ADAPTIVE ESTIMATION OF PLANAR CONVEX SETS

By T. Tony Cai[*,1], Adityanand Guntuboyina[†,2] and Yuting Wei[†]

University of Pennsylvania * *and University of California, Berkeley*[†]

In this paper, we consider adaptive estimation of an unknown planar compact, convex set from noisy measurements of its support function. Both the problem of estimating the support function at a point and that of estimating the whole convex set are studied. For pointwise estimation, we consider the problem in a general nonasymptotic framework, which evaluates the performance of a procedure at each individual set, instead of the worst case performance over a large parameter space as in conventional minimax theory. A data-driven adaptive estimator is proposed and is shown to be optimally adaptive to every compact, convex set. For estimating the whole convex set, we propose estimators that are shown to adaptively achieve the optimal rate of convergence. In both of these problems, our analysis makes no smoothness assumptions on the boundary of the unknown convex set.

1. Introduction. We study in this paper the problem of nonparametric estimation of an unknown planar compact, convex set from noisy measurements of its support function on a uniform grid. Before describing the details of the problem, let us first introduce the support function. For a compact, convex set K in $\mathbb{R}^2$, its support function is defined by

$$h_K(\theta) := \max_{(x_1, x_2) \in K} (x_1 \cos\theta + x_2 \sin\theta) \qquad \text{for } \theta \in \mathbb{R}.$$

Note that h_K is a periodic function with period 2π. It is useful to think about θ in terms of the direction $(\cos\theta, \sin\theta)$. The line $x_1 \cos\theta + x_2 \sin\theta = h_K(\theta)$ is a support line for K (i.e., it touches K and K lies on one side of it). Conversely, every support line of K is of this form for some θ. The convex set K is completely determined by the its support function h_K because $K = \bigcap_\theta \{(x_1, x_2) : x_1 \cos\theta + x_2 \sin\theta \leq h_K(\theta)\}$.

The support function h_K possesses the *circle-convexity* property (see, e.g., [33]): for every $\alpha_1 > \alpha > \alpha_2$ and $0 < \alpha_1 - \alpha_2 < \pi$,

$$(1) \qquad \frac{h_K(\alpha_1)}{\sin(\alpha_1 - \alpha)} + \frac{h_K(\alpha_2)}{\sin(\alpha - \alpha_2)} \geq \frac{\sin(\alpha_1 - \alpha_2)}{\sin(\alpha_1 - \alpha)\sin(\alpha - \alpha_2)} h_K(\alpha).$$

Moreover, the above inequality characterizes h_K, that is, any periodic function of period 2π satisfying the above inequality equals h_K for a unique compact, convex subset K in $\mathbb{R}^2$. The circle-convexity property (1) is clearly related to the usual convexity property. Indeed, replacing $\sin\alpha$ by α in (1) leads to the condition for convexity. In spite of this similarity, (1) is different from convexity as can be seen from the example of the function $h(\theta) = |\sin\theta|$, which satisfies (1) but is clearly not convex.

1.1. *The problem, motivations and background.* We are now ready to describe the problem studied in this paper. Let K^* be an unknown compact, convex set in $\mathbb{R}^2$. We study the problem of estimating K^* or h_{K^*} from noisy measurements of h_{K^*}. Specifically, we observe data $(\theta_1, Y_1), \ldots, (\theta_n, Y_n)$ drawn according to the model

$$(2) \qquad Y_i = h_{K^*}(\theta_i) + \xi_i \qquad \text{for } i = 1, \ldots, n,$$

where $\theta_1, \ldots, \theta_n$ are fixed grid points in $(-\pi, \pi]$ and $\xi_1, \ldots, \xi_n$ are i.i.d. Gaussian random variables with mean zero and known variance σ^2. We focus on the dual problems of estimating the scalar quantity $h_{K^*}(\theta_i)$ for each $1 \leq i \leq n$ as well as the convex set K^*. We propose data-driven adaptive estimators and establish their optimality for both of these problems.

The problem considered here has a range of applications in engineering. The regression model (2) was first proposed and studied by [29] who were motivated by an application to Computed Tomography. Lele, Kulkarni and Willsky [25] showed how solutions to this problem can be applied to target reconstruction from resolved laser-radar measurements in the presence of registration errors. Gregor and Rannou [15] considered application to Projection Magnetic Resonance Imaging. It is also a fundamental problem in geometric tomography; see [12]. Another application domain where this problem might plausibly arise is robotic tactical sensing as has been suggested by [29]. Finally, this is a natural shape constrained estimation problem and would fit right into the recent literature on shape constrained estimation. See, for example, [18].

Most proposed procedures for estimating K^* in this setting are based on least squares minimization. The least squares estimator $\hat{K}_{\text{ls}}$ is defined as any minimizer of $\sum_{i=1}^n (Y_i - h_K(\theta_i))^2$ as K ranges over all compact convex sets. The minimizer in this optimization problem is not unique and one can always take it to be a polytope. This estimator was first proposed by [29] who also proposed an algorithm for computing it based on quadratic programming. Further algorithms for computing $\hat{K}_{\text{ls}}$ were proposed in [13, 25, 29].

The theoretical performance of the least squares estimator was first considered by [14] who mainly studied its accuracy for estimating K^* under the natural fixed design loss:

$$(3) \qquad L_f(K^*, \hat{K}) := \frac{1}{n} \sum_{i=1}^n \big(h_{K^*}(\theta_i) - h_{\hat{K}}(\theta_i)\big)^2.$$

The key result of [14] (specialized to the planar case that we are studying) states that $L_f(K^*, \hat{K}_{ls}) = O(n^{-4/5})$ as $n \to \infty$ almost surely provided K^* is contained in a ball of bounded radius. This result is complemented by the minimax lower bound in [21] where it was shown that $n^{-4/5}$ is the minimax rate for this problem. These two results together imply minimax optimality of $\hat{K}_{ls}$ under the loss function L_f. No other theoretical results for this problem are available outside of those in [14] and [21].

As a result, the following basic questions are still unanswered:

1. How to optimally and adaptively estimate $h_{K^*}(\theta_i)$ for a fixed $i \in \{1, \ldots, n\}$? This is the pointwise estimation problem. In the literature on shape constrained estimation, pointwise estimation has been well studied. Prominent examples include [4, 7, 8, 16, 17, 24, 34] for monotonicity constrained estimation and [5, 19, 20, 23, 27] for convexity constrained estimation. For the problem considered here however, nothing is known about pointwise estimation. It may be noted that the result $L_f(K^*, \hat{K}_{ls}) = O(n^{-4/5})$ of [14] does not say anything about the accuracy of $h_{\hat{K}_{ls}}(\theta_i)$ as an estimator for $h_{K^*}(\theta_i)$.

2. How to construct minimax optimal estimators for the set K^* that also adapt to polytopes? Polytopes with a small number of extreme points have a much simpler structure than general convex sets. In the problem of estimating convex sets under more standard observation models different from the one studied here, it is possible to construct estimators that converge at faster rates for polytopes compared to the overall minimax rate (see [3] for a summary of this theory). Similar kinds of adaptation has been recently studied for monotonicity and convexity constrained estimation problems; see [2, 9, 22]. Based on these results, it is natural to expect minimax estimators that adapt to polytopes in this problem. This has not been addressed previously.

1.2. *Our contributions.* We answer both the above questions in the affirmative in this paper. The main contributions can be summarized as follows:

1. We study the pointwise adaptive estimation problem in detail in the decision theoretic framework where the focus is on the performance at every function, instead of the maximum risk over a large parameter space as in the conventional minimax theory in nonparametric estimation literature. This framework, first introduced in [6] and [5] for shape constrained regression, provides a much more precise characterization of the performance of an estimator than the conventional minimax theory does.

In the context of the present problem, the difficulty of estimating $h_{K^*}(\theta_i)$ at a given K^* and θ_i can be expressed by means of a benchmark $R_n(K^*, \theta)$, which is defined as follows [below $\mathbb{E}_L$ denotes expectation taken with respect to the joint distribution of $Y_1, \ldots, Y_n$ generated according to the model (2) with K^* replaced by L]:

$$(4) \qquad R_n(K^*, \theta) = \sup_L \inf_{\tilde{h}} \max\left(\mathbb{E}_{K^*}(\tilde{h} - h_{K^*}(\theta))^2, \mathbb{E}_L(\tilde{h} - h_L(\theta))^2 \right),$$

where the supremum above is taken over all compact, convex sets L while the infimum is over all estimators $\tilde{h}$. In our first result for pointwise estimation, we establish, for each $i \in \{1, \ldots, n\}$, a lower bound for the performance of every estimator for estimating $h_{K^*}(\theta_i)$. Specifically, it is shown that

$$
(5) \qquad\qquad R_n(K^*, \theta_i) \geq c \cdot \frac{\sigma^2}{k_*(i) + 1},
$$

where $k_*(i)$ is an integer for which an explicit formula can be given in terms of K^* and i; and c is a universal positive constant. It will turn out that $k_*(i)$ is related to the smoothness of $h_{K^*}(\theta)$ at $\theta = \theta_i$.

We construct a data-driven estimator, $\hat{h}_i$, of $h_{K^*}(\theta_i)$ based on local smoothing together with an optimization scheme for automatically choosing a bandwidth, and show that the estimator $\hat{h}_i$ satisfies

$$
(6) \qquad\qquad \mathbb{E}_{K^*}\big(\hat{h}_i - h_{K^*}(\theta_i)\big)^2 \leq C \cdot \frac{\sigma^2}{k_*(i) + 1}
$$

for a universal constant $C > 0$. Inequalities (5) and (6) [see also inequality (21)] together imply that $\hat{h}_i$ is, within a constant factor, an optimal estimator of $h_{K^*}(\theta_i)$ for every compact, convex set K^*. This optimality is much stronger than the traditional minimax optimality usually employed in nonparametric function estimation. The quantity $\sigma^2/(k_*(i) + 1)$ depends on the unknown set K^* in a similar way that the Fisher information depends on the unknown parameter in a regular parametric model. In contrast, the optimal rate in the minimax paradigm is given in terms of the worst case performance over a large parameter space and does not depend on individual parameter values.

2. Using the optimal adaptive point estimators $\hat{h}_1, \ldots, \hat{h}_n$, we construct two set estimators $\hat{K}$ and $\hat{K}'$. The details of this construction are given in Section 2.2. In Theorems 3.7 and 3.9, it is shown that $\hat{K}$ is minimax optimal for K^* under the loss function L_f while the estimator $\hat{K}'$ is minimax optimal under the integral squared loss function defined by

$$
(7) \qquad\qquad L(K^*, \hat{K}') := \int_{-\pi}^{\pi} \big(h_{\hat{K}'}(\theta) - h_{K^*}(\theta)\big)^2 \, d\theta.
$$

The square root of the above loss function is often referred to as the McClure–Vitale metric on the space of nonempty compact, convex sets (see, e.g., [28] and [10]). In Theorem 3.7, we prove that

$$
(8) \qquad\qquad \mathbb{E}_{K^*} L_f(K^*, \hat{K}) \leq C \left\{ \frac{\sigma^2}{n} + \left(\frac{\sigma^2 \sqrt{R}}{n} \right)^{4/5} \right\}
$$

provided K^* is contained in a ball of radius R. This, combined with the minimax lower bound in [21], proves the minimax optimality of $\hat{K}$. An analogous result is shown in Theorem 3.9 for $\mathbb{E}_{K^*} L(K^*, \hat{K}')$. For the pointwise estimation problem where the goal is to estimate $h_{K^*}(\theta_i)$, the optimal rate $\sigma^2/(k_*(i) + 1)$ can be as

large as $n^{-2/3}$; however, the bound (8) shows that the globally the risk is at most $n^{-4/5}$. The shape constraint given by convexity of K^* ensures that the points where pointwise estimation rate is $n^{-2/3}$ cannot be too many. Note that we make no smoothness assumptions for proving (8).

3. We show that our set estimators $\hat{K}$ and $\hat{K}'$ adapt to polytopes with bounded number of extreme points. Already inequality (8) implies that $\mathbb{E}_{K^*} L_f(K^*, \hat{K})$ is bounded from above by the parametric risk $C\sigma^2/n$ provided $R = 0$ (note that $R = 0$ means that K^* is a singleton). Because σ^2/n is much smaller than $n^{-4/5}$, the bound (8) shows that $\hat{K}$ adapts to singletons. Theorem 3.8 extends this adaptation phenomenon to polytopes and we show that $\mathbb{E}_{K^*} L_f(K^*, \hat{K})$ is bounded by the parametric rate (up to a logarithmic multiplicative factor of n) for all polytopes with bounded number of extreme points. An analogous result is also proved for $\mathbb{E}_{K^*} L(K^*, \hat{K}')$ in Theorem 3.9. It should be noted that the construction of our estimators $\hat{K}$ and $\hat{K}'$ (described in Section 2.2) does not involve any special treatment for polytopes; yet the estimators automatically achieve faster rates for polytopes.

We would like to stress two features of this paper: (a) we do not make any smoothness assumptions on the boundary of K^* throughout the paper; in particular, note that we obtain the $n^{-4/5}$ rate for the set estimators $\hat{K}$ and $\hat{K}'$ without any smoothness assumptions, and (b) we go beyond the traditional minimax paradigm by considering adaptive estimation in both the pointwise estimation problem and the problem of estimating the entire set K^*. In particular, pointwise estimation is studied in a general nonasymptotic framework, which evaluates the performance of a procedure at each individual set K^*, not the worst case performance over a large parameter space as in the conventional minimax theory.

1.3. *Organization of the paper.* The rest of the paper is structured as follows. The proposed estimators are described in detail in Section 2. The theoretical properties are analyzed in Section 3; Section 3.1 gives results for pointwise estimation while Section 3.2 deals with set estimators. Section 4 considers optimal estimation of some special compact convex sets K^* where we explicitly compute the associated rates of convergence. A simulation study is given in Section 5 where we compare the performance of our estimators to other existing estimators in the literature. In Section 6, we summarize our main results and discuss potential open problems for future work. The proofs of the main results are given in Section 7. Proofs of other results together with additional technical results as well as additional simulations are given in the Supplementary Material [32].

2. **Estimation procedures.** Recall the regression model (2), where we observe noisy measurements $(\theta_1, Y_1), \ldots, (\theta_n, Y_n)$ with $\theta_i = 2\pi i/n - \pi, i = 1, \ldots, n$ being fixed grid points in $(-\pi, \pi]$. In this section, we first describe in detail our estimate $\hat{h}_i$ for $h_{K^*}(\theta_i)$ for each i. Subsequently, we will put together these estimates $\hat{h}_1, \ldots, \hat{h}_n$ to yield set estimators for K^*.

2.1. *Estimators for $h_{K^*}(\theta_i)$ for each fixed i.* Fix $1 \le i \le n$. Our construction of the estimator $\hat{h}_i$ for $h_{K^*}(\theta_i)$ is based on the key circle-convexity property (1) of the function $h_{K^*}(\cdot)$. Let us define, for $\phi \in (0, \pi/2)$ and $\theta \in (-\pi, \pi]$, the following two quantities:

$$l(\theta, \phi) := \cos\phi\big(h_{K^*}(\theta + \phi) + h_{K^*}(\theta - \phi)\big) - \frac{h_{K^*}(\theta + 2\phi) + h_{K^*}(\theta - 2\phi)}{2}$$

and

$$u(\theta, \phi) := \frac{h_{K^*}(\theta + \phi) + h_{K^*}(\theta - \phi)}{2\cos\phi}.$$

The following lemma states that for every θ, the quantity $h_{K^*}(\theta)$ is sandwiched between $l(\theta, \phi)$ and $u(\theta, \phi)$ for every ϕ. This will be used crucially in defining $\hat{h}$. The proof of this lemma is a straightforward consequence of (1) and is given in the Supplementary Material [32], Section A.6.

LEMMA 2.1. *For every $0 < \phi < \pi/2$ and every $\theta \in (-\pi, \pi]$, we have* $l(\theta, \phi) \le h_{K^*}(\theta) \le u(\theta, \phi)$.

For a fixed $1 \le i \le n$, Lemma 2.1 implies that $l(\theta_i, \frac{2\pi j}{n}) \le h_{K^*}(\theta_i) \le u(\theta_i, \frac{2\pi j}{n})$ for every $0 \le j < \lfloor n/4 \rfloor$. Note that when $j = 0$, we have $l(\theta_i, 0) = h_{K^*}(\theta_i) = u(\theta_i, 0)$. Averaging these inequalities for $j = 0, 1, \ldots, k$ where k is a fixed integer with $0 \le k < \lfloor n/4 \rfloor$, we obtain

$$(9) \qquad L_k(\theta_i) \le h_{K^*}(\theta_i) \le U_k(\theta_i) \qquad \text{for every } 0 \le k < \lfloor n/4 \rfloor,$$

where

$$L_k(\theta_i) := \frac{1}{k+1} \sum_{j=0}^{k} l\left(\theta_i, \frac{2\pi j}{n}\right) \quad \text{and} \quad U_k(\theta_i) := \frac{1}{k+1} \sum_{j=0}^{k} u\left(\theta_i, \frac{2\pi j}{n}\right).$$

We are now ready to describe our estimator. Fix $1 \le i \le n$. Inequality (9) says that the quantity of interest, $h_{K^*}(\theta_i)$, is sandwiched between $L_k(\theta_i)$ and $U_k(\theta_i)$ for every k. Both $L_k(\theta_i)$ and $U_k(\theta_i)$ can naturally be estimated by unbiased estimators. Indeed, let

$$\hat{l}(\theta_i, 2j\pi/n) := \cos(2j\pi/n)(Y_{i+j} + Y_{i-j}) - \frac{Y_{i+2j} + Y_{i-2j}}{2},$$

$$\hat{u}(\theta_i, 2j\pi/n) := \frac{Y_{i+j} + Y_{i-j}}{2\cos(2j\pi/n)}$$

and take

$$(10) \quad \hat{L}_k(\theta_i) := \frac{1}{k+1} \sum_{j=0}^{k} \hat{l}(\theta_i, 2j\pi/n), \qquad \hat{U}_k(\theta_i) := \frac{1}{k+1} \sum_{j=0}^{k} \hat{u}(\theta_i, 2j\pi/n).$$

Obviously, in order for the above to be meaningful, we need to define Y_i even for $i \notin \{1, \ldots, n\}$. This is easily done in the following way: for any $i \in \mathbb{Z}$, let $s \in \mathbb{Z}$ be such that $i - sn \in \{1, \ldots, n\}$ and take $Y_i := Y_{i-sn}$.

As k increases, one averages more terms in (10), and hence the estimators $\hat{L}_k(\theta_i)$ and $\hat{U}_k(\theta_i)$ become more accurate. Let $\hat{\Delta}_k(\theta_i) := \hat{U}_k(\theta_i) - \hat{L}_k(\theta_i)$ which is the same as

$$(11) \qquad \hat{\Delta}_k(\theta_i) = \frac{1}{k+1} \sum_{j=0}^{k} \left(\frac{Y_{i+2j} + Y_{i-2j}}{2} - \frac{\cos(4j\pi/n)}{\cos(2j\pi/n)} \frac{Y_{i+j} + Y_{i-j}}{2} \right).$$

Because of (9), a natural strategy for estimating $h_{K^*}(\theta_i)$ is to choose k for which $\hat{\Delta}_k(\theta_i)$ is the smallest and then use either $\hat{L}_k(\theta_i)$ or $\hat{U}_k(\theta_i)$ at that k as the estimator. This is essentially our estimator with one small difference in that we also take into account the noise present in $\hat{\Delta}_k(\theta_i)$. Formally, our estimator for $h_{K^*}(\theta_i)$ is given by

$$(12) \qquad \hat{h}_i = \hat{U}_{\hat{k}(i)}(\theta_i) \qquad \text{where } \hat{k}(i) := \underset{k \in \mathcal{I}}{\operatorname{argmin}} \left\{ (\hat{\Delta}_k(\theta_i))_+ + \frac{2\sigma}{\sqrt{k+1}} \right\}$$

and $\mathcal{I} := \{0\} \cup \{2^j : j \geq 0 \text{ and } 2^j \leq \lfloor n/16 \rfloor\}$.

Our estimator $\hat{h}_i$ can be viewed as an angle-adjusted local averaging estimator. It is inspired by the estimator of [5] for convex regression. The number of terms averaged equals $\hat{k}(i) + 1$ and this is analogous to the bandwidth in kernel-based smoothing methods. Our $\hat{k}(i)$ is determined from an optimization scheme. Notice that unlike the least squares estimator $h_{\hat{K}_{ls}}(\theta_i)$, the construction of $\hat{h}_i$ for a fixed i does not depend on the construction of $\hat{h}_j$ for $j \neq i$.

2.2. *Set estimators for K^*.* We next present estimators for the set K^*. The point estimators $\hat{h}_1, \ldots, \hat{h}_n$ do not directly give an estimator for K^* because $(\hat{h}_1, \ldots, \hat{h}_n)$ is not necessarily a valid support vector, that is, $(\hat{h}_1, \ldots, \hat{h}_n)$ does not always belong to the following set:

$$\mathcal{H} := \{(h_K(\theta_1), \ldots, h_K(\theta_n)) : K \subseteq \mathbb{R}^2 \text{ is compact and convex}\}.$$

To get a valid support vector from $(\hat{h}_1, \ldots, \hat{h}_n)$, we need to project it onto $\mathcal{H}$ to obtain

$$(13) \qquad \hat{h}^P := (\hat{h}_1^P, \ldots, \hat{h}_n^P) := \underset{(h_1, \ldots, h_n) \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} (\hat{h}_i - h_i)^2.$$

The superscript P here stands for projection. An estimator for the set K^* can now be constructed immediately from $\hat{h}_1^P, \ldots, \hat{h}_n^P$ via

$$(14) \qquad \hat{K} := \{(x_1, x_2) : x_1 \cos\theta_i + x_2 \sin\theta_i \leq \hat{h}_i^P \text{ for all } i = 1, \ldots, n\}.$$

In Theorems 3.7 and 3.8, we prove upper bounds on the accuracy of $\hat{K}$ under the loss function L_f given in (3).

There is another reasonable way of constructing a set estimator for K^* based on the point estimators $\hat{h}_1, \ldots, \hat{h}_n$. We first interpolate $\hat{h}_1, \ldots, \hat{h}_n$ to define a function $\hat{h}' : (-\pi, \pi] \to \mathbb{R}$ as follows:

$$(15) \qquad \hat{h}'(\theta) := \frac{\sin(\theta_{i+1} - \theta)}{\sin(\theta_{i+1} - \theta_i)} \hat{h}_i + \frac{\sin(\theta - \theta_i)}{\sin(\theta_{i+1} - \theta_i)} \hat{h}_{i+1} \qquad \text{for } \theta_i \le \theta \le \theta_{i+1}.$$

Here, i ranges over $1, \ldots, n$ with the convention that $\theta_{n+1} = \theta_1 + 2\pi$ (and $\theta_n \le \theta \le \theta_{n+1}$ should be identified with $-\pi \le \theta \le -\pi + 2\pi/n$). Based on this function $\hat{h}'$, we can define our estimator $\hat{K}'$ of K^* by

$$(16) \qquad \hat{K}' := \underset{K}{\operatorname{argmin}} \int_{-\pi}^{\pi} \left(\hat{h}'(\theta) - h_K(\theta) \right)^2 d\theta.$$

The existence and uniqueness of $\hat{K}'$ can be justified in the usual way by the Hilbert space projection theorem. In Theorem 3.9, we prove bounds on the accuracy of $\hat{K}'$ as an estimator for K^* under the integral loss L given in (7).

Let us now briefly comment on the algorithms for computing our set estimators $\hat{K}$ and $\hat{K}'$. The expression (14) shows how to write $\hat{K}$ in terms of $\hat{h}_i^P$, $i = 1, \ldots, n$ and, therefore, we only need to be able to compute $\hat{h}_i^P$, $i = 1, \ldots, n$ for computing $\hat{K}$. This can be done via quadratic programming because the set $\mathcal{H}$ can explicitly written as $\{h \in \mathbb{R}^n : a_i^T h \le 0, i = 1, \ldots, n\}$ for some collection of vectors $a_1, \ldots, a_n$ in $\mathbb{R}^n$ (see, e.g., [29], Theorem 1). To compute $\hat{K}'$, we take a fine uniform grid of points $\alpha_1, \ldots, \alpha_M$ in $(-\pi, \pi]$ for a large value of M and approximate $\hat{K}'$ via

$$\underset{K}{\operatorname{argmin}} \sum_{i=1}^{M} (\hat{h}'(\alpha_i) - h_K(\alpha_i))^2.$$

More precisely, one can take $\hat{K}' := \{(x_1, x_2) : x_1 \cos \alpha_i + x_2 \sin \alpha_i \le \tilde{h}_i \text{ for all } i = 1, \ldots, M\}$ where

$$(\tilde{h}_1, \ldots, \tilde{h}_M) := \underset{(h_1, \ldots, h_M) \in \mathcal{H}^M}{\operatorname{argmin}} \sum_{i=1}^{M} (\hat{h}'(\alpha_i) - h_i)^2$$

with $\mathcal{H}^M := \{(h_K(\alpha_1), \ldots, h_K(\alpha_M)) : K \subseteq \mathbb{R}^2 \text{ is compact and convex}\}$. This estimator can then be computed in an analogous way as $\hat{K}$ by quadratic programming. We present simulation examples in Section 5 where one can see that there is often not much difference between $\hat{K}$ and $\hat{K}'$ in practice.

3. Main results. We now investigate the accuracy of the proposed point and set estimators. The proofs of these results are given in Section 7.

3.1. *Accuracy of the point estimator.* As mentioned in the Introduction, we evaluate the performance of the point estimator $\hat{h}_i$ at individual functions, not the worst case over a large parameter space. This provides a much more precise characterization of the accuracy of the estimator. Let us first recall inequality (9) where $h_{K^*}(\theta_i)$ is sandwiched between $L_k(\theta_i)$ and $U_k(\theta_i)$. Define $\Delta_k(\theta_i) := U_k(\theta_i) - L_k(\theta_i)$.

THEOREM 3.1. *Fix $i \in \{1, \ldots, n\}$. There exists a universal constant $C > 0$ such that the risk of $\hat{h}_i$ as an estimator of $h_{K^*}(\theta_i)$ satisfies the inequality,*

$$(17) \qquad \mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq C \cdot \frac{\sigma^2}{k_*(i) + 1},$$

where

$$(18) \qquad k_*(i) := \underset{k \in \mathcal{I}}{\operatorname{argmin}}\left(\Delta_k(\theta_i) + \frac{2\sigma}{\sqrt{k+1}}\right).$$

REMARK 3.1. It turns out that the bound in (17) is linked to the level of smoothness of the function h_{K^*} at θ_i. However, for this interpretation to be correct, one needs to regard h_{K^*} as a function on $\mathbb{R}^2$ instead of a subset of $\mathbb{R}$. This is further explained in Remark 4.1.

Theorem 3.1 gives an explicit bound on the risk of $\hat{h}_i$ in terms of the quantity $k_*(i)$ defined in (18). It is important to keep in mind that $k_*(i)$ depends on K^* even though this is suppressed in the notation. In the next theorem, we show that $\sigma^2/(k_*(i) + 1)$ also presents a lower bound on the accuracy of every estimator for $h_{K^*}(\theta_i)$. This implies, in particular, optimality of $\hat{h}_i$ as an estimator of $h_{K^*}(\theta_i)$.

One needs to be careful in formulating the lower bound in this setting. A first attempt might perhaps be to prove that, for a universal constant $c > 0$,

$$\inf_{\tilde{h}} \mathbb{E}_{K^*}(\tilde{h} - h_{K^*}(\theta_i))^2 \geq c \cdot \frac{\sigma^2}{k_*(i) + 1},$$

where the infimum is over all possible estimators $\tilde{h}$. This, of course, would not be possible because one can take $\tilde{h} = h_{K^*}(\theta_i)$, which would make the left hand side zero. A formulation of the lower bound which avoids this difficulty was proposed by [5] in the context of convex function estimation. Their idea, translated to our setting of estimating the support function h_{K^*} at a point θ_i, is to consider, instead of the risk at K^*, the maximum of the risk at K^* and the risk at L^* which is most difficult to distinguish from K^* in term of estimating $h_{K^*}(\theta_i)$. This leads to the benchmark $R_n(K^*, \theta_i)$ defined in (4).

THEOREM 3.2. *For any fixed $i \in \{1, \ldots, n\}$, we have*

$$(19) \qquad R_n(K^*, \theta_i) \geq c \cdot \frac{\sigma^2}{k_*(i) + 1}$$

for a universal constant $c > 0$.

Theorems 3.1 and 3.2 together imply that $\sigma^2/(k_*(i) + 1)$ is the optimal rate of estimation of $h_{K^*}(\theta_i)$ for a given compact, convex set K^*. The results show that our data driven estimator $\hat{h}_i$ for $h_{K^*}(\theta_i)$ performs uniformly within a constant factor of the ideal benchmark $R_n(K^*, \theta_i)$ for every i. This means that $\hat{h}_i$ adapts to every unknown set K^* instead of a collection of large parameter spaces as in the conventional minimax theory commonly used in nonparametric literature.

REMARK 3.2 (A stronger upper bound on the risk of $\hat{h}_i$).   From the proof of Theorem 3.2, it can be seen that the following statement is true: there exists a compact, convex set L^* such that

$$(20) \qquad \inf_{\tilde{h}} \max\left(\mathbb{E}_{K^*}(\tilde{h} - h_{K^*}(\theta_i))^2, \mathbb{E}_{L^*}(\tilde{h} - h_{L^*}(\theta_i))^2\right) \geq \frac{c\sigma^2}{k_*(i) + 1}$$

the infimum above being over all estimators $\tilde{h}$ of $h_{K^*}(\theta_i)$. In light of this, it is natural to ask whether the following inequality:

$$(21) \qquad \max\left(\mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2, \mathbb{E}_{L^*}(\hat{h}_i - h_{L^*}(\theta_i))^2\right) \leq \frac{C\sigma^2}{k_*(i) + 1}$$

holds for the same L^* where $\hat{h}_i$ refers to our estimator defined in (12) and C represents a universal constant. Note that this is a stronger inequality than (17). It turns out that (21) is indeed a true inequality and we provide a proof in the Supplementary Material [32], Section A.3.

Given a specific set K^* and $1 \leq i \leq n$, the quantity $k_*(i)$ is often straightforward to compute up to constant multiplicative factors. Several examples are provided in Section 4. From these examples, it will be clear that the size of $\sigma^2/(k_*(i) + 1)$ is linked to the level of smoothness of the function h_{K^*} at θ_i. However, for this interpretation to be correct, one needs to regard h_{K^*} as a function on $\mathbb{R}^2$ instead of a subset of $\mathbb{R}$. This is explained in Remark 4.1.

The following corollaries shed more light on the quantity $\sigma^2/(k_*(i) + 1)$. The proofs of these corollaries are given in the Supplementary Material [32], Section A.4. The first corollary below shows that $\sigma^2/(k_*(i) + 1)$ is at most $C(\sigma^2 R/n)^{-2/3}$ for every i and K^* (C is a universal constant) provided K^* is contained in a ball of radius R. In Example 4.3, we provide an explicit choice of i and K^* for which $\sigma^2/(k_*(i) + 1) \geq c(\sigma^2 R/n)^{-2/3}$ (c is a universal constant). This implies that the conclusion of the following corollary cannot in general be improved.

COROLLARY 3.3.   *Suppose K^* is contained in some closed ball of radius R. Then for every $i \in \{1, \ldots, n\}$, we have, for a universal constant $C > 0$,*

$$(22) \qquad \frac{\sigma^2}{k_*(i) + 1} \leq C\left\{\left(\frac{\sigma^2 R}{n}\right)^{2/3} + \frac{\sigma^2}{n}\right\}$$

and

$$(23) \qquad \mathbb{E}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq C\left\{\left(\frac{\sigma^2 R}{n}\right)^{2/3} + \frac{\sigma^2}{n}\right\}.$$

Note that the above corollary implies the consistency of $\hat{h}_i$ as an estimator for $h_{K^*}(\theta_i)$ for every i and K^*. It turns out that $\hat{h}_i$ is a minimax optimal estimator of $h_{K^*}(\theta_i)$ over the class of all compact convex sets K^* contained in some closed ball of radius R. This is proved in the next result.

PROPOSITION 3.4. *For $R \geq 0$, let $\mathcal{K}(R)$ denote the class of all compact, convex sets that are contained in some fixed closed ball of radius R. Then for every $i \in \{1, \ldots, n\}$, we have*

$$(24) \qquad \sup_{K^* \in \mathcal{K}(R)} \mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq C\left\{\frac{\sigma^2}{n} + \left(\frac{\sigma^2 R}{n}\right)^{2/3}\right\}$$

for a universal constant C. We further have

$$(25) \qquad \inf_{\tilde{h}} \sup_{K^* \in \mathcal{K}(R)} \mathbb{E}_{K^*}(\tilde{h} - h_{K^*}(\theta_i))^2 \geq c\left\{\frac{\sigma^2}{n} + \left(\frac{\sigma^2 R}{n}\right)^{2/3}\right\}$$

for a universal constant $c > 0$ where the infimum is taken over all possible estimators $\tilde{h}$ of $h_{K^*}(\theta_i)$.

It is clear from the definition (18) that $k_*(i) \leq n$ for all i and K_*. In the next corollary, we prove that there exist sets K_* and i for which $k_*(i) \geq cn$ for a constant c. For these sets, the optimal rate of estimating $h_{K^*}(\theta_i)$ is therefore parametric.

For a fixed i and K^*, let $\phi_1(i)$ and $\phi_2(i)$ be such that $\phi_1(i) \leq \theta_i \leq \phi_2(i)$ and such that there exists a single point $(x_1, x_2) \in K^*$ with

$$(26) \qquad h_{K^*}(\theta) = x_1 \cos\theta + x_2 \sin\theta \qquad \text{for all } \theta \in [\phi_1(i), \phi_2(i)].$$

The following corollary says that if the distance of θ_i to its nearest end-point in the interval $[\phi_1(i), \phi_2(i)]$ is large (i.e., of constant order), then the optimal rate of estimation of $h_{K^*}(\theta_i)$ is parametric. This situation happens usually for polytopes (polytopes are compact, convex sets with finitely many vertices); see Examples 4.1 and 4.3 for specific instances of this phenomenon. For nonpolytopes, it can often happen that $\phi_1(i) = \phi_2(i) = \theta_i$ in which case the conclusion of the next corollary is not useful.

COROLLARY 3.5. *For every $i \in \{1, \ldots, n\}$, we have*

$$(27) \qquad k_*(i) \geq cn \min(\theta_i - \phi_1(i), \phi_2(i) - \theta_i, \pi)$$

for a universal constant $c > 0$. Consequently,

$$(28) \qquad \mathbb{E}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq \frac{C\sigma^2}{1 + n \min(\theta_i - \phi_1(i), \phi_2(i) - \theta_i, \pi)}$$

for a universal constant $C > 0$.

From the above two corollaries, it is clear that the optimal rate of estimation of $h_{K^*}(\theta_i)$ can be as large as $n^{-2/3}$ and as small as the parametric rate n^{-1}. The rate $n^{-2/3}$ is achieved, for example, in the setting given in Example 4.3 while the parametric rate is achieved, for example, for polytopes.

The next corollary argues that bounding $k_*(i)$ in specific examples requires only bounding the quantity $\Delta_k(\theta_i)$ from above and below. This corollary will be useful in Section 4 while working out $k_*(i)$ in specific examples.

COROLLARY 3.6. *Fix $1 \leq i \leq n$. Let $\{f_k(\theta_i), k \in \mathcal{I}\}$ and $\{g_k(\theta_i), k \in \mathcal{I}\}$ be two sequences which satisfy $g_k(\theta_i) \leq \Delta_k(\theta_i) \leq f_k(\theta_i)$ for all $k \in \mathcal{I}$. Also let*

$$(29) \qquad \check{k}(i) := \max\left\{k \in \mathcal{I} : f_k(\theta_i) < \frac{(\sqrt{6} - 2)\sigma}{\sqrt{k+1}}\right\}$$

and

$$(30) \qquad \tilde{k}(i) := \min\left\{k \in \mathcal{I} : g_k(\theta_i) > \frac{6(\sqrt{2} - 1)\sigma}{\sqrt{k+1}}\right\}$$

as long as there is some $k \in \mathcal{I}$ for which $g_k(\theta_i) > 6(\sqrt{2} - 1)\sigma/\sqrt{k+1}$; otherwise take $\tilde{k}(i) := \max_{k \in \mathcal{I}} k$. We then have $\check{k}(i) \leq k_*(i) \leq \tilde{k}(i)$ and

$$(31) \qquad \mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq C \frac{\sigma^2}{\check{k}(i) + 1}$$

for a universal constant $C > 0$.

3.2. *Accuracy of set estimators.* We now turn to study the accuracy of the set estimators $\hat{K}$ [defined in (14)] and $\hat{K}'$ [defined in (16)]. The accuracy of $\hat{K}$ will be investigated under the loss function L_f [defined in (3)] while the accuracy of $\hat{K}'$ will be studied under the loss function L [defined in (7)].

In Theorem 3.7 below we prove that $\mathbb{E}_{K^*} L_f(K^*, \hat{K})$ is bounded from above by a constant multiple of $n^{-4/5}$ as long as K^* is contained in a ball of radius R. The discussions following the theorem shed more light on its implications.

THEOREM 3.7. *If K^* is contained in some closed ball of radius $R \geq 0$, then*

$$(32) \qquad \mathbb{E}_{K^*} L_f(K^*, \hat{K}) \leq C\left\{\frac{\sigma^2}{n} + \left(\frac{\sigma^2 \sqrt{R}}{n}\right)^{4/5}\right\}$$

for a universal constant $C > 0$. Note here that $R = 0$ is allowed (in which case K^* is a singleton).

Note that as long as $R > 0$, the right-hand side in (32) will be dominated by the $(\sigma^2 \sqrt{R}/n)^{-4/5}$ term for all large n. This would mean that

$$(33) \qquad \sup_{K^* \in \mathcal{K}(R)} \mathbb{E}_{K^*} L_f(K^*, \hat{K}) \leq C \left(\frac{\sigma^2 \sqrt{R}}{n} \right)^{4/5},$$

where $\mathcal{K}(R)$ denotes the set of all compact convex sets contained in some fixed closed ball of radius R.

The minimax rate of estimation over the class $\mathcal{K}(R)$ was studied in [21]. In [21], Theorems 3.1 and 3.2, it was proved that

$$(34) \qquad \inf_{\tilde{K}} \sup_{K^* \in \mathcal{K}(R)} \mathbb{E}_{K^*} L_f(K^*, \hat{K}) \asymp \left(\frac{\sigma^2 \sqrt{R}}{n} \right)^{4/5},$$

where $\asymp$ denotes equality up to constant multiplicative factors. From (33) and (34), it follows that $\hat{K}$ is a minimax optimal estimator of K^*. We should mention here that an inequality of the form (33) was proved for the least squares estimator $\hat{K}_{ls}$ by [14], which implies that $\hat{K}_{ls}$ is also a minimax optimal estimator of K^*.

The $n^{-4/5}$ minimax rate here is quite natural in connection with estimation of smooth functions. Indeed, this is the minimax rate for estimating twice differentiable one-dimensional functions. Although we have not made any smoothness assumptions here, we are working under a convexity-based constraint and convexity is associated, in a broad sense, with twice smoothness (see, e.g., [1]).

REMARK 3.3. Because of the formula (3) for the loss function L_f, the risk $\mathbb{E}_{K^*} L_f(K^*, \hat{K})$ can be seen as the average of the risk of $\hat{K}$ for estimating $h_{K^*}(\theta_i)$ over $i = 1, \dots, n$. We have seen in Section 3.1 that the optimal rate of estimating $h_{K^*}(\theta_i)$ can be as high as $n^{-2/3}$. Theorem 3.7, on the other hand, can be interpreted as saying that, on average over $i = 1, \dots, n$, the optimal rate of estimating $h_{K^*}(\theta_i)$ is at most $n^{-4/5}$. Indeed, the key to proving Theorem 3.7 is to establish the following inequality:

$$\frac{\sigma^2}{n} \sum_{i=1}^{n} \frac{1}{k_*(i) + 1} \leq C \left\{ \frac{\sigma^2}{n} + \left(\frac{\sigma^2 \sqrt{R}}{n} \right)^{4/5} \right\}.$$

Under the assumption that K^* is contained in a ball of radius R. Therefore, even though each term $\sigma^2/(k_*(i) + 1)$ can be as large as $n^{-2/3}$, on average, their size is at most $n^{-4/5}$.

REMARK 3.4. Theorem 3.7 provides different qualitative conclusions when K^* is a singleton. In this case, one can take $R = 0$ in (32) to get the parametric bound $C\sigma^2/n$ for $\mathbb{E}_{K^*} L_f(K^*, \hat{K})$. Because this is smaller than the nonparametric $n^{-4/5}$ rate, it means that $\hat{K}$ adapts to singletons. Singletons are simple examples of polytopes and one naturally wonders here if $\hat{K}$ also adapts to other polytopes as

well. This is, however, not implied by inequality (32), which gives the rate $n^{-4/5}$ for every K^* that is not a singleton. It turns out that $\hat{K}$ indeed adapts to other polytopes and we prove this in the next theorem. In fact, we prove that $\hat{K}$ adapts to any K^* that is well approximated by a polytope with not too many vertices. It is currently not known if the least squares estimator $\hat{K}_{\mathrm{ls}}$ has such adaptivity.

We next prove another bound for $\mathbb{E}_{K^*} L_f(K^*, \hat{K})$. This bound demonstrates adaptivity of $\hat{K}$ as described in the previous remark. Recall that polytopes are compact, convex sets with finitely many extreme points (or vertices). The space of all polytopes in $\mathbb{R}^n$ will be denoted by $\mathcal{P}$. For a polytope $P \in \mathcal{P}$, we denote by v_P, the number of extreme points of P. Also recall the notion of Hausdorff distance between two compact, convex sets K and L defined by

$$(35) \qquad \ell_H(K, L) := \sup_{\theta \in \mathbb{R}} |h_K(\theta) - h_L(\theta)|.$$

This is not the usual way of defining the Hausdorff distance. For an explanation of the connection between this and the usual definition, see, for example, [30], Theorem 1.8.11.

THEOREM 3.8. *There exists a universal constant $C > 0$ such that*

$$(36) \qquad \mathbb{E}_{K^*} L_f(K^*, \hat{K}) \le C \inf_{P \in \mathcal{P}} \left[\frac{\sigma^2 v_P}{n} \log\left(\frac{en}{v_P}\right) + \ell_H^2(K^*, P) \right].$$

REMARK 3.5 (Near-parametric rates for polytopes). The bound (36) implies that $\hat{h}$ has the parametric rate (up to a logarithmic factor of n) for estimating polytopes. Indeed, suppose that K^* is a polytope with v vertices. Then using $P = K^*$ in the infimum in (36), we have the risk bound

$$(37) \qquad \mathbb{E}_{K^*} L_f(K^*, \hat{K}) \le \frac{C\sigma^2 v}{n} \log\left(\frac{en}{v}\right).$$

This is the parametric rate $\sigma^2 v/n$ up to logarithmic factors and is smaller than the nonparametric rate $n^{-4/5}$ given in (32).

REMARK 3.6. When $v = 1$, inequality (37) has a redundant logarithmic factor. Indeed, when $v = 1$, we can use (32) with $R = 0$ which gives (37) without the additional logarithmic factor. We do not know if the logarithmic factor in (37) can be removed for values of v larger than one as well.

Now consider the second set estimator $\hat{K}'$. The next theorem gives an upper bound on its accuracy under the integral loss function L [defined in (7)].

THEOREM 3.9. *Suppose K^* is contained in some closed ball of radius $R \geq 0$. The risk $\mathbb{E}_{K^*} L(K^*, \hat{K}')$ satisfies both the following inequalities*:

$$(38) \qquad \mathbb{E}_{K^*} L(K^*, \hat{K}') \leq C \left\{ \frac{\sigma^2}{n} + \left(\frac{\sigma^2 \sqrt{R}}{n} \right)^{4/5} + \frac{R^2}{n^2} \right\}$$

and

$$(39) \qquad \mathbb{E}_{K^*} L(K^*, \hat{K}') \leq C \inf_{P \in \mathcal{P}} \left[\frac{\sigma^2 v_P}{n} \log\left(\frac{en}{v_P} \right) + \ell_H^2(K^*, P) + \frac{R^2}{n^2} \right].$$

The only difference between the inequalities (38) and (39) on one hand and (32) and (36) on the other is the presence of the R^2/n^2 term. This term is usually very small and does not change the qualitative behavior of the bounds. However, note that inequality (36) did not require any assumption on K^* being in a ball of radius R while this assumption is necessary for (39).

REMARK 3.7. The rate $(\sigma^2 \sqrt{R}/n)^{4/5}$ is the minimax rate for this problem under the loss function L. Although this has not been proved explicitly anywhere, it can be shown by modifying the proof of [21], Theorem 3.2, appropriately. Theorem 3.9 therefore shows that $\hat{K}'$ is a minimax optimal estimator of K^* under the loss function L.

4. Examples.

We now investigate the results given in the last section for specific choices of K^*. It is useful here to note that $\Delta_k(\theta_i) = U_k(\theta_i) - L_k(\theta_i)$ has the following alternative expression:

$$(40) \qquad \frac{1}{k+1} \sum_{j=0}^{k} \left(h_{K^*}(\theta_i \pm 4j\pi/n) - \frac{\cos(4j\pi/n)}{\cos(2j\pi/n)} h_{K^*}(\theta_i \pm 2j\pi/n) \right),$$

where we write $h_{K^*}(\theta_i \pm \phi)$ for $(h_{K^*}(\theta_i + \phi) + h_{K^*}(\theta_i - \phi))/2$ with $\phi = 2j\pi/n, 4j\pi/n$.

EXAMPLE 4.1 (Single point). Suppose $K^* := \{(x_1, x_2)\}$ for a fixed point $(x_1, x_2) \in \mathbb{R}^2$. In this case,

$$(41) \qquad h_{K^*}(\theta) = x_1 \cos \theta + x_2 \sin \theta \qquad \text{for all } \theta.$$

It can then be directly checked from (40) that $\Delta_k(\theta_i) = 0$ for every $k \in \mathcal{I}$ and $i \in \{1, \dots, n\}$. As a result, it follows that $k_*(i) = \max_{k \in \mathcal{I}} k \geq cn$ for a constant $c > 0$. Theorem 3.1 then says that the point estimator $\hat{h}_i$ satisfies

$$(42) \qquad \mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq \frac{C\sigma^2}{n}$$

for a universal constant $C > 0$. One therefore gets the parametric rate here.

Also, Theorem 3.7 and inequality (38) in Theorem 3.9 can both be used here with $R = 0$. This implies that the set estimators $\hat{K}$ and $\hat{K}'$ both converge to K^* at the parametric rate under the loss functions L_f and L, respectively.

EXAMPLE 4.2 (Ball). Suppose K^* is a ball centered at (x_1, x_2) with radius $R > 0$. It is then easy to verify that

$$(43) \qquad h_{K^*}(\theta) = x_1 \cos \theta + x_2 \sin \theta + R \qquad \text{for all } \theta.$$

As a result, for every $k \in \mathcal{I}$ and $i \in \{1, \ldots, n\}$, we have

$$(44) \qquad \Delta_k(\theta_i) = \frac{R}{k+1} \sum_{j=0}^{k} \left(1 - \frac{\cos \frac{4\pi j}{n}}{\cos \frac{2\pi j}{n}}\right) \le R\left(1 - \frac{\cos 4\pi k/n}{\cos 2\pi k/n}\right).$$

Because $k \le n/16$ for all $k \in \mathcal{I}$, it is easy to verify that $\Delta_k(\theta_i) \le 8R \sin^2(\pi k/n) \le 8R\pi^2 k^2/n^2$. Taking $f_k(\theta_i) = 8R\pi^2 k^2/n^2$ in Corollary 3.6, we obtain that $k_*(i) \ge c \min(n, (n^2\sigma/R)^{2/5})$ for a constant c. Also since the function $1 - \cos(2x)/\cos(x)$ is a strongly convex function on $[-\pi/4, \pi/4]$ with second derivative lower bounded by 3, we have

$$\Delta_k(\theta_i) = \frac{R}{k+1} \sum_{j=0}^{k} \left(1 - \frac{\cos \frac{4\pi j}{n}}{\cos \frac{2\pi j}{n}}\right) \ge \frac{R}{k+1} \sum_{j=0}^{k} \frac{3}{2}\left(\frac{2\pi j}{n}\right)^2 = \frac{R\pi^2 k(2k+1)}{n^2}.$$

This gives $k_*(i) \le C \min(n, (n^2\sigma/R)^{2/5})$ as well for a constant C. We thus have $k_*(i) \asymp (n^2\sigma/R)^{2/5}$ for every i. Theorem 3.1 then gives

$$(45) \qquad \mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \le C\left\{\frac{\sigma^2}{n} + \left(\frac{\sigma^2\sqrt{R}}{n}\right)^{4/5}\right\}$$

for every $i \in \{1, \ldots, n\}$. Theorem 3.7 and inequality (38) prove that the set estimators $\hat{K}$ and $\hat{K}'$ also converge to K^* at the $n^{-4/5}$ rate.

In the preceding examples, we saw that the optimal rate $\sigma^2/(k_*(i) + 1)$ for estimating $h_{K^*}(\theta_i)$ did not depend on i. Next, we consider *asymmetric* examples where the rate changes with i.

EXAMPLE 4.3 (Segment). Let K^* be the vertical line segment joining $(0, R)$ and $(0, -R)$ for a fixed $R > 0$. Then $h_{K^*}(\theta) = R|\sin \theta|$ for all θ. Assume that n is even and consider $i = n/2$ so that $\theta_{n/2} = 0$. It can be verified that

$$\Delta_k(\theta_{n/2}) = \Delta_k(0) = \frac{R}{k+1} \sum_{j=0}^{k} \tan \frac{2\pi j}{n} \qquad \text{for every } k \in \mathcal{I}.$$

Because $j \mapsto \tan(2\pi j/n)$ is increasing, it is straightforward to deduce from above that $3\pi Rk/(8n) \le \Delta_k(0) \le 4\pi Rk/n$. Corollary 3.6 then gives

$$(46) \qquad \frac{\sigma^2}{k_*(n/2) + 1} \asymp \frac{\sigma^2}{n} + \left(\frac{\sigma^2 R}{n}\right)^{2/3}.$$

It was shown in Corollary 3.3 that the right-hand side above represents the maximum possible value of $\sigma^2/(k_*(i) + 1)$ when K^* lies in a closed ball of radius R.

Therefore, this example presents the situation where estimation of $h_{K^*}(\theta_i)$ is the most difficult. See Remark 4.1 for the connection to smoothness of $h_{K^*}(\cdot)$ at θ_i.

Now suppose that $i = 3n/4$ (assume that $n/4$ is an integer for simplicity) so that $\theta_i = \pi/2$. Observe then that $h_{K^*}(\theta) = R \sin \theta$ (without the modulus) for $\theta = \theta_i \pm 4j\pi/n$ for every $0 \le j \le k$, $k \in \mathcal{I}$. Using (40), we have $\Delta_k(\theta_i) = 0$ for every $k \in \mathcal{I}$. This immediately gives $k_*(i) = \lfloor n/16 \rfloor$, and hence

$$(47) \qquad \frac{\sigma^2}{k_*(3n/4) + 1} \asymp \frac{\sigma^2}{n}.$$

In this example, the risk for estimating $h_{K^*}(\theta_i)$ changes with i. For $i = n/2$, we get the $n^{-2/3}$ rate while for $i = 3n/4$, we get the parametric rate. For other values of i, one gets a range of rates between $n^{-2/3}$ and n^{-1}.

Because K^* is a polytope with 2 vertices, Theorem 3.8 and inequality (39) imply that the set estimators $\hat{K}$ and $\hat{K}'$ converge at the near parametric rate $\sigma^2 \log n / n$. It is interesting to note here that even though for some θ_i, the optimal rate of estimation of $h_{K^*}(\theta_i)$ is $n^{-2/3}$, the entire set can be estimated at the near parametric rate.

EXAMPLE 4.4 (Half-ball). Suppose $K^* := \{(x_1, x_2) : x_1^2 + x_2^2 \le 1, x_2 \le 0\}$. One then has $h_K(\theta) = 1$ for $-\pi \le \theta \le 0$ and $h_K(\theta) = |\cos \theta|$ for $0 < \theta \le \pi$. Assume n is even and take $i = n/2$ so that $\theta_i = 0$. It can be checked that

$$\Delta_k(0) = \frac{1}{2(k+1)} \sum_{j=0}^{k} \left(1 - \frac{\cos 4\pi j/n}{\cos 2\pi j/n} \right).$$

This is exactly as in (44) with $R = 1$ and an additional factor of $1/2$. Arguing as in Example 4.2, we obtain that

$$\frac{\sigma^2}{k_*(n/2) + 1} \asymp \frac{\sigma^2}{n} + \left(\frac{\sigma^2}{n} \right)^{4/5}.$$

Now take $i = 3n/4$ (assume $n/4$ is an integer) so that $\theta_i = \pi/2$. Observe then that $h_{K^*}(\theta) = |\cos \theta|$ for $\theta = \theta_i \pm 4j\pi/n$ for every $0 \le j \le k$, $k \in \mathcal{I}$. The situation is therefore similar to (46) and we obtain

$$\frac{\sigma^2}{k_*(3n/4) + 1} \asymp \frac{\sigma^2}{n} + \left(\frac{\sigma^2}{n} \right)^{2/3}.$$

Similar to the previous example, the risk for estimating $h_{K^*}(\theta_i)$ changes with i and varies from $n^{-2/3}$ to $n^{-4/5}$. On the other hand, Theorem 3.7 states that the set estimator $\hat{K}$ still estimates K^* at the rate $n^{-4/5}$.

REMARK 4.1 (Connection between risk and smoothness). The reader may observe that the support functions (41) and (43) in the two examples above differ

only by the constant R. It might then seem strange that only the addition of a nonzero constant changes the risk of estimating $h_{K^*}(\theta_i)$ from n^{-1} to $n^{-4/5}$. It turns out that the function (41) is much more smoother than the function (43); the right way to view smoothness of $h_{K^*}(\cdot)$ is to regard it as a function on $\mathbb{R}^2$. This is done in the following way. Define, for each $z = (z_1, z_2) \in \mathbb{R}^2$,

$$h_{K^*}(z) = \max_{(x_1, x_2) \in K^*} (x_1 z_1 + x_2 z_2).$$

When $z = (\cos\theta, \sin\theta)$ for some $\theta \in \mathbb{R}$, this definition coincides with our definition of $h_{K^*}(\theta)$. A standard result (see, e.g., Corollary 1.7.3 and Theorem 1.7.4 in [30]) states that the subdifferential of $z \mapsto h_{K^*}(z)$ exists at every $z = (z_1, z_2) \in \mathbb{R}^2$ and is given by

$$F(K^*, z) := \{(x_1, x_2) \in K^* : h_{K^*}(z) = x_1 z_1 + x_2 z_2\}.$$

In particular, $z \mapsto h_{K^*}(z)$ is differentiable at z if and only if $F(K^*, z)$ is a singleton.

Studying h_{K^*} as a function on $\mathbb{R}^2$ sheds qualitative light on the risk bounds obtained in the examples. In the case of Example 4.1 when $K^* = \{(x_1, x_2)\}$, it is clear that $F(K^*, z) = \{(x_1, x_2)\}$ for all z. Because this set does not change with z, this provides the case of maximum smoothness (because the derivative is constant), and thus we get the n^{-1} rate.

In Example 4.2 when K^* is a ball centered at $x = (x_1, x_2)$ with radius R, it can be checked that $F(K^*, z) = \{x + Rz/\|z\|\}$ for every $z \neq 0$. Since $F(K^*, z)$ is a singleton for each $z \neq 0$, it follows that $z \mapsto h_{K^*}(z)$ is differentiable for every z. For $R \neq 0$, the set $F(K^*, z)$ changes with z and thus here h_{K^*} is not as smooth as in Example 4.1. This explains the slower rate in Example 4.2 compared to 4.1.

Finally, in Example 4.3, when K^* is the vertical segment joining $(0, R)$ and $(0, -R)$, it is easy to see that $F(K^*, z) = K^*$ when $z = (1, 0)$. Here, $F(K^*, z)$ is not a singleton, which implies that $h_{K^*}(z)$ is nondifferentiable at $z = (1, 0)$. This is why one gets the slow rate $n^{-2/3}$ for estimating $h_{K^*}(\theta_{n/2})$ in Example 4.3.

5. Simulations. In this section, we compare the performance of our estimators to other existing estimators for both the pointwise estimation and set estimation problems. We shall refer to our estimator $\hat{h}_i$ [defined in (12)] as the local averaging estimator (*LAE*). The set estimator $\hat{K}$ [defined in (14)] will be referred to as *LAE with projection* and the set estimator $\hat{K}'$ [defined in (16)] will be referred to as *LAE with infinite projection*.

Note that our estimators require knowledge of the noise level σ (which we have assumed to be known for our theoretical analysis). In practice, σ is typically unknown and needs to be estimated. Under the setting of the present paper, σ is easily estimable by using the median of the consecutive differences. Let

$\delta_i = Y_{2i} - Y_{2i-1}, i = 1, \ldots, \lfloor \frac{n}{2} \rfloor$. A simple robust estimator of the noise level σ is the following median absolute deviation (MAD) estimator:

$$(48) \quad \hat{\sigma} = \frac{\text{median}_i |\delta_i - \text{median}_j(\delta_j)|}{\sqrt{2}\Phi^{-1}(0.75)} \approx 1.05 \times \text{median}_i |\delta_i - \text{median}_j(\delta_j)|.$$

We use this estimate of σ in our simulations.

Let us now briefly describe the other estimators to which our estimators will be compared. The first of these is the least squares estimator [29], which we have already described in the paper. The other estimators come from [11], Section 2, where the authors propose four different estimators for K^*. These are: (A) a second-order local linear method; (B) a second-order Nadaraya–Watson kernel method; (C) a third-order local quadratic estimator and (D) a fourth-order Nadaraya–Watson kernel method. As remarked in [11], Section 3, their method (D) is always inferior to (C) [even when the smoothing parameters for (D) were chosen optimally]. Therefore, we only compare our estimators with the first three methods from [11]. We shall denote these estimators by *FHTW-A*, *FHTW-B* and *FHTW-C*. respectively (*FHTW* is an acronym for the author names of [11]). In our simulations, we allow these three estimators to have knowledge of the true noise level σ.

In total therefore, we evaluate the performance of seven estimators in this section: three estimators proposed in this paper (*LAE*, *LAE with projection* and *LAE with infinite projection*), the least squares estimator (*LSE*) and the three estimators from [11] (*FHTW-A*, *FHTW-B* and *FHTW-C*).

In the interest of space, we present simulation results here for only two cases: K^* is (a) the unit ball, and (b) the segment joining $(0, -3)$ to $(0, +3)$. Simulation results for other choices of K^* including square, ellipsoid and random polytope are given in the Supplementary Material [32], Section B.

5.1. *Pointwise estimation.* In this section, we evaluate the performances of the seven pointwise estimators $h_{K^*}(\theta_i)$ for fixed $1 \le i \le n$. We measure the performance of each estimator $\tilde{h}$ by the mean squared error (MSE) $\mathbb{E}_{K^*}(\tilde{h} - h_{K^*}(\theta_i))^2$. For every fixed n, we simulate 200 random ensembles according to the model (2) and then approximate the expectation by the average of error $(\tilde{h} - h_{K^*}(\theta_i))^2$. In simulations, $\sigma = 0.5$ and n ranges over $\{20, 50, 100, 200, 300, 500\}$. We plot the risk as a function of n.

Ball: We start with the case when K^* is a ball. Without loss of generality then, we can assume that the ball is the standard unit ball whose support function always equals one. By rotation invariance of the ball, it is enough to study the case when $\theta_i = 0$. In the following plot, we draw the mean squared errors of all the estimators against the sample size n.

From Figure 1, it is clear that the behaviors of *LSE* and both the *LAE projection* estimators (*LAE with projection* and *LAE with infinite projection*) are almost the
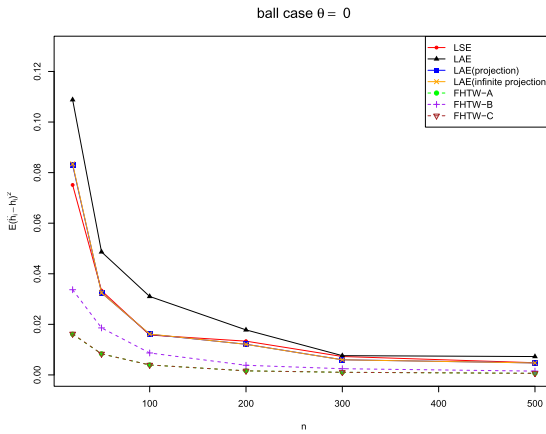
FIG. 1. *Point estimation error when K^* is a ball.*

same, while the performance of *LAE* is quite comparable. When n is large, the performance of *LAE* is as good as that of *LSE* and the *LAE projection* estimators, that is, in this case, projecting the *LAE* onto the support function space is unnecessary. Here, the *LAE*, which only uses local information, is quite similar to that of the *LSE*. Also note that the best performance in this setting is achieved by the three *FHTW* estimators.

Segment: Our second example is when K^* is the segment from $(0, -3)$ to $(0, +3)$ and we study the MSE when θ_i equal to $0, \pi/4, \pi/2$ (in this example, the performance of various estimators will vary with θ_i). The support function of K^* here equals $3|\sin \theta|$ (this function is plotted in the first plot of Figure 2); the three choices of θ_i are indicated in this plot in red. The mean squared errors of all estimators against n are plotted in the last three subplots of Figure 2 for each of the three choices of θ_i.

Observe that similar to the case of the ball, the behaviors of *LSE* and both the *LAE projection* estimators are almost the same. The *LAE* has comparable performance. An interesting fact is that if one looks at the range of the y-axis in the last three subplots of Figure 2, although the mean squared error is decreasing at each θ_i, the rate of decay varies with θ_i. It may be noted that this phenomenon is predicted in our theoretical analysis because the benchmark $R_n(K^*, \theta_i)$ is adaptive to the structure of h_{K^*} at θ_i.

Note that, in this example, the *FHTW* estimators perform poorly unlike the case of the ball. The reason is that in [11], the support function is assumed to be twice differentiable and so is the fitted $\hat{h}$. On the other hand, in this example, the true support function is nondifferentiable, which explains their poor performance. Note that, in contrast, our local averaging estimator requires no assumptions on the local smoothness and as we have seen, the estimator actually adapts to local smoothness.

Analogous plots for other choices of K^* are given in the Supplementary Material [32]. These plots reveal the same story as the previous two settings.
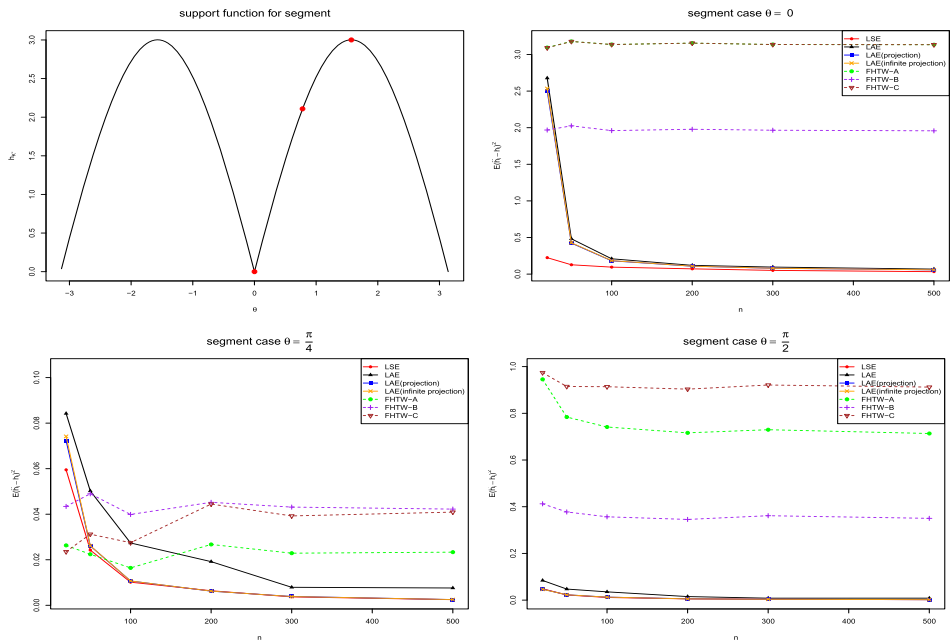
FIG. 2. *Point estimation error when K^* is a segment.*

5.2. *Set estimation.* We now turn to set estimation. Recall that we proposed two estimators for set estimation: the *LAE with projection* estimator $\hat{K}$ [defined in (14)] and the *LAE with infinite projection* estimator $\hat{K}'$ [defined in (16)]. We compare these two estimators to the *LSE* and the *FHTW* estimators from [11]. In our simulations, we found that *FHTW-B* works much better compared to *FHTW-A* and *FHTW-C*, which can also be seen from the simulations for point estimation above. So we only present the results for *FHTW-B* among all the three *FHTW* estimators.

For a set of specific choices of K^* and n, we compute the expected squared errors $\mathbb{E}_{K^*} L_f(\hat{K}, K^*)$ and $\mathbb{E}_{K^*} L(\hat{K}, K^*)$ for each of the estimators, where L_f and L are defined in (3) and (7), respectively. Similar to the point estimation case, these two expectations are approximated by the empirical average of 200 random ensembles according to the model (2). For our *LAE projection* estimators which require the value of σ, we estimate σ via (48). For the *FHTW-B* estimator which also requires σ, we take σ to be its true value.

We plot $\mathbb{E}_{K^*} L_f(\hat{K}, K^*)$ and $\mathbb{E}_{K^*} L(\hat{K}, K^*)$ for each estimator $\hat{K}$ as a function of n. For visualizing the set estimator, we picked an ensemble randomly from the 200 ensembles and plotted each estimator. Note that, for the LAE with infinite projection, as we mentioned before, we take a finer uniform grid of points $\alpha_1, \ldots, \alpha_M$ on $(-\pi, \pi]$ for a large value of M and approximate the set by the intersection of M hyperplanes. In this case, M is set to be 1000.
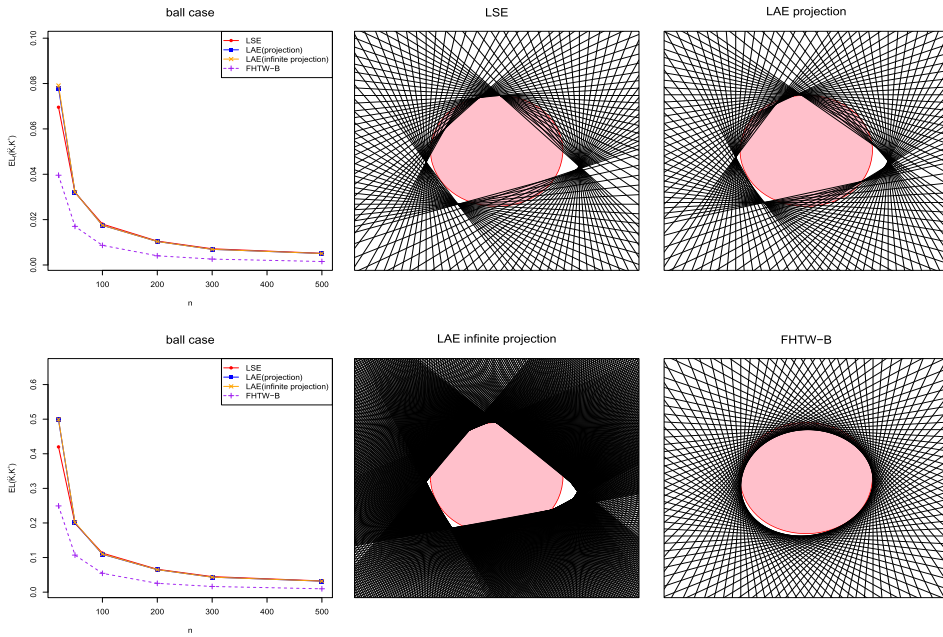
FIG. 3. *Set estimation when K* is a ball.*

Ball: Figure 3 presents the simulation results when K^* is the unit ball. It shows that the performance of the *LAE projection* estimator is almost identical to the that of the *LSE*. The three set estimators *LSE*, *LAE with projection* and *LAE with infinite projection* all look alike in the last subplot. Observe that for the *LAE with infinite projection* estimator, there are many more support lines compared to the *LAE with projection* estimator. This is because of the infinite nature of the projection that is used to define the *LAE with infinite projection* estimator. The best estimator in this example is the *FHTW-B* estimator because it captures the geometry of K^* exactly.

Segment: Our second example takes K^* to be the segment from $(0, -3)$ to $(0, +3)$. The plots are given in Figure 4. Similar to the ball case, our *LAE projection* estimators are comparable to that of the *LSE*. Note that the *FHTW-B* estimator, which assumes smoothness of the support function, becomes quite off (much higher risk) in this case.

From both these figures (as well as other set estimation figures in [32]), it is clear that both our set estimators ($\hat{K}$ and $\hat{K}'$) look quite similar and have near identical performance.

6. Discussions. In this paper, we study the problems of estimating both the support function at a point, $h_{K^*}(\theta_i)$, and the whole convex set K^*. Data-driven adaptive estimators are constructed and their optimality is established. For pointwise estimation, the quantity $k_*(i)$, which appears in both the upper bound (17)
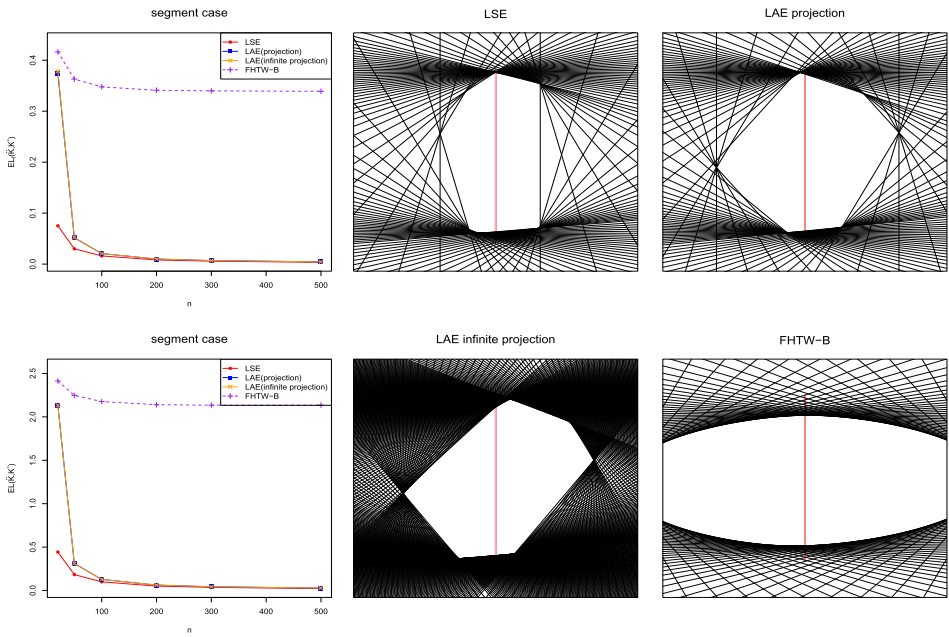
FIG. 4. *Set estimation when K^* is a segment.*

and the lower bound (19), is related to the smoothness of $h_{K^*}(\theta)$ at $\theta = \theta_i$. The construction of $\hat{h}_i$ is based on local smoothing together with an optimization algorithm for choosing the bandwidth. Smoothing methods for estimating the support function have previously been studied by [11]. Specifically, working under certain smoothness assumptions on the true support function $h_{K^*}(\theta)$, [11], estimated it using periodic versions of standard nonparametric regression techniques such as local regression, kernel smoothing and splines. They evade the problem of bandwidth selection however by assuming that the true support function is sufficiently smooth. Our estimator comes with a data-driven method for choosing the bandwidth automatically and we do not need any smoothness assumptions on the true convex set. The fact that our pointwise estimator uses only local information (i.e., for computing $\hat{h}_i$, we only use information on Y_j corresponding to θ_j near θ_i) is quite advantageous in that the computational complexity can be substantially reduced by parallelizing the computation.

It was noted that the construction of our estimators $\hat{K}$ and $\hat{K}'$ given in Section 2.2 does not involve any special treatment for polytopes; yet we obtain faster rates for polytopes. Such automatic adaptation to polytopes has been observed in other contexts: isotonic regression where one gets automatic adaptation for piecewise constant monotone functions (see [9]) and convex regression where one gets automatic adaptation for piecewise affine convex functions (see [22]).

Finally, we note that because $\sigma^2/(k_*(i)+1)$ gives the optimal rate in pointwise estimation, it can potentially be used as a benchmark to evaluate other estimators for $h_{K^*}(\theta_i)$ such as the least squares estimator $h_{\hat{K}_{ls}}(\theta_i)$. From our simulations in Section 5, it seems that the least squares estimator is also optimal in our strong sense for pointwise estimation. It is however difficult to prove accuracy results for the least squares estimator for pointwise estimation. The main difficulty comes from the fact that the least squares estimator is technically a nonlocal estimator [meaning that $h_{\hat{K}_{ls}}(\theta_i)$ can depend on the values of Y_j for θ_j far from θ_i]. This and the other fact that there is no closed-form expression for the least squares estimator makes it very hard to study its pointwise estimation properties. In the related problem of convex function estimation, pointwise properties of the least squares estimator have been studied in [19]. But their results are asymptotic in nature and, more importantly, they make certain smoothness assumptions on the true function. In the generality considered in the present paper, studying the least squares estimator seems difficult; it will probably require new techniques which are beyond the scope of the current paper. This is an interesting topic for future research.

7. Proofs of the main results. This section contains the proofs of the main theorems stated in Section 3. The proofs of the corollaries of Section 3.1 are given in the Supplementary Material [32], Section A.4. Some technical lemmas are required for the proofs given below. These lemmas are also given in the Supplementary Material [32], Section A.6.

Please note that because of space constraints, for the first three proofs given below (those of Theorem 3.1, Theorem 3.2 and Theorem 3.7), we only give a few details here and relegate the complete argument to the Supplementary Material.

7.1. *Proof of Theorem* 3.1. We provide the proof of Theorem 3.1 here. The proof uses three simple lemmas: Lemma A.2, A.3 and A.4, which are stated and proved in the Supplementary Material [32], Section A.6. Due to space constraints, we only provide the initial part of the proof here moving the rest to the Supplementary Material [32], Section A.1.

Fix $i = 1, \ldots, n$. Because $\hat{h}_i = \hat{U}_{\hat{k}(i)}(\theta_i)$, we write

$$(\hat{h}_i - h_{K^*}(\theta_i))^2 = \sum_{k \in \mathcal{I}} (\hat{U}_k(\theta_i) - h_{K^*}(\theta_i))^2 I\{\hat{k}(i) = k\},$$

where $I(\cdot)$ denotes the indicator function. Taking expectations on both sides and using Cauchy–Schwarz inequality, we obtain

$$\mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq \sum_{k \in \mathcal{I}} \sqrt{\mathbb{E}(\hat{U}_k(\theta_i) - h_{K^*}(\theta_i))^4} \sqrt{\mathbb{P}_{K^*}\{\hat{k}(i) = k\}}.$$

The random variable $\hat{U}_k - h_{K^*}(0)$ is normally distributed and we know that $\mathbb{E}Z^4 \leq 3(\mathbb{E}Z^2)^2$ for every Gaussian random variable Z. We therefore have

$$\mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq \sqrt{3} \sum_{k \in \mathcal{I}} \mathbb{E}(\hat{U}_k(\theta_i) - h_{K^*}(\theta_i))^2 \sqrt{\mathbb{P}_{K^*}\{\hat{k}(i) = k\}}.$$

Because $\mathbb{E}_{K^*}\hat{U}_k(\theta_i) = U_k(\theta_i)$ [defined in (9)], we have

$$\mathbb{E}_{K^*}(\hat{U}_k(\theta_i) - h_{K^*}(\theta_i))^2 = (U_k(\theta_i) - h_{K^*}(\theta_i))^2 + \text{var}(\hat{U}_k(\theta_i)).$$

Because $L_k(\theta_i) \leq h_{K^*}(\theta_i) \leq U_k(\theta_i)$, it is clear that $U_k(\theta_i) - h_{K^*}(\theta_i) \leq U_k(\theta) - L_k(\theta_i) = \Delta_k(\theta_i)$. Also, Lemma A.4 states that the variance of $\hat{U}_k$ is at most $\sigma^2/(k+1)$. Putting these together, we obtain

$$\mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2 \leq \sqrt{3} \sum_{k \in \mathcal{I}} \left(\Delta_k^2(\theta_i) + \frac{\sigma^2}{k+1}\right) \sqrt{\mathbb{P}_{K^*}\{\hat{k}(i) = k\}}.$$

The proof of (17) will therefore be complete if we show that

$$(49) \qquad \sum_{k \in \mathcal{I}} \left(\Delta_k^2(\theta_i) + \frac{\sigma^2}{k+1}\right) \sqrt{\mathbb{P}_{K^*}\{\hat{k}(i) = k\}} \leq C \frac{\sigma^2}{k_*(i)+1}$$

for a universal positive constant C. The proof of this inequality is technical and we have moved it to the Supplementary Material [32], Section A.1.

7.2. *Proof of Theorem* 3.2. This subsection is dedicated to the proof of Theorem 3.2. The proof is again long and we have moved most to the Supplementary Material (see [32], Section A.2). The basic idea is presented below and is based on a classical inequality due to [26], which states that for every estimator $\tilde{h}$ and compact, convex set L^*, the quantity

$$\max\left[\mathbb{E}_{K^*}(\tilde{h} - h_{K^*}(\theta_i))^2, \mathbb{E}_{L^*}(\tilde{h} - h_{L^*}(\theta_i))^2\right]$$

is bounded from above by

$$(50) \qquad \geq \frac{1}{4}(h_{K^*}(\theta_i) - h_{L^*}(\theta_i))^2(1 - \|P_{K^*} - P_{L^*}\|_{\text{TV}}).$$

Here, P_{L^*} is the product of the Gaussian probability measures with mean $h_{L^*}(\theta_i)$ and variance σ^2 for $i = 1, \ldots, n$. Also $\|P - Q\|_{\text{TV}}$ denotes the total variation distance between P and Q.

For ease of notation, we assume, without loss of generality, that $\theta_i = 0$. We also write Δ_k for $\Delta_k(\theta_i)$ and k_* for $k_*(i)$.

Suppose first that K^* satisfies the following condition: There exists some $\alpha \in (0, \pi/4)$ such that

$$(51) \qquad \frac{h_{K^*}(\alpha) + h_{K^*}(-\alpha)}{2\cos\alpha} - h_{K^*}(0) > \frac{\sigma}{\sqrt{n_\alpha}},$$

where n_α denotes the number of integers i for which $-\alpha < 2i\pi/n < \alpha$. This condition will not be satisfied, for example, when K^* is a singleton. We shall handle such K^* later. Observe that $n_\alpha \geq 1$ for all $0 < \alpha < \pi/4$ because we can take $i = 0$.

Let us define, for each $\alpha \in (0, \pi/4)$,

$$(52) \qquad a_{K^*}(\alpha) := \left(\frac{h_{K^*}(\alpha) + h_{K^*}(-\alpha)}{2\cos\alpha}, \frac{h_{K^*}(\alpha) - h_{K^*}(-\alpha)}{2\sin\alpha} \right),$$

and let $L^* = L^*(\alpha)$ be defined as the smallest convex set that contains both K^* and the point $a_{K^*}(\alpha)$. In other words, L^* is the convex hull of $K^* \cup \{a_{K^*}(\alpha)\}$.

We now use Le Cam's bound (50) with this choice of L^*. Details are given in [32], Section A.2.

7.3. Proof of Theorem 3.7.

Recall the definition of $\tilde{h}^P$ in (13) and the definition of the estimator $\hat{K}$ in (14). The first thing to note is that

$$(53) \qquad h_{\hat{K}}(\theta_i) = \hat{h}_i^P \qquad \text{for every } i = 1, \ldots, n.$$

To see this, observe first that, because $\hat{h}^P = (\hat{h}_1^P, \ldots, \hat{h}_n^P)$ is a valid support vector, there exists a set $\tilde{K}$ with $h_{\tilde{K}}(\theta_i) = \hat{h}_i^P$ for every i. It is now trivial (from the definition of $\hat{K}$) to see that $\tilde{K} \subseteq \hat{K}$ which implies that $h_{\hat{K}(\theta_i)} \geq h_{\tilde{K}}(\theta_i) = \hat{h}_i^P$. On the other hand, the definition of $\hat{K}$ immediately gives $h_{\hat{K}}(\theta_i) \leq \hat{h}_i^P$.

The observation (53) immediately gives

$$\mathbb{E}_{K^*} L_f(K^*, \hat{K}) = \mathbb{E}_{K^*} \frac{1}{n} \sum_{i=1}^n (h_{K^*}(\theta_i) - \hat{h}_i^P)^2.$$

It will be convenient here to introduce the following notation. Let $h_{K^*}^{\text{vec}}$ denote the vector $(h_{K^*}(\theta_1), \ldots, h_{K^*}(\theta_n))$. Also, for $u, v \in \mathbb{R}^n$, let $\ell(u, v)$ denote the scaled Euclidean distance defined by $\ell^2(u, v) := \sum_{i=1}^n (u_i - v_i)^2/n$. With this notation, we have

$$(54) \qquad \mathbb{E}_{K^*} L_f(K^*, \hat{K}) = \mathbb{E}_{K^*} \ell^2(h_{K^*}^{\text{vec}}, \hat{h}^P).$$

Recall that $\hat{h}^P$ is the projection of $\hat{h} := (\hat{h}_1, \ldots, \hat{h}_n)$ onto $\mathcal{H}$. Because $\mathcal{H}$ is a closed convex subset of $\mathbb{R}^n$, it follows that (see, e.g., [31])

$$\ell^2(h, \hat{h}) \geq \ell^2(\hat{h}, \hat{h}^P) + \ell^2(h, \hat{h}^P) \qquad \text{for every } h \in \mathcal{H}.$$

In particular, with $h = h_{K^*}^{\text{vec}}$, we obtain $\ell^2(h_{K^*}^{\text{vec}}, \hat{h}^P) \leq \ell^2(h_{K^*}^{\text{vec}}, \hat{h})$. Combining this with (54), we obtain

$$(55) \qquad \mathbb{E}_{K^*} L_f(K^*, \hat{K}) \leq \mathbb{E}_{K^*} \ell^2(h_{K^*}^{\text{vec}}, \hat{h}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{K^*} (\hat{h}_i - h_{K^*}(\theta_i))^2.$$

In Theorem 3.1, we proved that

$$\mathbb{E}_{K^*} (\hat{h}_i - h_{K^*}(\theta_i))^2 \leq \frac{C\sigma^2}{k_*(i) + 1} \qquad \text{for every } i = 1, \ldots, n.$$

This implies that

$$\mathbb{E}_{K^*} L_f\left(K^*, \hat{K}\right) \le \frac{C\sigma^2}{n} \sum_{i=1}^{n} \frac{1}{k_*(i)+1}.$$

For inequality (32), it is therefore enough to prove that

(56)
$$\sum_{i=1}^{n} \frac{1}{k_*(i)+1} \le C\left\{1 + \left(\frac{R\sqrt{n}}{\sigma}\right)^{2/5}\right\}.$$

Proving the above inequality is the main part of the proof of Theorem 3.7. Because of space constraints, we have moved this proof to [32], Section A.5. Our proof of (56) is inspired by an argument due to [35], proof of Theorem 2.1, in a very different context.

7.4. *Proof of Theorem* 3.8. Let us start with some notation. For every compact, convex set P and $i = 1, \ldots, n$, let $k_*^P(i)$ denote the quantity $k_*(i)$ with K^* replaced by P. More precisely,

(57)
$$k_*^P(i) := \underset{k \in \mathcal{I}}{\operatorname{argmin}}\left(\Delta_k^P(\theta_i) + \frac{2\sigma}{\sqrt{k+1}}\right),$$

where $\Delta_k^P(\theta_i)$ is defined as in (40) with K^* replaced by P. Lemma A.6 (stated and proved in the Supplementary Material [32], Section A.6) will be used crucially in the proof below. This lemma states that for every $i = 1, \ldots, n$, the risk $\mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2$ can be bounded from above by a combination of $k_*^P(i)$ and how well K^* can be approximated by P. This result holds for every P. The approximation of K^* by P is measured in terms of the Hausdorff distance [defined in (35)].

We are now ready to prove Theorem 3.8. We first use inequality (55) from the proof of Theorem 3.7 which states

$$\mathbb{E}_{K^*} L_f\left(K^*, \hat{K}\right) \le \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{K^*}\left(\hat{h}_i - h_{K^*}(\theta_i)\right)^2.$$

An application of Lemma A.6, specifically inequality (147) for $i = 1, \ldots, n$, now implies the existence of a universal positive constant C such that

$$\mathbb{E}_{K^*} L_f\left(K^*, \hat{K}\right) \le C\left(\frac{\sigma^2}{n} \sum_{i=1}^{n} \frac{1}{k_*^P(i)+1} + \ell_H^2(K^*, P)\right)$$

for every compact, convex set P. By restricting P to be in the class of polytopes, we get

$$\mathbb{E}_{K^*} L_f\left(K^*, \hat{K}\right) \le C \inf_{P \in \mathcal{P}}\left(\frac{\sigma^2}{n} \sum_{i=1}^{n} \frac{1}{k_*^P(i)+1} + \ell_H^2(K^*, P)\right).$$

For the proof of (36), it is therefore enough to show that

$$(58) \qquad \sum_{i=1}^{n} \frac{1}{k_*^P(i) + 1} \leq C v_P \log \frac{en}{v_P} \qquad \text{for every } P \in \mathcal{P},$$

where v_P denotes the number of extreme points of P and C is a universal positive constant. Fix a polytope P with $v_P = k$. Let the extreme points of P be $z_1, \ldots, z_k$. Let $S_1, \ldots, S_k$ denote a partition of $\{\theta_1, \ldots, \theta_n\}$ into k nonempty sets such that for each $j = 1, \ldots, m$, we have

$$h_P(\theta_i) = z_j(1) \cos \theta_i + z_j(2) \sin \theta_i \qquad \text{for all } \theta_i \in S_j,$$

where $z_j = (z_j(1), z_j(2))$. For (58), it is enough to prove that

$$(59) \qquad \sum_{i:\theta_i \in S_j} \frac{1}{k_*^P(i) + 1} \leq C \log(en_j) \qquad \text{for every } j = 1, \ldots, k,$$

where n_j is the cardinality of S_j. This is because we can write

$$\sum_{i=1}^{n} \frac{1}{k_*^P(i) + 1} = \sum_{j=1}^{k} \sum_{i:\theta_i \in S_j} \frac{1}{k_*^P(i) + 1} \leq C \sum_{j=1}^{k} \log(en_j) \leq Ck \log \frac{en}{k},$$

where we used the concavity of $x \mapsto \log(ex)$. We prove (59) below. Fix $1 \leq j \leq k$. The inequality is obvious if S_j is a singleton because $k_*^P(i) \geq 0$. So suppose that $n_j = m \geq 2$. Without loss of generality, assume that $S_j = \{\theta_{u+1}, \ldots, \theta_{u+m}\}$ where $0 \leq u \leq n - m$. The definition of S_j implies that

$$h_P(\theta) = z_j(1) \cos \theta + z_j(2) \sin \theta \qquad \text{for all } \theta \in [\theta_{u+1}, \theta_{u+m}].$$

We can therefore apply inequality (27) to claim the existence of a positive constant c such that

$$k_*^P(i) \geq cn \min(\theta_i - \theta_{u+1}, \theta_{u+m} - \theta_i) \qquad \text{for all } u + 1 \leq i \leq u + m.$$

The minimum with π in (27) is redundant here because $\theta_{u+m} - \theta_{u+1} < 2\pi$. Because $\theta_i = 2\pi i / n - \pi$, we get

$$k_*^P(i) \geq 2\pi c \min(i - u - 1, u + m - i) \qquad \text{for all } u + 1 \leq i \leq u + m.$$

Therefore, there exists a universal constant C such that

$$\sum_{i:\theta_i \in S_j} \frac{1}{k_*^P(i) + 1} \leq C \sum_{i=1}^{m} \frac{1}{1 + \min(i - 1, m - i)} \leq C \sum_{i=1}^{m} \frac{1}{i} \leq C \log(em).$$

This proves (59), thereby completing the proof of Theorem 3.8.

7.5. *Proof of Theorem* 3.9. Recall the definition (16) of the estimator $\hat{K}'$ and that of the interpolating function (15). Following an argument similar to that used at the beginning of the proof of Theorem 3.7, we observe that

$$
\begin{aligned}
\mathbb{E}_{K^*} L\big(K^*, \hat{K}'\big) &\leq \int_{-\pi}^{\pi} \mathbb{E}_{K^*}\big(h_{K^*}(\theta) - \hat{h}'(\theta)\big)^2 \, d\theta \\
&= \sum_{i=1}^{n} \int_{\theta_i}^{\theta_{i+1}} \mathbb{E}_{K^*}\big(h_{K^*}(\theta) - \hat{h}'(\theta)\big)^2 \, d\theta.
\end{aligned}
$$

(60)

Now fix $1 \leq i \leq n$, $\theta_i \leq \theta \leq \theta_{i+1}$ and let $u(\theta) := \mathbb{E}_{K^*}(h_{K^*}(\theta) - \hat{h}'(\theta))^2$. Using the expression (15) for $\hat{h}'(\theta)$, we get that

$$
u(\theta) = \mathbb{E}_{K^*}\left(h_{K^*}(\theta) - \frac{\sin(\theta_{i+1} - \theta)}{\sin(\theta_{i+1} - \theta_i)} \hat{h}_i - \frac{\sin(\theta - \theta_i)}{\sin(\theta_{i+1} - \theta_i)} \hat{h}_{i+1} \right)^2.
$$

We now write $\hat{h}_i = \hat{h}_i - h_{K^*}(\theta_i) + h_{K^*}(\theta_i)$ and a similar expression for $\hat{h}_{i+1}$. The elementary inequality $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ along with $\max(\sin(\theta - \theta_i), \sin(\theta_{i+1} - \theta)) \leq \sin(\theta_{i+1} - \theta_i)$ then imply that

$$
u(\theta) \leq 3\mathbb{E}_{K^*}\big(\hat{h}_i - h_{K^*}(\theta_i)\big)^2 + 3\mathbb{E}_{K^*}\big(\hat{h}_{i+1} - h_{K^*}(\theta_{i+1})\big)^2 + 3b^2(\theta),
$$

where

$$
b(\theta) := h_{K^*}(\theta) - \frac{\sin(\theta_{i+1} - \theta)}{\sin(\theta_{i+1} - \theta_i)} h_{K^*}(\theta_i) - \frac{\sin(\theta - \theta_i)}{\sin(\theta_{i+1} - \theta_i)} h_{K^*}(\theta_{i+1}).
$$

Therefore, from (60) (remember that $|\theta_{i+1} - \theta_i| = 2\pi/n$), we deduce

$$
\mathbb{E}_{K^*} L\big(K^*, \hat{K}'\big) \leq \frac{12\pi}{n} \sum_{i=1}^{n} \mathbb{E}_{K^*}\big(\hat{h}_i - h_{K^*}(\theta_i)\big)^2 + 3 \int_{-\pi}^{\pi} b^2(\theta) \, d\theta.
$$

Now, to bound $\sum_{i=1}^{n} \mathbb{E}_{K^*}(\hat{h}_i - h_{K^*}(\theta_i))^2$ we can simply use the arguments from the proofs of Theorems 3.7 and 3.8. Therefore, to complete the proof of Theorem 3.9, we only need to show that

(61)
$$
\big|b(\theta)\big| \leq \frac{CR}{n} \qquad \text{for every } \theta \in (-\pi, \pi]
$$

for some universal constant C. For this, we use the hypothesis that K^* is contained in a ball of radius R. Suppose that the center of the ball is (x_1, x_2). Define $K' := K^* - \{(x_1, x_2)\} := \{(y_1, y_2) - (x_1, x_2) : (y_1, y_2) \in K^*\}$ and note that $h_{K'}(\theta) = h_{K^*}(\theta) - x_1 \cos\theta - x_2 \sin\theta$. It is then easy to see that $b(\theta)$ is the same for both K^* and K'. It is therefore enough to prove (61) assuming that $(x_1, x_2) = (0, 0)$. In this case, it is straightforward to see that $|h_{K^*}(\theta)| \leq R$ for all θ and also that h_{K^*} is Lipschitz with constant R. Now, because $\max(\sin(\theta - \theta_i), \sin(\theta_{i+1} - \theta)) \leq \sin(\theta_{i+1} - \theta_i)$, it can be checked that $|b(\theta)|$ is bounded from above by

$$
\big|h_{K^*}(\theta)\big|\left| 1 - \frac{\sin(\theta_{i+1} - \theta)}{\sin(\theta_{i+1} - \theta_i)} - \frac{\sin(\theta - \theta_i)}{\sin(\theta_{i+1} - \theta_i)} \right| + \sum_{j=i}^{i+1} \big|h_{K^*}(\theta_j) - h_{K^*}(\theta)\big|.
$$

Because h_{K^*} is R-Lipschitz and bounded by R, it is clear that we only need to show

$$\left| 1 - \frac{\sin(\theta_{i+1} - \theta)}{\sin(\theta_{i+1} - \theta_i)} - \frac{\sin(\theta - \theta_i)}{\sin(\theta_{i+1} - \theta_i)} \right| \leq \frac{C}{n}$$

in order to prove (61). For this, write $\alpha = \theta_{i+1} - \theta$ and $\beta = \theta - \theta_i$ so that the above expression becomes

$$\left| 1 - \frac{\sin \alpha + \sin \beta}{\sin(\alpha + \beta)} \right| \leq |1 - \cos \alpha| + |1 - \cos \beta| \leq \frac{\alpha^2 + \beta^2}{2} \leq \frac{C}{n^2} \leq \frac{C}{n}.$$

This completes the proof of Theorem 3.9.

SUPPLEMENTARY MATERIAL

Supplement to "Adaptive estimation of planar convex sets"
(DOI: 10.1214/17-AOS1576SUPP; .pdf). Technical Appendix. Contains proofs of some results in the main paper as well as additional technical results and simulations.

REFERENCES

[1] ALEXANDROFF, A. D. (1939). Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. *Leningrad State Univ. Annals* [*Uchenye Zapiski*] *Math. Ser.* **6** 3–35. MR0003051

[2] BARAUD, Y. and BIRGÉ, L. (2015). Rates of convergence of rho-estimators for sets of densities satisfying shape constraints. Preprint. Available at arXiv:1503.04427.

[3] BRUNEL, V.-E. (2014). Non-parametric estimation of convex bodies and convex polytopes. Ph.D. thesis, Univ. Pierre et Marie Curie–Paris VI; Univ. of Haifa.

[4] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos.*, *Indiana Univ.*, *Bloomington, Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070

[5] CAI, T. T. and LOW, M. G. (2015). A framework for estimation of convex functions. *Statist. Sinica* **25** 423–456. MR3379081

[6] CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119

[7] CAROLAN, C. and DYKSTRA, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canad. J. Statist.* **27** 557–566. MR1745821

[8] CATOR, E. (2011). Adaptivity and optimality of the monotone least-squares estimator. *Bernoulli* **17** 714–735. MR2787612

[9] 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

[10] DEZA, M. M. and DEZA, E. (2009). *Encyclopedia of Distances*. Springer, Berlin. MR2538177

[11] FISHER, N. I., HALL, P., TURLACH, B. A. and WATSON, G. S. (1997). On the estimation of a convex set from noisy data on its support function. *J. Amer. Statist. Assoc.* **92** 84–91. MR1436100

[12] GARDNER, R. J. (2006). *Geometric Tomography*, 2nd ed. *Encyclopedia of Mathematics and Its Applications* **58**. Cambridge Univ. Press, New York. MR2251886

[13] GARDNER, R. J. and KIDERLEN, M. (2009). A new algorithm for 3D reconstruction from support functions. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **31** 556–562. MR2527431

[14] 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

[15] GREGOR, J. and RANNOU, F. R. (2002). Three-dimensional support function estimation and application for projection magnetic resonance imaging. *Int*. *J*. *Imaging Syst*. *Technol*. **12** 43–50.

[16] GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann*. *Probab*. **11** 1016–1027. MR0714964

[17] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, *Vol*. *II* (*Berkeley*, *Calif*., 1983). 539–555. Wadsworth, Belmont, CA. MR0822052

[18] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators*, *Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293

[19] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann*. *Statist*. **29** 1620–1652. MR1891741

[20] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann*. *Statist*. **29** 1653–1698. MR1891742

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

[22] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab*. *Theory Related Fields* **163** 379–411. MR3405621

[23] HANSON, D. L. and PLEDGER, G. (1976). Consistency in concave regression. *Ann*. *Statist*. **4** 1038–1050. MR0426273

[24] JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann*. *Statist*. **42** 625–653. MR3210981

[25] LELE, A. S., KULKARNI, S. R. and WILLSKY, A. S. (1992). Convex-polygon estimation from support-line measurements and applications to target reconstruction from laser-radar data. *Journal of the Optical Society of America*, *Series A* **9** 1693–1714.

[26] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411

[27] MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann*. *Statist*. **19** 741–759. MR1105842

[28] MCCLURE, D. E. and VITALE, R. A. (1975). Polygonal approximation of plane convex bodies. *J*. *Math*. *Anal*. *Appl*. **51** 326–358. MR0385714

[29] PRINCE, J. L. and WILLSKY, A. S. (1990). Reconstructing convex sets from support line measurements. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **12** 377–389.

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

[31] STARK, H. and YANG, Y. (1998). *Vector Space Projections*. Wiley, New York.

[32] CAI, T. T., GUNTUBOYINA, A. and WEI, Y. (2018). Supplement to "Adaptive estimation of planar convex sets." DOI:10.1214/17-AOS1576SUPP.

[33] VITALE, R. A. (1979). Support functions of plane convex sets. Technical report, Claremont Graduate School, Claremont, CA.

[34] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann*. *Statist*. **9** 443–448. MR0606630

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

T. T. CAI
DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: tcai@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~tcai/

A. GUNTUBOYINA
UNIVERSITY OF CALIFORNIA, BERKELEY
423 EVANS HALL
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: aditya@stat.berkeley.edu

Y. WEI
UNIVERSITY OF CALIFORNIA, BERKELEY
451 EVANS HALL
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: ytwei@berkeley.edu