OPTIMAL RATES OF CONVERGENCE FOR CONVEX SET ESTIMATION FROM SUPPORT FUNCTIONS

BY ADITYANAND GUNTUBOYINA

University of Pennsylvania

We present a minimax optimal solution to the problem of estimating a compact, convex set from finitely many noisy measurements of its support function. The solution is based on appropriate regularizations of the least squares estimator. Both fixed and random designs are considered.

1. Introduction. We study the nonparametric estimation problem of estimating a compact, convex set in Euclidean space from noisy support function measurements. The support function h_K of a compact, convex subset K of $\mathbb{R}^d$ ($d \geq 2$) is defined for u in the unit sphere, $S^{d-1} := \{x \in \mathbb{R}^d : x_1^2 + \cdots + x_d^2 = 1\}$ by

$$h_K(u) := \sup_{x \in K}(x \cdot u) \qquad \text{where } x \cdot u := x_1 u_1 + \cdots + x_d u_d.$$

The function h_K is called the support function of K because it provides information on support hyperplanes and halfspaces of K. Indeed, every support halfspace of K is of the form $\{x : x \cdot u \leq h_K(u)\}$ for some $u \in S^{d-1}$ and since K equals the intersection of all its support halfspaces, the function h_K uniquely determines K. For a proof of this and other elementary properties of the support function, see Schneider [26, Section 1.7] or Rockafellar [25, Section 13].

We consider the problem of estimating an unknown compact, convex set K from observations $(u_1, Y_1), \ldots, (u_n, Y_n)$ drawn according to the model

$$(1) \qquad Y_i = h_K(u_i) + \xi_i \qquad \text{for } i = 1, \ldots, n,$$

where $u_1, \ldots, u_n$ are unit vectors and $\xi_1, \ldots, \xi_n$ are independent normally distributed random variables with mean zero and variance σ^2. We work with both fixed and random design settings for $u_1, \ldots, u_n$.

Direct motivation for a theoretical study of this problem comes from applications. For example, Prince and Willsky [24], who were the first to propose

the regression model (1) for this problem, were motivated by an application to Computed Tomography. Lele, Kulkarni, and Willsky [19] 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 [13] considered applications to Projection Magnetic Resonance Imaging. Another application domain where this problem might plausibly arise is robotic tactile sensing as has been suggested by Prince and Willsky [24]. Under an observation model that is different from the one considered here, Goldenshluger and Zeevi [12] studied the estimation of the support function of the convex support of an unknown intensity function in the context of image analysis.

Additional reasons for analyzing this estimation problem arise from the fact that it has a similar flavour to other well-studied regression problems:

1. It is essentially a nonparametric function estimation problem where the true function is assumed to be the support function of a compact, convex set i.e., there is an implicit convexity-based constraint on the true regression function. Regression and density esimation problems with explicit such constraints e.g., log-concave density estimation and convex regression have received much attention. Some examples of work in this general area include Balabdaoui and Wellner [1], Balabdaoui, Rufibach, and Wellner [2], Cule, Samworth, and Stewart [7], Dümbgen and Rufibach [8], Groeneboom, Jongbloed, and Wellner [14, 15], Mammen [20], Seijo and Sen [27].

2. Our model $Y_i = \max_{x \in K}(x \cdot u_i) + \xi_i$ can also be viewed as a variant of the usual linear regression model where the dependent variable is modeled as the maximum of linear combinations of the explanatory variables over a set of parameter values and where the interest lies in estimating the convex hull of the set of parameters. While we do not know if this maximum regression model has been used outside the context of convex set estimation, the idea of combining linear functions of independent variables into nonlinear algorithmic prediction models for the response variable is familiar (as in neural networks).

Let us now briefly describe the previous work on this problem. The least squares estimator has been the most commonly used. It is defined as

$$(2) \qquad \hat{K}_{ls} := \operatorname*{argmin}_{L} \sum_{i=1}^{n} (Y_i - h_L(u_i))^2 \,,$$

where the minimum is taken over all compact, convex subsets L. The minimizer here is not unique and one can always take it to be a polytope (convex

set with finitely many corners; more carefully defined in the next section). This estimator, for $d = 2$, was first proposed by Prince and Willsky [24], who assumed that $u_1, \ldots, u_n$ are evenly spaced on the unit circle and that the error variables $\xi_1, \ldots, \xi_n$ are normal with zero mean. They also proposed an algorithm for computing it based on quadratic programming. Lele et al. [19] extended this algorithm to include the case of non-evenly spaced two-dimensional $u_1, \ldots, u_n$ as well. Recently, Gardner and Kiderlen [10] proposed an algorithm for computing a minimizer of the least squares criterion for every dimension $d \geq 2$ and every sequence $u_1, \ldots, u_n$.

In addition to the least squares estimator, Prince and Willsky [24] and Lele et al. [19] also proposed estimators (in the case $d = 2$) designed to take advantage of certain forms of prior knowledge, when available, about the true compact, convex set. These estimators are all based on a least squares minimization.

Fisher, Hall, Turlach, and Watson [9] proposed estimators for $d = 2$ that are not based on the least squares criterion. They made smoothness assumptions on the true support function h_K (viewed as a function on the unit circle or on the interval $(-\pi, \pi]$) and estimated it using periodic versions of standard nonparametric regression techniques such as local regression, kernel smoothing and splines. They suggested a way to convert the estimator of h_K into an estimator for K using a formula, which works for smooth h_K, for the boundary of K in terms of h_K. Hall and Turlach [18] added a corner-finding technique to the method of Fisher et al. [9] to estimate two-dimensional convex sets with certain types of corners.

There are relatively fewer theoretical results in the literature. Fisher et al. [9, Theorem 4.1] stated a theorem without proof which appears to imply consistency and certain rates of convergence for their estimator under certain smoothness assumptions on the true support function. Gardner, Kiderlen, and Milanfar [11] proved consistency of the least squares estimator and also derived rates of convergence. They worked with the following assumptions:

1. $u_1, u_2, \ldots$ are deterministic satisfying

$$\max_{u \in S^{d-1}} \min_{1 \leq i \leq n} \mathfrak{d}_d(u, u_i) = O(n^{-1/(d-1)}) \qquad \text{as } n \to \infty,$$

   where $\mathfrak{d}_d$ denotes the usual Euclidean distance on $\mathbb{R}^d$ (see (8)).
2. $\xi_1, \xi_2, \ldots$ are independently distributed according to the normal distribution with mean zero and variance σ^2.
3. K is contained in the ball of radius Γ centered at the origin with $\Gamma \geq \sigma^{15/2}$.

For the loss function ℓ_f^2 defined in (5) below, Gardner et al. [11, Corollary

5.7] showed that $\ell_f^2(K, \hat{K}_{ls}) = O_{d,\sigma,\Gamma}(\beta_n)$ as n approaches ∞ almost surely, where

$$(3) \qquad \beta_n := \begin{cases} n^{-4/(d+3)} & \text{when } d = 2, 3, 4 \\ n^{-1/2} (\log n)^2 & \text{when } d = 5 \\ n^{-2/(d-1)} & \text{when } d \geq 6. \end{cases}$$

Here $O_{d,\sigma,\Gamma}$ is the usual big-O notation where the constant involved depends on d, σ and Γ. Gardner et al. [11, Corollary 5.7] provided explicit expressions for the dependence of the constant with respect to σ and Γ (but not d) which we have suppressed here because our interest only lies in the dependence on n. Also see Gardner et al. [11, Lemma 3.2] for implications of Assumption 1 on the unit vector sequence $\{u_i\}$.

Such a strange dependence of the rates of convergence of the least squares estimation on dimension has also been observed in other situations (see, for example, Birgé and Massart [3], Van de Geer [29], Seregin and Wellner [28]).

Our results in this paper, described below, imply that the rates (3) proved by Gardner et al. [11] for the least squares estimator are optimal when $d \leq 4$ and suboptimal when $d \geq 5$. We show how estimators can be constructed that converge at the rate $n^{-4/(d+3)}$ for all dimensions $d \geq 2$. Our estimators are based on regularizing $\hat{K}_{ls}$ by minimizing the least squares criterion on certain well-chosen subsets of the parameter space. In contrast to Gardner et al. [11], we took the more customary approach in nonparametric statistics by proving rates for the expected loss or risk instead of almost sure bounds for the loss. An advantage is that this results in bounds for a finite (though large) n thereby circumventing the need to let n approach infinity.

We establish an optimal minimax theory for this problem in both fixed-design and random-design frameworks.

In the fixed-design framework, we assume that $u_1, \ldots, u_n$ are deterministic. We define the minimax risk in this setting as (the subscript f below stands for fixed-design):

$$(4) \qquad R_f(n) = R_f(n; \sigma, \Gamma) := \inf_{\hat{K}} \sup_{K \in \mathcal{K}^d(\Gamma)} \mathbb{E}_K \ell_f^2(K, \hat{K})$$

with

$$(5) \qquad \ell_f^2(K, K') := \frac{1}{n} \sum_{i=1}^n \left(h_K(u_i) - h_{K'}(u_i) \right)^2,$$

where $\mathcal{K}^d(\Gamma)$ denotes the set of all compact, convex sets contained in the ball of radius Γ centered at the origin and $\mathbb{E}_K$ denotes expectation taken

when the true compact, convex set equals K. We assume that σ and Γ are known. The infimum in the definition of $R_f(n)$ is over all possible estimators $\hat{K}$ where estimators are defined to be functions of $(u_1, Y_1), \ldots, (u_n, Y_n)$ as well as of σ and Γ taking values in the space of all compact, convex sets.

For every deterministic set of unit vectors $u_1, \ldots, u_n$, we show that $R_f(n)$ is bounded from above by a constant (which is independent of n) multiple of $n^{-4/(d+3)}$. Under a specific assumption on $u_1, \ldots, u_n$, we also prove that a constant multiple of $n^{-4/(d+3)}$ is a lower bound for $R_f(n)$. The upper bound is proved by considering least squares estimators on appropriate subsets of the set of all compact, convex sets in $\mathbb{R}^d$. The lower bound is proved by the application of Assouad's lemma to a special finite collection of convex sets.

We also study the random-design setting where we assume that $u_1, \ldots, u_n$ are independently distributed according to a fixed probability measure, ν, on S^{d-1} and that the errors $\xi_1, \ldots, \xi_n$ are independent of $u_1, \ldots, u_n$. Here we define the minimax risk as (the subscript r below stands for random-design)

$$(6) \qquad R_r(n) = R_n(n; \sigma, \Gamma) := \inf_{\hat{K}} \sup_{K \in \mathcal{K}^d(\Gamma)} \mathbb{E}_K \ell_r^2(K, \hat{K})$$

with

$$(7) \qquad \ell_r^2(K, K') := \int_{S^{d-1}} (h_K(u) - h_{K'}(u))^2 \, d\nu(u).$$

For every probability measure ν on S^{d-1}, we show that $R_r(n)$ is bounded from above by a constant (which is independent on n) multiple of $n^{-4/(d+3)}$. The proof techniques here are similar to the fixed-design setting. When ν equals ν_{unif}, the uniform probability measure on S^{d-1}, we prove that a constant multiple of $n^{-4/(d+3)}$ is also a lower bound for $R_r(n)$. We use a different lower bound proof here from the one used in the fixed-design setting.

We would like to remark here that the rate $n^{-4/(d+3)}$ is quite natural in connection to the minimax estimation of smooth functions. Indeed, the unit sphere has dimension $d - 1$, and the class of smooth functions on a space of dimension $d - 1$ with smoothness γ allows the minimax rate $n^{-2\gamma/(2\gamma+d-1)}$. Our problem here has a convexity constraint and convexity is associated, in a broad sense, with the smoothness $\gamma = 2$, which explains the rate $n^{-4/(d+3)}$.

After setting up the necessary notation in the next section, we prove the fixed-design bounds in Section 3 and the random-design bounds in Section 4. Some auxiliary results that are needed for the proofs of the main theorems are collected in the three Appendices.

2. Notation. This section will fix notation and introduce some standard notions that are used in the paper.

$\mathbb{P}_K$ denotes the probability distribution of the observations when the true compact, convex set equals K. In other words, $\mathbb{P}_K$ is the joint distribution of $(Y_1, \ldots, Y_n)$ in the fixed-design setting and the joint distribution of $(u_1, Y_1), \ldots, (u_n, Y_n)$ in the random-design setting. We use the same notation in both cases as the setting will be clear from the context. Expectation under $\mathbb{P}_K$ is denoted by $\mathbb{E}_K$.

For a real-valued function f on $S^{d-1} \times \cdots \times S^{d-1}$, let $\mathbb{E}_\nu f(u_1, \ldots, u_n)$ denotes expectation taken under the assumption that $u_1, \ldots, u_n$ are independently distributed according to ν.

We denote the usual Euclidean distance on $\mathbb{R}^d$ by $\mathfrak{d}_d$ i.e., for $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ in $\mathbb{R}^d$,

$$(8) \qquad \mathfrak{d}_d(x, y) := \left(\sum_{i=1}^d (x_i - y_i)^2 \right)^{1/2}.$$

The closed ball in $\mathbb{R}^d$ of radius $\alpha > 0$ centered at $a \in \mathbb{R}^d$ is denoted by $B_d(a, \alpha)$ i.e., $B_d(a, \alpha) := \{x \in \mathbb{R}^d : \mathfrak{d}_d(x, a) \leq \alpha\}$.

The uniform probability measure on S^{d-1} is denoted by ν_{unif}.

The convex hull of a subset S of $\mathbb{R}^d$, defined as the intersection of all convex sets containing S, is denoted by $\text{conv}(S)$. Convex hulls of finite subsets of $\mathbb{R}^d$ are called polytopes. The set of all polytopes in $\mathbb{R}^d$ with at most m extreme points (corners) is denoted by $\mathcal{P}_m$, i.e.,

$$\mathcal{P}_m := \Big\{ \text{conv}(S) : S \subseteq \mathbb{R}^d \text{ with cardinality at most } m \Big\}.$$

For $\Gamma > 0$, let $\mathcal{P}_m(\Gamma)$ denote the set of all polytopes in $\mathcal{P}_m$ that are contained in $B_d(0, \Gamma)$:

$$(9) \qquad \mathcal{P}_m(\Gamma) := \{K \in \mathcal{P}_m : K \subseteq B_d(0, \Gamma)\}.$$

For two compact, convex subsets K and K' of $\mathbb{R}^d$, the Hausdorff distance between them is defined as

$$(10) \qquad \ell_H(K, K') := \sup_{u \in S^{d-1}} |h_K(u) - h_{K'}(u)|.$$

It is apparent that both $\ell_f^2(K, K')$ and $\ell_r^2(K, K')$ are less than or equal to $\ell_H^2(K, K')$. The Hausdorff distance has the following alternative expression:

$$(11) \qquad \ell_H(K, K') = \max \left(\sup_{x \in K} \inf_{y \in K'} \mathfrak{d}_d(x, y), \sup_{x \in K'} \inf_{y \in K} \mathfrak{d}_d(x, y) \right)$$

A simple proof of the equivalence of (10) and (11) can be found in Schneider [26, Theorem 1.8.11].

The standard notions of packing and covering numbers will be frequently used and we have collected their definitions below for the convenience of the reader. Let Θ be an arbitrary set and let ρ be a nonnegative function on $\Theta \times \Theta$ (ρ does not necessarily have to be a metric).

1. **Packing**: By an η-packing subset of (Θ, ρ), we mean a subset $S \subseteq \Theta$ for which $\rho(\theta, \theta') \geq \eta$ for all $\theta, \theta' \in S$ with $\theta \neq \theta'$. By a maximal η-packing subset, we mean an η-packing subset that is not a proper subset of any other η-packing subset. The packing number $N(\Theta, \eta; \rho)$ is defined as the maximum of the cardinalities of all η-packing subsets of Θ.

2. **Covering**: By an ϵ-covering subset of (Θ, ρ), we mean a subset $S \subseteq \Theta$ such that $\min_{s \in S} \rho(t, s) \leq \epsilon$ for every $t \in \Theta$. The ϵ-covering number $M(\Theta, \epsilon; \rho)$ is defined as the minimum of the cardinalities of all ϵ-covering subsets of Θ.

The cardinality of a finite set F is denoted by $|F|$.

We use the following notions of distance between probability measures P and Q having densities p and q with respect to a common measure μ:

1. Total variation distance: $||P - Q||_{TV} := \int (|p - q|/2) d\mu$.
2. Kullback-Leibler divergence: $D(P||Q) := \int p \log(p/q) d\mu$ if P is absolutely continuous with respect to Q and ∞ otherwise.
3. Chi-squared divergence: $\chi^2(P||Q) := \int (p^2/q) d\mu - 1$ if P is absolutely continuous with respect to Q and ∞ otherwise. We also write $\chi(P, Q) := \left(\chi^2(P||Q) \right)^{1/2}$.

Pinsker's inequality states

$$(12) \quad D(P||Q) \geq 2||P - Q||_{TV}^2 \qquad \text{for all probability measures } P \text{ and } Q.$$

We use the symbols c, C etc. to denote positive constants depending only on the dimension d. Their value may change with every occurence.

3. Fixed Design Setting. In this section, we assume that $u_1, \ldots, u_n$ are deterministic. The errors $\xi_1, \ldots, \xi_n$ are assumed to be independently distributed according to the normal distribution with mean zero and variance σ^2. We consider the loss function ℓ_f^2 (defined in (5)) and prove upper and lower bounds for the corresponding minimax risk $R_f(n)$ over $\mathcal{K}^d(\Gamma)$ (see the definition (4)).

3.1. *Upper bound for $R_f(n)$.* The following result shows that $R_f(n)$ is at most $n^{-4/(d+3)}$ up to a multiplicative constant that is independent of n. We make no assumptions on the deterministic design unit vectors $u_1, \ldots, u_n$ and they are completely arbitrary. It should be noted that the loss function ℓ_f^2 is naturally associated with this fixed design setup enabling the following theorem to hold with no assumptions whatsoever on $u_1, \ldots, u_n$. On the other hand, such assumptions would be unavoidable if one is interested in proving risk bounds for other loss functions under the fixed design setting. A similar remark also applies to Theorem 4.1 where the natural loss function is ℓ_r^2.

THEOREM 3.1. *There exist positive constants c and C depending only on the dimension d such that*

$$(13) \qquad R_f(n) \leq c \; \sigma^{8/(d+3)} \Gamma^{2(d-1)/(d+3)} n^{-4/(d+3)} \qquad \text{if } n \geq C(\sigma/\Gamma)^2.$$

PROOF. For each finite subset F of $\mathcal{K}^d(\Gamma)$, let us define the least squares estimator $\hat{K}_F$ by

$$(14) \qquad \hat{K}_F := \operatorname*{argmin}_{L \in F} \sum_{i=1}^n (Y_i - h_L(u_i))^2 \,.$$

We show that, if F is chosen appropriately, then $\mathbb{E}_K \ell_f^2(K, \hat{K}_F)$ is bounded from above by the right hand side of (13) for every $K \in \mathcal{K}^d(\Gamma)$.

Fix $K \in \mathcal{K}^d(\Gamma)$. We start with the following trivial inequality which holds for every nonnegative function G on F and every $\alpha > 0$:

$$G(\hat{K}_F) \leq \sum_{L \in F} G(L) \exp \left(\alpha \sum_i (Y_i - h_{\hat{K}_F}(u_i))^2 - \alpha \sum_i (Y_i - h_L(u_i))^2 \right),$$

the reason being that the term for $L = \hat{K}_F$ in the sum on the right hand side equals $G(\hat{K}_F)$.

Because $\hat{K}_F$ is the least squares estimator over F, we can replace it in the right hand side above by an arbitrary $L' \in F$. Taking expectation on both sides of the resulting inequality, we obtain

$$(15) \qquad \mathbb{E}_K G(\hat{K}_F) \leq \sum_{L \in F} G(L) \mathbb{E}_K e^{\alpha \sum_i (Y_i - h_{L'}(u_i))^2 - \alpha \sum_i (Y_i - h_L(u_i))^2},$$

for every $L' \in F$. The expectation term in the right hand side equals $\exp \left(-\alpha n \ell_f^2(K, L) + \alpha n \ell_f^2(K, L') + 2\alpha^2 \sigma^2 n \ell_f^2(L, L') \right)$. We then use the elementary inequality $\ell_f^2(L, L') \leq 2\ell_f^2(K, L) + 2\ell_f^2(K, L')$ and the fact that

$\ell_f^2(K, L') \leq \ell_H^2(K, L')$ to obtain the following upper bound for $\mathbb{E}_K G(\hat{K}_F)$:

$$\min_{L' \in F} \sum_{L \in F} G(L) \exp\left((-\alpha + 4\alpha^2\sigma^2)n\ell_f^2(K, L) + (\alpha + 4\alpha^2\sigma^2)n\ell_H^2(K, L')\right).$$

The choices
$$G(L) = \exp\left(\frac{n\ell_f^2(K, L)}{16\sigma^2}\right) \text{ and } \alpha = \frac{1}{8\sigma^2}$$

lead to the following risk bound:

$$(16) \qquad \mathbb{E}_K \exp\left(\frac{n\ell_f^2(K, \hat{K}_F)}{16\sigma^2}\right) \leq |F| \exp\left(\frac{3n}{16\sigma^2} \min_{L' \in F} \ell_H^2(K, L')\right)$$

where $|F|$ denotes the cardinality of F. Using Jensen's inequality on the left hand side and taking logarithms on both sides, we deduce

$$\mathbb{E}_K \ell_f^2(K, \hat{K}_F) \leq \frac{16\sigma^2}{n} \log |F| + 3 \min_{L' \in F} \ell_H^2(K, L').$$

Since $K \in \mathcal{K}^d(\Gamma)$ was arbitrary, we get

$$(17) \qquad R_f(n) \leq \frac{16\sigma^2}{n} \log |F| + 3 \sup_{K \in \mathcal{K}^d(\Gamma)} \min_{L' \in F} \ell_H^2(K, L').$$

We now use a classical result on the covering numbers of $(\mathcal{K}^d(\Gamma), \ell_H)$ due to Bronshtein [4, Theorem 3 and Remark 1] which states that there exist postive constants c and ϵ_0 which depend on d alone such that for every $\epsilon \leq \Gamma\epsilon_0$, there exists a finite subset $F \subseteq \mathcal{K}^d(\Gamma)$ satisfying

$$(18) \qquad \log |F| \leq c \left(\frac{\Gamma}{\epsilon}\right)^{(d-1)/2} \text{ and } \sup_{K \in K^d(\Gamma)} \min_{L' \in F} \ell_H^2(K, L') \leq \epsilon^2.$$

Combining (17) and (18), we get

$$R_f(n) \leq \left[16c\frac{\sigma^2}{n}\left(\frac{\Gamma}{\epsilon}\right)^{(d-1)/2} + 3\epsilon^2\right] \qquad \text{for every } 0 < \epsilon \leq \Gamma\epsilon_0$$

If we now choose $\epsilon := \sigma^{4/(d+3)}\Gamma^{(d-1)/(d+3)}n^{-2/(d+3)}$, then $\epsilon \leq \Gamma\epsilon_0$ provided $n \geq C(\sigma/\Gamma)^2$ for a large enough constant C depending only on d and the required inequality (13) follows. $\qquad \square$

3.2. *Lower bound for $R_f(n)$.* We show that $n^{-4/(d+3)}$ is also a lower bound for $R_f(n)$ up to a multiplicative constant that is independent of n. We make the assumption that the fixed unit vectors $u_1, \ldots, u_n$ form a maximal ϵ-packing subset (under the Euclidean metric $\mathfrak{d}_d$) of S^{d-1} for some $\epsilon \in (0, 1]$. The definition of a maximal packing set was given in Section 2. Note that it is impossible to proved the lower bound $n^{-4/(d+3)}$ for $R_f(n)$ without any assumptions on $u_1, \ldots, u_n$. For example, if $u_1 = \cdots = u_n$, then $R_f(n)$ is of the order $1/n$.

A standard argument (sketched in the appendix, see inequality (41)) shows that our assumption on $u_1, \ldots, u_n$ implies that

$$(19) \qquad\qquad c\epsilon^{1-d} \leq n \leq C\epsilon^{1-d}$$

for two constants c and C depending only on d. The following is the main theorem of this section.

THEOREM 3.2. *Suppose $u_1, \ldots, u_n$ form a maximal ϵ-packing subset of S^{d-1} for some $\epsilon \in (0, 1]$. There exist positive constants c and C depending only on d such that*

$$(20) \qquad\qquad R_f(n) \geq c\ \sigma^{8/(d+3)}\Gamma^{2(d-1)/(d+3)}n^{-4/(d+3)}$$

whenever $n \geq C \max\left((\sigma/\Gamma)^2, (\Gamma/\sigma)^{(d-1)/2}\right)$.

Our proof is based on the application of Assouad's lemma to an explicitly constructed finite subset of $\mathcal{K}^d(\Gamma)$. The following version of Assouad's lemma is taken from van der Vaart [30, Page 347]. Recall that $\mathbb{P}_K$ denotes the probability distribution of the observations when the true compact, convex set is K.

LEMMA 3.3 (Assouad). *Let m be a positive integer and suppose that, for each $\tau \in \{0, 1\}^m$, there is an associated set $K(\tau)$ in $\mathcal{K}^d(\Gamma)$. Then the following inequality holds:*

$$R_f(n) \geq \frac{m}{8} \min_{\tau \neq \tau'} \frac{\ell_f^2(K(\tau), K(\tau'))}{\Upsilon(\tau, \tau')} \min_{\Upsilon(\tau, \tau')=1} \left(1 - ||\mathbb{P}_{K(\tau)} - \mathbb{P}_{K(\tau')}||_{TV}\right),$$

where $\Upsilon(\tau, \tau') := \sum_i \{\tau_i \neq \tau_i'\}$.

PROOF OF THEOREM 3.2. We apply Assouad's lemma to the following construction. For a fixed positive $\eta \leq 1/8$, we consider unit vectors $v_1, \ldots, v_m$

that form a maximal $2\sqrt{2\eta}$-packing subset of the unit sphere (under $\mathfrak{d}_d$) and define

$$K(\tau) := D_1(\tau_1) \cap \cdots \cap D_m(\tau_m) \qquad \text{for } \tau \in \{0,1\}^m,$$

where

$$D_i(0) := B_d(0,\Gamma) \cap \{x : x \cdot v_j \leq \Gamma(1-\eta)\} \quad \text{and} \quad D_i(1) := B_d(0,\Gamma).$$

One consequence of the assumption on $v_1, \ldots, v_m$ is that $m \geq c\eta^{(1-d)/2}$ for a constant c. Another consequence is that the sets $B_d(0,\Gamma) \cap \{x \cdot v_j > 1-\eta\}$ are disjoint which implies that

$$\ell_f^2(K(\tau), K(\tau')) = \sum_{i:\tau_i \neq \tau_i'} \ell_f^2(D_i(0), D_i(1)) = \Upsilon(\tau,\tau')\ell_f^2(D_1(0), D_1(1))$$

for every $\tau, \tau' \in \{0,1\}^m$. In Lemma 5.1 (stated and proved in Appendix I), we show that there exist constants c and C such that

$$c\Gamma^2\eta^{(d+3)/2} \leq \ell_f^2(D_1(0), D_1(1)) \leq C\Gamma^2\eta^{(d+3)/2}$$

provided $0 < \eta \leq 1/8$ and $\eta \geq C\epsilon^2$. Therefore,

$$(21) \quad \min_{\tau \neq \tau'} \frac{\ell_f^2(K(\tau), K(\tau'))}{\Upsilon(\tau,\tau')} \geq c\Gamma^2\eta^{(d+3)/2} \qquad \text{if } 0 < \eta \leq 1/8 \text{ and } \eta \geq C\epsilon^2.$$

To bound $||\mathbb{P}_{K(\tau)} - \mathbb{P}_{K(\tau')}||_{TV}$, we use Pinsker's inequality (12) because the Kullback-Leibler divergence $D(P_{K(\tau)}||P_{K(\tau')})$ has a simple expression in terms of $\ell_f^2(K(\tau), K(\tau'))$:

$$\begin{aligned}||\mathbb{P}_{K(\tau)} - \mathbb{P}_{K(\tau')}||_{TV}^2 &\leq \frac{1}{2}D(\mathbb{P}_{K(\tau)}||\mathbb{P}_{K(\tau')}) \\ &= \frac{n}{4\sigma^2}\ell_f^2(K(\tau), K(\tau')) = \frac{n}{4\sigma^2}\Upsilon(\tau,\tau')\ell_f^2(D_1(0), D_1(1)).\end{aligned}$$

By a second application of Lemma 5.1, we obtain

$$(22) \qquad \min_{\Upsilon(\tau,\tau')=1} \left(1 - ||\mathbb{P}_{K(\tau)} - \mathbb{P}_{K(\tau')}||_{TV}\right) \geq 1 - C\frac{\Gamma\sqrt{n}}{\sigma}\eta^{(d+3)/4}$$

if $0 < \eta \leq 1/8$ and $\eta \geq C\epsilon^2$. Therefore, applying Assouad's Lemma with the inequalities (21), (22) and $m \geq c\eta^{-(d-1)/2}$, we obtain

$$(23) \qquad\qquad R_f(n) \geq c\Gamma^2\eta^2 \left(1 - C\frac{\Gamma\sqrt{n}}{\sigma}\eta^{(d+3)/4}\right)$$

if $0 < \eta \le 1/8$ and $\eta \ge C\epsilon^2$. We now make the choice

$$\eta := \left(\frac{1}{2C} \frac{\sigma}{\Gamma \sqrt{n}} \right)^{4/(d+3)}.$$

Then $\eta \le 1/8$ provided $n \ge C(\sigma/\Gamma)^2$. Also, from (19), $\epsilon \le Cn^{-1/(d-1)}$ and thus, for $\eta \ge C\epsilon^2$, it is enough to ensure that $\eta \ge Cn^{-2/(d-1)}$ which, upon simplification, reduces to $n \ge C(\Gamma/\sigma)^{(d-1)/2}$. Inequality (23) with this choice of η then implies (20) which completes the proof. $\square$

3.3. *A more natural estimator.* In Theorem 3.1, we used the least squares estimator on an appropriate finite subset of $\mathcal{K}^d(\Gamma)$ to prove the optimal upper bound for $R_f(n)$. This estimator can be viewed as a regularized version of the full least squares estimator $\hat{K}_{ls} := \operatorname{argmin}_L \sum_i (Y_i - h_L(u_i))^2$ for which the rates (see (3)) proved by Gardner et al. [11] are suboptimal for $d \ge 5$.

We remarked (just after (2)) that, for the full least squares estimator, the set L which minimizes $\sum_i (Y_i - h_L(u_i))^2$ is not unique. Gardner and Kiderlen [10] observed that a minimizer can always be chosen to be a polytope with at most n extreme points and provided an algorithm for computing such a minimizer. In light of this observation of Gardner and Kiderlen [10], we consider the following estimator which is a more intuitive regularization of $\hat{K}_{ls}$ compared to $\hat{K}_F$:

$$(24) \qquad \breve{K}_m := \operatorname*{argmin}_{L \in \mathcal{P}_m(\Gamma)} \sum_{i=1}^{n} (Y_i - h_L(u_i))^2.$$

The set $\mathcal{P}_m(\Gamma)$ was defined in (9). The best risk achievable by $\breve{K}_m$ is defined as

$$\breve{R}_f(n) := \inf_{m \ge 1} \sup_{K \in \mathcal{K}^d(\Gamma)} \mathbb{E}_K \ell_f^2(K, \breve{K}_m).$$

It is not too hard to see that $\breve{K}_m$ equals the least squares estimator $\hat{K}_{ls}$ whenever $m \ge n$. On the other hand, for $m < n$, they can be quite different.

In this section, we prove the following theorem which shows that $\breve{R}_f(n)$ is bounded from above by $n^{-4/(d+3)}$ up to a multiplicative factor that is logarithmic in n. No assumptions on $u_1, \ldots, u_n$ are necessary.

THEOREM 3.4. *There exist positive constants c and C that depend only on the dimension d such that*

$$(25) \qquad \breve{R}_f(n) \le c\, \sigma^{8/(d+3)} \Gamma^{2(d-1)/(d+3)} n^{-4/(d+3)} \log\left(cn\Gamma^2/\sigma^2 \right) \log(cn)$$

if $n \ge C(\sigma/\Gamma)^2$.

For the proof of this theorem, we use the following result which is a special instance of a result on convergence rates of sieved least squares estimators from Van de Geer [29, Pages 184-185]. For a polytope $P \in \mathcal{P}_m$ and $\omega > 0$, let

$$S_m(P, \omega) := \left\{ L \in \mathcal{P}_m : \ell_f^2(P, L) \leq \omega^2 \right\}$$

and let $M(S_m(P, \omega), \epsilon; \ell_f)$ denote the ϵ-covering number of $S_m(P, \omega)$ under ℓ_f.

THEOREM 3.5. *Fix a polytope $P \in \mathcal{P}_m(\Gamma)$. Suppose Ψ is a function on $(0, \infty)$ such that*

$$\Psi(\omega) \geq \int_0^\omega \sqrt{\log M(S_m(P, \omega), \epsilon; \ell_f)} d\epsilon \qquad \text{for every } \omega > 0$$

and such that $\Psi(\omega)/\omega^2$ is decreasing on $(0, \infty)$. Then there exists a universal constant C such that

$$(26) \qquad \mathbb{P}_K \left(\ell_f^2(\check{K}_m, P) > \delta \right) \leq C \sum_{s \geq 0} \exp \left(\frac{-n 2^{2s} \delta}{C^2 \sigma^2} \right)$$

for every δ satisfying $\delta \geq 8 \ell_f^2(K, P)$ and $\sqrt{n} \delta \geq C \sigma \Psi \left(\sqrt{\delta} \right)$.

The application of this theorem for the proof of (25) requires an upper bound on $M(S_m(P, \omega), \epsilon; \ell_f)$. Such a bound is provided in Lemma 6.1 (stated and proved in Appendix II).

PROOF OF THEOREM 3.4. Fix $m \geq 1$ and an arbitrary polytope $P \in \mathcal{P}_m(\Gamma)$. In Lemma 6.1, we show that

$$M(S_m(P, \omega), \epsilon; \ell_f) \leq \left(4 + \frac{2\sqrt{n}\omega}{\epsilon} \right)^{b_1 m d \log(b_2 m)}$$

for universal positive constants b_1 and b_2. This implies that

$$\int_0^\omega \sqrt{\log M(S_m(P, \omega), \epsilon; \ell_f)} d\epsilon \leq \sqrt{b_1 m d \log(b_2 m)} \int_0^\omega \sqrt{\log \left(4 + \frac{2\sqrt{n}\omega}{\epsilon} \right)} d\epsilon$$

$$= \omega \sqrt{b_1 m d \log(b_2 m)} \int_1^\infty \frac{\sqrt{\log(4 + 2\sqrt{n}x)}}{x^2} dx$$

$$\leq C \omega \sqrt{m d \log(b_2 m) \log(4 + 2\sqrt{n})}.$$

As a result, the function $\Psi(\omega)$ appearing in Theorem 3.5 can be taken to be

$$\Psi(\omega) := C\omega\sqrt{md\log(b_2 m)\log(4 + 2\sqrt{n})} \qquad \text{for every } \omega > 0$$

and then (26) gives

$$(27) \quad \mathbb{P}_K\left(\ell_f^2(\breve{K}_m, P) > \delta\right) \le C\sum_{s \ge 0}\exp\left(\frac{-n2^{2s}\delta}{C^2\sigma^2}\right) \qquad \text{whenever } \delta \ge \delta_0,$$

where

$$\delta_0 := C\left(\ell_f^2(K, P) + \frac{\sigma^2}{n}md\log(b_2 m)\log(4 + 2\sqrt{n})\right).$$

Because $n\delta_0 \ge C\sigma^2$ for a constant C, the sum on the right hand side of (27) can be bounded from above by a constant multiple of the first term, and we deduce

$$\mathbb{P}_K\left(\ell_f^2(\breve{K}_m, P) > \delta\right) \le C\exp\left(\frac{-n\delta}{C^2\sigma^2}\right) \qquad \text{whenever } \delta \ge \delta_0.$$

Integrating both sides of the above inequality with respect to $\delta \in [\delta_0, \infty)$, we get

$$\mathbb{E}_K\left(\ell_f^2(\breve{K}_m, P) - \delta_0\right)_+ \le \frac{C\sigma^2}{n}\exp\left(\frac{-n\delta_0}{C^2\sigma^2}\right) \le \frac{C\sigma^2}{n},$$

where $x_+ := \max(x, 0)$. Because $\sigma^2/n \le C\delta_0$, we get the expectation bound $\mathbb{E}_K\ell_f^2(\breve{K}_m, P) \le C\delta_0$. The elementary inequality $\ell_f^2(\breve{K}_m, K) \le 2\ell_f^2(\breve{K}_m, P) + 2\ell_f^2(K, P)$ and the fact that $\delta_0 \ge \ell_f^2(K, P)$ yield the following risk bound:

$$(28) \qquad \mathbb{E}_K\ell_f^2(\breve{K}_m, K) \le C\left(\ell_f^2(K, P) + \frac{\sigma^2}{n}md\log(b_2 m)\log(4 + 2\sqrt{n})\right).$$

Since $K \in \mathcal{K}^d(\Gamma)$ and $P \in \mathcal{P}_m(\Gamma)$ were arbitrary in the above analysis, we have proved the following bound for $\breve{R}_f(n)$:

$$\breve{R}_f(n) \le C\inf_{m \ge 1}\left[\sup_{K \in \mathcal{K}^d(\Gamma)}\inf_{P \in \mathcal{P}_m(\Gamma)}\ell_H^2(K, P) + \frac{\sigma^2}{n}md\log(b_2 m)\log(4 + 2\sqrt{n})\right],$$

where we have also used $\ell_f^2(K, P) \le \ell_H^2(K, P)$.

Bronshteyn and Ivanov [6] (see also Bronshtein [5]) proved that there exist positive constants C_1 and C_2 depending only on the dimension d such that

$$(29) \quad \sup_{K \in \mathcal{K}^d(\Gamma)}\inf_{P \in \mathcal{P}_m(\Gamma)}\ell_H^2(K, P) \le C_1\Gamma^2 m^{-4/(d-1)} \qquad \text{whenever } m \ge C_2.$$

From this result, we have

$$\breve{R}_f(n) \leq C \inf_{m \geq C_2} \left[\Gamma^2 m^{-4/(d-1)} + \frac{\sigma^2}{n} m d \log(b_2 m) \log(4 + 2\sqrt{n}) \right].$$

If we now choose $m := \sigma^{-2(d-1)/(d+3)} \Gamma^{2(d-1)/(d+3)} n^{(d-1)/(d+3)}$, then $m \geq C_2$ provided $n \geq C(\sigma/\Gamma)^2$ for a large enough constant C depending only on d and the required inequality (25) follows.   □

REMARK 3.1.   *From the above proof, it can be seen that Theorem 3.4 also holds for the following estimator:*

$$\tilde{K}_m := \operatorname*{argmin}_{L \in \mathcal{P}_m} \sum_{i=1}^{n} (Y_i - h_L(u_i))^2.$$

The only difference between $\tilde{K}_m$ and $\breve{K}_m$ is that the argmin in the definition of $\tilde{K}_m$ is taken over all polytopes in $\mathcal{P}_m$ while that in the definition of $\breve{K}_m$ is only over those polytopes in $\mathcal{P}_m$ that are contained in the ball of radius Γ centered at the origin. Theorem 3.4 also holds for $\tilde{K}_m$ because we have not used this boundedness property of sets in $\mathcal{P}_m(\Gamma)$ in our covering number calculations in Lemma 6.1.

REMARK 3.2.   *The above proof also reveals that the value of m for which $\breve{K}_m$ achieves the optimal rate up to logarithmic terms is of the order $n^{(d-1)/(d+3)}$. Since this is much smaller than n, the estimator $\breve{K}_m$ for this m is quite different from the full least squares estimator $\hat{K}_{ls}$.*

4. Random Design Setting.   In this section, we assume that $u_1, \ldots, u_n$ are independently distributed according to a fixed probability measure ν on the unit sphere, S^{d-1}. The measurement errors $\xi_1, \ldots, \xi_n$ are independent normal random variables with mean zero and variance σ^2. We also assume that $\xi_1, \ldots, \xi_n$ are independent of $u_1, \ldots, u_n$. We consider the loss function ℓ_r^2 (defined in (7)) and prove upper and lower bounds for the corresponding minimax risk $R_r(n)$ (defined in (6)).

4.1. *Upper bound for $R_r(n)$.*   The following result is the random-design analogue of Theorem 3.1. We show that $R_r(n)$ is bounded from above by $n^{-4/(d+3)}$ up to multiplication by a constant that is independent of n. No assumptions on ν are required and it is completely arbitrary.

THEOREM 4.1. *There exist positive constants c and C depending only on the dimension d such that*

$$(30) \qquad R_r(n) \le \frac{c(\Gamma^2/\sigma^2)}{1 - e^{-\Gamma^2/(4\sigma^2)}} \sigma^{8/(d+3)} \Gamma^{2(d-1)/(d+3)} n^{-4/(d+3)}$$

if $n \ge C(\sigma/\Gamma)^2$.

REMARK 4.1. *Our proof below also shows that if one works with the smaller loss function:*

$$(31) \qquad \ell^2_{new}(K, K') := -16\sigma^2 \log \int \exp\left(-\frac{(h_K(u) - h_{K'}(u))^2}{16\sigma^2}\right) d\nu(u)$$

instead of ℓ^2_r, then the factor $(\Gamma^2/\sigma^2)(1 - e^{-\Gamma^2/(4\sigma^2)})^{-1}$ in the minimax risk bound (30) can be removed.

PROOF OF THEOREM 4.1. As in the proof of Theorem 3.1, we consider the least squares estimator $\hat{K}_F$ (defined in (14)) over a finite subset F of $\mathcal{K}^d(\Gamma)$ for which inequality (15), reproduced below, holds for every $K, L' \in F$ and $\alpha > 0$:

$$\mathbb{E}_K G(\hat{K}_F) \le \sum_{L \in F} G(L) \mathbb{E}_K \exp\left(\alpha \sum_i (Y_i - h_{L'}(u_i))^2 - \alpha \sum_i (Y_i - h_L(u_i))^2\right).$$

Under the random-design setting, the expectation in the right hand side above equals

$$\mathbb{E}_\nu \exp\left(-\alpha n \ell^2_f(K, L) + \alpha n \ell^2_f(K, L') + 2\alpha^2 \sigma^2 n \ell^2_f(L, L')\right),$$

where, as explained in Section 2, the expectation $\mathbb{E}_\nu$ is taken under the assumption that $u_1, \dots, u_n$ are independently distributed according to ν (note that ℓ^2_f depends on $u_1, \dots, u_n$). Using the inequalities $\ell^2_f(L, L') \le 2\ell^2_f(K, L) + 2\ell^2_f(K, L')$ and $\ell^2_f(K, L') \le \ell^2_H(K, L')$, we obtain the following upper bound for $\mathbb{E}_K G(\hat{K}_F)$:

$$\sum_{L \in F} \exp\left((\alpha + 4\alpha^2\sigma^2) n \ell^2_H(K, L')\right) G(L) \mathbb{E}_\nu \exp\left((-\alpha + 4\alpha^2\sigma^2) n \ell^2_f(K, L)\right),$$

for every $L' \in F$. It may be helpful to note that $\ell^2_H(K, L')$ does not depend on $u_1, \dots, u_n$ and is non-random. We apply this inequality to the following choices of G and α:

$$G(L) = \left(\mathbb{E}_\nu \exp\left(-\frac{n \ell^2_f(K, L)}{16\sigma^2}\right)\right)^{-1} \quad \text{and} \quad \alpha = \frac{1}{8\sigma^2}.$$

A straightforward calculation reveals that this function $G(L)$ has the following alternative expression

$$G(L) = \exp\left(\frac{n}{16\sigma^2}\ell_{new}^2(K, L)\right),$$

where ℓ_{new}^2 is defined as in (31). Specializing the upper bound for $\mathbb{E}_K G(\hat{K}_F)$ to these choices of G and α, we deduce

$$\mathbb{E}_K \exp\left(\frac{n\ell_{new}^2(K, \hat{K}_F)}{16\sigma^2}\right) \le |F| \exp\left(\frac{3n}{16\sigma^2} \min_{L' \in F} \ell_H^2(K, L')\right).$$

This is the same inequality as (16) with the loss function ℓ_f^2 replaced by ℓ_{new}^2. Thus, following the same steps as in the proof of Theorem 3.1, we deduce the existence of positive constants c and C depending only on d such that

$$(32) \qquad R_{new}(n) \le c\ \sigma^{8/(d+3)} \Gamma^{2(d-1)/(d+3)} n^{-4/(d+3)} \qquad \text{if } n \ge C(\sigma/\Gamma)^2,$$

where

$$R_{new}(n) := \inf_{\hat{K}} \sup_{K \in \mathcal{K}^d(\Gamma)} \mathbb{E}_K \ell_{new}^2(K, \hat{K}).$$

This proves the claim made in Remark 4.1. We now give a simple inequality relating the loss functions $\ell_{new}^2(K, K')$ and $\ell_r^2(K, K')$ which enables us to convert this bound for $R_{new}(n)$ into the required inequality (30) for $R_r(n)$. For every $K \in \mathcal{K}^d(\Gamma)$ and $u \in S^{d-1}$, we have $|h_K(u)| \le \Gamma$. Therefore,

$$\frac{(h_K(u) - h_{K'}(u))^2}{16\sigma^2} \le \frac{\Gamma^2}{4\sigma^2} \qquad \text{for all } K, K' \in \mathcal{K}^d(\Gamma).$$

Since the convex function $x \mapsto e^{-x}$ lies below the chord joining the points $(0, 1)$ and $(\Gamma^2/(4\sigma^2), \exp(-\Gamma^2/(4\sigma^2)))$, we have

$$e^{-x} \le 1 + \frac{4\sigma^2}{\Gamma^2}\left(e^{-\Gamma^2/(4\sigma^2)} - 1\right)x \qquad \text{for } 0 \le x \le \Gamma^2/(4\sigma^2).$$

Using this with $x = (h_K(u) - h_{K'}(u))^2/(16\sigma^2)$ and integrating both sides of the resulting expression with respect to ν, we get

$$\int \exp\left(-\frac{(h_K(u) - h_{K'}(u))^2}{16\sigma^2}\right) d\nu(u) \le 1 + \frac{e^{-\Gamma^2/(4\sigma^2)} - 1}{4\Gamma^2}\ell_r^2(K, K').$$

Taking logarithms on both sides, we obtain

$$\log \int \exp\left(-\frac{(h_K(u) - h_{K'}(u))^2}{16\sigma^2}\right) d\nu(u) \le \frac{e^{-\Gamma^2/(4\sigma^2)} - 1}{4\Gamma^2}\ell_r^2(K, K'),$$

where, on the right hand side, we have used $\log(1 + y) \leq y$. The above inequality can be rewritten as

$$\ell_r^2(K, K') \leq \frac{\Gamma^2/(4\sigma^2)}{1 - e^{-\Gamma^2/(4\sigma^2)}} \ell_{new}^2(K, K').$$

The proof is complete because the required bound (30) follows by combining the above inequality with (32). □

4.2. *Lower Bound for $R_r(n)$.* The following theorem is the random design analogue of Theorem 3.2. We assume that $\nu = \nu_{\mathrm{unif}}$ is the uniform probability measure on S^{d-1} and prove that $R_r(n)$ is bounded from below by a constant multiple of $n^{-4/(d+3)}$. Note that the lower bound of $n^{-4/(d+3)}$ can not be proved for $R_r(n)$ for arbitrary ν. For example, when ν is concentrated at a single point, $R_r(n)$ is of order $1/n$.

THEOREM 4.2.   *Consider the random-design setting where ν equals the uniform probability measure ν_{unif} on S^{d-1}. Then there exist positive constants c and C depending only on d such that*

(33) $R_r(n) \geq c\, \sigma^{8/(d+3)} \Gamma^{2(d-1)/(d+3)} n^{-4/(d+3)}$ *whenever $n \geq C(\sigma/\Gamma)^2$.*

It is possible to prove this theorem by an appropriate modification of the proof of Theorem 3.2. We however give a different proof using a global metric entropy minimax lower bound from Guntuboyina [16]. This proof has an interesting implication that is described in Remark 4.2. A version of this proof appeared in Guntuboyina [16, Section V] although the result there has a different assumption on $u_1, \ldots, u_n$ and is also slightly less precise.

PROOF OF THEOREM 4.2. Let $\mathfrak{P} := \{\mathbb{P}_K : K \in \mathcal{K}^d(\Gamma)\}$. We use the following minimax lower bound from Guntuboyina [16, Inequality (22)]:

(34) $R_r(n) \geq \dfrac{\eta^2}{4} \left(1 - \dfrac{1}{N(\mathcal{K}^d(\Gamma), \eta; \ell_r)} - \sqrt{\dfrac{(1 + \epsilon^2)M(\mathfrak{P}, \epsilon; \chi)}{N(\mathcal{K}^d(\Gamma), \eta; \ell_r)}} \right)$

for all $\eta > 0$ and $\epsilon > 0$. The notation was set in Section 2: $N(\mathcal{K}^d(\Gamma), \eta; \ell_r)$ denotes the maximal η-packing number of $\mathcal{K}^d(\Gamma)$ under the ℓ_r-metric (the square-root of the loss function ℓ_r^2) and $M(\mathfrak{P}, \epsilon; \chi)$ denotes the ϵ-covering number of $\mathfrak{P}$ when distances are measured by the square-root of the chi-squared divergence i.e., it is the smallest integer M for which there exist probability measures $\mathbb{Q}_1, \ldots, \mathbb{Q}_M$ satisfying $\min_{1 \leq i \leq M} \chi^2(\mathbb{P}||\mathbb{Q}_i) \leq \epsilon^2$ for every $\mathbb{P} \in \mathfrak{P}$.

The application of (34) requires a lower bound on $N(\mathcal{K}^d(\Gamma), \eta; \ell_r)$ and an upper bound on $M(\mathfrak{P}, \epsilon; \chi)$. Guntuboyina [16, Theorem VII.1], building on a result of Bronshtein [4], showed the existence of positive constants η_0 and c' depending only on d such that

$$\log N(\mathcal{K}^d(\Gamma), \eta; \ell_r) \geq c' \left(\frac{\Gamma}{\eta} \right)^{(d-1)/2} \qquad \text{whenever } \eta \leq \Gamma \eta_0.$$

The above bound uses crucially the fact that ν equals ν_{unif}. It is not true for arbitrary probability measures on S^{d-1}.

For $M(\mathfrak{P}, \epsilon; \chi)$, we note that the chi-squared divergence $\chi^2(\mathbb{P}_K || \mathbb{P}_{K'})$ satisfies

$$1 + \chi^2(\mathbb{P}_K || \mathbb{P}_{K'}) = \left(\int \exp \left(\frac{(h_K(u) - h_{K'}(u))^2}{\sigma^2} \right) d\nu(u) \right)^n$$
$$\leq \exp \left(\frac{n \ell_H^2(K, K')}{\sigma^2} \right).$$

As a result,

$$\chi^2(\mathbb{P}_K || \mathbb{P}_{K'}) \leq \epsilon^2 \qquad \text{whenever } \ell_H(K, K') \leq \epsilon' := \sigma \sqrt{\log(1 + \epsilon^2)} / \sqrt{n}$$

and $M(\mathfrak{P}, \epsilon; \chi) \leq M(\mathcal{K}^d(\Gamma), \epsilon'; \ell_H)$. Upper bound for the covering number $M(\mathcal{K}^d(\Gamma), \epsilon'; \ell_H)$ has been proved by Bronshtein [4, Theorem 3 and Remark 1]. We already used this result (inequality (18)) in the proof of Theorem 3.1. We use it again to obtain

$$\log M(\mathfrak{P}, \epsilon; \chi) \leq c \left(\frac{\Gamma \sqrt{n}}{\sigma \sqrt{\log(1 + \epsilon^2)}} \right)^{(d-1)/2} \qquad \text{if } \log(1 + \epsilon^2) \leq n \Gamma^2 \epsilon_0^2 / \sigma^2,$$

for positive constants c and ϵ_0. Let us now define

$$\eta(n) := c_1 \sigma^{4/(d+3)} \Gamma^{(d-1)/(d+3)} n^{-2/(d+3)} \quad \& \quad \alpha(n) := \left(\frac{\Gamma \sqrt{n}}{\sigma} \right)^{(d-1)/(d+3)},$$

where c_1 is a positive constant that depends on d alone and will be specified shortly. Also let $\epsilon^2(n) := \exp(\alpha^2(n)) - 1$. We then have

$$\log N(\mathcal{K}^d(\Gamma), \eta(n); \ell_r) \geq c' c_1^{-(d-1)/2} \alpha^2(n) \quad \& \quad \log M(\mathfrak{P}, \epsilon(n); \chi) \leq c \alpha^2(n)$$

provided

$$(35) \qquad \eta(n) \leq \Gamma \eta_0 \quad \& \quad \alpha^2(n) \leq n \Gamma^2 \epsilon_0^2 / \sigma^2.$$

Inequality (34) with $\eta = \eta(n)$ and $\epsilon = \epsilon(n)$ gives the following lower bound for $R_r(n)$:

$$\frac{\eta^2(n)}{4} \left[1 - \exp\left(-\alpha^2(n)c'c_1^{-(d-1)/2}\right) - \exp\left(\frac{\alpha^2(n)}{2}(1 + c - c'c_1^{-(d-1)/2})\right) \right].$$

If we choose c_1 so that $c'c_1^{-(d-1)/2} = 2(1+c)$, then

$$R_r(n) \geq \frac{\eta^2(n)}{4} \left(1 - 2\exp\left(-\frac{1+c}{2}\alpha^2(n)\right) \right).$$

If the condition $(1 + c)\alpha^2(n) \geq 2\log 4$ holds, then the above inequality implies $R_r(n) \geq \eta^2(n)/8$ which yields (33). This condition as well as (35) hold provided $n \geq C(\sigma/\Gamma)^2$ for a large enough C. The proof is complete. $\quad\square$

REMARK 4.2. *In the above proof, the random design assumption on the unit vectors $u_1, \ldots, u_n$ was used only in*

$$\chi^2(\mathbb{P}_K || \mathbb{P}_{K'}) \leq \exp\left(\frac{n\ell_H^2(K, K')}{\sigma^2}\right) - 1.$$

This inequality is easily seen to be true for every joint distribution of $(u_1, \ldots, u_n)$ as long as they are independent of the errors $\xi_1, \ldots, \xi_n$. Consequently, $n^{-4/(d+3)}$, up to multiplicative constants, is a lower bound for the minimax risk (observe that the integral below is with respect to the uniform probability measure ν_{unif}):

$$\inf_{\hat{K}} \sup_{K \in \mathcal{K}^d(\Gamma)} \mathbb{E}_K \int_{S^{d-1}} \left(h_K(u) - h_{\hat{K}}(u)\right)^2 \nu_{unif}(du)$$

for every arbitrary choice of the design unit vectors (deterministic or random) as long as they are independent of $\xi_1, \ldots, \xi_n$ (this independence assumption is only relevant in a random-design setting).

4.3. *Least Squares on Polytopes.* We prove the random-design analogue of Theorem 3.4. Let

$$\breve{R}_r(n) := \inf_{m \geq 1} \sup_{K \in \mathcal{K}^d(\Gamma)} \mathbb{E}_K \ell_r^2(K, \breve{K}_m).$$

be the best achievable risk by $\breve{K}_m$ (defined in (24)) in the random design setting under the loss function ℓ_r^2. The following theorem shows that $\breve{R}_r(n)$ is bounded from above by $n^{-4/(d+3)}$ up to a multiplicative factor that is logarithmic in n. No assumptions on ν are necessary.

THEOREM 4.3. *There exist positive constants c and C that depend only on the dimension d such that*

$$(36) \qquad \breve{R}_r(n) \leq c \max(\sigma^2, \Gamma^2) \; n^{-4/(d+3)} \left(\log(cn)\right)^2 \qquad \text{if } n \geq C.$$

The proof strategy is to use the fixed design bound (25) along with Lemma 7.1 (stated and proved in Appendix III) which relates the risks under the two loss functions ℓ_f^2 and ℓ_r^2.

PROOF OF THEOREM 4.3. We start with the inequality

$$\ell_r^2(\breve{K}_m, K) \leq 2 \left(\ell_r(\breve{K}_m, K) - 2\ell_f(\breve{K}_m, K) \right)_+^2 + 8\ell_f^2(\breve{K}_m, K)$$

which implies that

$$\mathbb{E}_K \ell_r^2(\breve{K}_m, K) \leq 2\mathbb{E}_\nu \sup_{L \in \mathcal{P}_m(\Gamma)} \left(\ell_r(L, K) - 2\ell_f(L, K) \right)_+^2 + 8\mathbb{E}_K \ell_f^2(\breve{K}_m, K).$$

The first expectation in the right hand side is bounded using Lemma 7.1 where it is shown that

$$(37) \qquad \mathbb{E}_\nu \sup_{L \in \mathcal{P}_m(\Gamma)} \left(\ell_r(L, K) - 2\ell_f(L, K) \right)_+^2 \leq c \frac{\Gamma^2}{n} md \log(cn)$$

for a universal positive constant c. For the second expectation, we use ideas from the proof of Theorem 3.4. Indeed, the same argument which led to the inequality (28) gives, for every $P \in \mathcal{P}_m$,

$$\mathbb{E}_K \left(\ell_f^2(\breve{K}_m, K) | u_1, \ldots, u_n \right) \leq c \left(\ell_f^2(K, P) + \frac{\sigma^2}{n} md \log(b_2 m) \log(cn) \right).$$

for every $u_1, \ldots, u_n$. Taking expectation with respect to $u_1, \ldots, u_n$ independently distributed according to ν, we get

$$(38) \qquad \mathbb{E}_K \ell_f^2(\breve{K}_m, K) \leq c \left(\ell_r^2(K, P) + \frac{\sigma^2}{n} md \log(b_2 m) \log(cn) \right).$$

Putting (37) and (38) together, we obtain

$$\mathbb{E}_K \ell_r^2 \left(\breve{K}_m, K \right) \leq c \left(\ell_r^2(K, P) + \frac{\sigma^2}{n} md \log(b_2 m) \log(cn) + \frac{\Gamma^2}{n} md \log(cn) \right)$$

$$\leq c \left(\ell_H^2(K, P) + \frac{\max(\sigma^2, \Gamma^2)}{n} md \log(b_2 m) \log(cn) \right)$$

Because $K \in \mathcal{K}^d(\Gamma)$ and $P \in \mathcal{P}_m(\Gamma)$ are arbitrary, we have shown

$$\breve{R}_r(n) \leq c \inf_{m \geq 1} \left[\sup_{K \in \mathcal{K}^d(\Gamma)} \inf_{P \in \mathcal{P}_m(\Gamma)} \ell_H^2(K, P) + \frac{\max(\sigma^2, \Gamma^2)}{n} md \log(b_2 m) \log(cn) \right].$$

Just as in the proof of Theorem 3.4, we use the result (29) due to Bronshteyn and Ivanov [6] to get

$$\breve{R}_r(n) \leq C \inf_{m \geq C_2} \left[\Gamma^2 m^{-4/(d-1)} + \frac{\max(\sigma^2, \Gamma^2)}{n} md \log(b_2 m) \log(cn) \right]$$

$$\leq C \max(\sigma^2, \Gamma^2) \inf_{m \geq C_2} \left[m^{-4/(d-1)} + \frac{md}{n} \log(b_2 m) \log(cn) \right].$$

If we now choose $m := n^{(d-1)/(d+3)}$, then $m \geq C_2$ provided $n \geq C$ for a large enough constant C depending only on d and the required inequality (36) follows. $\qquad\square$

5. Appendix I. In this section, we shall prove the following result which was used in the proof of Theorem 3.2. We assume that $u_1, \ldots, u_n$ form a maximal ϵ-packing subset of S^{d-1} for some $\epsilon \in (0, 1]$.

LEMMA 5.1. *For a fixed $0 < \eta \leq 1/8$ and a unit vector v, consider the following two subsets of the ball $B_d(0, \Gamma)$:*

$$D(0) := B_d(0, \Gamma) \cap \{x : x \cdot v \leq 1 - \eta\} \quad and \quad D(1) := B_d(0, \Gamma).$$

Then there exists constants c and C such that the following inequality holds whenever $\eta \geq C\epsilon^2$

$$(39) \qquad\qquad c\eta^{(d+3)/2} \leq \ell_f^2(D(0), D(1)) \leq C\eta^{(d+3)/2}.$$

We need some elementary results on spherical caps for the proof of this lemma. For a unit vector u and a real number $0 < \delta \leq 1$, consider the spherical cap $\mathcal{S}(u; \delta) := S^{d-1} \cap B_d(u, \delta)$. It can be checked that this spherical cap consists of precisely those unit vectors which form an angle of at most α with u, where α is related to δ through

$$\cos \alpha = 1 - \frac{\delta^2}{2} \text{ and } \sin \alpha = \frac{\delta \sqrt{4 - \delta^2}}{2}.$$

A standard result is that $\nu_{\text{unif}}(\mathcal{S}(x; \delta))$ equals $C \int_0^\alpha \sin^{d-2} t \, dt$ (recall that ν_{unif} denotes the uniform probability measure on the unit sphere). This

integral can be bounded from above and below in the following simple way. For a lower bound, we write

$$\int_0^\alpha \sin^{d-2} t \ dt \geq \int_0^\alpha \sin^{d-2} t \cos t \ dt \geq \frac{\sin^{d-1} \alpha}{d-1},$$

and for an upper bound, we note

$$\int_0^\alpha \sin^{d-2} t \ dt \leq \int_0^\alpha \frac{\cos t}{\cos \alpha} \sin^{d-2} t \ dt \leq \frac{\sin^{d-1} \alpha}{(d-1) \cos \alpha}.$$

We thus have $c \sin^{d-1} \alpha \leq \nu(\mathcal{S}(u; \delta)) \leq C \sin^{d-1} \alpha / \cos \alpha$. Writing $\cos \alpha$ and $\sin \alpha$ in terms of δ and using the assumption that $0 < \delta \leq 1$, we obtain that

$$(40) \qquad c\delta^{d-1} \leq \nu_{\mathrm{unif}}(\mathcal{S}(u; \delta)) \leq C\delta^{d-1} \qquad \text{if } \delta \in (0, 1].$$

This inequality can be combined with a simple volumetric argument to show that

$$(41) \qquad c\delta^{1-d} \leq N(S^{d-1}, \delta; \mathfrak{d}_d) \leq C\delta^{1-d} \qquad \text{if } \delta \in (0, 1].$$

In particular, since $u_1, \ldots, u_n$ form a maximal ϵ-packing subset of S^{d-1}, we have (19).

The following lemma is used in the proof of Lemma 5.1.

LEMMA 5.2. *Fix positive ϵ, δ such that $\delta \leq 4/5$ and $\epsilon \leq \delta/2$. Let $u_1, \ldots u_n$ be a maximal ϵ-packing subset of the unit sphere and v be an arbitrary unit vector. Let $V(\epsilon, \delta)$ denote the number of points $u_1, \ldots, u_n$ that are contained in $\mathcal{S}(v, \delta)$. Then*

$$(42) \qquad c \left(\frac{\delta}{\epsilon}\right)^{d-1} \leq V(\epsilon, \delta) \leq C \left(\frac{\delta}{\epsilon}\right)^{d-1}.$$

PROOF. For the lower bound on $V(\epsilon, \delta)$, we observe that

$$(43) \qquad \mathcal{S}(v, \delta/2) \subseteq \bigcup_{i: u_i \in \mathcal{S}(v, \delta)} \mathcal{S}(u_i, \epsilon).$$

Indeed, because $\epsilon \leq \delta/2$, for every $w \in \mathcal{S}(v, \delta/2)$, we can find u_i such that $\mathfrak{d}_d(u_i, w) < \epsilon \leq \delta/2$ because $u_1, \ldots, u_n$ form a maximal ϵ-packing subset of S^{d-1}. Thus $u_i \in \mathcal{S}(v, \delta)$ by triangle inequality which proves (43).

It follows from (43) that $V(\epsilon, \delta) \geq \nu_{\mathrm{unif}}(\mathcal{S}(v, \delta/2))/\nu_{\mathrm{unif}}(\mathcal{S}(u_1, \epsilon))$ from which the lower bound in (42) follows by use of (40).

For the upper bound on $V(\epsilon, \delta)$, we use the inequality

$$\mathcal{S}(v, \delta + \epsilon/2) \supseteq \bigcup_{i:u_i \in \mathcal{S}(v,\delta)} \mathcal{S}(u_i, \epsilon/2)$$

and then, noting that the spherical caps on the right hand side of the above inequality are disjoint as $u_1, \ldots, u_n$ form a ϵ-packing subset, we obtain $V(\epsilon, \delta) \leq \nu_{\text{unif}}(\mathcal{S}(v, \delta + \epsilon/2))/\nu_{\text{unif}}(\mathcal{S}(u_1, \epsilon/2))$. The upper bound in (42) again follows from (40). □

We are now ready to prove Lemma 5.1.

PROOF OF LEMMA 5.1. It can be checked that the support functions of $D(0)$ and $D(1)$ differ only for unit vectors in the spherical cap $\mathcal{S}(v, \sqrt{2\eta})$. This spherical cap consists of all unit vectors which form an angle of at most α with v where $\cos \alpha = 1 - \eta$. In fact, if θ denotes the angle between an arbitrary unit vector u and v, it can be verified by elementary trigonometry that

$$(44) \qquad h_{D(0)}(u) - h_{D(1)}(u) = \begin{cases} \Gamma \left(1 - \cos\left(\alpha - \theta\right)\right) & \text{if } 0 \leq \theta \leq \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

As a result, it follows that $\ell_f^2(D(0), D(1)) \leq \Gamma^2 \eta^2 V(\epsilon, \sqrt{2\eta})/n$ where V is as defined in Lemma 5.2. Thus (42) gives $\ell_f^2(D(0), D(1)) \leq C\eta^{(d+3)/2}\epsilon^{1-d}/n$. The conditions in Lemma 5.2 are satisfied if $0 < \eta \leq 1/8$ and $\eta \geq C\epsilon^2$ for a large enough C. Moreover, by (19), we have $c \leq n\epsilon^{d-1} \leq C$ which implies that $\ell_f^2(D(0), D(1)) \leq C\eta^{(d+3)/2}$.

For a lower bound, fix $0 < b \leq 1$ and let $0 \leq \beta \leq \alpha$ denote the angle for which $1 - \cos(\alpha - \beta) = b\eta$. It follows from (44) that the difference in the support functions of $D(0)$ and $D(1)$ is at least $b\Gamma\eta$ for all unit vectors in the spherical cap consisting of all unit vectors forming an angle of at most β with v. This spherical cap is $\mathcal{S}(v, t)$ where t is given by $t^2 := 2(1 - \cos \beta)$. Therefore $\ell_f^2(D(0), D(1)) \geq b^2 \Gamma^2 \eta^2 V(\epsilon, t)/n$. The inequality $t^2 \leq 2(1 - \cos \alpha) \leq 2\eta$ is easily checked. Also, $t \geq \sin \beta$ and $\sin \beta$ can be bounded from below in the following way

$$1 - b\eta = \cos(\alpha - \beta) \leq \cos \alpha + \sin \alpha \sin \beta \leq 1 - \eta + \sqrt{2\eta} \sin \beta.$$

Thus $t \geq \sin \beta \geq (1 - b)\sqrt{\eta/2}$ and from (42), it follows that

$$\ell_f^2(D(0), D(1)) \geq cb^2 \frac{\Gamma^2 \eta^2}{n} \left(\frac{t}{\epsilon}\right)^{d-1} \geq cb^2(1-b)^{d-1} \frac{\Gamma^2 \eta^{(d+3)/2}}{n\epsilon^{d-1}}$$

for all $0 < b \leq 1$. Note that we have used $\eta \geq C\epsilon^2$ here to satisfy the conditions in Lemma 5.2. We now use (19) and choose $b = 1/2$ to get $\ell_f^2(D(0), D(1)) \geq c\eta^{(d+3)/2}$ provided $\eta \geq C\epsilon^2$. The proof is complete. $\square$

6. Appendix II. In this section, we prove the following lemma which was used in the proof of Theorem 3.4.

LEMMA 6.1.   *Fix $m \geq 1$ and $\omega, \epsilon > 0$. The following bound holds for every $P \in \mathcal{P}_m$,*

$$M(S_m(P, \omega), \epsilon; \ell_f) \leq \left(4 + \frac{2\sqrt{n}\omega}{\epsilon}\right)^{b_1 md \log(b_2 m)},$$

where b_1 and b_2 are universal positive constants.

For the proof of this lemma, we use available techniques for bounding covering numbers using combinatorial notions of dimension. Specifically, we use the notion of pseudodimension, introduced by Pollard [23, Chapter 4] as a generalization of the Vapnik-Červonenkis dimension to classes of real valued functions. The pseudodimension of a subset A of $\mathbb{R}^n$ is defined as the maximum cardinality of a subset $\sigma \subseteq \{1, \ldots, n\}$ for which there exists $(h_1, \ldots, h_n) \in \mathbb{R}^n$ such that: for every $\sigma' \subseteq \sigma$, one can find $(a_1, \ldots, a_n) \in A$ with $a_i < h_i$ for $i \in \sigma'$ and $a_i > h_i$ for $i \in \sigma \setminus \sigma'$. The following theorem is a special case of results in Pollard [23, Chapter 4] and gives an upper bound for the covering number (with respect to the Euclidean metric) of a subset of $\mathbb{R}^n$ in terms of its pseudodimension. Stronger results of this kind have been proved by Mendelson and Vershynin [21] and the following theorem is also a special case of Mendelson and Vershynin [21, Theorem 1].

THEOREM 6.2.   *Let A be a subset of $\mathbb{R}^n$ with $\max_i |a_i| \leq B$ for all $a = (a_1, \ldots, a_n) \in A$. If the pseudodimension of A is at most V, then, for every $t > 0$, we have*

$$M(A, t; \mathfrak{d}_n) \leq \left(4 + \frac{2B\sqrt{n}}{t}\right)^{bV}$$

where b is a universal positive constant.

PROOF OF LEMMA 6.1. Fix a polytope $P \in \mathcal{P}_m$ and let $x_0 \in \mathbb{R}^n$ denote the point $(h_P(u_1), \ldots, h_P(u_n))$. Also, for $m \geq 1$, let

$$H_m := \left\{x \in \mathbb{R}^n : x = (h_L(u_1), \ldots, h_L(u_n)) \text{ for some } L \in \mathcal{P}_m\right\}.$$

Clearly

$$M(S_m(P, \omega), \epsilon; \ell_f) = M(B_n(0, \sqrt{n}\omega) \cap H_m - x_0, \sqrt{n}\epsilon; \mathfrak{d}_n),$$

where $H_m - x_0 := \{x - x_0 : x \in H_m\}$.

We now show that the pseudodimension of H_m, which clearly equals the pseudodimension of $H_m - x_0$, is less than or equal to $\varsigma md \log(\varsigma m)$ where $\varsigma = 2/\log 2$. An application of Theorem 6.2 would then complete the proof. Note that the quantity B in the statement of Theorem 6.2 can be taken to be $\sqrt{n}\omega$ because $|a_i| \leq \sqrt{n}\omega$ for every $(a_1, \ldots, a_n) \in B_n(0, \sqrt{n}\omega)$.

Every L in $\mathcal{P}_m$ is of the form $\mathrm{conv}(S)$ for some $S \subset \mathbb{R}^d$ with cardinality at most m and thus $h_L(u) = \max_{x \in S}(x \cdot u)$. Therefore H_1 is a linear subspace of $\mathbb{R}^n$ with dimension at most d which implies (see Pollard [23, Page 15]) that the pseudodimension of H_1 is at most d. The pseudodimension of H_m, which consists of coordinatewise maxima of at most m points in H_1, can then be bounded from above following the argument in Pollard [23, Proof of Lemma (5.1)]. Indeed, that argument shows that the pseudodimension of H_m is bounded from above by the smallest positive integer k for which

$$(45) \qquad \binom{k}{0} + \cdots + \binom{k}{d} < 2^{k/m}.$$

If $\mathfrak{B}$ denotes a binomial random variable with parameters k and $1/2$, then the left hand side of the above inequality equals $2^k \mathbb{P}\{\mathfrak{B} \geq k - d\}$ and can be bounded, again following Pollard [23, Proof of Lemma (5.1)], as shown below. For every $\alpha > 0$,

$$(46) \quad 2^k \mathbb{P}\{\mathfrak{B} \geq k - d\} \leq 2^k \mathbb{P}\{\alpha^{\mathfrak{B}} \geq \alpha^{k-d}\} \geq 2^k \alpha^{d-k} \mathbb{E}\alpha^{\mathfrak{B}} = (1+\alpha)^k \alpha^{d-k}.$$

If we now choose $\alpha = (2^{1/(2m)} - 1)^{-1}$, then $(1+\alpha)/\alpha = 2^{1/(2m)}$ and also, applying the inequality $x - 1 > \log x$ to $x = 2^{1/(2m)}$, we get that $\alpha < 2m/(\log 2)$. Therefore, from (46), we have

$$2^k \mathbb{P}\{\mathfrak{B} \geq k - d\} < \left(\frac{2m}{\log 2}\right)^d 2^{k/(2m)}.$$

The following inequality therefore ensures that k satisfies (45):

$$\left(\frac{2m}{\log 2}\right)^d 2^{k/(2m)} \leq 2^{k/m} \text{ or, equivalently, } k \geq \varsigma md \log(\varsigma m),$$

where $\varsigma = 2/(\log 2)$. It follows, therefore, that the pseudodimension of H_m is at most $\varsigma md \log(\varsigma m)$. The proof of Lemma 6.1 is complete. $\square$

7. Appendix III. In this section, we prove the following result which was used in the proof of Theorem 4.3. It relates the risks under the loss functions ℓ_f^2 and ℓ_r^2. Its proof is based on standard empirical process arguments due to Pollard [22]. We have also borrowed ideas from Györfi, Kohler, Krzyżak, and Walk [17, Chapter 11].

LEMMA 7.1.   *There exists a universal constant c such that the following inequality holds for every $m \geq 1$, $\Gamma > 0$ and $K \in \mathcal{P}_m(\Gamma)$,*

$$(47) \qquad \mathbb{E}_\nu \sup_{L \in \mathcal{P}_m(\Gamma)} (\ell_r(L, K) - 2\ell_f(L, K))_+^2 \leq c\frac{\Gamma^2}{n} md \log(cn).$$

PROOF. For $x > 0$, let

$$\varrho_x := \mathbb{P}_\nu \left\{ \sup_{L \in \mathcal{P}_m(\Gamma)} (\ell_r(L, K) - 2\ell_f(L, K))_+^2 > x \right\}$$

Letting $\mathcal{F} := \{h_L - h_K : L \in \mathcal{P}_m(\Gamma)\}$, we have the trivial inequality

$$\varrho_x \leq \mathbb{P}_\nu \left\{ \sup_{f \in \mathcal{F}} \left[\left(\int f^2(u) d\nu(u) \right)^{1/2} - 2 \left(\frac{1}{n} \sum_{i=1}^n f^2(u_i) \right)^{1/2} \right] > \sqrt{x} \right\}.$$

Each function in $\mathcal{F}$ is bounded in absolute value by 2Γ. We now use Györfi et al. [17, Theorem 11.2], which is a slight variation of Pollard [22, Lemma 33 and Problem 24]), to obtain

$$(48) \qquad \varrho_x \leq 4\mathbb{E}_\nu \min \left(1, M(\mathcal{P}_m(\Gamma), \sqrt{x}/24; \ell_f) \exp \left(-\frac{nx}{2304\Gamma^2} \right) \right).$$

Lemma 6.1 with $\omega = \Gamma$, $\epsilon = \sqrt{x}/24$ and $P = \{0\}$ provides an upper bound for the covering number $M(\mathcal{P}_m(\Gamma), \sqrt{x}/24; \ell_f)$. The set $\mathcal{P}_m(\Gamma)$ is much smaller than $S_m(P, \omega)$ however and the following direct argument gives the simpler upper bound:

$$(49) \quad M(\mathcal{P}_m(\Gamma), \sqrt{x}/24; \ell_f) \leq M(\mathcal{P}_m(\Gamma), \sqrt{x}/24; \ell_H) \leq \left(1 + \frac{48\Gamma}{\sqrt{x}} \right)^{md}.$$

To see this, let $\mathfrak{C}$ be an $\epsilon := \sqrt{x}/24$-covering subset of $B_d(0, \Gamma)$ under the Euclidean metric, $\mathfrak{d}_d$. Then $\mathfrak{H} := \{\text{conv}(S) : S \subseteq \mathfrak{C} \text{ with } |S| \leq m\}$ forms an ϵ-covering subset of $\mathcal{P}_m(\Gamma)$ under ℓ_H. Indeed, if $K = \text{conv}\{a_1, \ldots, a_m\}$ with $a_1, \ldots, a_m \in B_d(0, \Gamma)$ is an arbitrary element set in $\mathcal{P}_m(\Gamma)$, then we can choose $a_1', \ldots, a_m' \in \mathfrak{C}$ with $\mathfrak{d}_d(a_i, a_i') \leq \epsilon$ for each i. It is then easy to see

(using (11)) that the Hausdorff distance between K and $\text{conv}\{a_1', \ldots, a_m'\}$ is at most ϵ. It is evident that the cardinality of $\mathfrak{H}$ is at most $|\mathfrak{C}|^m$. A standard volumetric argument shows that the cardinality of $\mathfrak{C}$ can be taken to be smaller than $(1 + (2\Gamma/\epsilon))^d$ which proves (49).

Combining (48) and (49), we get

$$\varrho_x \leq 4 \min\left(1, \left(1 + \frac{48\Gamma}{\sqrt{x}}\right)^{md} \exp\left(-\frac{nx}{2304\Gamma^2}\right)\right).$$

The left hand side of (47) can now be bounded, for every $\Lambda > 0$, by

$$\int_0^\infty \varrho_x dx = \int_0^\Lambda \varrho_x dx + \int_\Lambda^\infty \varrho_x dx$$

$$\leq 4\Lambda + 4 \int_\Lambda^\infty \left(1 + \frac{48\Gamma}{\sqrt{x}}\right)^{md} \exp\left(-\frac{nx}{2304\Gamma^2}\right) dx$$

$$\leq 4\Lambda + 9216\frac{\Gamma^2}{n} \left(1 + \frac{48\Gamma}{\sqrt{\Lambda}}\right)^{md} \exp\left(-\frac{n\Lambda}{2304\Gamma^2}\right).$$

The required bound (47) is now easily deduced by using the above inequality with

$$\Lambda := 2304\frac{\Gamma^2}{n} md \log\left(1 + 48\sqrt{n}\right).$$

$\square$

References.

[1] F. Balabdaoui and J. A. Wellner. Estimation of a k-monotone density: limit distribution theory and the spline connection. *Annals of Statistics*, 35:2536–2564, 2007.

[2] F. Balabdaoui, K. Rufibach, and J. A. Wellner. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of Statistics*, 37:1229–1331, 2009.

[3] L. Birgé and P. Massart. Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields*, 97:113–150, 1993.

[4] E. M. Bronshtein. ϵ-entropy of convex sets and functions. *Siberian Mathematical Journal*, 17:393–398, 1976.

[5] E. M. Bronshtein. Approximation of convex sets by polytopes. *Journal of Mathematical Sciences*, 153:727–762, 2008.

[6] E. M. Bronshteyn and L. D. Ivanov. The approximation of convex sets by polyhedra. *Siberian Mathematical Journal*, 16:852–853, 1975.

[7] M. L. Cule, R. J. Samworth, and M. I. Stewart. Maximum likelihood estimation of a multi-dimensional log-concave density (with discussion). *Journal of the Royal Statistical Society, Series B*, 72:545–600, 2010.

[8] L. Dümbgen and K. Rufibach. Maximum likelihood estimation of a log-concave density and its distribution function. *Bernoulli*, 15:40–68, 2009.

[9] N. I. Fisher, P. Hall, B. A. Turlach, and G. S. Watson. On the estimation of a convex set from noisy data on its support function. *Journal of the American Statistical Association*, 92:84–91, 1997.

[10] R. J. Gardner and M. Kiderlen. A new algorithm for 3D reconstruction from support functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31: 556–562, 2009.

[11] R. J. Gardner, M. Kiderlen, and P. Milanfar. Convergence of algorithms for reconstructing convex bodies and directional measures. *Annals of Statistics*, 34:1331–1374, 2006.

[12] A. Goldenshluger and A. Zeevi. Recovering convex boundaries from blurred and noisy observations. *Annals of Statistics*, 34:1375–1394, 2006.

[13] J. Gregor and F. R. Rannou. Three-dimensional support function estimation and application for projection magnetic resonance imaging. *International Journal of Imaging Systems and Technology*, 12:43–50, 2002.

[14] P. Groeneboom, G. Jongbloed, and J. A. Wellner. A canonical process for estimation of convex functions: The "invelope" of integrated brownian motion $+t^4$. *Annals of Statistics*, 29:1620–1652, 2001a.

[15] P. Groeneboom, G. Jongbloed, and J. A. Wellner. Estimation of convex functions: characterizations and asymptotic theory. *Annals of Statistics*, 29:1653–1698, 2001b.

[16] A. Guntuboyina. Lower bounds for the minimax risk using f divergences, and applications. *IEEE Transactions on Information Theory*, 57:2386–2399, 2011.

[17] László Györfi, Michael Kohler, Adam Krzyżak, and Harro Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York, 2002.

[18] P. Hall and B. A. Turlach. On the estimation of a convex set with corners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:225–234, 1999.

[19] A. S. Lele, S. R. Kulkarni, and A. S. Willsky. 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, 1992.

[20] Enno Mammen. Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics*, 19:741–759, 1991.

[21] S. Mendelson and R. Vershynin. Entropy and the combinatorial dimension. *Inventiones mathematicae*, 152:37–55, 2003.

[22] D. Pollard. *Convergence of Stochastic Processes*. Springer, New York, 1984.

[23] D. Pollard. *Empirical Processes: Theory and Applications*, volume 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA, 1990.

[24] Jerry L. Prince and Alan S. Willsky. Reconstructing convex sets from support line measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:377–389, 1990.

[25] R. T. Rockafellar. *Convex Analysis*. Princeton Univ. Press, Princeton, New Jersey, 1970.

[26] Rolf Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge Univ. Press, Cambridge, 1993.

[27] E. Seijo and B. Sen. Nonparametric least squares estimation of a multivariate convex regression function. *Annals of Statistics*, 39:1633–1657, 2011.

[28] Arseni Seregin and Jon A. Wellner. Nonparametric estimation of multivariate convex-transformed densities. *Annals of Statistics*, 38:3751–3781, 2010.

[29] S. Van de Geer. *Applications of Empirical Process Theory*. Cambridge University Press, 2000.

[30] Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.