-dimensional face of $P$, $k = 0, \ldots, d$. For example, $F(P) = 2^d d!$ if $P$ is the $d$-dimensional hypercube, or $F(P) = d!$ if $P$ is the $(d-1)$-dimensional simplex.

THEOREM 3 ([4]). *Let $G = P$ be a polytope. Then*

$$\lim_{n \to \infty} \frac{n}{(\ln n)^{d-1}} \mathbb{E}_P \left[ \frac{|P \setminus \hat{K}_n|}{|P|} \right] = \frac{T(P)}{(d+1)^{d-1}(d-1)!}.$$

In particular, if $G$ is a polytope, then there is a significant gain in the speed of convergence of $\hat{K}_n$, which becomes nearly parametric up to logarithmic factors. In other words, $\hat{K}_n$ adapts to polytopal supports. However, its rate still suffers the curse of dimensionality because of the $(\ln n)^{d-1}$ factor. In [16], it is shown that this rate is not optimal over subclasses of polytopes with given number of vertices in a minimax sense. The idea is that $\hat{K}_n$ maximizes the likelihood function over the class of all convex bodies, which would too rich if it was known in advance that $G$ is a polytope with a given number of vertices. If $G$ has at most $r$ vertices, where $r \geq d+1$ is known *a priori*, [16] considers the maximum likelihood estimator over the corresponding subclass of polytopes. Namely, denote by

$\mathcal{P}_r$ the class of all polytopes with at most $r$ vertices. The maximum likelihood estimator of $P$ in the class $\mathcal{P}_r$ is defined as $\hat{P}_n^{(r)} \in \arg\max_{Q \in \mathcal{P}_r} |P|^{-n} \mathbb{1}_{X_i \in P, \forall i = 1, \ldots, n}$: It is a polytope with at most $r$ vertices that contains $X_1, \ldots, X_n$ and has minimum volume. Note that, unlike $\hat{K}_n$, the maximum likelihood estimator $\hat{P}_n^{(r)}$ may not be uniquely defined. However, the rate of this estimator no longer suffers the curse of dimensionality when $G \in \mathcal{P}_r$.

THEOREM 4 ([16]). *Let $r \geq d+1$. Then*

$$\frac{1}{n} \lesssim_d \inf_{\tilde{G}_n} \sup_{P \in \mathcal{P}_n^{(r)}} \mathbb{E}_P \left[ \frac{|P \triangle \tilde{G}_n|}{|P|} \right]$$

$$\leq \sup_{P \in \mathcal{P}_n^{(r)}} \mathbb{E}_P \left[ \frac{|P \triangle \hat{P}_n^{(r)}|}{|P|} \right] \lesssim_d \frac{r \ln n}{n},$$

*where the infimum is taken over all estimators $\tilde{G}_n$ based on $n$ i.i.d. observations.*

In [16], a better lower bound is proven when $d = 2$, namely,

$$\inf_{\tilde{G}_n} \sup_{P \in \mathcal{P}_n^{(r)}} \mathbb{E}_P \left[ \frac{|P \triangle \tilde{G}_n|}{|P|} \right] \gtrsim \frac{r}{n}.$$

The proof of the upper bound in Theorem 4 builds on a simple discretization of the class $\mathcal{P}_n^{(r)}$, obtained by considering polytopes with vertices on a finite grid in $[0, 1]^d$, and applying similar methods to those presented in Section 2.4.

The estimator $\hat{P}_n^{(r)}$ is not computable in practice, but it gives a benchmark for the optimal rate in estimation of $P \in \mathcal{P}_r$, under the Nikodym metric. It is still not clear whether the logarithmic factor could be dropped in the upper bound (see [16], Section 3.2). A drawback of $\hat{P}_n^{(r)}$ is that it requires the knowledge of $r$, whereas $\hat{K}_n$ is completely agnostic to the facial structure of $G$. In order to fix this issue, [16] proposes a fully adaptive procedure and defines an estimator $\hat{P}_n^{\text{adapt}}$ that is agnostic of the facial structure of $G$ and yet performs at the same rate as $\hat{P}_n^{(r)}$ when $G \in \mathcal{P}_r$ for some integer $r \geq d+1$, and as $\hat{K}_n$ for general supports $G$ (see [16] and [15] for more details). However, the estimators $\hat{P}_n^{(r)}$ and $\hat{P}_n^{\text{adapt}}$ are not computationally tractable, and when the dimension $d$ is not too large, the convex hull $\hat{K}_n$ is a more realistic estimator of $G$.

## 2.3 More Results on Random Polytopes

Even though this survey focuses on the statistical aspects of random polytopes, it is worth mentioning

many works that have tackled other probabilistic and geometric properties, which are indirectly related to the statistical estimation of the support and pose new statistical challenges.

In [74, 75], the expected number of vertices of $\hat{K}_n$ is computed in the planar case, up to some negligible terms as $n \to \infty$. [35] shows a very elegant identity which relates the missing volume of $\hat{K}_n$ and its number of vertices. It can be stated in a very general setup as follows. Given a sequence of i.i.d. random points $X_1, X_2, \ldots$ from some arbitrary probability measure $\mu$ in $\mathbb{R}^d$, let $\hat{K}_n$ be the convex hull of $X_1, \ldots, X_n$ and $N_n$ be the number of vertices of $\hat{K}_n$, for $n \geq 1$. Then, for all $n \geq 1$,

$$\mathbb{E}[1 - \mu(\hat{K}_n)] = \frac{\mathbb{E}[N_{n+1}]}{n+1}.$$

When $\mu$ is the uniform probability measure on a convex body $G \in \mathcal{K}_d$, this identity becomes $\mathbb{E}_G[\frac{|G \setminus \hat{K}_n|}{|G|}] = \frac{\mathbb{E}_G[N_{n+1}]}{n+1}$. Extensions of this inequality to higher moments of $|G \setminus \hat{K}_n|$ can be found in [20].

In [71], more results about the random polytope $\hat{K}_n$, involving variance bounds, are proven using Efron–Stein jackknife inequalities [36]. Very importantly, [71] compares the random polytope $\hat{K}_n$ to best polytopal approximations of smooth convex bodies. Let $G \in \mathcal{K}_d$ be a smooth convex body and let $G_N^*$ be a polytope with at most $N$ vertices, included in $G$, with minimum missing volume $|G \setminus G_N^*|$. With probability one,

$$\frac{|G \setminus G_{N_n}^*|}{|G \setminus \hat{K}_n|} \longrightarrow c_d, \quad n \to \infty,$$

where $c_d \leq 1$ is a positive constant that only depends on the dimension $d$. Moreover, [71] shows that $c_d \longrightarrow 1$ as $d \to \infty$. This shows that in high dimensions, with probability 1, $\hat{K}_n$ performs nearly as well as the best approximating inscribed polytope with same number of vertices, as $n$ becomes large.

Central limit theorems for the volume, number of vertices, or more generally, number of $k$-dimensional faces for $k \leq d - 1$, of random polytopes are proven in [73, 67, 68]. A worth mentioning technique that is used in the proofs of these central limit theorems could be called *Poissonization–de-Poissonization*. The idea is to first consider a Poisson polytope, defined as the convex hull of a Poisson point process [6] supported on a convex body, with growing intensity. These are somewhat easier to work with, and it is shown that their behavior is close enough to that of the random polytope $\hat{K}_n$.

Hence, the central limit theorems are first proven for the Poisson polytope, and the results are transferred to the random polytope by a *depoissonization step*. At a high level, this idea relies on the fact that if $G \in \mathcal{K}_d$ has volume one and if $\mathcal{X} = \{X_1, \ldots, X_N\}$ is a Poisson point process with constant intensity $n$ supported on $G$, then $N$ is a Poisson random variable with parameter $n$, hence, $\mathbb{E}[N] = \mathsf{Var}(N) = n$ and $N \approx n$ with high probability, and conditional on $N = n$, $X_1, \ldots, X_N$ are $n$ i.i.d. random points uniformly distributed in $G$.

Asymptotic properties of the intrinsic volumes of the random polytope are studied in [3, 72, 10] under different assumptions on the boundary of the underlying convex body. The intrinsic volumes of a convex body can be defined through Steiner formula [80], Section 4.1. For $G \in \mathcal{K}_d$ and $\varepsilon > 0$, let $G^\varepsilon = G + \varepsilon B(0, 1)$ be the set of all points $x \in \mathbb{R}^d$ that are within a distance at most $\varepsilon$ of $G$. Steiner formula states that $|G^\varepsilon|$ is a degree $d$ polynomial in $\varepsilon$. Namely, one can write, for all $\varepsilon > 0$,

$$(2) \qquad |G^\varepsilon| = \sum_{j=0}^{d} \beta_{d-j} v_j(G) \varepsilon^j,$$

where $v_j(G) \geq 0$ is called the $j$th intrinsic volume of $G$, for $j = 0, \ldots, d$. For instance, $v_0(G) = |G|$ is the volume of $G$, $v_1(G)$ is its surface area, $v_2(G)$ is its mean width and $v_d(G) = 1$. In [10], it is shown that if $G$ is a smooth convex body satisfying the $r$-rolling ball condition, then for all $j = 0, \ldots, d - 1$, $n^{2/(d+1)} \mathbb{E}_G[v_j(G) - v_j(\hat{K}_n)] \longrightarrow c(d, G)$ as $n \to \infty$, where $c(d, G)$ is a positive constant that depends on both the dimension and $G$. In particular, the plug-in estimator $v_j(\hat{K}_n)$ is a consistent estimator of $v_j(G)$, and it converges at the same rate as the rate of convergence of $\hat{K}_n$ in the Nikodym metric. Whether the plug-in estimator $v_j(\hat{K}_n)$ is an optimal estimator of $v_j(G)$ in a minimax sense is not known in general, except when $j = 0$, when the answer is negative. [41] considers the general problem of minimax estimation of the volume of the support of an unknown density, not necessarily uniform. In the particular case of the uniform density on an unknown convex body $G \in \mathcal{K}_d$, a sample splitting procedure is applied in order to correct the plug-in estimator $|\hat{K}_{n/2}|$. It is shown that the minimax risk for the estimation of the volume of $G \in \mathcal{K}_d^{(1)}$ is of order $n^{-\frac{d+3}{2d+2}}$, and this rate of convergence is attained by the explicit estimator given in [41]. The estimation of the volume of $G$ is also tackled in [1], where the same Poissonization–de-Poissonization procedure as mentioned above is used in order to obtain an estimator of $|G|$ based on a dilation of the random polytope $\hat{K}_n$.

## 2.4 Convex Bodies and Covering Numbers

Covering numbers provide a powerful tool to describe the complexity of a class. In empirical process theory, they are often used in order to bound the statistical performance of an estimator in expectation or with high probability, when the estimator is obtained by optimizing a criterion, such as the likelihood function.

Consider the problem of estimating the support $G$ of a uniform distribution, with $G \in \mathcal{K}_d$. Because the support of the likelihood function (see Section 2.2) depends on the unknown parameter itself, it is not valid to take its logarithm and it cannot be approached through the lens of empirical process theory. However, tools such as covering numbers can still be borrowed from that theory in order to prove deviation inequalities for $\hat{K}_n$.

Without loss of generality, one can assume that $B(a, d^{-1}) \subseteq G \subseteq B(0, 1)$ for some $a \in B(0, 1)$. This guarantees that $G \in \mathcal{K}_d^{(1)}$, which is a bounded class of convex bodies, and that $|G|$ is uniformly bounded from below. This is due to John's theorem (e.g., see [2]) and affine equivariance of $\hat{K}_n$. John's theorem (e.g., see [2]) implies the existence an invertible affine transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ and a point $a \in B(0, 1)$ with $B(a, d^{-1}) \subseteq TG \subseteq B(0, 1)$. Moreover, if we rather denote by $\hat{K}_n(X_1, \ldots, X_n)$ the convex hull of $X_1, \ldots, X_n$, then $\hat{K}_n(X_1, \ldots, X_n) = T^{-1}\hat{K}_n(TX_1, \ldots, TX_n)$. Since $X_1, \ldots, X_n$ are i.i.d. uniform random points in $G$, $TX_1, \ldots, TX_n$ are i.i.d. uniform random points in $TG$, and $\frac{|G \setminus \hat{K}_n(X_1, \ldots, X_n)|}{|G|} = \frac{|TG \setminus \hat{K}_n(TX_1, \ldots, TX_n)|}{|TG|}$. As a consequence, the rescaled risk $\frac{|G \setminus \hat{K}_n|}{|G|}$ is bounded from above by $\frac{|G \setminus \hat{K}_n|}{\beta_d}$ and we only need to bound $|G \setminus \hat{K}_n|$ uniformly on $\mathcal{K}_d^{(1)}$ instead of the whole unbounded class $\mathcal{K}_d$.

Let $\varepsilon > 0$ and let $\mathsf{d}(\cdot, \cdot)$ be a metric on $\mathcal{K}_d^{(1)}$ (e.g., Nikodym or Hausdorff distance). An $\varepsilon$-net of $\mathcal{K}_d^{(1)}$ with respect to the metric $\mathsf{d}(\cdot, \cdot)$ is a set $\mathcal{N} \subseteq \mathcal{K}_d^{(1)}$ such that for all $G \in \mathcal{K}_d^{(1)}$, there is $G^* \in \mathcal{N}$ with $\mathsf{d}(G, G^*) \leq \varepsilon$. The $\varepsilon$-covering number of $\mathcal{K}_d^{(1)}$ with respect to $\mathsf{d}(\cdot, \cdot)$ is the minimum cardinality of an $\varepsilon$-net of $\mathcal{K}_d^{(1)}$. The following theorem is an upper bound for the $\varepsilon$-covering number of $\mathcal{K}_d^{(1)}$ with respect to the Hausdorff distance. By [13], Lemma 2, the Nikodym distance is dominated by the Hausdorff distance uniformly on $\mathcal{K}_d^{(1)}$: $\mathsf{d}_\triangle(G_1, G_2) \leq \alpha \mathsf{d}_\mathsf{H}(G_1, G_2)$, for all $G_1, G_2 \in \mathcal{K}_d^{(1)}$, where $\alpha$ is a positive constant that depends on $d$ only.

This result is a direct consequence of Steiner formula for convex bodies (see Lemma 2). Hence, the following theorem also implies an upper bound for the $\varepsilon$-covering number of $\mathcal{K}_d^{(1)}$ with respect to the Nikodym metric.

THEOREM 5 ([12]). *Let $\varepsilon \in (0, 1)$. The $\varepsilon$-covering number of $\mathcal{K}_d^{(1)}$ with respect to the Hausdorff distance is at most $c_1 e^{c_2 \varepsilon^{-(d-1)/2}}$, for some positive constants $c_1$ and $c_2$ that depend on $d$.*

We also refer to Section 8.4 in [32] for more details on metric entropy for classes of convex sets. Building on this theorem combined with standard techniques from M-estimation and empirical processes (see, e.g., [90, 89]), [13] proves the following deviation inequality for $\hat{K}_n$, which holds uniformly for all $G \in \mathcal{K}_d$.

THEOREM 6 ([13]). *There exist positive constants $a_1, a_2$ and $a_3$ such that the following holds. Let $x \geq 0$ and $n \geq 2$ be an integer. For all $G \in \mathcal{K}_d$,*

$$\frac{|G \setminus \hat{K}_n|}{|G|} \leq a_1 n^{-\frac{2}{d+1}} + \frac{x}{n}$$

*with $\mathbb{P}_G$-probability at least $1 - a_2 e^{-a_3 x}$.*

Using the same techniques, more general deviation inequalities are proven in [13], when the density of the $X_i$'s is not uniform, but only supported on a convex body $G$. For all measurable sets $G, G' \subseteq \mathbb{R}^d$ and all densities $f$ on $\mathbb{R}^d$, denote by $\mathsf{d}_f(G, G') = \int_{G \setminus G'} f(x)\, dx$. Note that $\mathsf{d}_f(G, G') = \frac{|G \setminus G'|}{|G|}$ when $f$ is the uniform density on $G$.

THEOREM 7 ([13]). *There exist positive constants $C_1$ and $C_2$, that depend on $d$ only, such that the following holds. Let $x \geq 0$ and $n \geq 2$ be an integer. Let $G \in \mathcal{K}_d^{(1)}$ and $f$ be a density supported in $G$, with $f \leq M$ almost everywhere, for some positive number $M$. Let $X_1, \ldots, X_n$ be i.i.d. random points with density $f$ and $\hat{K}_n$ be their convex hull. Then*

$$\mathsf{d}_f(G, \hat{K}_n) \leq C_1(M + 1)n^{-2/(d+1)} + \frac{x}{n}$$

*with probability at least $1 - C_2 e^{-x}$.*

It is not known whether a similar upper bound would hold without the assumption that $f \leq M$ almost everywhere. This open problem amounts to the following open question. Let $\mu$ be any probability measure supported in a convex body $G \in \mathcal{K}_d^{(1)}$. Do there exist positive constants $c_1$ and $c_2$ that only depend on $d$, such that the $\varepsilon$-covering of $\mathcal{K}_d^{(1)}$ with respect to the metric $d(G_1, G_2) = \mu(G_1 \triangle G_2)$, $G_1, G_2 \in \mathcal{K}_d^{(1)}$ is bounded

from above by $c_1 e^{c_2 \varepsilon^{-(d-1)/2}}$, for all $\varepsilon \in (0, 1)$? If $\mu$ has a bounded density $f$ with respect to the Lebesgue measure, the answer is positive, and it is a consequence of Theorem 8 below.

In the uniform case, concentration inequalities for $\hat{K}_n$ were proven in [91], using geometric techniques. However, constants were not explicit and depended on the support $G$, hence, could not be used in a minimax approach.

## 2.5 Application of Empirical Process Theory to the Estimation of Density Level Sets

In this section, we show how similar ideas as in Section 2.4 can be used to estimate density level sets under a convexity restriction, in the Nikodym metric. The level sets of a density $f$ in $\mathbb{R}^d$ are the sets $G_\lambda = \{x \in \mathbb{R}^d : f(x) \geq \lambda\}$, for $\lambda > 0$. Estimation of density level sets and, more specifically, of convex level sets, has been tackled, for example, in [51, 70, 86]. As pointed by [51], estimation of density level sets may be useful in cluster analysis. It arises as a natural tool in testing for multimodality [66] and, more recently, it has been explored under the lens of topological data analysis [94, 25]. Notice that the 0-level set of a density $f$ is its support, so support estimation is a particular case of density level set estimation. However, in this section we only treat the case of positive levels $\lambda$, where empirical process theory has proven to be a successful tool.

Let $\lambda > 0$ such that $G_\lambda \neq \varnothing$. The excess mass of a measurable set $C \subseteq \mathbb{R}^d$ is defined as $\mathcal{M}_\lambda(C) = \int_C f(x)\, dx - \lambda|C|$. Simple algebra shows that $\mathcal{M}_\lambda(C) \leq \mathcal{M}_\lambda(G_\lambda)$, for all measurable sets $C \subseteq \mathbb{R}^d$. The empirical excess mass of a set $C$, given a sample $X_1, \ldots, X_n$, is naturally defined as $\widehat{\mathcal{M}}_\lambda(C) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in C} - \lambda|C|$. Hence, the main idea to estimate $G_\lambda$ is to maximize $\widehat{\mathcal{M}}_\lambda(C)$ over $C \in \mathcal{C}$, where $\mathcal{C}$ is a given class of measurable sets. In this section, we assume that $G_\lambda \in \mathcal{K}_d^{(1)}$ and we take $\mathcal{C} = \mathcal{K}_d^{(1)}$. For instance, convexity of $G_\lambda$ is ensured if $f$ is log-concave or, more generally, quasiconcave. If $f$ is the uniform density on a convex body $G \in \mathcal{K}_d^{(1)}$, then $G = G_\lambda$ for all $\lambda \in (0, \beta_d^{-1})$: In that case, support estimation is equivalent to level set estimation, for small levels $\lambda$ and the methods presented here could be applied to estimate $G$ itself. In what follows, $\lambda > 0$ is a fixed number and we define the estimator $\hat{G}_n \in \operatorname{argmax}_{G \in \mathcal{K}_d^{(1)}} \widehat{\mathcal{M}}_\lambda(G)$.

In order to achieve consistency, an assumption is usually made about the behavior of $f$ around the boundary of its level set $G_\lambda$. Namely, $f$ should not be too flat near the boundary of $G_\lambda$. The assumption proposed in [70] takes the following form, where $\mu$ is the continuous probability measure on $\mathbb{R}^d$ with density $f$.

ASSUMPTION 1. There exist positive constants $c$ and $\gamma$ such that

$$\mu(\{x \in \mathbb{R}^d : |f(x) - \lambda| < \eta\}) \leq c\eta^\gamma,$$

for all $\eta > 0$ small enough.

Assumption 1, also known as *margin condition*, is usually imposed for discriminant analysis [62, 60], statistical learning [87], level set estimation (a stronger assumption is proposed in [86]; see Assumption 2 below) or density support estimation [13].

In [70], the notion of covering number with inclusion, slightly different from that of covering number, is used to prove the main results.

DEFINITION 1 (Covering number with inclusion). Let $\mathcal{C}$ be a class of measurable subsets of $B(0, 1)$, $\mu$ a probability distribution in $\mathbb{R}^d$ and $\varepsilon > 0$. The $\varepsilon$-covering number of $\mathcal{C}$ with inclusion with respect to $\mu$ is the smallest integer $N$ such that there exists a collection $\mathcal{N}$ of measurable sets, with $\#\mathcal{N} = N$, satisfying the following: For all $C \in \mathcal{C}$, there exist $C_*, C^* \in \mathcal{N}$ with $C_* \subseteq C \subseteq C^*$ and $\mu(C^* \setminus C_*) \leq \varepsilon$. It is denoted by $N_I(\varepsilon, \mathcal{C}, \mu)$ and $\ln N_I(\varepsilon, \mathcal{C}, \mu)$ is called the *metric entropy with inclusion* of the class $\mathcal{C}$ with respect to $\mu$.

Note that in this definition, $\mathcal{N}$ need not be included in $\mathcal{C}$. Also note that a similar notion, called metric entropy with bracketing, is widely used in function estimation, especially in empirical process theory (e.g., see [90], Section 19.2). Let $(\mathcal{E}, \|\cdot\|)$ be a normed space of real-valued functions defined on a set $X$ and let $\mathcal{F} \subseteq \mathcal{E}$. For any two functions $l, r \in \mathcal{E}$, the bracket $[l, r]$ is defined as the set of all functions $f \in \mathcal{F}$ satisfying $l(x) \leq f(x) \leq r(x)$ for all $x \in X$. For all $\varepsilon > 0$, the $\varepsilon$-bracketing number of $F$ with respect to $\|\cdot\|$ is the smallest numbers of brackets $[l, r]$ with $\|r - l\| \leq \varepsilon$ needed to cover $\mathcal{F}$. It is denoted by $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ and $\ln N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ is called the *metric entropy with bracketing* of the class $\mathcal{F}$ with respect to $\|\cdot\|$. It is easy to see that for all class of measurable sets $\mathcal{C}$, if we let $\mathcal{F}_\mathcal{C} = \{\mathbb{1}_C : C \in \mathcal{C}\}$, then $\ln N_I(\varepsilon, \mathcal{C}, \mu)$ and $\ln N_{[]}(\varepsilon, \mathcal{F}_\mathcal{C}, \|\cdot\|_{1,\mu})$ differ by at most a factor 2, where $\|\phi\|_{1,\mu} = \int_{\mathbb{R}^d} |\phi(x)|\, d\mu(x)$, for all measurable, bounded functions $\phi : \mathbb{R}^d \to \mathbb{R}$. The following estimate is available for the class $\mathcal{K}_d^{(1)}$.

THEOREM 8 ([32]). *Let $\mu$ be a continuous probability measure on $B(0, 1)$ with a density $f$ with respect to the Lebesgue measure. Assume that $f \leq M$ almost everywhere, where $M > 0$ is a given number. Then, as $\varepsilon \to 0$,*

$$\ln N_I\left(\varepsilon, \mathcal{K}_d^{(1)}, \mu\right) \lesssim_{d,M} \varepsilon^{-\frac{d-1}{2}}.$$

Together with this estimate, [70], Theorem 3.7, yields the following result.

THEOREM 9 ([70]). *Assume that $d \geq 2$. There exists a constant $c(d)$ such that the following holds with probability tending to one, as $n$ goes to infinity. Let $\mu$ be a probability measure on $B(0, 1)$ with a bounded density $f$ with respect to the Lebesgue measure and let Assumption 1 hold. Let $\lambda > 0$ and let $G_\lambda \in \mathcal{K}_d^{(1)}$. Then*

$$\mu(\hat{G}_n \triangle G_\lambda) \leq \begin{cases} c(2)n^{-\frac{2\gamma}{3\gamma+4}} & \text{if } d = 2, \\ c(3)n^{-\frac{\gamma}{2\gamma+2}} \ln n & \text{if } d = 3, \\ c(d)n^{-\frac{2\gamma}{(\gamma+1)(d+1)}} & \text{if } d \geq 4. \end{cases}$$

In fact, this theorem is stated under more general assumptions than convexity of the level sets. If the level sets belong to a class of sets with metric entropy with inclusion of order $\varepsilon^{-r}$, for some exponent $r > 0$ (e.g., $r = (d-1)/2$ for the class $\mathcal{K}_d^{(1)}$), the rates given in the theorem depend on $d$, $\gamma$ and $r$. It is noticeable that the exponent $r = (d-1)/2$ in the metric entropy with inclusion of the class $\mathcal{K}_d^{(1)}$ matches that of the class of sets with twice differentiable boundaries in some sense (see [32]).

Note that the estimator $\hat{G}_n$ defined in [70] is a polytope, and its vertices are sample points. Indeed, for all $C \in \mathcal{K}_d^{(1)}$, $\widehat{\mathcal{M}}_\lambda(C) \leq \widehat{\mathcal{M}}_\lambda(C^*)$, where $C^*$ is the convex hull of the sample points contained in $C$. In the two-dimensional case, [51] designs an algorithm to compute $\hat{G}_n$. To the best of our knowledge, there is no algorithm to compute $\hat{G}_n$ or an approximation of $\hat{G}_n$ in higher dimensions.

Optimality of the upper bounds in the above theorem is not proven in [70]. However, [86] proves lower bounds for the minimax risk in both Nikodym and Hausdorff metrics. In the Nikodym metric, the lower bounds proven by [86] match the upper bounds given in the above theorem only for $d = 2, 3$ (up to a logarithmic factor when $d = 3$), and they are faster when $d \geq 4$. The estimation of convex level sets in the Hausdorff metric requires completely different techniques. It has been tackled in [79, 86]. In [79], the author considers both level sets corresponding to a given level

and level sets with given probability content (see also [22] for the estimation of level sets with given probability content); the results are then applied to the estimation of the mode of the density, by considering the smallest estimated level set. Note that a control of the estimated level sets in the Nikodym metric could not yield consistent estimation of the mode, since two sets can have a very small Nikodym distance if they both have very small volumes, even if they are far apart from each other in the space. Optimal rates in estimation of convex density level sets in both Nikodym and Hausdorff metrics are given in [86] when $d = 2$ and they are extended to higher dimensions. More generally, [86] proves optimal rates for density level sets whose boundaries satisfy some smoothness condition. In fact, it is noticed that if $G \in \mathcal{K}_d$ satisfies $B(0, r) \subseteq G \subseteq B(0, R)$ for some $0 < r < R$, then the boundary of $G$ is Lipschitz, in the sense that the radial function of $G$, defined as $r_G(u) = \max\{\lambda \geq 0 : \lambda u \in G\}$, is Lipschitz. For completeness, we include the precise statement and its proof here.

LEMMA 1. *Let $G \in \mathcal{K}_d$ satisfies $B(0, r) \subseteq G \subseteq B(0, R)$ for some $0 < r < R$. Then the radial function $r_G$ satisfies $|r_G(u) - r_G(u')| \leq R/(2r)\|u - u'\|$, for all $u, u' \in \mathbb{S}^{d-1}$.*

PROOF. Let $G^\circ$ be the polar body of $G$, defined as $G^\circ = \{x \in \mathbb{R}^d : \langle x, y \rangle \leq 1, \forall y \in K\}$. By standard properties of polar bodies (see [80], Chapter 1), one has $B(0, R^{-1}) \subseteq G^\circ \subseteq B(0, r^{-1})$ and the radial function $r_G$ is the inverse of the support function of $G^\circ$: $r_G(u) = (h_{G^\circ}(u))^{-1}$, for all $u \in \mathbb{S}^{d-1}$. Subadditivity of support functions yield $|h_{G^\circ}(u) - h_{G^\circ}(u')| \leq \max(h_{G^\circ}(u - u'), h_{G^\circ}(u' - u)) = \|u - u'\| \max(h_{G^\circ}(\frac{u-u'}{\|u-u'\|}), h_{G^\circ}(\frac{u'-u}{\|u-u'\|}))$, for all $u, u' \in \mathbb{S}^{d-1}$ with $u \neq u'$. Since $G^\circ \subseteq B(0, r^{-1})$, $h_{G^\circ}(v) \leq r^{-1}$, for all $v \in \mathbb{S}^{d-1}$. This proves that $h_{G^\circ}$ is $r^{-1}$-Lipschitz. Now, since we also have that $B(0, R^{-1}) \subseteq G^\circ$, $h_{G^\circ}(v) \geq R^{-1}$, for all $v \in \mathbb{S}^{d-1}$. Hence, $(h_{G^\circ})^{-1}$ is $(R/(2r))$-Lipschitz. $\square$

First, [86] computes the optimal rates for star shaped density level sets with smooth radial functions. Standard techniques from functional estimation are used, such as local polynomial approximations. Then the author tackles the problem of estimating convex level sets. As shown in Lemma 1, the case of convex level sets is included in the case of star shaped level sets with Lipschitz radial functions. Hence, the optimal rates for convex sets are not larger than the ones corresponding to Lipschitz radial functions. Perhaps surprisingly, in

the Hausdorff metric, convexity of the level set does not make the problem easier than just the Lipschitz property of its radial function, since [86] shows that the optimal rate under convexity matches the optimal rate under just the Lipschitz assumption, up to logarithmic factors. In the Nikodym metric, the situation is very different: [86] proves that at least in dimension 2, the optimal rate for convex sets is actually much faster than in the case of Lipschitz radial functions: This is a consequence of Theorem 9 above. It can be seen easily that the same holds when $d = 3$, and [86] suggests that this holds in arbitrary dimension, without a giving proof. Hence, in the Nikodym metric, convexity does contribute and improve the optimal rate from the Lipschitz assumption.

[86] does not exactly use the same margin condition as [70], but makes the following assumption. Let $f$ be a density in $\mathbb{R}^d$ and let $G_\lambda$ be its level set with level $\lambda > 0$. Assume that $G_\lambda$ is star shaped around the origin, and let $r_\lambda$ be its radial function.

ASSUMPTION 2. Let $b_1, b_2 > 0$ with $b_1 < b_2$, $\nu, \delta_0 > 0$. Then, for all $u \in \mathbb{S}^{d-1}$ and $r > 0$ such that $|f(ru) - \lambda| \le \delta_0$,

$$b_1 \le \frac{|f(ru) - \lambda|}{|r - r_\lambda(u)|^\nu} \le b_2.$$

Roughly, Assumption 2 is stronger than Assumption 1 if one takes $\gamma = 1/\nu$. Under Assumption 2, [86] characterizes the optimal rates for the estimation of a convex level set $G_\lambda$ that satisfies $B(0, r) \subseteq G_\lambda \subseteq B(0, R)$ with $0 < r < R$ when $d = 2$ and suggest the following extensions to higher dimensions: $n^{-2/(4\nu+d+1)}$ in the Nikodym metric and $n^{-1/(2\nu+d)}$ (up to a logarithmic factor) in the Hausdorff metric. In the Nikodym metric, the upper bound follows directly from [70] when $d = 2$ but [86] does not give a proof for larger $d$. For arbitrary $d$, the rates suggested in [86] are actually faster than the upper bounds given in [70]. In the Hausdorff metric and for any $d$, as explained above, the upper bound follows directly from the Lipschitz case, by Lemma 1.

When dealing with level sets with smooth radial functions in arbitrary dimension, [86] proves that the minimax rates are exactly given by $n^{-\beta/((2\nu+1)\beta+d-1)}$ in the Nikodym metric and $(n/\ln n)^{-\beta/((2\nu+1)\beta+d-1)}$ in the Hausdorff metric, where $\beta$ is a smoothness parameter that roughly corresponds to the number of bounded derivatives of the radial function (e.g., $\beta = 1$ corresponds to the Lipschitz case). It is noticeable that for convex level sets, the minimax rate $n^{-1/(2\nu+d)}$ in the Hausdorff metric matches the one that corresponds

to smoothness $\beta = 1$, as discussed above (and as predicted by Lemma 1), whereas in the Nikodym metric, the minimax rate $n^{-2/(4\nu+d+1)}$ for convex level sets matches the rate that corresponds to smoothness $\beta = 2$. This complements the remark we made earlier: The exponent $r = (d - 1)/2$ in the metric entropy with inclusion for convex bodies is the same as for sets with twice differentiable boundary (see [70] and [32] for more details), and the corresponding minimax rates match. However, note that even though the boundary of any convex body is twice differentiable almost everywhere, the class of convex bodies $G \in \mathcal{K}_d$ with $B(0, r) \subseteq G \subseteq B(0, R)$, where $0 < r < R$, contains polytopes with arbitrarily many vertices, which have very non-smooth boundaries, together with convex bodies with smooth boundaries and positive curvature everywhere, which yet can take arbitrarily large values.

Finally, note that the rates $n^{-2/(4\nu+d+1)}$ and $n^{-1/(2\nu+d)}$ given in [86] match (up to logarithmic factors) those obtained in the estimation of the support of a uniform distribution, that is, at the limit $\nu = 0$. In the Nikodym metric, the minimax rate of estimation of convex bodies is $n^{-2/(d+1)}$ (see Theorem 2 above), whereas in the Hausdorff metric, it is $(n/\ln n)^{-1/d}$, as shown in Theorem 11 (with $\alpha = d$); see [17].

## 2.6 Convex Support Estimation in Nonparametric Regression

Let the following model hold:

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \ldots, n,$$

where $X_1, \ldots, X_n$ are deterministic or random points in $[0, 1]^d$, $\xi_1, \ldots, \xi_n$ are i.i.d. random variables, with mean zero, independent of $X_1, \ldots, X_n$ and $f : [0, 1] \to [0, \infty)$. In this section, we are interested in the estimation of the support $G$ of $f$, that is, the closure of the set $\{x \in [0, 1]^d : f(x) > 0\}$. Throughout the section, we assume that $G$ is a convex body included in $[0, 1]^d$. In [14], the function $f$ is the indicator function of $G$: $f(x) = 1$ for $x \in G$, $f(x) = 0$ otherwise. The design points $X_1, \ldots, X_n$ are i.i.d., uniformly distributed in $[0, 1]^d$ and the $\xi_i$'s are sub-Gaussian, that is, $\mathbb{E}[e^{t\xi_1}] \le e^{\frac{\sigma^2 t^2}{2}}$, for all $t \in \mathbb{R}$, where $\sigma > 0$ need not be known. [14] considers a least squares estimator $\hat{G}_n \in \operatorname{argmin}_{C \in \mathcal{N}} \mathcal{A}(C)$, where $\mathcal{N}$ is a $n^{-2/(d+1)}$-net of $\mathcal{K}_d^{(1)}$, with respect to the Nikodym metric and $\mathcal{A}(C) = \sum_{i=1}^n (1 - 2Y_i) \mathbb{1}_{X_i \in C}$. The following upper bound is shown in [14], in which $\mathbb{P}_G$ stands for the joint distribution of the sample with $f(\cdot) = \mathbb{1}_{\cdot \in G}$.

THEOREM 10 ([14]). *There exist three positive constants $C_1$, $C_2$, $C_3$ that depend on $d$ and $\sigma^2$ and a positive integer $n_0$ that depends on $d$ only, such that the following holds:*

*For all $G \in \mathcal{K}_d^{(1)}$, all $n \geq n_0$ and all $x \geq 0$,*

$$|\hat{G}_n \triangle G| \leq C_1 n^{-\frac{2}{d+1}} + \frac{x}{n}$$

*with $\mathbb{P}_G$ probability at least $1 - C_2 e^{-C_3 x}$.*

Of course, the estimator $\hat{G}_n$ is not computable in practice, since $\varepsilon$-nets of $\mathcal{K}_d^{(1)}$ are not available. We believe that maximizing the functional $\mathcal{A}$ over the whole class $\mathcal{K}_d^{(1)}$ would yield a similar upper bound, whose rate is proven to be minimax optimal. However, it would remain unclear how to compute the resulting estimator $\tilde{G}_n$. Nonetheless, note that for all $C \in \mathcal{K}_d^{(1)}$, $\mathcal{A}(C) = \mathcal{A}(C^*)$, where $C^*$ is the convex hull of all the design points $X_i \in C$, implying that $\tilde{G}_n$ can be chosen to be a polytope whose vertices are design points. Another open question is whether $\hat{G}_n$ (or $\tilde{G}_n$) is adaptive to polytopal supports. In [14], it is shown that the minimax rate on the class $\mathcal{P}_r$ of polytopes with at most $r$ vertices is of the order $(\ln n)/n$: Like in density support estimation, is the error of $\hat{G}_n$ (or $\tilde{G}_n$) of that order when the true support $G$ is a polytope, up to logarithmic factors?

Here, we have only discussed the case when $f$ is an indicator function, but more general models for $f$ are considered in [58]. All these models, though, impose a sharp separation condition on $f$, for example, boundedness away from zero on its support, which essentially reduces to the case of indicator functions. To the best of our knowledge, harder cases, for example, when $f$ satisfies a margin type condition, that is, $|\{x \in [0,1]^d : f(x) \leq \eta\}| \leq c\eta^\gamma$ for all $\eta > 0$ small enough, where $c$ and $\gamma$ are positive constants, have not been tackled in the literature.

## 3. ESTIMATION OF CONVEX SETS UNDER THE HAUSDORFF METRIC

### 3.1 Support Functions and Polyhedral Representations of Convex Bodies

Support functions play a central role in estimation of convex bodies under the Hausdorff metric. Indeed, the Hausdorff distance between two convex bodies $G_1$ and $G_2$ can be written in terms of their support functions $h_{G_1}$ and $h_{G_2}$.

LEMMA 2. *For all convex bodies $G_1, G_2 \in \mathcal{K}_d$,*

$$\mathsf{d}_H(G_1, G_2) = \sup_{u \in \mathbb{S}^{d-1}} |h_{G_1}(u) - h_{G_2}(u)|.$$

Here, we state a few results about support functions that are useful in estimation of convex sets, and we refer to [80] for more details on their account.

Note that a convex body is completely determined by its support function and $G = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq h_G(u), \forall u \in \mathbb{S}^{d-1}\}$, for all $G \in \mathcal{K}_d$.

A function $h : \mathbb{S}^{d-1} \to \mathbb{R}$ is the support function of a convex set if and only if it is subadditive, in the following sense.

DEFINITION 2. Let $h : \mathbb{S}^{d-1} \to \mathbb{R}$. Define the function $\tilde{h}$ as $\tilde{h}(v) = \|v\| h(v/\|v\|)$ if $v \neq 0$, $\tilde{h}(0) = 0$. We say that $h$ is subadditive if $\tilde{h}$ is convex.

If $h : \mathbb{S}^{d-1} \to \mathbb{R}$ is subadditive, then it is the support function of the convex set $\{x \in \mathbb{R}^d : \langle u, x \rangle \leq h(u), \forall u \in \mathbb{S}^{d-1}\}$.

A polyhedral representation of a convex body is a way of writing it as the intersection of closed half-spaces or, equivalently, as a collection of affine constraints. For $\phi : \mathbb{S}^{d-1} \to \mathbb{R}$, we let $G_\phi = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq \phi(u), \forall u \in \mathbb{S}^{d-1}\}$. It is easy to see that $h_{G_\phi}(u) \leq \phi(u)$, for all $u \in \mathbb{S}^{d-1}$. In general, the two functions are not equal, since $\phi$ is not necessarily subadditive. Subadditivity is actually a necessary and sufficient condition for $\phi$ to be a support function (see [18], Proposition 1).

LEMMA 3. *Let $\phi : \mathbb{S}^{d-1} \to \mathbb{R}$. Then, $\phi = h_{G_\phi}$ if and only if $\phi$ is subadditive.*

Actually, an interesting consequence of this result is the following.

LEMMA 4. *Let $\phi : \mathbb{S}^{d-1} \to \mathbb{R}$. Then $h_{G_\phi}$ is the largest subadditive function that is smaller or equal to $\phi$.*

PROOF. Let $g : \mathbb{S}^{d-1} \to \mathbb{R}$ be a subadditive function with $g(u) \leq \phi(u)$, for all $u \in \mathbb{S}^{d-1}$. Then $G_g \subseteq G_\phi$. As a consequence, $h_{G_g}(u) \leq h_{G_\phi}(u)$, for all $u \in \mathbb{S}^{d-1}$. By Lemma 3, since $g$ is subadditive, $h_{G_g} = g$, yielding $g(u) \leq h_{G_\phi}(u)$ for all $u \in \mathbb{S}^{d-1}$. $\square$

By Lemma 2, the Hausdorff distance between two convex sets $G_\phi$ and $G_\psi$ can be written in terms of $h_{G_\phi}$ and $h_{G_\psi}$. However, these support functions may not be easy to compute in terms of $\phi$ and $\psi$. The following lemma provides a partial solution to this issue when both $\phi$ and $\psi$ are continuous.

LEMMA 5 ([18]). *Let $\phi, \psi : \mathbb{S}^{d-1} \to \mathbb{R}$ be two continuous functions. Assume that $G_\phi$ and $G_\psi$ have nonempty interiors. Moreover, let $R > r > 0$ and assume that $B(a, r) \subseteq G_\phi \subseteq B(a, R)$, for some $a \in$*

$\mathbb{R}^d$. *Let* $\eta = \max_{u \in \mathbb{S}^{d-1}} |\psi(u) - \phi(u)|$. *If* $\eta < r$, *then* $\mathsf{d}_H(G_\psi, G_\phi) \leq \frac{\eta R}{r} \frac{1+\eta/r}{1-\eta/r}$.

In this lemma, the mapping $\psi$ plays the role of an estimate of $\phi$. Moreover, a control of the estimation error of $\phi$ in sup-norm yields a control of the estimation error of $G_\phi$ in the Hausdorff metric. However, in certain cases, it may not be easy to control the accuracy of the estimation of $\phi(u)$ for all $u \in \mathbb{S}^{d-1}$ simultaneously, but instead, only on a finite subset of $\mathbb{S}^{d-1}$. Recall that for $\varepsilon \in (0, 1)$, an $\varepsilon$-net of $\mathbb{S}^{d-1}$ is a subset $\mathcal{N}$ of $\mathbb{S}^{d-1}$ such that for all $u \in \mathbb{S}^{d-1}$, there is $u^* \in \mathcal{N}$ with $\|u - u^*\| \leq \varepsilon$. For a subset $\mathcal{N}$ of $\mathbb{S}^{d-1}$ and a function $\psi : \mathcal{N} \to \mathbb{R}$, let $G_\psi^{\mathcal{N}} = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq \psi(u), \forall u \in \mathcal{N}\}$. Then the following result is complementary to the previous lemma, when $\phi = h_{G_\phi}$ (i.e., by Lemma 3, when $\phi$ is subadditive) and when $\psi$ is only defined on a (fine enough) discretization of $\mathbb{S}^{d-1}$.

LEMMA 6 ([18]). *Let* $G \in \mathcal{K}_d$ *and* $\psi : \mathbb{S}^{d-1} \to \mathbb{R}$. *Let* $\varepsilon \in (0, 1)$ *and* $\mathcal{N}$ *be an* $\varepsilon$-net of $\mathbb{S}^{d-1}$. *Let* $R > r > 0$ *and assume that* $B(a, r) \subseteq G \subseteq B(a, R)$, *for some* $a \in \mathbb{R}^d$. *Let* $\eta = \max_{u \in \mathcal{N}} |\psi(u) - h_G(u)|$. *Then, if* $\eta < r$, $\mathsf{d}_H(G_\psi^{\mathcal{N}}, G) \leq \frac{\eta R}{r} \frac{1+\eta/r}{1-\eta/r} + \frac{2R\varepsilon}{1-\varepsilon}$.

Finally, the next lemma provides a polyhedral representation of the convex hull of a finite collection of points in the space. It is straightforward, given that a linear function defined on a convex body is necessarily maximized at an extreme point, but it yields a useful representation of random polytopes.

LEMMA 7. *Let* $x_1, \ldots, x_n \in \mathbb{R}^d$ *and let* $K_n$ *be their convex hull. Then*

$$h_{K_n}(u) = \max_{1 \leq i \leq n} \langle u, x_i \rangle \quad \forall u \in \mathbb{S}^{d-1}.$$

*In particular*, $K_n = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq \max_{1 \leq i \leq n} \langle u, x_i \rangle, \forall u \in \mathbb{S}^{d-1}\}$.

Of course, this polyhedral representation is not optimal: since $K_n$ is a polytope, the Minkowski–Weyl theorem for polyhedra states that only a finite number of affine constraints should be sufficient to describe the set $K_n$. Moreover, these constraints correspond to the normal vectors of the $(d-1)$-dimensional faces of $K_n$: All other affine constraints are redundant. However, in practice, finding the $(d-1)$-dimensional faces of the convex hull of a given finite collection of points is a hard problem (see, e.g., [11]).

## 3.2 Estimation of Support Functions

Sometimes, in order to estimate a convex body $G$, it can be natural to estimate its support function $h_G$. The previous lemmas indicate that if the estimation error are measured with respect to the Hausdorff metric, then it is enough to bound the pointwise or sup-norm error of the estimation of $h_G$.

In this section, $G$ is an unknown convex body and we assume that an estimator $\hat{h}_n$ of $h_G$ has been extracted from available data. Then $G_{\hat{h}_n}$ is a natural estimator of $G$. We review examples and show how the results and properties stated in the previous section apply, by distinguishing two cases: when $\hat{h}_n$ is subadditive, hence, when it is the support function $G_{\hat{h}_n}$, and when it is not.

3.2.1 *When* $\hat{h}_n$ *is subadditive.* This is the case, for instance, in density support estimation: A natural estimator of the support $G$ of i.i.d. random points $X_1, \ldots, X_n$ is $\hat{h}_n(u) = \max_{1 \leq i \leq n} \langle u, X_i \rangle, u \in \mathbb{S}^{d-1}$. By Lemma 7, $\hat{h}_n$ is actually the support function of $\hat{K}_n$, hence, it is subadditive.

LEMMA 8. *Let* $G \in \mathcal{K}_d^{(1)}$. *Let* $\hat{h}_n$ *be a subadditive estimator of* $h_G$ *and let* $\hat{G}_n = G_{\hat{h}_n}$. *Let* $\delta, \eta \in (0, 1)$ *and assume that* $|\hat{h}_n(u) - h_G(u)| \leq \eta$ *with probability at least* $1 - \delta$, *for all* $u \in \mathbb{S}^{d-1}$. *Then* $\mathsf{d}_H(\hat{G}_n, G) \leq 2\eta$ *with probability at least* $1 - 2(18/\eta)^d \delta$.

PROOF. The proof of this lemma relies on two fundamental facts. First, by a standard volumetric argument, there exists an $\varepsilon$-net of $\mathbb{S}^{d-1}$ of cardinality at most $(3/\varepsilon)^d$, for all $\varepsilon \in (0, 1)$. Second, the following result, which is Lemma 5.2 in [39], is a very elegant tool to work with approximations of the unit sphere.

LEMMA 9 ([39]). *Let* $\varepsilon \in (0, 1)$ *and let* $\mathcal{N}$ *be an* $\varepsilon$-net of $\mathbb{S}^{d-1}$. *Then, for all* $u \in \mathbb{S}^{d-1}$, *there are sequences* $(u_k)_{k \geq 0} \subseteq \mathcal{N}$ *and* $(\varepsilon_k)_{k \geq 1} \subseteq \mathbb{R}$ *such that* $u = u_0 + \sum_{k=1}^{\infty} \varepsilon_k u_k$, *with* $0 \leq \varepsilon_k \leq \varepsilon^k, \forall k \geq 1$.

Let $\varepsilon = \eta/6$ and let $\mathcal{N}$ be an $\varepsilon$-net of $\mathbb{S}^{d-1}$. Denote by $-\mathcal{N} = \{-u : u \in \mathcal{N}\}$. Let $\mathcal{A}$ be the event when $|\hat{h}_n(u) - h_G(u)| \leq \eta$, simultaneously for all $u \in \mathcal{N} \cup (-\mathcal{N})$. By a union bound,

$$\mathbb{P}[\mathcal{A}] \geq 1 - \sum_{u \in \mathcal{N} \cup (-\mathcal{N})} \mathbb{P}[|\hat{h}_n(u) - h_G(u)| \leq \eta]$$

$$\geq 1 - 2(\#\mathcal{N})\delta \geq 1 - 2(18/\eta)^d \delta.$$

Now, assume that $\mathcal{A}$ is satisfied and let $u \in \mathbb{S}^{d-1}$. Then, with the notation of Lemma 9, using subadditivity of

support functions and the fact that $h_G \leq 1$ (since $G \subseteq B(0,1)$),

$$\hat{h}_n(u) \geq \hat{h}_n(u_0) - \sum_{k \geq 1} \varepsilon^k \hat{h}_n(-u_k)$$

$$\geq h_G(u_0) - \eta - \sum_{k \geq 1} \varepsilon^k \big(h_G(-u_k) + \eta\big)$$

$$\geq h_G(u_0) - \eta - 2\sum_{k \geq 1} \varepsilon^k$$

$$= h_G(u_0) - \eta - \frac{2\varepsilon}{1 - \varepsilon}$$

$$\geq h_G(u) - \sum_{k \geq 1} \varepsilon^k h_G(u_k) - \eta - \frac{2\varepsilon}{1 - \varepsilon}$$

$$\geq h_G(u) - \eta - \frac{3\varepsilon}{1 - \varepsilon} \geq h_G(u) - 2\eta$$

and similarly, $h_G(u) \geq \hat{h}_n(u) - 2\eta$. Thus, if $\mathcal{A}$ is satisfied, then $\mathsf{d}_\mathsf{H}(\hat{G}_n, G) = \sup_{u \in \mathbb{S}^{d-1}} |\hat{h}_n(u) - h_G(u)| \leq 2\eta$, which completes the proof of the lemma. $\square$

This technique is applied in [17] for support estimation under smoothness conditions (see Assumption 3 below), where ideas from [33] are refined in order to obtain nonasymptotic deviation inequalities. The general framework considered in that work is the following. Let $G \in \mathcal{K}_d^{(1)}$ and $\mu$ a probability measure on $\mathbb{R}^d$, satisfying the following assumption.

ASSUMPTION 3. For all $u \in \mathbb{R}^d$ and all $t \in [0, r]$, $\mu(C_G(u, t)) \geq Lt^\alpha$, where:

- $r, L, \alpha$ are given positive numbers with $r < 1$;
- $C_G(u, t)$ is the cap of $G$ in the direction $u$ and with height $t$, that is,

$$C_G(u, t) = \{x \in G : \langle u, x \rangle \geq h_G(u) - t\}.$$

Particular cases of such pairs $(G, \mu)$ include uniform distributions on general convex bodies ($\alpha = d$) or smooth convex bodies ($\alpha = (d + 1)/2$), uniform distributions on the boundary of smooth convex bodies ($\alpha = (d - 1)/2$), linear projections of uniform distributions that are supported on higher dimensional smooth convex bodies, distributions with densities supported on a convex body with polynomial decay near the boundary, etc.

It is easy to show that under Assumption 3, for all $u \in \mathbb{S}^{d-1}$ and all $t \in [0, r]$, $|\hat{h}_n(u) - h_G(u)| \leq t$ with probability at least $1 - e^{-Lnt^\alpha}$. Hence, the following theorem is shown, using Lemma 8. Set

$$\tau_\alpha = \begin{cases} 1 & \text{if } \alpha \geq 1, \\ 2^{\alpha-1} & \text{if } 0 < \alpha < 1. \end{cases}$$

THEOREM 11 ([17]). *Let $r \in (0, 1)$ and $L, \alpha > 0$. There exists a positive constant $C$ such that the following holds. Set $a_n = (\frac{C \ln n}{n})^{\frac{1}{\alpha}}$ and $b_n = n^{\frac{-1}{\alpha}}$. Let $G \in \mathcal{K}_d^{(1)}$ and $\mu$ be a probability measure on $\mathbb{R}^d$ and let Assumption 3 hold. Then, for all $x \geq 0$ such that $a_n + b_n x \leq r$,*

$$\mathsf{d}_\mathsf{H}(\hat{K}_n, G) \leq 2a_n + 2b_n x$$

*with $\mu$-probability at least $1 - 12^d \exp(-C_\alpha L x^\alpha)$.*

As a consequence, under Assumption 3, $\hat{K}_n$ satisfies $\mathsf{d}_\mathsf{H}(\hat{K}_n, G) = O_\mathbb{P}((\frac{\ln n}{n})^{-1/\alpha})$.

REMARK 1. [17] uses a different technique in order to bound the error of $\hat{K}_n$ when $G$ is a polytope that satisfies a standardness condition. Using the technique shown above would yield the rate $((\ln n)/n)^{1/d}$ ($\alpha = d$), which is suboptimal. Indeed, the following inequality is shown when $G$ is a polytope with at most $p$ vertices ($p \geq d + 1$) and satisfies a $\nu$-standardness condition, for some $\nu \in (0, 1)$:

$$\mathbb{P}_G[n^{1/d}\mathsf{d}_\mathsf{H}(K, \hat{K}_n) \geq x] \leq pe^{-\nu\beta_d x^d} \quad \forall x \geq 0.$$

Surprisingly, the adaptive feature of $\hat{K}_n$ disappears under the Hausdorff metric. As discussed in Section 2, $\hat{K}_n$ adapts to polytopal supports under the Nikodym metric, and its rate of convergence in that metric is the fastest when $G$ is a polytope. On the contrary, the rate of $\hat{K}_n$ in the Hausdorff metric is the worst when the support $G$ is a polytope.

Another model where the same ideas could be applied has been considered in [40, 48]. Given independent observations $Y_i = h_G(u_i) + \varepsilon_i$, where $u_i \in \mathbb{S}^{d-1}$ is either deterministic or random and $\varepsilon_i$ is a zero mean error term, independent of $u_i$, [48] estimates $h_G$ by the least squares estimator on the class of all subadditive functions $h : \mathbb{S}^{d-1} \to \mathbb{R}$. This produces an estimator $\hat{h}_n$ that is a support function itself. In general, $L^2$-type metrics are natural to measure the performance of least squares estimators. [40, 48] prove rates of convergence of their estimator with respect to the $L^2$-distance, with minimax optimality proven in [48]. However, the $L^2$-distance between the support functions of convex bodies does not translate into a natural and geometric measurement of a distance between the convex bodies themselves. [40] uses elegant norm inequalities for subadditive functions, which show that the $L^\infty$ distance is dominated by the $L^2$ distance in some sense (see [47], Proposition 2.3.1, and [40], Proposition 2.2). In turn, they obtain error bounds for the $L^\infty$ metric between the support functions, and hence, for the Hausdorff distance between the convex sets. Nevertheless,

as remarked in [48], these error bounds are very loose, and the optimal rate of estimation of $G \in \mathcal{K}_d^{(1)}$ with respect to the Hausdorff distance in that model is still an unsolved problem.

3.2.2 *When the estimator $\hat{h}_n$ is not subadditive.* There are many problems where an estimator of an unknown convex body $G$ is defined as $\hat{G}_n = G_{\hat{h}_n}$ for some estimator $\hat{h}_n$ of $h_G$, that is not necessarily subadditive. Lemmas 5 and 6 are two key results to deal with this case. Lemma 5 is useful when $\hat{h}_n$ is continuous almost surely and its error can be bounded uniformly on the unit sphere $\mathbb{S}^{d-1}$. Indeed, if $|\hat{h}_n(u) - h_G(u)| \leq \eta$ simultaneously for all $u \in \mathbb{S}^{d-1}$ with high probability, then, if $G$ satisfies the assumptions of Lemma 5 and $\eta$ is small enough, $\mathsf{d}_{\mathsf{H}}(G_{\hat{h}_n}, G) \leq \eta'$ with high probability, where $\eta'$ is a small number that depends on $\eta$ and other parameters. We apply Lemma 6 in two cases: when the true support function $h_G$ cannot be described simply, that is, when $G = G_\phi$ for some function $\phi$ that is not necessarily subadditive, and when the error of $\hat{h}_n$ can only be controlled pointwise, that is, for each $u \in \mathbb{S}^{d-1}$ separately. By a union bound, pointwise deviations transfer to uniform deviations on $\varepsilon$-nets of $\mathbb{S}^{d-1}$. We give two examples, borrowed from [18] and [19], where these results are used.

*Multivariate quantile estimation*: *Level sets of the Tukey depth* [18]. As already pointed out in [38, 37, 39], the study of the convex hull of a cloud of points is a multivariate extension of extreme value theory. Indeed, by Lemma 7, the support function of $\hat{K}_n$ in any direction $u \in \mathbb{S}^{d-1}$ is given by the maximum of i.i.d. real random variables. In this regard, the definition of the random polytope $\hat{K}_n$ can be extended to that of the $k$-hull of the sample, for $0 \leq k \leq n - 1$. For each direction $u \in \mathbb{S}^{d-1}$, instead of considering the extreme statistic $\max_{i=1,\dots,n}\langle u, X_i \rangle$, take the $k$th order statistics $X_{(k)}(u)$, where $X_i(u) = \langle u, X_i \rangle$ and $X_{(1)}(u) \geq X_{(2)}(u) \geq \cdots \geq X_{(n)}(u)$ is the reordered list of $X_1(u), \dots, X_n(u)$ in nonincreasing order. The $k$-hull of $X_1, \dots, X_n$ is the convex set $\hat{K}_n^{(k)} = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq X_{(k)}(u), \forall u \in \mathbb{S}^{d-1}\}$ [28]. The convex hull is the 0-hull. In [18], $k$-hulls are used as estimators of the Tukey depth level sets of probability measures in $\mathbb{R}^d$.

Let $\mu$ be a probability measure on $\mathbb{R}^d$. The Tukey depth $D_\mu(x)$ of a point $x \in \mathbb{R}^d$ with respect to $\mu$ is defined as the smallest probability mass of a closed half-space containing $x$:

$$D_\mu(x) = \inf_{x \in H \in \mathcal{H}_d} \mu(H),$$

where $\mathcal{H}_d$ is the collection of closed half-spaces in $\mathbb{R}^d$. For $\alpha \in (0, 1)$, the $\alpha$-level set of $D_\mu$ is the set $G = \{x \in \mathbb{R}^d : D_\mu(x) \geq \alpha\}$. It is a closed convex set and it has the following polyhedral representation (see [54], Theorem 2, and [18], Lemma 1):

$$(3) \qquad G = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq q_u, \forall u \in \mathbb{S}^{d-1}\},$$

where $q_u$ is the upper $(1 - \alpha)$-quantile of $\langle u, X \rangle$ with $X \sim \mu$. Let $X_1, \dots, X_n$ be i.i.d. random points with distribution $\mu$ and let $\mu_n$ be the corresponding empirical measure: $\mu_n(A) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i \in A}$, for all Borel sets $A \subseteq \mathbb{R}^d$. The empirical $\alpha$-level set of the Tukey depth is $\hat{G}_n = \{x \in \mathbb{R}^d : D_{\mu_n}(x) \geq \alpha\}$, which also has a simple polyhedral representation: $\hat{G}_n = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq \hat{q}_u, \forall u \in \mathbb{S}^{d-1}\}$, where $\hat{q}_u = \sup\{t \in \mathbb{R} : \#\{i = 1, \dots, n : \langle u, X_i \rangle \geq t\} \geq n\alpha\}$ is the empirical upper $(1-\alpha)$-quantile of the sample $\langle u, X_1 \rangle, \dots, \langle u, X_n \rangle$, for all $u \in \mathbb{S}^{d-1}$. With the above notation, $\hat{q}_u = X_{(k)}(u)$, for $k = \lceil n\alpha \rceil$, where, for $x \in \mathbb{R}$, $\lceil x \rceil$ stands for the smallest integer larger or equal to $x$. Hence, the empirical level set $\hat{G}_n$ of the Tukey depth coincides with the $k$-hull of $X_1, \dots, X_n$. At the population level, there is another very important connection to the theory of random polytopes: As shown in [18], if $\mu$ is the uniform distribution on a convex body $K \in \mathcal{K}_d$, the population level set $G$ coincides with the $\alpha$-floating body of $K$, as defined in Section 2.1.

In general, neither $u \mapsto q_u$ nor $u \mapsto \hat{q}_u$ are subadditive, and hence, they are not the support functions of the sets $G$ and $\hat{G}_n$, respectively. However, $u \mapsto \hat{q}_u$ is always continuous, as a consequence of [18], Lemma 15, and $u \mapsto q_u$ is continuous under some weak assumptions on $\mu$ (see [18], Lemma 11). Building on standard results from empirical process theory, [18] shows that under some assumption on $\mu$, $\sup_{u \in \mathbb{S}^{d-1}} |\hat{q}_u - q_u|$ is small, with high probability. Hence, using Lemma 6, this yields the following result.

THEOREM 12 ([18], Corollary 2). *Let $\mu$ have a density $f$ with respect to the Lebesgue measure and let $0 < \alpha < \max_{x \in \mathbb{R}^d} D_\mu(x)$. Assume that $f$ satisfies one of the following conditions*:

1. *$f$ is continuous and positive everywhere and there exist $C > 0$ and $\nu > d - 1$ such that $|f(x)| \leq C(1 + \|x\|)^{-\nu}, \forall x \in \mathbb{R}^d$.*

2. *$f$ is supported on a bounded convex set and is uniformly continuous on its support.*

*Then $\mathsf{d}_{\mathsf{H}}(\hat{G}_n, G) = O_{\mathbb{P}}(n^{-1/2})$.*

More precise deviation inequalities are proven in [18] but here, we only state the result in this form

for simplicity of the development. Note that any log-concave distribution satisfies one of the two conditions imposed on $f$ in this theorem and, as a particular case, the uniform distribution on a convex body $K \in \mathcal{K}_d$ satisfies the second condition. Computational questions are tackled in Section 4.

*Estimation of the convex support of a density from noisy observations* [19]. The last model that we present here is an extension to density support estimation, where the observations are contaminated with some additive noise. Consider a convex body $G \in \mathcal{K}_d$ and let $X_1, \ldots, X_n$ be i.i.d. random points uniformly distributed in $G$. Assume that only $Y_i = X_i + \xi_i$, $i = 1, \ldots, n$, are observed, where the $\xi_i$'s are i.i.d. centered Gaussian vectors with covariance matrix $\sigma^2 I_d$ ($I_d$ is the identity matrix and $\sigma^2 > 0$ is known), independent of the $X_i$'s. For $u \in \mathbb{S}^{d-1}$, a consistent estimator of $h_G(u)$ is given by $\hat{h}_n(u) = \max_{1 \le i \le n} \langle u, Y_i \rangle - b_n$, where $b_n \approx \sqrt{2\sigma^2 \ln n}$ is a deterministic debiasing term, and the set estimator of $G$ is defined as $\hat{G}_n = G_{\hat{h}_n}$. By extending some results on extreme value statistics [45, 19] proves that $|\hat{h}_n(u) - h_G(u)|$ is small with high probability, for every single $u \in \mathbb{S}^{d-1}$. However, due to the constant term $b_n$ in its definition, $\hat{h}_n$ is not subadditive in general. In [19], Lemma 6 is used in order to show that

$$\mathsf{d}_\mathsf{H}(\hat{G}_n, G) = O_{\mathbb{P}}\left(\frac{\ln \ln n}{\sqrt{\ln n}}\right).$$

A stronger deviation inequality is proven in [19], which uses the same ideas described above. Moreover, building on techniques developed in [43], it is also shown that the minimax rate of estimation of $G$ in this model is very slow and that $\hat{G}_n$ is nearly rate optimal in a minimax sense.

## 4. APPROXIMATING CONVEX SET ESTIMATORS: A STATISTICAL/COMPUTATIONAL TRADE-OFF

The computational complexity of a set estimator can be very large, and some estimators may not even be computable in practice. For instance, computing the convex hull of $n$ points in $\mathbb{R}^d$ requires approximately $n^{d/2}$ operations (see, e.g., [26], and [7] for a description of the *Quickhull* algorithm, implemented by the function `convhulln` in the **R** package `geometry`), which is not doable if the dimension is too large. Note, though, that in many applications that require the estimation of a set, the dimension is typically small: In spatial data analysis such as home range estimation, convex hulls or local convex hulls [44], or $\alpha$-hulls [21]

are computed with $d = 2$; In econometrics, the efficient boundary problem deals with the feasible productivity domain of a firm, and $d$ is roughly the number of inputs at the firm [83]; In image reconstruction, for example, from satellite data [78], $d = 2$, typically. As for the empirical level sets of the Tukey depth, aka $k$-hulls [28], there is no available algorithm to compute them in high dimensions. As we have seen above (Lemma 7 and (3)), $k$-hulls (including convex hulls, for $k = 0$) can be easily written using infinitely many affine constraints. When $k = 0$, those constraints define a subadditive function, which turns out to be the support function of the random polytope, by Lemma 3. n Section 3.2.2, when $k \ge 1$, we saw that in general, the function defined by these constraints is not subadditive, hence, the support function of the $k$-hull is not simply determined by the univariate order statistics $X_{(k)}(u)$, $u \in \mathbb{S}^{d-1}$.

The idea exploited in [17] for convex hulls and in [18] for $k$-hulls ($k \ge 1$) is to select a finite number of these constraints, leading to a larger approximating polytope, using a discretization of the unit sphere. One must bear in mind the purpose of approximating these sets, which is the estimation of an underlying set: The support of a density, or a population level set of the Tukey depth. Hence, an acceptable approximation error may be as large as the statistical error of the initial estimator.

In order to discretize $\mathbb{S}^{d-1}$, the idea is to sample enough independent random unit vectors. The following lemma, proven in [18], shows that with high probability this procedure provides an $\varepsilon$-net of $\mathbb{S}^{d-1}$, where $\varepsilon$ depends on the number of sampled unit vectors.

LEMMA 10 ([18]). *Let $M$ be a positive integer and $U_1, \ldots, U_M$ be i.i.d. uniform vectors in $\mathbb{S}^{d-1}$. Let $\varepsilon \in (0, 1)$ and let $\mathcal{C}$ be the event satisfied when the collection $\{U_1, \ldots, U_M\}$ is an $\varepsilon$-net of $\mathbb{S}^{d-1}$. Then*

$$\mathbb{P}[\mathcal{C}] \ge 1 - 6^d \exp\left(-\frac{M\varepsilon^{d-1}}{2d8^{\frac{d-1}{2}}} + d\ln\left(\frac{1}{\varepsilon}\right)\right).$$

Together with Lemma 6, this result is used in [17] in order to compute an approximate convex hull in $O(d^2 8^{\frac{d}{2}} n^{2-1/d} (\ln n)^{1/d})$ steps if the data are uniformly distributed in a general convex body and $O(d^2 8^{\frac{d}{2}} n^{\frac{3d-1}{d+1}} (\ln n)^{-\frac{d-3}{d+1}})$ steps if the data are uniformly distributed in a smooth convex body. In [18], an approximate $k$-hull is computed in $O(8^{\frac{d}{2}} d^3 n^d \ln n)$ operations, under some extra assumptions on the population level set, when $k = \alpha n$ for some fixed number $\alpha \in (0, 1)$. Even though [18] provides an algorithm to

approximate the $k$-hull, it still suffers the curse of dimensionality, because of the factor $(\sqrt{8}n)^d$ in its complexity.

# REFERENCES

[1] BALDIN, N. and REISS, M. (2016). Unbiased estimation of the volume of a convex body. *Stochastic Process. Appl.* **126** 3716–3732. MR3565474

[2] BALL, K. (1992). Ellipsoids of maximal volume in convex bodies. *Geom. Dedicata* **41** 241–250. MR1153987

[3] BÁRÁNY, I. (1992). Random polytopes in smooth convex bodies. *Mathematika* **39** 81–92. MR1176473

[4] BÁRÁNY, I. and BUCHTA, C. (1993). Random polytopes in a convex polytope, independence of shape, and concentration of vertices. *Math. Ann.* **297** 467–497. MR1245400

[5] BÁRÁNY, I. and LARMAN, D. G. (1988). Convex bodies, economic cap coverings, random polytopes. *Mathematika* **35** 274–291. MR0986636

[6] BÁRÁNY, I. and REITZNER, M. (2010). Poisson polytopes. *Ann. Probab.* **38** 1507–1531. MR2663635

[7] BARBER, C. B., DOBKIN, D. P. and HUHDANPAA, H. (1996). The quickhull algorithm for convex hulls. *ACM Trans. Math. Software* **22** 469–483. MR1428265

[8] BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. MR3782383

[9] BLASCHKE, W. (1923). *Vorlesungen über Differentialgeometrue II*. Springer, Berlin.

[10] BÖRÖCZKY, K. J. JR., HOFFMANN, L. M. and HUG, D. (2008). Expectation of intrinsic volumes of random polytopes. *Period. Math. Hungar.* **57** 143–164. MR2469601

[11] BREMNER, D., FUKUDA, K. and MARZETTA, A. (1998). Primal-dual methods for vertex and facet enumeration. *Discrete Comput. Geom.* **20** 333–357. MR1649776

[12] BRONSHTEIN, E. M. (1976). $\varepsilon$-entropy of convex sets and functions. *Sib. Math. J.* **17** 303–398.

[13] BRUNEL, V.-E. Uniform deviation and moment inequalities for random polytopes with general densities in arbitrary convex bodies. Preprint. Submitted. Available at arXiv:1704.01620.

[14] BRUNEL, V.-E. (2013). Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron. J. Stat.* **7** 1301–1327. MR3063609

[15] BRUNEL, V.-E. (2014). Non parametric estimation of convex bodies and convex polytopes. Ph.D. thesis, Univ. Pierre et Marie Curie and Univ. Haifa. Available at https://tel.archives-ouvertes.fr/tel-01066977/document.

[16] BRUNEL, V.-E. (2016). Adaptive estimation of convex and polytopal density support. *Probab. Theory Related Fields* **164** 1–16. MR3449384

[17] BRUNEL, V.-E. (2018). Uniform behaviors of random polytopes under the Hausdorff metric. *Bernoulli*. To appear.

[18] BRUNEL, V.-E. (2018). Concentration of the empirical level sets of Tukey's halfspace depth. *Probab. Theory Related Fields*. To appear.

[19] BRUNEL, V.-E., KLUSOWSKI, J. and YANG, D. Estimation of convex supports from noisy measurements. Preprint. Submitted. Available at arXiv:1804.09879.

[20] BUCHTA, C. (2005). An identity relating moments of functionals of convex hulls. *Discrete Comput. Geom.* **33** 125–142. MR2105754

[21] BURGMAN, M. A. and FOX, J. C. (2003). Bias in species range estimates from minimum convex polygons: Implications for conservation and options for improved planning. In *Animal Conservation Forum* **6** 19–28. Cambridge Univ. Press, Cambridge.

[22] CADRE, B., PELLETIER, B. and PUDLO, P. (2013). Estimation of density level sets with a given probability content. *J. Nonparametr. Stat.* **25** 261–272. MR3039981

[23] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878

[24] CHATTERJEE, S. and LAFFERTY, J. (2015). Adaptive risk bounds in unimodal regression. Preprint. Available at arXiv:1512.02956.

[25] CHAZAL, F. and MICHEL, B. (2017). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. Preprint. Available at arXiv:1710.04019.

[26] CHAZELLE, B. (1993). An optimal convex hull algorithm in any fixed dimension. *Discrete Comput. Geom.* **10** 377–409. MR1243335

[27] CHEVALIER, J. (1976). Estimation du support et du contour du support d'une loi de probabilité. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **12** 339–364. MR0451491

[28] COLE, R., SHARIR, M. and YAP, C. K. (1984). On $k$-hulls and related problems. In *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing* 154–166. ACM, New York.

[29] CUEVAS, A. (2009). Set estimation: Another bridge between statistics and geometry. *Bol. Estad. Investig. Oper.* **25** 71–85. MR2750781

[30] CUEVAS, A., FRAIMAN, R. and PATEIRO-LÓPEZ, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.* **44** 311–329. MR2977397

[31] DEVROYE, L. and WISE, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.* **38** 480–488. MR0579432

[32] DUDLEY, R. M. (2014). *Uniform Central Limit Theorems*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **142**. Cambridge Univ. Press, New York. MR3445285

[33] DÜMBGEN, L. and WALTHER, G. (1996). Rates of convergence for random approximations of convex sets. *Adv. in Appl. Probab.* **28** 384–393. MR1387882

[34] DUPIN, C. (1822). *Applications de Géométrie et de Méchanique à la Marine*. Ponts et Chaussées, Paris.

[35] EFRON, B. (1965). The convex hull of a random set of points. *Biometrika* **52** 331–343. MR0207004

[36] EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596. MR0615434

[37] FISHER, L. (1969). Limiting sets and convex hulls of samples from product measures. *Ann. Math. Stat.* **40** 1824–1832. MR0253391

[38] FISHER, L. D. JR. (1966). The convex hull of a sample. *Bull. Amer. Math. Soc.* **72** 555–558. MR0192526

[39] FRESEN, D. J. and VITALE, R. A. (2014). Concentration of random polytopes around the expected convex hull. *Electron. Commun. Probab.* **19** no. 59, 1–8. MR3254738

[40] GARDNER, R. J., KIDERLEN, M. and MILANFAR, P. (2006). Convergence of algorithms for reconstructing convex bodies and directional measures. *Ann. Statist.* **34** 1331–1374. MR2278360

[41] GAYRAUD, G. (1997). Estimation of functionals of density support. *Math. Methods Statist.* **6** 26–46. MR1456645

[42] GEFFROY, J. (1964). Sur un problème d'estimation géométrique. *Publ. Inst. Stat. Univ. Paris* **13** 191–210. MR0202237

[43] GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40** 941–963. MR2985939

[44] GETZ, W. M. and WILMERS, C. C. (2004). A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography* **27** 489–505.

[45] GOLDENSHLUGER, A. and TSYBAKOV, A. (2004). Estimating the endpoint of a distribution in the presence of additive observation errors. *Statist. Probab. Lett.* **68** 39–49. MR2064684

[46] GROEMER, H. (1974). On the mean value of the volume of a random polytope in a convex set. *Arch. Math.* (*Basel*) **25** 86–90. MR0341286

[47] GROEMER, H. (1996). *Geometric Applications of Fourier Series and Spherical Harmonics. Encyclopedia of Mathematics and Its Applications* **61**. Cambridge Univ. Press, Cambridge. MR1412143

[48] GUNTUBOYINA, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist.* **40** 385–411. MR3014311

[49] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2017). Isotonic regression in general dimensions. Preprint. Available at arXiv:1708.09468.

[50] HAN, Q. and WELLNER, J. A. (2016). Multivariate convex regression: Global risk bounds and adaptation. Preprint. Available at arXiv:1601.06844.

[51] HARTIGAN, J. A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267–270. MR0883354

[52] KIM, A. K. H., GUNTUBOYINA, A. and SAMWORTH, R. J. (2018). Adaptation in log-concave density estimation. *Ann. Statist.* **46** 2279–2306. MR3845018

[53] KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560

[54] KONG, L. and MIZERA, I. (2012). Quantile tomography: Using quantiles with multivariate data. *Statist. Sinica* **22** 1589–1610. MR3027100

[55] KOROSTELËV, A. P., SIMAR, L. and TSYBAKOV, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489. MR1332577

[56] KOROSTELEV, A. P., SIMAR, L. and TSYBAKOV, A. B. (1995). On estimation of monotone and convex boundaries. *Publ. Inst. Stat. Univ. Paris* **39** 3–18. MR1744393

[57] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). Estimating the support of a density and functionals of it. *Problemy Peredachi Informatsii* **29** 3–18. MR1215185

[58] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statistics* **82**. Springer, New York. MR1226450

[59] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1994). Asymptotic efficiency in estimation of a convex set. *Probl. Inf. Transm.* **30** 317–327.

[60] LOUSTAU, S. and MARTEAU, C. (2015). Noisy discriminant analysis with boundary assumptions. *J. Nonparametr. Stat.* **27** 425–441. MR3406320

[61] MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. MR1332579

[62] MAMMEN, E. and TSYBAKOV, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618

[63] MANI-LEVITSKA, P. (1993). Characterizations of convex sets. In *Handbook of Convex Geometry, Vol. A, B* 19–41. North-Holland, Amsterdam. MR1242975

[64] MOLCHANOV, I. (2005). *Theory of Random Sets*. Springer London, Ltd., London. MR2132405

[65] MOORE, M. (1984). On the estimation of a convex set. *Ann. Statist.* **12** 1090–1099. MR0751296

[66] MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746. MR1147099

[67] PARDON, J. (2011). Central limit theorems for random polygons in an arbitrary convex set. *Ann. Probab.* **39** 881–903. MR2789578

[68] PARDON, J. (2012). Central limit theorems for uniform model random polygons. *J. Theoret. Probab.* **25** 823–833. MR2956214

[69] PATEIRO-LOPEZ, B. (2008). Set estimation under convexity type restrictions. Ph.D. thesis. Available at http://eio.usc.es/pub/pateiro/files/THESIS_BeatrizPateiroLopez.pdf.

[70] POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—An excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204

[71] REITZNER, M. (2003). Random polytopes and the Efron–Stein jackknife inequality. *Ann. Probab.* **31** 2136–2166. MR2016615

[72] REITZNER, M. (2004). Stochastic approximation of smooth convex bodies. *Mathematika* **51** 11–29. MR2220208

[73] REITZNER, M. (2005). Central limit theorems for random polytopes. *Probab. Theory Related Fields* **133** 483–507. MR2197111

[74] RÉNYI, A. and SULANKE, R. (1963). Über die konvexe Hülle von *n* zufällig gewählten Punkten. *Z. Wahrsch. Verw. Gebiete* **2** 75–84. MR0156262

[75] RÉNYI, A. and SULANKE, R. (1964). Über die konvexe Hülle von *n* zufällig gewählten Punkten. II. *Z. Wahrsch. Verw. Gebiete* **3** 138–147. MR0169139

[76] RIPLEY, B. D. and RASSON, J.-P. (1977). Finding the edge of a Poisson forest. *J. Appl. Probab.* **14** 483–491. MR0451339

[77] RODRÍGUEZ CASAL, A. (2007). Set estimation under convexity type assumptions. *Ann. Inst. Henri Poincaré Probab. Stat.* **43** 763–774. MR3252430

[78] RODRÍGUEZ-CASAL, A. and SAAVEDRA-NIEVES, P. (2016). A fully data-driven method for estimating the shape of a point cloud. *ESAIM Probab. Stat.* **20** 332–348. MR3557598

[79] SAGER, T. W. (1979). An iterative method for estimating a multivariate mode and isopleth. *J. Amer. Statist. Assoc.* **74** 329–339. MR0548023

[80] SCHNEIDER, R. (1993). *Convex Bodies*: *The Brunn–Minkowski Theory*. *Encyclopedia of Mathematics and Its Applications* **44**. Cambridge Univ. Press, Cambridge. MR1216521

[81] SCHÜTT, C. (1994). Random polytopes and affine surface area. *Math. Nachr.* **170** 227–249. MR1302377

[82] SCHÜTT, C. and WERNER, E. (1990). The convex floating body. *Math. Scand.* **66** 275–290. MR1075144

[83] SIMAR, L. and WILSON, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. *J. Product. Anal.* **13** 49–78.

[84] THÄLE, C. (2008). 50 years sets with positive reach—A survey. *Surv. Math. Appl.* **3** 123–165. MR2443192

[85] TSYBAKOV, A. B. (1994). Multidimensional change-point problems and boundary estimation. In *Change-Point Problems* (*South Hadley, MA*, 1992). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23** 317–329. IMS, Hayward, CA. MR1477933

[86] TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. MR1447735

[87] TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002

[88] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

[89] VAN DE GEER, S. (1998). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

[90] VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[91] VU, V. H. (2005). Sharp concentration of random polytopes. *Geom. Funct. Anal.* **15** 1284–1318. MR2221249

[92] WALTHER, G. (1997). Granulometric smoothing. *Ann. Statist.* **25** 2273–2299. MR1604445

[93] WALTHER, G. (1999). On a generalization of Blaschke's rolling theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.* **22** 301–316. MR1671447

[94] WASSERMAN, L. (2018). Topological data analysis. *Annu. Rev. Stat. Appl.* **5** 501–535. MR3774757

[95] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. MR1902898

[96] ZIEGLER, G. M. (1995). *Lectures on Polytopes*. *Graduate Texts in Mathematics* **152**. Springer, New York. MR1311028