

Submitted to Bernoulli

Estimation of convex supports from noisy measurements

VICTOR-EMMANUEL BRUNEL¹, JASON M. KLUSOWSKI², and DANA YANG³

¹*Department of Statistics
CREST UMR 9194, ENSAE
5 avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau cedex, France
E-mail: victor.emmanuel.brunel@ensae.fr*

²*Department of Statistics
Rutgers University
110 Frelinghuysen Road
Piscataway, New Jersey 08854-8019, USA
E-mail: jason.klusowski@rutgers.edu*

³*The Fuqua School of Business
Duke University
100 Fuqua Drive
Durham, North Carolina 27708, USA
E-mail: xiaoqian.yang@duke.edu*

A popular class of problems in statistics deals with estimating the support of a density from n observations drawn at random from a d -dimensional distribution. In the one-dimensional case, if the support is an interval, the problem reduces to estimating its end points. In practice, an experimenter may only have access to a noisy version of the original data. Therefore, a more realistic model allows for the observations to be contaminated with additive noise.

In this paper, we consider estimation of convex bodies when the additive noise is distributed according to a multivariate Gaussian (or nearly Gaussian) distribution, even though our techniques could easily be adapted to other noise distributions. Unlike standard methods in deconvolution that are implemented by thresholding a kernel density estimate, our method avoids tuning parameters and Fourier transforms altogether. We show that our estimator, computable in $(O(\log n))^{(d-1)/2}$ time, converges at a rate of $O_d(\log \log n / \sqrt{\log n})$ in Hausdorff distance, in accordance with the polylogarithmic rates encountered in Gaussian deconvolution problems. Part of our analysis also involves the optimality of the proposed estimator. We provide a lower bound for the minimax rate of estimation in Hausdorff distance that is $\Omega_d(1/\log^2 n)$.

AMS 2000 subject classifications: Primary 62H12; secondary 62G30.

Keywords: Convex bodies, order statistics, support estimation, support function.

1. Preliminaries

1.1. Introduction

The problem of estimating the support of a distribution, given i.i.d. samples, poses both statistical and computational questions. When the support of the distribution is known to be convex, geometric methods have been borrowed from stochastic and convex geometry with the use of random polytopes since the seminal works [16, 17]. When the distribution of the samples is uniform on a convex body, estimation in a minimax setup has been tackled in [14] (see also the references therein). There, the natural estimator defined as the convex hull of the samples is shown to attain the minimax rate of convergence on the class of convex bodies, under the Nikodym metric.

When the samples are still supported on a convex body but their distribution is no longer uniform, [2] studies the performance of the convex hull of the samples as an estimator of the convex support under the Nikodym metric, whereas [4] focuses on the Hausdorff metric. In the latter, computational issues are addressed in higher dimensions. In particular, determining the list of vertices of the convex hull of n points in dimension $d \geq 2$ is very expensive, namely, exponential in $d \log n$ (see [5]). Consequently, in [4], a randomized algorithm produces an approximation of the convex hull of the samples that achieves a trade-off between computational cost and statistical accuracy.

Both of the aforementioned works [2, 4] assume that one has access to direct samples. Here, we are interested in the case when the samples are contaminated, more specifically, subject to measurement errors. In [15], a closely related problem is studied, where two independent contaminated samples are observed, and one wants to estimate the set where $f - g$ is positive, where f and g are the respective densities that generated the batches of samples. In that work, the contamination is modeled as an additive noise with known distribution, and some techniques borrowed from inverse problems are used. The main drawback is that the estimator is not tractable and it only gives a theoretical benchmark for minimax estimation.

Goldenshluger and Tsybakov [10] study the problem of estimating the endpoint of a univariate distribution, given samples contaminated with additive noise. Their analysis suggests that their estimator is optimal in a minimax sense and its computation is straightforward. In our work, we first extend their result in one dimension to handle a broader class of data distributions, and then we lift it to a higher dimensional setup—that of estimating the convex support of a uniform distribution, given samples that are contaminated with additive noise. Our method relies on projecting the data points along a finite collection of unit vectors. Unlike in [15], we give an explicit form for our estimator. In addition, our estimator is tractable when the ambient dimension is not too large. If the dimension is too high, the number of steps required to compute a membership oracle for our estimator (i.e., evaluate the indicator function of the estimator at any given point of the space) becomes exponentially large in the dimension, namely, of order

$(O(\log n))^{(d-1)/2}$. When the additive noise is Gaussian, we prove upper and lower bound on the minimax risk of estimation of the support, both of which are polylogarithmic in the sample size.

1.2. Notation

In this work, $d \geq 2$ is a fixed integer standing for the dimension of the ambient Euclidean space \mathbb{R}^d . The Euclidean ball with center $a \in \mathbb{R}^d$ and radius $r \geq 0$ is denoted by $B_d(a, r)$. The Euclidean inner product between two vectors u and v in \mathbb{R}^d is denoted by $\langle u, v \rangle$. The unit sphere in \mathbb{R}^d is denoted by \mathbb{S}^{d-1} and κ_d stands for the volume of the unit Euclidean ball.

We refer to convex and compact sets with nonempty interior in \mathbb{R}^d as convex bodies. The collection of all convex bodies in \mathbb{R}^d is denoted by \mathcal{K}_d . Let $\sigma^2 > 0$ and $n \geq 1$. If X_1, \dots, X_n are i.i.d. random uniform points in a convex body G and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d., we denote by \mathbb{P}_G the joint distribution of $X_1 + \varepsilon_1, \dots, X_n + \varepsilon_n$ and by \mathbb{E}_G the corresponding expectation operator (we omit the dependency on n for simplicity).

The support function of a convex set $G \subseteq \mathbb{R}^d$ is defined as $h_G(u) = \sup_{x \in G} \langle u, x \rangle$, $u \in \mathbb{R}^d$, where $\langle \cdot, \cdot \rangle$ is the canonical scalar product in \mathbb{R}^d —it is the largest signed distance between the origin and a supporting hyperplane of G orthogonal to u .

The Hausdorff distance between two sets $A, B \subseteq \mathbb{R}^d$ is $\mathbf{d}_H(A, B) = \inf\{\epsilon > 0 : G_1 \subseteq G_2 + \epsilon B_d(0, 1) \text{ and } G_2 \subseteq G_1 + \epsilon B_d(0, 1)\}$. If A and B are convex bodies, it can be written in terms of their support functions: $\mathbf{d}_H(A, B) = \sup_{u \in \mathbb{S}^{d-1}} |h_A(u) - h_B(u)|$.

For f in $L^1(\mathbb{R}^d)$, let $\mathcal{F}[f](t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} f(x) dx$ denote the Fourier transform of f .

The total variation distance between two distributions P and Q having densities p and q with respect to a dominating measure μ is defined by $\text{TV}(P, Q) = \frac{1}{2} \int |p - q| d\mu$.

The Lebesgue measure of a measurable, bounded set A in \mathbb{R}^d is denoted by $|A|$. For a vector $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, we define $\|x\| = (\sum_{i=1}^d |x_i|^2)^{1/2}$ and $\|x\|_\infty = \sup_{1 \leq i \leq d} |x_i|$. For a function, f defined on a set A , let $\|f\|_\infty = \sup_{x \in A} |f(x)|$. The Nikodym distance between two measurable, bounded sets A and B is defined by $\mathbf{d}_\Delta(A, B) = |A \Delta B|$, where $A \Delta B$ is the symmetric difference between A and B . We let ϕ_σ denote the Gaussian density with mean zero and variance σ^2 , i.e., $\phi_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}$ for all $x \in \mathbb{R}$.

We use standard big- O notations: For any positive sequences $\{a_n\}$ and $\{b_n\}$, $a_n = O(b_n)$ or $a_n \lesssim b_n$ if $a_n \leq C b_n$ for some absolute constant $C > 0$, $a_n = o(b_n)$ or $a_n \ll b_n$ if $\lim a_n/b_n = 0$. Finally, we write $a_n \asymp b_n$ or $a_n = \Theta(b_n)$ when both $a_n \gtrsim b_n$ and $a_n \lesssim b_n$ hold. Furthermore, the subscript in $a_n = O_r(b_n)$ means $a_n \leq C_r b_n$ for some constant C_r .

depending on the parameter r only. We write $a_n \propto b_n$ when $a_n = Cb_n$ for some absolute constant C .

1.3. Model and outline

In what follows, we consider the problem of estimating a convex body from noisy observations. More formally, suppose we have access to independent observations

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n, \quad (1)$$

where X_1, \dots, X_n are i.i.d. uniform random points in an unknown convex body G and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. Gaussian random vectors with zero mean and covariance matrix $\sigma^2 I$, independent of X_1, \dots, X_n . In the sequel, we assume that σ^2 is a fixed and known positive number. The goal is to estimate G using Y_1, \dots, Y_n . This can be seen as an inverse problem: the object of interest is a special feature (here, the support) of a density that is observed up to a convolution with a Gaussian distribution. Our approach will not follow the usual path of inverse problems, but instead, will be essentially based on geometric arguments. As we will see, our method is robust to some type of misspecification of both the uniform distribution on G and the distribution of the noise.

The error of an estimator \hat{G} of G is defined as $\mathbb{E}_G [\mathbf{d}_H(\hat{G}, G)]$. Let $\mathcal{C} \subseteq \mathcal{K}_d$ be a subclass of the class of all convex bodies in \mathbb{R}^d . The risk of an estimator \hat{G} on the class \mathcal{C} is $\sup_{G \in \mathcal{C}} \mathbb{E}_G [\mathbf{d}_H(\hat{G}, G)]$ and the minimax risk on \mathcal{C} is defined as

$$\mathcal{R}_n(\mathcal{C}) = \inf_{\hat{G}} \sup_{G \in \mathcal{C}} \mathbb{E}_G [\mathbf{d}_H(\hat{G}, G)],$$

where the infimum is taken over all estimators \hat{G} based on Y_1, \dots, Y_n . The minimax rate on the class \mathcal{C} is the speed at which $\mathcal{R}_n(\mathcal{C})$ goes to zero.

As we mentioned earlier, our strategy for estimating G avoids standard methods from inverse problems that would require Fourier transforms and tuning parameters. To give intuition for our procedure, first observe that a convex set can be represented in terms of its support function via

$$G = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq h_G(u) \text{ for all } u \in \mathbb{S}^{d-1}\}.$$

If we can find a suitable way of estimating h_G , say by \hat{h}_n , then there is hope that an estimator of the form

$$\hat{G} = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq \hat{h}_n(u) \text{ for all } u \in \mathbb{S}^{d-1}\} \quad (2)$$

will perform well. This is the core idea of our procedure: We project the data points Y_1, \dots, Y_n along unit vectors and for all such $u \in \mathbb{S}^{d-1}$, we estimate the endpoint of the distribution of $\langle u, X_1 \rangle$, given the one-dimensional sample $\langle u, Y_1 \rangle, \dots, \langle u, Y_n \rangle$.

Section 2 is devoted to the study of the one-dimensional case, where we extend the results proven in [10]. The one-dimensional case reduces to estimating the end-point of a univariate density. This problem has been extensively studied in the noiseless case [6, 11] and more recently as an inverse problem [10, 12]. In [10], it is assumed that the density of the (one-dimensional) X_j 's is *exactly* equal to a polynomial in a neighborhood of the endpoint of the support. We extend their results to the case when the distribution function is only bounded by two polynomials whose degrees may differ in the vicinity of the endpoint.

In Section 3, we use these one-dimensional results to provide theory for our estimator of the support G , if the X_j 's are drawn uniformly from a convex set, and to bound its risk on a certain subclass of \mathcal{K}_d . The main case of interest is when the noise terms are nearly Gaussian. When the noise distribution is exactly Gaussian, we show that our estimator nearly attains the minimax rate on that class, up to logarithmic factors.

Intermediate lemmas and proofs of corollaries are deferred to the supplementary material [3].

2. Estimation of the endpoint of a distribution with contaminated samples

2.1. Preliminary bounds

Let $\varepsilon_1, \dots, \varepsilon_n$ be i.i.d. random variables with some density ϕ on \mathbb{R} . Then, under some assumption on the right tail of ϕ , the maximum $\max_{1 \leq j \leq n} \varepsilon_j$ concentrates around some large value: For instance, if ϕ is a Gaussian density with variance $\sigma^2 > 0$, that value is $\sqrt{2\sigma^2 \log n}$. In this section, we prove that adding i.i.d. nonpositive random variables to the ε_j 's does not affect this concentration, as long as their cumulative distribution function increases polynomially to 1 near zero. As a byproduct, we get a guarantee for the estimation of the endpoint of a distribution, given contaminated samples.

Theorem 1. *Let X be a random variable on $(-\infty, 0)$ with its cdf F satisfying*

$$L^{-1}t^\alpha \leq 1 - F(-t) \leq Lt^\beta,$$

for all $t \in [0, r]$, where $L, r > 0$ and $\alpha \geq \beta > 0$. Let ε be a real valued random variable, independent of X , with density ϕ with respect to the Lebesgue measure. Let G be the cdf of $X + \varepsilon$. In each of the following cases, C_1 and C_2 are positive constants that depend on the stated parameters.

(i) *If, for some positive constants A, γ, x_0, c, C ,*

$$ce^{-Ax^\gamma} \leq \phi(x) \leq Ce^{-Ax^\gamma}, \quad \forall x \geq x_0,$$

then for all $x \geq \max(x_0, r)$,

$$C_1 \frac{e^{-Ax^\gamma}}{x^{(\gamma-1)(\alpha+1)}} \leq 1 - G(x) \leq C_2 \frac{e^{-Ax^\gamma}}{x^{(\gamma-1)(\beta+1)}}.$$

(ii) If, for some positive constants x_0, c, C , and $\gamma > 1$,

$$cx^{-\gamma} \leq \phi(x) \leq Cx^{-\gamma}, \quad \forall x \geq x_0,$$

then for all $x \geq \max(x_0, r)$,

$$\frac{C_1}{x^{\gamma-1}} \leq 1 - G(x) \leq \frac{C_2}{x^{\gamma-1}}.$$

(iii) If there is a real number Q , a nonnegative constant $\gamma \geq 0$ and positive constants c, C , and x_0 such that $\phi(x) = 0$ for all $x \geq Q$ and

$$c(Q-x)^\gamma \leq \phi(x) \leq C(Q-x)^\gamma, \quad \forall x \in [Q-x_0, Q],$$

then, $G(x) = 1$ for all $x \geq Q$ and, if $\max(Q-x_0, Q-r) \leq x \leq Q$, then

$$C_1(Q-x)^{\gamma+\alpha+1} \leq 1 - G(x) \leq C_2(Q-x)^{\gamma+\beta+1}.$$

In Case (i), a distribution for ε with $\gamma = 2$ has a *nearly Gaussian* right tail, whereas for $\gamma = 1$, its right tail is *nearly exponential*. In (ii), a distribution with $\gamma = 2$ has a *nearly Cauchy* right tail. Case (iii) with $\gamma = 0$ includes uniform, or nearly uniform distributions on segments of the form $[\tilde{Q}, Q]$ for some $\tilde{Q} \leq Q$.

Proof.

Case (i) Let $x \geq \max(x_0, r)$. Then,

$$\begin{aligned} 1 - G(x) &= \int_0^\infty (1 - F(-t)) \phi(x+t) dt \\ &= \underbrace{\int_0^r (1 - F(-t)) \phi(x+t) dt}_I + \underbrace{\int_r^\infty (1 - F(-t)) \phi(x+t) dt}_{II}. \end{aligned} \quad (3)$$

In both integrals, we can use the assumption on ϕ , since $x+t$ is always larger or equal to x_0 . In the first integral, we can also use the assumption on F :

$$\begin{aligned} I &\leq LC \int_0^r t^\beta e^{-A(x+t)^\gamma} dt \\ &= LCe^{-Ax^\gamma} \int_0^r t^\beta e^{-A((x+t)^\gamma - x^\gamma)} dt. \end{aligned}$$

Now, for all $x \geq r$ and $t \in (0, r)$, a Taylor expansion yields the existence of $u \in (0, t)$ (hence, $u < x$), such that $(x+t)^\gamma - x^\gamma = \gamma t(x+u)^{\gamma-1}$. Thus, if $\gamma \geq 1$, then

$$\gamma t x^{\gamma-1} \leq (x+t)^\gamma - x^\gamma \leq 2^{\gamma-1} \gamma t x^{\gamma-1}, \quad (4)$$

and if $\gamma < 1$, these inequalities are reversed, i.e.,

$$2^{\gamma-1} \gamma t x^{\gamma-1} \leq (x+t)^\gamma - x^\gamma \leq \gamma t x^{\gamma-1}. \quad (5)$$

In any case, for some $c_1 > c_0 > 0$, we get that

$$c_0 t x^{\gamma-1} \leq (x+t)^\gamma - x^\gamma \leq c_1 t x^{\gamma-1}. \quad (6)$$

$$\begin{aligned} I &\leq LC e^{-Ax^\gamma} \int_0^r t^\beta e^{-c_0 t x^{\gamma-1}} dt \\ &\leq LC e^{-Ax^\gamma} \int_0^\infty t^\beta e^{-c_0 t x^{\gamma-1}} dt \\ &= LC \frac{e^{-Ax^\gamma}}{(c_0 x^{\gamma-1})^{\beta+1}} \int_0^\infty t^\beta e^{-t} dt \\ &\leq \frac{LC \Gamma(\beta+1)}{c_0^{\beta+1}} \frac{e^{-Ax^\gamma}}{x^{(\gamma-1)(\beta+1)}}, \end{aligned} \quad (7)$$

where $\Gamma(\cdot)$ is Euler's Gamma function.

On the other hand, a similar sequence of inequalities yields that

$$\begin{aligned} I &\geq L^{-1} c \int_0^r t^\beta e^{-A(x+t)^\gamma} dt \\ &\geq L^{-1} c \frac{e^{-Ax^\gamma}}{(c_1 x^{\gamma-1})^{\alpha+1}} \int_0^{c_1 r x^{\gamma-1}} t^\alpha e^{-t} dt \\ &\geq \frac{L^{-1} c}{c_1^{\alpha+1}} \left(\int_0^{c_1 r^\gamma} t^\alpha e^{-t} dt \right) \frac{e^{-Ax^\gamma}}{x^{(\gamma-1)(\alpha+1)}}, \end{aligned} \quad (8)$$

where c_1 was defined in (6). Now, let us bound the second integral in (8). Since it is already nonnegative, we only bound it from above. Since $0 \leq F(u) \leq 1$, for all $u \in \mathbb{R}$,

$$\begin{aligned} II &\leq \int_r^\infty \phi(x+t) dt \\ &\leq C \int_r^\infty e^{-A(x+t)^\gamma} dt. \end{aligned} \quad (10)$$

Assume that $\gamma \geq 1$. Then, for all $x, t \geq r$, a Taylor expansion yields that $(x+t)^\gamma - x^\gamma \geq$

$\gamma tx^{\gamma-1}$. Thus,

$$\begin{aligned}
II &\leq Ce^{-Ax^\gamma} \int_r^\infty e^{-A\gamma tx^{\gamma-1}} dt \\
&= Ce^{-Ax^\gamma} \frac{e^{-A\gamma rx^{\gamma-1}}}{A\gamma x^{\gamma-1}} \\
&= \frac{C}{A\gamma} \frac{e^{-Ax^\gamma}}{x^{(\gamma-1)(\beta+1)}} e^{-A\gamma rx^{\gamma-1} + (\gamma-1)\beta \log x} \\
&\leq \frac{Ce^{-\beta + \beta \log \frac{\beta}{A\gamma r}}}{A\gamma} \frac{e^{-Ax^\gamma}}{x^{(\gamma-1)(\beta+1)}}.
\end{aligned} \tag{11}$$

Now, if $\gamma < 1$, let $III = \int_r^\infty e^{-A(x+t)^\gamma} dt$. Then, by an integration by parts,

$$\begin{aligned}
III &= \int_{x+r}^\infty e^{-At^\gamma} dt \\
&= \frac{e^{-A(x+r)^\gamma}}{\gamma A(x+r)^{\gamma-1}} + \frac{1-\gamma}{\gamma A} \int_{x+r}^\infty \frac{1}{t^\gamma} e^{-At^\gamma} dt \\
&\leq \frac{e^{-A(x+r)^\gamma}}{\gamma A(x+r)^{\gamma-1}} + \frac{1-\gamma}{\gamma A(x+r)^\gamma} III \\
&\leq \frac{e^{-A(x+r)^\gamma}}{\gamma A(x+r)^{\gamma-1}} + (1/2)III,
\end{aligned}$$

when x is large enough. This yields

$$III \leq \frac{2e^{-A(x+r)^\gamma}}{\gamma A(x+r)^{\gamma-1}} \leq \frac{\tilde{C}e^{-Ax^\gamma}}{x^{\gamma-1}} \leq \frac{\tilde{C}e^{-Ax^\gamma}}{x^{(\gamma-1)(\beta+1)}}, \tag{12}$$

for some positive constant \tilde{C} , and (10) yields

$$II \leq C' \frac{e^{-Ax^\gamma}}{x^{(\gamma-1)(\beta+1)}}, \tag{13}$$

for some positive constant C' . All together, (3), (10), (12) and (13) imply the desired result.

Case (ii) For all $x \geq \max(x_0, r)$,

$$\begin{aligned}
1 - G(x) &= \int_0^\infty (1 - F(-t)) \phi(x+t) dt \\
&\leq \int_0^\infty \phi(x+t) dt \\
&\leq C \int_0^\infty (x+t)^{-\gamma} \\
&\leq \frac{C}{\gamma-1} \frac{1}{x^{\gamma-1}}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
 1 - G(x) &\geq \int_r^\infty (1 - F(-t)) \phi(x+t) dt \\
 &\geq (1 - F(-r)) \int_r^\infty \phi(x+t) dt \\
 &\geq L^{-1} r^{-\alpha} c \int_r^\infty (x+t)^{-\gamma} \\
 &= \frac{L^{-1} r^{-\alpha} c}{\gamma - 1} \frac{1}{(x+r)^{\gamma-1}} \\
 &\geq \frac{C_\gamma L^{-1} r^{-\alpha} c}{\gamma - 1} \frac{1}{(x+r)^{\gamma-1}},
 \end{aligned}$$

where $C_\gamma = 1$ if $\gamma < 1$ and $C_\gamma = 2^{\gamma-1}$ if $\gamma \geq 1$.

Case (iii) Let $x \geq \max(Q - x_0, Q - r)$.

$$\begin{aligned}
 1 - G(x) &= \int_0^{Q-x} (1 - F(-t)) \phi(x+t) dt \\
 &\leq LC \int_0^{Q-x} t^\beta (Q-x-t)^\gamma dt \\
 &= LC(Q-x)^{\gamma+\beta+1} \int_0^1 u^\beta (1-u)^\gamma du
 \end{aligned}$$

and similarly,

$$1 - G(x) \geq L^{-1} c (Q-x)^{\gamma+\alpha+1} \int_0^1 u^\alpha (1-u)^\gamma du.$$

□

As a consequence of Theorem 1, we get the following deviation bounds for the extreme statistics of i.i.d. samples.

Theorem 2. *Let F be a cdf on \mathbb{R} such that $L^{-1}t^\alpha \leq 1 - F(-t) \leq Lt^\beta$, for all $t \in [0, r]$, where $L, r > 0$ and $\alpha \geq \beta > 0$. Let ϕ be a density on \mathbb{R} with $ce^{-Ax^\gamma} \leq \phi(x) \leq Ce^{-Ax^\gamma}$ for all $x \geq x_0$, where A, x_0, c, C are positive constants and $\gamma \geq 1$. Let ε be a random variable with density ϕ , independent of X and let $Y = X + \varepsilon$. Let Y_1, Y_2, \dots be a sequence of i.i.d. copies of Y . For $n \geq 1$, let $M_n = \max(Y_1, \dots, Y_n)$ and $b_n = (A^{-1} \log n)^{1/\gamma}$. Then, there exist positive constants c_0, c_1, c_2, c_3 and a positive integer n_0 , which only depend on $L, c, \alpha, \beta, \gamma$, such that for all $n \geq n_0$ and for all $t \geq 0$,*

$$\mathbb{P} \left[|M_n - b_n| > \frac{t + c_0 \log(b_n)}{b_n^{\gamma-1}} \right] \leq c_1 e^{-c_2 t} + e^{-c_3 n}.$$

Proof. Let $n \geq 1$ such that $b_n \geq \max(x_0, r)$. Then, by Theorem 1, for all $x \geq 0$,

$$\begin{aligned} G(b_n + x) &\geq 1 - C_2 \frac{e^{-A(b_n+x)^\gamma}}{(b_n+x)^{(\gamma-1)(\beta+1)}} \\ &\geq 1 - \frac{C_2}{b_n^{(\gamma-1)(\beta+1)}} e^{-Ab_n^\gamma} e^{-A((b_n+x)^\gamma - b_n^\gamma)} \\ &\geq 1 - \frac{C_2}{n} e^{-A\gamma x b_n^{\gamma-1}}, \end{aligned} \quad (14)$$

hence, so long as $C_2/n \geq 1/2$,

$$G(b_n + x)^n \geq 1 - (2 \log 2) C e^{-A\gamma x b_n^{\gamma-1}}, \quad (15)$$

where we used the fact that for $u \in [0, 1/2]$, $\log(1-u) \leq 2(\log 2)u$ and for $v \in \mathbb{R}$, $e^v \geq 1+v$.

Now, if $b_n - x \geq \max(x_0, r)$, Theorem 1 yields

$$\begin{aligned} G(b_n - x) &\leq 1 - C_1 \frac{e^{-A(b_n-x)^\gamma}}{(b_n-x)^{(\gamma-1)(\alpha+1)}} \\ &\leq 1 - \frac{C_1}{b_n^{(\gamma-1)(\alpha+1)}} e^{-Ab_n^\gamma} e^{A(b_n^\gamma - (b_n-x)^\gamma)} \\ &= 1 - \frac{C_1}{n b_n^{(\gamma-1)(\alpha+1)}} e^{Ab_n^\gamma (1 - (1-x/b_n)^\gamma)} \\ &\leq 1 - \frac{C_1}{n b_n^{(\gamma-1)(\alpha+1)}} e^{A x b_n^{\gamma-1}} \\ &= 1 - \frac{C_1}{n} e^{A x b_n^{\gamma-1} - (\gamma-1)(\alpha+1) \log b_n}. \end{aligned} \quad (16)$$

Therefore, using $(1-u/n)^n \leq e^{-u} \leq 1/u$, for all $u > 0$,

$$G(b_n - x)^n \leq \frac{1}{C_1} e^{-A x b_n^{\gamma-1} + (\gamma-1)(\alpha+1) \log b_n}. \quad (17)$$

If $b_n - x \leq \max(x_0, r) =: M$, then since G is nondecreasing, $G(b_n - x)^n \leq G(M)^n \leq e^{-c_3 n}$, where $c_3 = -\log \left(1 - \frac{C_1 e^{-AM^\gamma}}{M^{(\gamma-1)(\alpha+1)}} \right)$. Finally, for all $x \geq 0$, using (15) and (17),

$$\begin{aligned} \mathbb{P}[|M_n - b_n| > x] &= \mathbb{P}[M_n > b_n + x] + \mathbb{P}[M_n < b_n - x] \\ &= 1 - G(b_n + x)^n + G(b_n - x)^n \\ &\leq (2 \log 2) C e^{-A\gamma x b_n^{\gamma-1}} + \frac{1}{C_1} e^{-A x b_n^{\gamma-1} + (\gamma-1)(\alpha+1) \log b_n} + e^{-c_3 n} \\ &\leq (2C \log 2 + C_1^{-1}) e^{-A x b_n^{\gamma-1} + (\gamma-1)(\alpha+1) \log b_n} + e^{-c_3 n}. \end{aligned} \quad (18)$$

Take x of the form $x = \frac{t + A^{-1}(\gamma-1)(\alpha+1) \log b_n}{b_n^{\gamma-1}}$, for $t \geq 0$. Then, (18) implies the desired result, with $c_0 = A^{-1}(\gamma-1)(\alpha+1)$, $c_1 = 2C \log 2 + C_1^{-1}$ and $c_2 = A$. \square

When $\alpha = \beta$, it is possible to improve the bound of Theorem 2 by a log-log factor.

Theorem 3. *Let the assumptions of Theorem 2 hold with $\alpha = \beta$. Set $\tilde{b}_n = (A^{-1} \log n)^{\frac{1}{\gamma}} \left(1 - \frac{(\gamma-1)(\alpha+1) \log \log n}{\gamma^2 \log n}\right)$. Then, there exist $n_0 \geq 1$ and $c_1, c_2 > 0$ that depend on α, L, r, A and x_0 only, such that for all $n \geq n_0$ and $t > 0$,*

$$\mathbb{P} \left[|M_n - \tilde{b}_n| > \frac{t}{\tilde{b}_n^{\gamma-1}} \right] \leq c_1 e^{-\frac{t}{2\sigma^2}} + e^{-c_2 n}.$$

Proof. The proof of Theorem 3 follows the same lines as that of Theorem 2, where b_n is replaced with \tilde{b}_n . In the lower bound of $G(\tilde{b}_n + x)$, (14) becomes

$$G(b_n + x) \geq 1 - C_2 e^{-A\gamma x \tilde{b}_n^{\gamma-1} - Ab_n^\gamma + (\gamma-1)(\alpha+1) \log \tilde{b}_n}$$

and in the upper bound of $G(\tilde{b}_n - x)$, (16) becomes

$$G(\tilde{b}_n - x) \leq 1 - C_1 e^{A\gamma x \tilde{b}_n^{\gamma-1} - Ab_n^\gamma + (\gamma-1)(\alpha+1) \log \tilde{b}_n}.$$

The next step consists of showing that $\log n - B \leq Ab_n^\gamma + (\gamma-1)(\alpha+1) \log \tilde{b}_n \leq \log n + B$, for some positive number B , and for all n larger than some given integer n_0 . Let $u = \frac{(\gamma-1)(\alpha+1) \log \log n}{\gamma^2 \log n}$ and let n_0 be the smallest integer such that $u \leq 1/2$, for all $n \geq n_0$ (since $u \leq \frac{(\gamma-1)(\alpha+1)}{\gamma^2 \sqrt{\log n}}$, one can take any $n_0 \geq e^{4(\gamma-1)^2(\alpha+1)^2/\gamma^4}$). Then, $-\log 2 \leq \log(1-u) \leq 0$ and, by a Taylor expansion, $1 - \gamma u - C_\gamma u^2 \leq (1-u)^\gamma \leq 1 - \gamma u - C_\gamma u^2$, where $C_\gamma = \frac{\gamma(\gamma-1)}{2} \max(1, 2^{2-\gamma})$. The result follows easily.

Finally, similarly to (18), we get:

$$\begin{aligned} \mathbb{P} [|M_n - b_n| > x] &= \mathbb{P} [M_n > b_n + x] + \mathbb{P} [M_n < b_n - x] \\ &\leq C e^{-A\gamma x b_n^{\gamma-1}} + e^{-c'n} \end{aligned}$$

for some positive constants C, c and $n \geq n_0$, hence, taking x of the form $x = \frac{t}{\tilde{b}_n^{\gamma-1}}$, for $t \geq 0$, yields Theorem 3. \square

When the noise distribution has lighter tails, the extreme statistics concentrates faster, which is quantified in the next bound.

Theorem 4. *Let F be a cdf on \mathbb{R} such that $L^{-1}t^\alpha \leq 1 - F(-t) \leq Lt^\beta$, for all $t \in [0, r]$, where $L, r > 0$ and $\alpha \geq \beta > 0$. Let ϕ be a density on \mathbb{R} with $\phi(x) = 0$ for all $x \geq Q$ and $c(Q-x)^\gamma \leq \phi(x) \leq C(Q-x)^\gamma$ for all $x \in [Q-x_0, Q]$, where γ, Q, x_0, c, C are positive constants. Let ε be a random variable with density ϕ , independent of X and let $Y = X + \varepsilon$.*

Let Y_1, Y_2, \dots be a sequence of i.i.d. copies of Y . Then, for all positive integers n and $t \in [0, n^{1/(\gamma+\alpha+1)} \min(r, x_0)]$,

$$\mathbb{P}\left[M_n < Q - tn^{-\frac{1}{\gamma+\alpha+1}}\right] \leq e^{-C_1 t^{\gamma+\alpha+1}},$$

where C_1 is the constant appearing in Case (iii) of Theorem 1.

Proof. The proof is straightforward, using Case (iii) of Theorem 1. Indeed, for all $x \in [0, \min(r, x_0)]$,

$$\begin{aligned} \mathbb{P}[M_n < Q - x] &= G(Q - x)^n \\ &\leq (1 - C_1 x^{\gamma+\alpha+1})^n \\ &\leq e^{-C_1 n x^{\gamma+\alpha+1}}. \end{aligned}$$

□

Remark 1. Note that when the noise distribution satisfies Case (ii) of Theorem 1, the extreme statistic M_n does not concentrate around its expected value. Indeed, it is easy to see that its variance will explode as $n \rightarrow \infty$; more precisely, it will be of order $n^{2/(\gamma-1)}$.

2.2. Statistical consequences

Here, we use the previous bounds in order to estimate the endpoint of an unknown distribution, given that its decay near the endpoint is polynomial.

For any cdf F on \mathbb{R} , let θ_F be its endpoint, i.e., $\theta_F = \inf\{t \in \mathbb{R} : F(t) = 1\} \in \mathbb{R} \cup \{\infty\}$. For $r, L > 0$, let $\mathcal{F}(r, L)$ be the class of all cdf's F with $\theta_F < \infty$ and such that $L^{-1}t^\alpha \leq 1 - F(\theta_F - t) \leq Lt^\beta$, $\forall t \in [0, r]$, for some (unknown) $\alpha \geq \beta \geq 0$. For a fixed $\alpha \geq 0$, we also denote by $\mathcal{F}(r, L, \alpha)$ the class of all cdf's F with $\theta_F < \infty$ and such that $L^{-1}t^\alpha \leq 1 - F(\theta_F - t) \leq Lt^\alpha$, $\forall t \in [0, r]$.

Let ϕ be a density on \mathbb{R} and F be a cdf on \mathbb{R} . Let X be a random variable with cdf F and ε a random variable with density ϕ . Set $Y = X + \varepsilon$, and let Y_1, \dots, Y_n be i.i.d. copies of Y . Based on the observation of Y , let us estimate θ_F , given that $F \in \mathcal{F}(r, L)$ (resp. $F \in \mathcal{F}(r, L, \alpha)$), for some known $r, L > 0$ (resp. for some known $r, L > 0$ and $\alpha \geq 0$).

We assume that it is known that ϕ satisfies either Case (i) with $\gamma > 1$ or Case (iii) of Theorem 1, with known parameters A, γ, R . We define the estimator $\hat{\theta} = M_n - b_n$, where $M_n = \max(Y_1, \dots, Y_n)$ and

$$b_n = \begin{cases} (A^{-1} \log n)^{1/\gamma} & \text{in Case (i) and when } F \in \mathcal{F}(r, L); \\ (A^{-1} \log n)^{1/\gamma} \left(1 - \frac{(\gamma-1)(\alpha+1) \log \log n}{\gamma^2 \log n}\right) & \text{in Case (i) and when } F \in \mathcal{F}(r, L, \alpha); \\ Q & \text{in Case (iii)}. \end{cases}$$

In the next corollary, we denote by \mathbb{E}_F the expectation taken over the random variables Y_1, \dots, Y_n that are copies of $Y = X + \varepsilon$, when X has cdf F .

Corollary 1.

1. Using the first case of b_n described above,

$$\sup_{F \in \mathcal{F}(r, L)} \mathbb{E}_F[|\hat{\theta} - \theta_F|] \lesssim \frac{\log \log n}{(\log n)^{\frac{\gamma-1}{\gamma}}}.$$

2. Using the second case of b_n described above,

$$\sup_{F \in \mathcal{F}(r, L, \alpha)} \mathbb{E}_F[|\hat{\theta} - \theta_F|] \lesssim \frac{1}{(\log n)^{\frac{\gamma-1}{\gamma}}}.$$

3. Using the third case of b_n described above,

$$\sup_{F \in \mathcal{F}(r, L, \alpha)} \mathbb{E}_F[|\hat{\theta} - \theta_F|] \lesssim n^{-\frac{1}{\gamma+\alpha+1}}.$$

When ϕ is a centered Gaussian density, Theorem 2 in [10] suggests that the upper bound in Corollary 1 is optimal, up to a sublogarithmic factor. However, their result is only for a modified version of the model (where the hypothesis space consists of densities defined in terms of the convolution $Y = X + \varepsilon$ instead of X) and hence does not actually show a lower bound for a hypothesis space that is consistent with the one used for their upper bound.

As a conclusion, our results suggest that in the presence of unbounded errors, the endpoint θ_F of the distribution of the contaminated data can only be estimated at a polylogarithmic rate, in a minimax sense. In fact, in Section 3.3, we prove a lower bound in a multivariate setup for the Gaussian error case, whose rate is polylogarithmic in the sample size.

3. Application to convex support estimation from noisy data

In this section, we apply Theorems 2 and 4 to the problem of estimating a convex body from noisy observations of independent uniform random points. Let G be a convex body in \mathbb{R}^d and let X be uniformly distributed on G . Let $\varepsilon \in \mathbb{R}^d$ be a random vector independent of X and let $Y = X + \varepsilon$. From n i.i.d. copies Y_1, \dots, Y_n of Y , our goal is to recover G . Here, we only consider two different distributions for the noise ε . First, we consider the case when ε is a centered Gaussian random variable with covariance matrix $\sigma^2 I$, where I is the

$d \times d$ identity matrix and $\sigma^2 > 0$ is known. By Theorem 1, it is straightforward to extend the results to nearly Gaussian noise ε , i.e., when the densities of its one-dimensional projections have a nearly Gaussian right tail. Second, we consider the case when ε is uniformly (or nearly uniformly) distributed in some known Euclidean ball. Again, by Theorem 1, this can be very easily extended to an error distribution supported on a known Euclidean ball, with a density that decays at some known rate to zero near the boundary.

Our estimation scheme consists in reducing the d -dimensional estimation problem to a one-dimensional one, based on the following observation. Let $u \in \mathbb{S}^{d-1}$. Then, $\langle u, Y \rangle = \langle u, X \rangle + \langle u, \varepsilon \rangle$ and:

- $\langle u, \varepsilon \rangle$ is a one-dimensional noise distribution, which is easy to characterize;
- $h_G(u)$ is the endpoint of the distribution of $\langle u, X \rangle$.

In the sequel, we denote by F_u the cumulative distribution function of $\langle u, X \rangle$. Consider the following assumption, which entails the next lemma.

Assumption 1. $B_d(a, r) \subseteq G \subseteq B_d(0, R)$, for some $a \in \mathbb{R}^d$.

For brevity, the next lemma is proved in the supplementary material [3].

Lemma 1. *Let G satisfy Assumption 1. Then, for all $u \in \mathbb{S}^{d-1}$, $\theta_{F_u} = h_G(u)$ and $F_u \in \mathcal{F}(r, L)$, where $L = (2R)^{d-1} r^d \kappa_d \max\left(1, \frac{d}{r^{d-1} \kappa_{d-1}}\right)$.*

More precisely, the values of α and β in the definition of $\mathcal{F}(r, L)$ corresponding to F_u are $\alpha = d$ and $\beta = 1$.

As a consequence of Lemma 1, projecting the data Y_j , $1 \leq j \leq n$, on any direction brings us back to the one-dimensional setup studied in Section 2, where the end point of the corresponding distribution is the value of the support function of G in the direction of the projection.

We now state two assumptions on the noise distribution, which broadly correspond to the two cases that we consider in this section: (nearly) Gaussian and (nearly) uniform on a Euclidean ball.

Assumption 2. *There exist $\sigma^2 > 0$ and positive constants $c, C > 0$ such that the following holds. The noise term ε has a density ϕ on \mathbb{R}^d which satisfies, for all $x \in \mathbb{R}^d$,*

$$ce^{-\frac{\|x\|_2^2}{2\sigma^2}} \leq \phi(x) \leq Ce^{-\frac{\|x\|_2^2}{2\sigma^2}}.$$

Note that we never require c, C to be known, whereas we will always assume that σ^2 is known.

Assumption 3. *There exist $Q > 0, \gamma \geq 0$ and positive constants $c, C > 0$ such that the following holds. The noise term ε has a density ϕ on \mathbb{R}^d which satisfies, for all $x \in B_d(0, Q)$,*

$$c(Q - \|x\|_2)^\gamma \leq \phi(x) \leq C(Q - \|x\|_2)^\gamma$$

and $\phi(x) = 0$ for $x \notin B_d(0, Q)$.

Here as well, note that we never require c, C to be known, whereas we will always assume that Q and γ are known. The case $\gamma = 0$ corresponds to nearly uniform distributions on $B_d(0, Q)$.

We are now in a position to define an estimator of G . For $u \in \mathbb{R}^d$, let $\hat{h}(u)$ be the estimator of $h_G(u)$ defined as $\hat{h}(u) = \max_{1 \leq j \leq n} \langle u, Y_j \rangle - b_n$, where b_n is given in the next theorem.

3.1. Nearly Gaussian noise

3.1.1. A general estimator

Consider the random convex set $\hat{G} = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq \hat{h}(u), \forall u \in \mathbb{S}^{d-1}\}$. This estimator satisfies the following deviation inequality.

Theorem 5. *Let Assumption 2 hold. Set $n_0 = \max(e^{e^2/(2\sigma^2)}, e^{1/\sigma^4}, e^{8R^2/\sigma^2})$ and let $b_n = \sqrt{2\sigma^2 \log n}$. Then, for all convex bodies G that satisfy Assumption 1, for all integer $n \geq n_0$ and for all positive number x with $x \leq rb_n/2 - (2d/c_2 + c_0) \log b_n$,*

$$d_H(\hat{G}, G) \leq \frac{3R}{rb_n} (2x + (2d/c_2 + c_0 + 3) \log b_n)$$

with probability at least $1 - 2 \cdot 3^d (c_1 e^{-c_2 x} + \sqrt{2\sigma^2} e^{-c_3 n + d \log \log n})$, where c_0, c_1, c_2, c_3 are the positive constants given in Theorem 2.

Proof. Let $b_n = \sqrt{2\sigma^2 \log n}$ and let $z > 0$ and $\epsilon = (\log b_n)/b_n^2$. Let \mathcal{N} be an ϵ -net of \mathbb{S}^{d-1} and consider the event $\mathcal{A} = \{|\hat{h}(u) - h(u)| \leq z, \forall u \in \mathcal{N} \cup -\mathcal{N}\}$, where $-\mathcal{N} = \{-u : u \in \mathcal{N}\}$.

Assume that the event \mathcal{A} is satisfied. We use the following lemma (see [8]).

Lemma 2. *Let $\epsilon > 0$ and \mathcal{N} be an ϵ -net of \mathbb{S}^{d-1} . Then, for all $u \in \mathbb{S}^{d-1}$, there are sequences $(u_k)_{k \geq 0} \subseteq \mathbb{S}^{d-1}$ and $(\epsilon_k)_{k \geq 1} \subseteq [0, \infty)$ such that $u = u_0 + \sum_{k=1}^{\infty} \epsilon_k u_k$, with $0 \leq \epsilon_k \leq \epsilon^k$ for all $k \geq 1$.*

Let $u \in \mathbb{S}^{d-1}$, that we write as in Lemma 2, $u = u_0 + \sum_{k=1}^{\infty} \epsilon_k u_k$. Note that \hat{h} is not necessarily subadditive, however, the function $\tilde{h}(v) := \hat{h}(v) + b_n |v| = \max_{i=1, \dots, n} \langle v, Y_i \rangle$ is continuous and subadditive, since it is the support function of the convex hull of Y_1, \dots, Y_n . Therefore, since $h(v) \leq R$, for all $v \in \mathbb{S}^{d-1}$,

$$\begin{aligned}
\hat{h}(u) &= \tilde{h}(u) - b_n \leq \tilde{h}(u_0) + \sum_{k=1}^{\infty} \epsilon_k \tilde{h}(u_k) - b_n = \hat{h}(u_0) + \sum_{k=1}^{\infty} \epsilon_k (\hat{h}(u_k) + b_n) \\
&\leq h(u_0) + z + \sum_{k=1}^{\infty} \epsilon_k (h(u_k) + z + b_n) \leq h(u_0) + \frac{z}{1-\epsilon} + \frac{\epsilon b_n}{1-\epsilon} + \sum_{k=1}^{\infty} \epsilon_k h(u_k) \\
&\leq h(u) + \sum_{k=1}^{\infty} \epsilon_k (h(-u_k) + h(u_k)) + \frac{z}{1-\epsilon} + \frac{\epsilon b_n}{1-\epsilon} \\
&\leq h(u) + \frac{2R\epsilon}{1-\epsilon} + \frac{z}{1-\epsilon} + \frac{\epsilon b_n}{1-\epsilon},
\end{aligned} \tag{19}$$

where we used sub-additivity together with the fact that $u_0 = u + \sum_{k=1}^{\infty} \epsilon_k (-u_k)$.

On the other hand, by noting that since the event \hat{A} holds, it is true that $\hat{h}(-u_k) \leq h(-u_k) + z \leq R + z$, for all $k \geq 0$,

$$\begin{aligned}
h(u) &\leq h(u_0) + \sum_{k=1}^{\infty} \epsilon_k h(u_k) \\
&\leq \hat{h}(u_0) + z + \sum_{k=1}^{\infty} \epsilon_k R \leq \tilde{h}(u_0) - b_n + z + \frac{R\epsilon}{1-\epsilon} \\
&\leq \tilde{h}(u) + \sum_{k=1}^{\infty} \epsilon_k \tilde{h}(-u_k) - b_n + z + \frac{R\epsilon}{1-\epsilon} \\
&= \hat{h}(u) + \sum_{k=1}^{\infty} \epsilon_k (\hat{h}(-u_k) + b_n) + z + \frac{R\epsilon}{1-\epsilon} \\
&\leq \hat{h}(u) + \sum_{k=1}^{\infty} \epsilon_k (R + z + b_n) + z + \frac{R\epsilon}{1-\epsilon} \\
&\leq \hat{h}(u) + \frac{2R\epsilon}{1-\epsilon} + \frac{z}{1-\epsilon} + \frac{\epsilon b_n}{1-\epsilon}
\end{aligned} \tag{20}$$

It follows from (19) and (20) that, if \mathcal{A} holds, then, $\sup_{u \in \mathbb{S}^{d-1}} |\hat{h}(u) - h(u)| \leq \frac{2R\epsilon}{1-\epsilon} + \frac{z}{1-\epsilon} + \frac{\epsilon b_n}{1-\epsilon}$. If $n \geq n_0$, then $\epsilon \leq b_n^{-1} = (2\sigma^2 \sqrt{\log n})^{-1} \leq 1/2$ and $4R \leq b_n$, so $\sup_{u \in \mathbb{S}^{d-1}} |\hat{h}(u) - h(u)| \leq 2z + 3\epsilon b_n$.

Finally, we use [1, Lemma 7], which we state here in a simplified form.

Lemma 3. *Let $\delta \in (0, 1/2]$ and \mathcal{N} be a δ -net of \mathbb{S}^{d-1} . Let G be a convex body in \mathbb{R}^d and h_G its support function. Let $a \in \mathbb{R}^d$ and $0 < r \leq R$ such that $B_d(a, r) \subseteq G \subseteq B_d(a, R)$. Let $\hat{h} : \mathbb{S}^{d-1} \rightarrow \mathbb{R}$ and $\hat{G}_{\mathcal{N}} = \{x \in \mathbb{R}^d : \langle u, x \rangle \leq \hat{h}(u), \forall u \in \mathcal{N}\}$. Let $t = \max_{u \in \mathcal{N}} |\hat{h}(u) - h_G(u)|$. If $t \leq r/2$, then $d_H(\hat{G}_{\mathcal{N}}, G) \leq \frac{3tR}{2r} + 4R\delta$.*

Noting that Assumption 1 implies that $B(a, r) \subseteq G \subseteq B(a, 2R)$, by Lemma 3 with $\delta = 0$, the event \mathcal{A} implies that

$$d_H(\hat{G}, G) \leq \frac{6Rz}{r} + \frac{9R\epsilon b_n}{r},$$

so long as $2z + 3\epsilon b_n \leq r/2$.

Therefore,

$$\mathbb{P}_G \left[d_H(\hat{G}, G) \geq \frac{6Rz}{r} + \frac{9R\epsilon b_n}{r} \right] \leq 1 - \mathbb{P}_G[\mathcal{A}]. \quad (21)$$

Now, let us control the probability of the complement of \mathcal{A} .

Combining Lemma 1 and Theorem 2 yields, for all $u \in \mathbb{S}^{d-1}$, and all $t \geq 0$,

$$\mathbb{P}_G \left[|\hat{h}(u) - h_G(u)| > \frac{t + c_0 \log(b_n)}{b_n} \right] \leq c_1 e^{-\frac{b_n t}{2\sigma^2}} + e^{-c_2 n}, \quad (22)$$

where the constants c_1, c_2 and c_3 are given in Theorem 2 with $\alpha = d$, $\beta = 1$, $\gamma = 2$ and $A = (2\sigma^2)^{-1}$.

By a volumetric argument, $\#\mathcal{N} \leq (3/\epsilon)^d$. Hence, for z of the form $z = \frac{t + c_0 \log b_n}{b_n}$, a union bound yields that the event \mathcal{A} holds with probability at least $1 - 3^d \delta$, where

$$\delta = 2c_1 e^{-c_2 t - d \log \epsilon} + 2c_2 e^{-c_3 n - d \log \epsilon}.$$

Finally, taking $t = x + \frac{2d}{c_2} \log b_n$, for $x \geq 0$, yields the desired result, by noting that $\log \log b_n \geq 0$, since $n \geq e^{\epsilon^2/(2\sigma^2)}$. □

Corollary 2. *Let the assumptions of Theorem 5 hold. Then, the estimator \hat{G} satisfies*

$$\sup_{G \in \mathcal{K}_{r,R}} \mathbb{E}_G [d_H(\hat{G}, G)] = O \left(\frac{\log \log n}{\sqrt{\log n}} \right).$$

Proof. The proof is similar to Corollary 3. □

3.1.2. A computable estimator

The estimator \hat{G} that we defined in (2) is given by a polyhedral representation with infinitely many constraints and, hence, it is not clear how to compute it in practice. Here, we propose a modified version which is based on a discretization of the set of polyhedral constraints.

Let M be a positive integer and U_1, \dots, U_M be independent uniform random vectors on the sphere \mathbb{S}^{d-1} and define

$$\hat{G}_M = \{x \in \mathbb{R}^d : \langle U_j, x \rangle \leq \hat{h}(U_j), \forall j = 1, \dots, M\}. \quad (23)$$

We also define a truncated version of \hat{G}_M . Let $\hat{\mu}_n = \frac{1}{n} \sum_{j=1}^n Y_j$. Define

$$\tilde{G}_M = \begin{cases} \hat{G}_M \cap B_d(\hat{\mu}_n, \log n) & \text{if } \hat{G}_M \neq \emptyset \\ \{\hat{\mu}_n\} & \text{otherwise.} \end{cases} \quad (24)$$

First, we give a deviation inequality for the estimator \hat{G}_M . As a corollary, we obtain an upper bound on the risk of the truncated estimator \tilde{G}_M for some prescribed value of M .

Theorem 6. *Let Assumption 2 hold. Let $n > 3$, $b_n = \sqrt{2\sigma^2 \log n}$ and M be a positive integer with $(\log M)/b_n \leq \min(r/(4\sigma^2), 1/2)$. Then, there exist positive constants c_0, c_1, c_2 and c_3 such that the following holds. For all convex bodies G that satisfy Assumption 1, for all positive x with $x \leq \frac{rb_n}{4\sigma^2} - \log M$,*

$$d_H(\hat{G}_M, G) \leq c_0 \frac{x + \log M}{b_n}$$

with probability at least $1 - c_1 M e^{-x} - M e^{-c_2 n} - (6b_n)^d e^{-c_3 M (\log M)^{d-1} b_n^{-(d-1)}}$.

Proof. Let G satisfy Assumption 1. Combining Lemma 1 and Theorem 2, we have that for all $u \in \mathbb{S}^{d-1}$, and all $t \geq 0$,

$$\mathbb{P}_G \left[|\hat{h}(u) - h_G(u)| > t \right] \leq c_1 e^{-\frac{b_n t}{2\sigma^2}} + e^{-c_2 n}, \quad (25)$$

with c_1 and c_2 as in Theorem 2 with $\alpha = d$, $\beta = 1$, $\gamma = 2$ and $A = (2\sigma^2)^{-1}$. Hence, by a union bound,

$$\mathbb{P}_G \left[\max_{1 \leq j \leq M} |\hat{h}(U_j) - h_G(U_j)| > t \right] \leq c_1 M e^{-\frac{b_n t}{2\sigma^2}} + M e^{-c_2 n}. \quad (26)$$

Let $t < r/2$. Consider the event \mathcal{A} where U_1, \dots, U_M form a δ -net of \mathbb{S}^{d-1} , where $\delta \in (0, 1/2)$. By Lemma 3, if \mathcal{A} holds and if $|\hat{h}(U_j) - h_G(U_j)| \leq t$ for all $j = 1, \dots, M$, then $d_H(\hat{G}_M, G) \leq \frac{3tR}{r} + 4R\delta$. Hence, by (26) and Lemma 10 in [1],

$$\begin{aligned} \mathbb{P}_G \left[d_H(\hat{G}_M, G) > \frac{3tR}{r} + 4R\delta \right] \\ \leq c_1 M e^{-\frac{b_n t}{2\sigma^2}} + M e^{-c_2 n} + 6^d \exp \left(-c_3 M \delta^{d-1} + d \log \left(\frac{1}{\delta} \right) \right), \end{aligned} \quad (27)$$

where $c_3 = (2d8^{(d-1)/2})^{-1}$. Taking $\delta = (\log M)/b_n$ ends the proof of Theorem 6. \square

Theorem 6 yields a uniform upper bound on the risk of \tilde{G}_M , which we derive for a special choice of M and state next as a corollary. For brevity, we defer the proof to the supplementary material [3].

Corollary 3. *Denote by $\mathcal{K}_{r,R}$ the collection of all convex bodies satisfying Assumption 1. Let the assumptions of Theorem 6 hold and set $M = \lfloor 2d(d+1)8^{(d-1)/2}b_n^{d-1}(\log b_n)^{-(d-2)} \rfloor$. Then, the truncated estimator \tilde{G}_M satisfies*

$$\sup_{G \in \mathcal{K}_{r,R}} \mathbb{E}_G[d_H(\tilde{G}_M, G)] = O\left(\frac{\log \log n}{\sqrt{\log n}}\right).$$

Remark 2. *Suppose that for all $x \in \partial G$, there exist $a, b \in \mathbb{R}^d$ such that $B_d(a, r) \subseteq G \subseteq B_d(b, R)$, $x \in B_d(a, r)$ and $x \in \partial B_d(b, R)$. In particular, this means that the complement of G has reach at least r , i.e., one can roll a Euclidean ball of radius r inside G along its boundary (see, e.g., [19, Definition 11]). In addition, G can roll freely inside a Euclidean ball of radius R , along its boundary. This ensures that for all $u \in \mathbb{S}^{d-1}$, the random variable $\langle u, X \rangle - h_G(u)$ satisfies the assumption of Theorem 3 with $\alpha = \beta = (d+1)/2$ and some $L > 0$ that depends on r and R only. Hence, we are in the case where $\alpha = \beta$ in Theorem 3, which shows that the rate of estimation of the support function of G at a single unit vector can be improved by a sublogarithmic factor. However, a close look at the proof of Theorem 6 suggests that a sublogarithmic factor is still unavoidable with our current proof technique, because of the union bound on a covering of the unit sphere.*

Remark 3. *Theorem 6 can be easily extended to cases where the X_j 's are not uniformly distributed on G . What matters for the proof is that, uniformly over unit vectors u , the cumulative distribution function F_u of $\langle u, X \rangle - h_G(u)$ increases polynomially near zero. Examples of such distributions are given in [4].*

Remark 4. *Note that in general, the estimate \hat{h} defined above is not a support function. In particular, it is not enough to control the differences $\hat{h}(U_j) - h_G(U_j)$, $j = 1, \dots, M$ in order to obtain a bound on the Hausdorff distance between \hat{G}_M and G .*

3.2. Nearly uniform noise

Recall that a crucial aspect of our analysis is that $\langle u, \varepsilon \rangle$ (i.e., the projection of the noise along a direction) has a density with known decay (e.g., Gaussian densities) that we can analyze. For example, a desirable property of Gaussian errors ε is that $\langle u, \varepsilon \rangle$ is also distributed Gaussian and is independent of the direction u . Other simplifications hold for other spherically symmetric or stable distributions, but a separate analysis is required to extract the correct asymptotic bias b_n . Hence, the general techniques that we have developed so far can also be extended to other noise distributions, provided they are amenable to analysis. Another case of interest is when the noise terms are bounded and

nearly uniform on a ball. More specifically, let G satisfy Assumption 1 and suppose that ϵ satisfies Assumption 3, where $Q > r$ is known. We can take $b_n = Q$ and set

$$\hat{h}_n(u) = \max_{1 \leq j \leq n} \langle u, Y_j \rangle - Q$$

to form the estimator \hat{G} . The next set of results provide statistical guarantees for this estimator.

Theorem 7. *Let Assumption 3 hold. Let $n \geq 1$ and $b_n = Q$. Then, there exists a positive constant C_1 such that the following holds. For all convex bodies G that satisfy Assumption 1, for all positive x , if $x \leq r/8$,*

$$d_H(\hat{G}, G) \leq \frac{12Rx}{r}$$

with probability at least $1 - 2 \cdot 3^d (2 + Q/R)^d e^{-C_1 n x^{\gamma+\alpha+1} - d \log x}$.

Proof. The proof follows the same lines as Theorem 3. Let \mathcal{A} and \mathcal{N} be the sets defined in the proof of Theorem 5. From (19) and (20), if \mathcal{A} holds, then, $\sup_{u \in \mathbb{S}^{d-1}} |\hat{h}(u) - h_G(u)| \leq \frac{2R\epsilon}{1-\epsilon} + \frac{z}{1-\epsilon} + \frac{\epsilon Q}{1-\epsilon}$. Furthermore, if $\epsilon \leq 1/2$, then $\sup_{u \in \mathbb{S}^{d-1}} |\hat{h}(u) - h_G(u)| \leq 2z + 2(2R + Q)\epsilon$. Next, by Lemma 3 with $\delta = 0$, after noting that Assumption 1 implies that $B(a, r) \subseteq G \subseteq B(a, 2R)$, the event \mathcal{A} implies that

$$d_H(\hat{G}, G) \leq \frac{6Rz}{r} + \frac{6(2R + Q)\epsilon}{r},$$

so long as $2z + 2(2R + Q)\epsilon \leq r/2$. Combining Lemma 1 and Theorem 4 yields, for all $u \in \mathbb{S}^{d-1}$, and all $t \geq 0$,

$$\mathbb{P}_G \left[|\hat{h}(u) - h_G(u)| > t n^{-\frac{1}{\gamma+\alpha+1}} \right] \leq e^{-C_1 t^{\gamma+\alpha+1}},$$

where the constant C_1 is given in Theorem 4 with $\alpha = d$ and $\beta = 1$.

By a volumetric argument, $\#\mathcal{N} \leq (3/\epsilon)^d$. Hence, for z of the form $z = t n^{-\frac{1}{\gamma+\alpha+1}}$, a union bound yields that the event \mathcal{A} holds with probability at least $1 - 3^d \delta$, where

$$\delta = 2e^{-C_1 n z^{\gamma+\alpha+1} - d \log \epsilon}.$$

Finally, taking $\epsilon = R(2R + Q)^{-1} z$ yields the desired result. \square

Corollary 4. *Suppose the assumptions of Theorem 7 hold. Then, the estimator \hat{G} satisfies*

$$\sup_{G \in \mathcal{K}_{r,R}} \mathbb{E}_G [d_H(\hat{G}, G)] = O((n/\log n)^{-\frac{1}{\gamma+\alpha+1}}).$$

Proof. The proof follows from a simple truncation argument:

$$\begin{aligned} \mathbb{E}_G[\mathbf{d}_H(\hat{G}, G)] &= \mathbb{E}_G[\mathbf{d}_H(\hat{G}, G)\mathbb{1}_{\mathbf{d}_H(\hat{G}, G) \leq x}] + \mathbb{E}_G[\mathbf{d}_H(\hat{G}, G)\mathbb{1}_{\mathbf{d}_H(\hat{G}, G) > x}] \\ &\leq x + 2R\mathbb{P}_G[\mathbf{d}_H(\hat{G}, G) > x]. \end{aligned}$$

Finally, apply Theorem 7 with $x \asymp \left(\frac{\log n}{n}\right)^{\frac{1}{\gamma+\alpha+1}}$. \square

Remark 5. Note that when ε is exactly uniform on a ball of radius Q , then $\gamma = (d-1)/2$. If the convex set G also satisfies the rolling ball assumption (see Remark 2), then $\alpha = (d+1)/2$ and hence the minimax upper bound above becomes $O((n/\log n)^{-1/(d+1)})$. Compare this with the minimax rate of $\Theta((n/\log(n))^{-2/(d+1)})$ [4]. Thus, this worse rate is the price to pay for having contaminated observations.

Remark 6. Similar convergence guarantees also hold for the computable estimator \hat{G}_M (23), though due to space constraints, we do not include them here.

3.3. Minimax lower bound for Gaussian errors

The next theorem gives a lower bound for the minimax risk of estimation $G \in \mathcal{K}_{r,R}$ in the case that the error distribution is Gaussian. As with Corollary 2, it is also polylogarithmic in the sample size.

Theorem 8. Suppose the error distribution is a centered Gaussian with covariance matrix $\sigma^2 I$, where I is the $d \times d$ identity matrix with $d > 1$. Let r and R be any two positive real numbers satisfying $R/r = 2\sqrt{d}$. For each τ in $(0, 1)$, there exist positive constants c and C depending only on d, σ, τ, r , and R such that

$$\inf_{\hat{G}} \sup_{G \in \mathcal{K}_{r,R}} \mathbb{P}_G[\mathbf{d}_H(G, \hat{G}) > c(\log n)^{-2/\tau}] \geq C,$$

and

$$\inf_{\hat{G}} \sup_{G \in \mathcal{K}_{r,R}} \mathbb{E}_G[\mathbf{d}_H(G, \hat{G})] \geq C(\log n)^{-2/\tau},$$

where the infimum runs over all estimators \hat{G} of G based on Y_1, \dots, Y_n .

Remark 7. The assumption $d > 1$ in Theorem 8 is important. For example, if $d = 1$ and the points X_1, \dots, X_n are uniformly distributed on a bounded interval, then a method of moments estimator yields the optimal (polynomial) $\Theta(1/\sqrt{n})$ rate of convergence for the endpoints.

Proof. In the following, we assume that c and C are generic positive constants, depending only on d, σ, τ , and δ .

Let $\delta > 0$ be fixed and m be a positive integer. Let ψ be chosen as in Lemma 7 in the supplementary material [3] and $\gamma_m = (4/3)\delta^{-1}\pi m$. Replacing ψ by $x \mapsto (\psi(-x/\delta) + \psi(x/\delta))/2$, we can assume without loss of generality that ψ is symmetric about the origin, supported in the interval $[-\delta, \delta]$, and $\inf_{|x| \leq \delta(3/4)} \psi(x) > 0$. Note that this assumption does not alter the decay rate of its Fourier transform.

Define $h_m(x) = \psi(x) \sin(\gamma_m x)$, $H_m(x_1, \dots, x_{d-1}) = \prod_{k=1}^{d-1} h_m(x_k)$, and for $L > 0$ and $\omega \in \{-1, +1\}$, let

$$b_\omega(x_1, \dots, x_{d-1}) = \sum_{k=1}^{d-1} g(x_k) + \omega (L/\gamma_m^2) H_m(x_1, \dots, x_{d-1}),$$

where g satisfies:

$$\max_{x \in [-\delta, \delta]} g''(x) < 0, \quad \text{and} \quad (28)$$

$$|\mathcal{F}[g](t)| \leq C e^{-c|t|^\tau}, \quad \text{for some positive constants } c \text{ and } C \quad (29)$$

For concreteness, one can take an appropriately scaled Cauchy density, $g(x) \propto \frac{1}{1+x^2/\delta_0^2}$, $x \in \mathbb{R}$, which is strictly concave in the region where $|x| < \delta_0/\sqrt{3}$ (thereby satisfying (28)) and satisfies (29) with $\tau = 1$. In fact, from the inequality $1 + |t| \geq |t|^\tau$, we have that (29) is satisfied for all $\tau \in (0, 1)$.

By (28) and Lemma 8, if L is chosen small enough, we ensure that the Hessian of b_ω , i.e., $\nabla^2 b_\omega$, is negative-semidefinite and hence the sets

$$G_\omega = \{(x_1, \dots, x_d) \in [-\delta, \delta]^{d-1} \times \mathbb{R} : -\delta \leq x_d \leq b_\omega(x_1, \dots, x_{d-1})\}$$

are convex. By scaling g and choosing L small enough, we can ensure that $0 \leq b_\omega \leq \delta$. This means that $G_\omega \subset [-\delta, \delta]^d$, and since $[-\delta, \delta]^d \subset B_d(0, \sqrt{d}\delta)$, we may take $R = \sqrt{d}\delta$. Finally, observe that $B_d(-\delta/2, \delta/2) \subset G_\omega$, since the cube $[-\delta, 0]^d$ is (trivially) contained in G_ω and $B_d(0, \delta) \subset [-\delta, \delta]^d$. Thus, we may take $r = \delta/2$. With these choices of r and R , we have $G_\omega \in \mathcal{K}_{r,R}$.

Note that h_m is an odd function about the origin. Thus $\int_{[-\delta, \delta]^{d-1}} H_m(x) dx = 0$ because we are integrating an odd function about the origin. Therefore, $|G_\omega| = \delta(2\delta)^{d-1} + (d-1) \int_{[-\delta, \delta]} g(x) dx$. Also, note that

$$\begin{aligned} \mathbf{d}_\Delta(G_{+1}, G_{-1}) &= \int_{[-\delta, \delta]^{d-1}} |b_{+1}(x) - b_{-1}(x)| dx \\ &= \frac{2L}{\gamma_m^2} \int_{[-\delta, \delta]^{d-1}} |H_m(x)| dx \\ &= \frac{2L}{\gamma_m^2} \prod_{k=1}^{d-1} \int_{[-\delta, \delta]} |\sin(\gamma_m x_k) \psi(x_k)| dx_k. \end{aligned}$$

The factor $\prod_{k=1}^{d-1} \int_{[-\delta, \delta]} |\sin(\gamma_m x_k) \psi(x_k)| dx_k$ in the above expression can be lower bounded by a constant, independent of m . In fact,

$$\begin{aligned} \int_{[-\delta, \delta]} |\sin(\gamma_m x_k) \psi(x_k)| dx_k &\geq \int_{|x_k| \leq \delta(3/4)} |\sin(\gamma_m x_k) \psi(x_k)| dx_k \\ &\geq 3\delta/4 \inf_{|x| \leq \delta(3/4)} |\psi(x)| \int_{|x_k| \leq 1} |\sin(\pi m x_k)| dx_k \\ &= 3\delta/\pi \inf_{|x| \leq \delta(3/4)} |\psi(x)| \\ &> 0. \end{aligned}$$

Here, we used the fact that

$$\begin{aligned} \int_{[-1, 1]} |\sin(\pi m x)| dx &= 4m \int_{[0, 1/(2m)]} |\sin(\pi m x)| dx \\ &= (4/\pi) \int_{[0, \pi/2]} \sin(x) dx \\ &= 4/\pi, \end{aligned}$$

for any non-zero integer m . Thus, there exists a constant $C_1 > 0$, independent of m , such that

$$d_{\Delta}(G_{+1}, G_{-1}) \geq \frac{C_1}{m^2}. \quad (30)$$

For $\omega = \pm 1$, define $f_{\omega} = \mathbb{1}_{G_{\omega}}/|G_{\omega}|$. Let $\phi_{\sigma}(x) = (1/(2\pi\sigma^2))^{d/2} e^{-\|x\|^2/(2\sigma^2)}$ for $x \in \mathbb{R}^d$. Note that for all $y > 0$,

$$\begin{aligned} \text{TV}(\mathbb{P}_{G_{+1}}, \mathbb{P}_{G_{-1}}) &= \frac{1}{2} \int_{\mathbb{R}^d} |(f_{+1} - f_{-1}) * \phi_{\sigma}(x)| dx \\ &= \frac{1}{2} \int_{\|x\| > y} |(f_{+1} - f_{-1}) * \phi_{\sigma}(x)| dx + \frac{1}{2} \int_{\|x\| \leq y} |(f_{+1} - f_{-1}) * \phi_{\sigma}(x)| dx \\ &\leq \int_{\|x\| > y} \sup_{z \in [-\delta, \delta]^d} \phi_{\sigma}(x - z) dx + \\ &\quad \frac{1}{2} \sqrt{|B_d(0, y)|} \sqrt{\int_{\mathbb{R}^d} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_{\sigma}](t)|^2 dt} \\ &\leq C_2 e^{-c_2 y^2} + C_2 y^{d/2} \sqrt{\int_{\mathbb{R}^d} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_{\sigma}](t)|^2 dt}, \end{aligned}$$

for some positive constants c_2 and C_2 that depend only on δ , σ , and d . Set $y \propto \sqrt{\log(\frac{1}{\int_{\mathbb{R}^d} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_{\sigma}](t)|^2 dt})}$ so that $\text{TV}(\mathbb{P}_{G_{+1}}, \mathbb{P}_{G_{-1}})$ can be bounded by a fixed power of $\int_{\mathbb{R}^d} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_{\sigma}](t)|^2 dt$.

Split $\int_{\mathbb{R}^d} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_{\sigma}](t)|^2 dt$ into two integrals with domains of integration $\|t\|_{\infty} \leq am^{\tau}$ and $\|t\|_{\infty} > am^{\tau}$, where a is as in Lemma 5. Using the fact that $\mathcal{F}[\phi_{\sigma}](t) = \sigma^d e^{-\sigma^2 \|t\|_2^2/2}$, we have

$$\int_{\|t\|_{\infty} > am^{\tau}} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_{\sigma}](t)|^2 dt \leq C_3 e^{-c_3 m^{2\tau}}. \quad (31)$$

By Lemma 5, we have

$$|\mathcal{F}[f_{+1} - f_{-1}](t)| = \frac{1}{|G_{+1}|} |\mathcal{F}[\mathbb{1}_{G_{+1}} - \mathbb{1}_{G_{-1}}](t)| \leq C e^{-cm^\tau},$$

whenever $\|t\|_\infty \leq am^\tau$. Thus,

$$\int_{\|t\|_\infty \leq am^\tau} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_\sigma](t)|^2 dt \leq C e^{-cm^\tau} \int_{\mathbb{R}^d} |\mathcal{F}[\phi_\sigma](t)|^2 dt. \quad (32)$$

Recall that $\text{TV}(\mathbb{P}_{G_{+1}}, \mathbb{P}_{G_{-1}})$ is bounded by a fixed power of $\int_{\mathbb{R}^d} |\mathcal{F}[f_{+1} - f_{-1}](t) \mathcal{F}[\phi_\sigma](t)|^2 dt$, which in turn is bounded by the sum of (31) and (32). This shows that

$$\text{TV}(\mathbb{P}_{G_{+1}}, \mathbb{P}_{G_{-1}}) \leq C_4 e^{-c_4 m^\tau},$$

for some positive constants c_4 and C_4 that depend only on d, σ, τ , and δ .

In summary, we have shown that $\mathbf{d}_\Delta(G_{+1}, G_{-1}) \geq \frac{C_1}{m^2}$ and $\text{TV}(\mathbb{P}_{G_{+1}}, \mathbb{P}_{G_{-1}}) \leq C_4 e^{-c_4 m^\tau}$, where the constants depend only on d, σ, τ, δ .

The minimax probability lower bound is constructed from a simple two point statistical hypothesis test. That is, choose $m \asymp (\log n)^{1/\tau}$ and apply Theorem 2.2(i) in [20] to lower bound the minimax probability:

$$\inf_{\hat{G}} \sup_{G \in \mathcal{K}_{\tau, R}} \mathbb{P}_G[\mathbf{d}_H(G, \hat{G}) > c_5 (\log n)^{-2/\tau}] \geq \inf_{\hat{G}} \sup_{G \in \mathcal{K}_{\tau, R}} \mathbb{P}_G[\mathbf{d}_\Delta(G, \hat{G}) > c_6 (\log n)^{-2/\tau}], \quad (33)$$

for some positive constants c_5 and c_6 that depend only on d, σ, τ, δ . Note that in establishing (33), we used Lemma 4 to upper bound \mathbf{d}_Δ by \mathbf{d}_H . The second conclusion of the theorem involving the lower bound on the minimax risk is a direct consequence of Markov's inequality. \square

3.4. Gap between the lower and upper bounds

Note that our upper (Corollary 2) and lower (Theorem 8) bounds do not match. However, as with all Gaussian or ill-posed deconvolution problems [7], the rate is very slow (logarithmic), and the difference between $O(\log \log n / \sqrt{\log n})$ and $\Omega(1/\log^2 n)$ is immaterial. In fact, Gaussian deconvolution in the context of manifold estimation was considered previously in [9, Section 5], where they also obtain *non-matching* upper and lower bounds on the minimax Hausdorff distance of $O(1/\sqrt{\log n})$ and $\Omega(1/\log n)$, respectively. Importantly, these authors did not have to deal with the additional shape constraints imposed by convex bodies and difficulties that arise with Fourier transforms of compactly supported functions (we address this further in the next paragraph). As with those authors (see the remarks at the end of [9, Section 5]), we do not know how to close the gap in the rates at the moment and we leave such work for future consideration.

The proof of our lower bound is different than other lower bounds in deconvolution problems. In standard density deconvolution [7], the hypothesis space is rich enough to ensure the existence of a function whose Fourier transform vanishes on a compact interval. These functions are then used to construct densities that are well-separated yet induce statistically indistinguishable models. The uncertainty principle for Fourier transforms [18] make such constructions impossible in our setting. That, is our function class consists of compactly supported densities of the form $\mathbb{1}_G/|G|$ for compact, convex sets G . The Fourier transform of an infinitely differentiable, compactly supported function decays faster than any polynomial. Indeed, such a function was used in constructing our two-point test. For the proof of the lower bound, to ensure that the total variation between the two induced distributions was small, we needed to be able to match the exponential decay of the Fourier transform of the Gaussian density. However, there is a limit on how fast the Fourier transform of a compactly supported function can decay. For example, a result in [13] shows that a decay of $O(e^{-|t|\epsilon(t)})$ is possible if and only if $\int_1^\infty \frac{\epsilon(t)}{t} dt$ is finite. Unfortunately, $O(e^{-c|t|^2})$ decay is needed to match the decay of the Fourier transform of a Gaussian density.

Supplementary Material

Supplement A: Additional Proofs

(doi: [COMPLETED BY THE TYPESETTER](#); .pdf). We provide the proofs of the corollaries and intermediate lemmas omitted from the main paper.

References

- [1] V.-E. Brunel. Concentration of the empirical level sets of tukeys halfspace depth. *Probability Theory and Related Fields*, 173(3-4):1165–1196, 2019.
- [2] V.-E. Brunel. Uniform deviation and moment inequalities for random polytopes with general densities in arbitrary convex bodies. *To appear, Bernoulli*, 2019.
- [3] V.-E. Brunel, J. M. Klusowski, and D. Yang. Supplement to “Estimation of convex supports from noisy measurements”. 2020.
- [4] V.-E. Brunel et al. Uniform behaviors of random polytopes under the Hausdorff metric. *Bernoulli*, 25(3):1770–1793, 2019.
- [5] B. Chazelle. An optimal convex hull algorithm in any fixed dimension. *Discrete Computational Geometry*, 10:377–409, 1993.
- [6] S. Csörgő and D. M. Mason. Simple estimators of the endpoint of a distribution. In *Extreme value theory (Oberwolfach, 1987)*, volume 51 of *Lecture Notes in Statist.*, pages 132–147. Springer, New York, 1989. . URL http://dx.doi.org/10.1007/978-1-4612-3634-4_12.

- [7] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272, 1991. ISSN 0090-5364. . URL <http://dx.doi.org/10.1214/aos/1176348248>.
- [8] D. Fresen et al. A multivariate Gnedenko law of large numbers. *The Annals of Probability*, 41(5):3051–3080, 2013.
- [9] C. R. Genovese, M. Perone-Pacifco, I. Verdinelli, and L. Wasserman. Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.*, 40(2): 941–963, 2012. ISSN 0090-5364. . URL <http://dx.doi.org/10.1214/12-AOS994>.
- [10] A. Goldenshluger and A. Tsybakov. Estimating the endpoint of a distribution in the presence of additive observation errors. *Statist. Probab. Lett.*, 68(1):39–49, 2004. ISSN 0167-7152. . URL <http://dx.doi.org/10.1016/j.spl.2004.01.013>.
- [11] P. Hall. On estimating the endpoint of a distribution. *Ann. Statist.*, 10(2): 556–568, 1982. ISSN 0090-5364. URL [http://links.jstor.org/sici?sici=0090-5364\(198206\)10:2<556:0ETEOA>2.0.CO;2-V&origin=MSN](http://links.jstor.org/sici?sici=0090-5364(198206)10:2<556:0ETEOA>2.0.CO;2-V&origin=MSN).
- [12] P. Hall and L. Simar. Estimating a changepoint, boundary, or frontier in the presence of observation error. *J. Amer. Statist. Assoc.*, 97(458):523–534, 2002. ISSN 0162-1459. . URL <http://dx.doi.org/10.1198/016214502760047050>.
- [13] A. E. Ingham. A Note on Fourier Transforms. *J. London Math. Soc.*, S1-9(1):29, 1934. . URL <http://dx.doi.org/10.1112/jlms/s1-9.1.29>.
- [14] A. P. Korostel'ev and A. B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993. ISBN 0-387-94028-6. URL <https://doi.org/10.1007/978-1-4612-2712-0>.
- [15] S. Loustau and C. Marteau. Minimax fast rates for discriminant analysis with errors in variables. *Bernoulli*, 21(1):176–208, 2015. ISSN 1350-7265. . URL <http://dx.doi.org/10.3150/13-BEJ564>.
- [16] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten. *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 2:75–84, 1963.
- [17] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten II. *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 3:138–147, 1964.
- [18] E. M. Stein and R. Shakarchi. *Fourier analysis*, volume 1 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2003. ISBN 0-691-11384-X. An introduction.
- [19] C. Thäle. 50 years sets with positive reach—a survey. *Surv. Math. Appl.*, 3:123–165, 2008.
- [20] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. . URL <http://dx.doi.org/10.1007/b13794>. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.