The Annals of Statistics
2020, Vol. 48, No. 6, 3251–3282
https://doi.org/10.1214/19-AOS1928

LIMIT DISTRIBUTION THEORY FOR BLOCK ESTIMATORS IN MULTIPLE ISOTONIC REGRESSION

BY QIYANG HAN[1,*] AND CUN-HUI ZHANG[1,†]

[1]*Department of Statistics, Rutgers University,* [*]*qh85@stat.rutgers.edu;* [†]*czhang@stat.rutgers.edu*

We study limit distributions for the tuning-free max–min block estimator originally proposed in (Fokianos, Leucht and Neumann (2017)) in the problem of multiple isotonic regression, under both fixed lattice design and random design settings. We show that, if the regression function f_0 admits vanishing derivatives up to order α_k along the kth dimension ($k = 1, \ldots, d$) at a fixed point $x_0 \in (0, 1)^d$, and the errors have variance σ^2, then the max–min block estimator $\hat{f}_n$ satisfies

$$(n_*/\sigma^2)^{\frac{1}{2+\sum_{k\in\mathcal{D}_*}\alpha_k^{-1}}}\left(\hat{f}_n(x_0) - f_0(x_0)\right) \rightsquigarrow \mathbb{C}(f_0, x_0).$$

Here, $\mathcal{D}_*, n_*$, depending on $\{\alpha_k\}$ and the design points, are the set of all "effective dimensions" and the size of "effective samples" that drive the asymptotic limiting distribution, respectively. If furthermore either $\{\alpha_k\}$ are relative primes to each other or all mixed derivatives of f_0 of certain critical order vanish at x_0, then the limiting distribution can be represented as $\mathbb{C}(f_0, x_0) =_d K(f_0, x_0) \cdot \mathbb{D}_{\boldsymbol{\alpha}}$, where $K(f_0, x_0)$ is a constant depending on the local structure of the regression function f_0 at x_0, and $\mathbb{D}_{\boldsymbol{\alpha}}$ is a nonstandard limiting distribution generalizing the well-known Chernoff distribution in univariate problems. The above limit theorem is also shown to be optimal both in terms of the local rate of convergence and the dependence on the unknown regression function whenever such dependence is explicit (i.e., $K(f_0, x_0)$), for the full range of $\{\alpha_k\}$ in a local asymptotic minimax sense.

There are two interesting features in our local theory. First, the max–min block estimator automatically adapts to the local smoothness and the intrinsic dimension of the isotonic regression function at the optimal rate. Second, the optimally adaptive local rates are in general not the same in fixed lattice and random designs. In fact, the local rate in the fixed lattice design case is no slower than that in the random design case, and can be much faster when the local smoothness levels of the isotonic regression function or the sizes of the lattice differ substantially along different dimensions.

1. Introduction.

1.1. *Overview.* Limit distribution theory for shape-restricted estimators is of fundamental importance in the area of statistical inference under shape restrictions. There are two main types of limit distribution theories so far available in the literature.

One line starts from the seminal contribution of [57], who showed that the limiting distribution of the maximum likelihood estimator (MLE) of a decreasing density on $[0, \infty)$ (known as Grenander estimator) at a fixed point is given by the following: Suppose the true density f_0 is decreasing on $[0, \infty)$ and continuously differentiable around $x_0 \in (0, \infty)$ with $f_0'(x_0) < 0$. Then the MLE $\hat{f}_n$ satisfies

$$n^{1/3}\left(\hat{f}_n(x_0) - f_0(x_0)\right) \rightsquigarrow \left|f_0'(x_0) f_0(x_0)/2\right|^{1/3}\mathbb{Z}. \tag{1.1}$$

Received February 2019; revised September 2019.
MSC2020 subject classifications. Primary 60E15; secondary 62G05.
Key words and phrases. Limit distribution theory, multiple isotonic regression, Gaussian process, shape constraints, adaptation.

Here, $\mathbb{Z}$, known as the Chernoff distribution, is the slope at zero of the least concave majorant of the process $t \mapsto \mathbb{B}(t) - t^2$, where $\mathbb{B}$ is the standard Brownian motion starting at 0. Later on, [31] gives an exact analytic characterization of the limiting Chernoff distribution, whereas [30] suggests the "switching relation" that quickly becomes popular as a powerful proof technique in univariate problems with monotonicity shape restrictions. The limiting Chernoff distribution arises in a number of different problems with univariate monotonicity shape restrictions, for example, (1) estimation of a regression function [16, 62, 64], (2) estimation of a monotone failure rate [45, 46, 58], (3) estimation in interval censoring models [32, 39], etc. We refer the reader to the recent survey [27] for extensive references in this direction.

Another line of limit theorems for shape restricted estimators is pioneered by [37, 38], who studied limit distribution for the MLE of a convex decreasing density on $[0, \infty)$ and the least squares estimator (LSE) of a convex regression function at a fixed point. In the density setting, if the true density f_0 is convex decreasing on $[0, \infty)$ and twice continuously differentiable in a neighborhood of x_0 with $f_0''(x_0) > 0$, [38] showed that the MLE $\hat{f}_n$ satisfies

$$n^{2/5}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \big(f_0''(x_0) f_0^2(x_0)/24\big)^{1/5} \mathbb{H}''(0), \tag{1.2}$$

where $\mathbb{H}$ is a particular upper invelope of an integrated two-sided Brownian motion plus t^4; cf. [37]. The process $\mathbb{H}$ appears in several other problems involving univariate convexity shape restrictions, for example, for the MLE of a log-concave density on $\mathbb{R}$ (cf. [9]), for the MLE of a convex bathtub-shaped hazard function (cf. [49]), for the Rényi-divergence estimators for s-concave densities on $\mathbb{R}$ (cf. [43]), etc. A generalized version of $\mathbb{H}$ appears in [10] in the context of k-monotone density estimation.

Limit theorems of types (1.1)–(1.2) are not only interesting from a statistical point of view, but are of theoretical value in their own rights. Indeed, these limit theorems and the proof techniques used therein serve as fundamental building blocks for numerous further developments, including likelihood based inferential methods [11, 13, 23, 36], bootstrap in nonstandard problems [53, 60], estimation and inference with dependence structures [2, 4, 5], limit theory for global loss functions and functionals [25, 26, 33, 47, 54], limit distribution theory for shape-restricted estimators of discrete functions [6–8, 48], limit distribution theory for split points in decision trees [12, 17], cube-root asymptotics in more general settings [4, 52], just to name a few.

Despite the wealth of limit distribution theories for univariate shape-restricted problems, much less is known in multidimensional settings. The only exception we are aware of is the recent work [3], in which asymptotic distributions for isotonized estimators are derived in the settings of multidimensional discrete isotonic regression and probability mass function estimation. The goal of this paper is to study limit theorems for shape-restricted estimators, with a focus on the problem of multiple isotonic regression in a continuous setting.

Here is our setup. Consider the regression model

$$Y_i = f_0(X_i) + \xi_i, \quad i = 1, \ldots, n, \tag{1.3}$$

where $X_1, \ldots, X_n$ are design points which can be either fixed or random, and $\xi_1, \ldots, \xi_n$ are random errors. By multiple isotonic regression, we assume that the regression function $f_0 \in \mathcal{F}_d$, where $\mathcal{F}_d$ denotes the class of coordinate-wise nondecreasing functions on $[0, 1]^d$:

$$\mathcal{F}_d \equiv \big\{ f : [0, 1]^d \to \mathbb{R}, f(x) \le f(y) \text{ if } x_i \le y_i \text{ for all } i = 1, \ldots, d \big\}.$$

In addition to the aforementioned importance of having a limit distribution theory for shape restricted estimators beyond univariate settings, there is one further consideration for such a theory, related to one distinct attractive feature of shape-constrained models: the MLE/LSE often exists and automatically adapts to certain structures of the underlying truth without the need of any tuning. This automatic adaptation property has attracted a lot of recent attention, mostly from a global perspective. Indeed, adaptation of shape constrained MLEs/LSEs to piecewise simple structures in global metrics is confirmed extensively in various univariate models; cf. [14, 19, 40, 51, 65]. Typically, these tuning-free estimators adapt to piecewise constant/linear signals in univariate models with monotonicity/convexity shape constraints. For the multiple isotonic regression model (1.3), global adaptation of the LSE to piecewise constant signals is proved for the bivariate case $d = 2$ in [20], and for the case of general dimensions in [42]. See also [41]. Despite these very positive adaptation results, there remain two important drawbacks for considering global adaptation of the natural LSE in the multiple isotonic regression model:

(D1) The isotonic regression function is of global smoothness level 1 or ∞, so the LSE can adapt, if at all possible, to limited global structures. In fact, piecewise constancy is the only known global structure to which the LSE is confirmed to adapt, cf. [20, 42].

(D2) The LSE does not adapt at the optimal rate to constant signals when $d \geq 3$: it is shown in [42] that the LSE adapts to the global constant structures at a *strictly sub-minimax rate* $n^{-1/d}$ (up to logarithmic factors) in L_2-type losses.

The reasons for these limitations, however, lie in very different places: the drawback in (D1) is due to the perspective of considering global adaptation, while the drawback in (D2) is due to the use of the LSE.

In view of these limitations, in this paper we consider the local behavior of the following alternative max–min block estimator originally proposed by [28]: for any $x_0 \in [0, 1]^d$,

$$\begin{aligned}\hat{f}_n(x_0) &\equiv \max_{x_u \leq x_0} \min_{\substack{x_v \geq x_0 \\ [x_u, x_v] \cap \{X_i\} \neq \varnothing}} \frac{\sum_{i: x_u \leq X_i \leq x_v} Y_i}{|\{i : x_u \leq X_i \leq x_v\}|} \\ &= \max_{x_u \leq x_0} \min_{\substack{x_v \geq x_0 \\ [x_u, x_v] \cap \{X_i\} \neq \varnothing}} \bar{Y}|_{[x_u, x_v]} \\ &= \max_{x_u \leq x_0} \min_{\substack{x_v \geq x_0 \\ [x_u, x_v] \cap \{X_i\} \neq \varnothing}} \bar{Y}|_{[x_u, 1] \cap [0, x_v]}.\end{aligned} \tag{1.4}$$

Here, $[x_u, x_v] = \{x \in \mathbb{R}^d : x_u \leq x \leq x_v\}$, $\bar{Y}|_A$ is the average of $\{Y_i : X_i \in A\}$ as in (1.6), and for any $x, y \in \mathbb{R}^d$, $x \leq y$ if and only if $x_j \leq y_j$ for all $1 \leq j \leq d$, and the similar definition applies to $\geq$. It is easy to see that $\hat{f}_n \in \mathcal{F}_d$ and is tuning-free. The computation for (1.4) is exact and requires at most $\mathcal{O}(n^2)$ for each design point, so the total computational complexity is at most $\mathcal{O}(n^3)$, independent of the dimension d.

The max–min block estimator (1.4) above is closely related to the LSE studied in [42], in the sense that the LSE also admits a max–min representation [59], but with the rectangles $[x_u, 1], [0, x_v]$ replaced by all *upper sets* and *lower sets* containing x_0. Since the upper and lower sets reduce to intervals in dimension one, (1.4) coincides with the standard univariate isotonic LSE in $d = 1$. The representation through upper and lower sets is also observed in a related monotone density estimation problem in $d = 2$; cf. [56].

The max–min representation gives one heuristic explanation for the difficulty of the LSE in the sense of (D2): the class of upper and lower sets is too large for the partial sum process to remain tight in the large sample limit as soon as $d \geq 3$ (cf., [24]). On the other hand, using the smaller class of rectangles as in (1.4), it is shown in [22] that (1.4) does adapt to constant signals at a nearly optimal parametric rate in all dimensions, as opposed to the slower rate $n^{-1/d}$ for the LSE; cf. (D2). For the same reason, it is hard to expect a limiting distribution theory for the LSE.

The main contribution of this paper is to develop a limit distribution theory for the max–min block estimator (1.4). We show that, the limiting distribution of $\hat{f}_n$, depending on the local structure of f_0 at x_0, takes the following general form: Suppose f_0 admits vanishing derivatives up to order α_k along the kth dimension ($k = 1, \ldots, d$) at a fixed point $x_0 \in (0, 1)^d$, and the errors $\{\xi_i\}$ have variance σ^2. Then

$$(1.5) \qquad \big(n_*/\sigma^2\big)^{\frac{1}{2+\sum_{k\in\mathcal{D}_*}\alpha_k^{-1}}}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \mathbb{C}(f_0, x_0).$$

Here, $\mathcal{D}_*$ and n_*, determined by the value of $(\alpha_1, \ldots, \alpha_d)$ and the design of $\{X_i\}$, are the set of all "effective dimensions" and the size of "effective samples" that drive the asymptotic limiting distribution, the exact meaning of which will be clarified in Section 2. The dependence of the limiting distribution $\mathbb{C}(f_0, x_0)$ on the local properties of f_0 at x_0 cannot be in general expressed by a simple factor, due to possible existence of nonzero mixed derivatives of critical order $\boldsymbol{j} = (j_1, \ldots, j_d)$ satisfying $\sum_{k=1}^d j_k/\alpha_k = 1$ and $\|\boldsymbol{j}\|_0 > 1$. However, in situations where $\{\alpha_k\}$ are relative primes to each other (so that any such index vector $\boldsymbol{j}$ must have $\|\boldsymbol{j}\|_0 = 1$), or all mixed derivatives of f_0 of the critical order vanish at x_0, the limiting distribution $\mathbb{C}(f_0, x_0)$ can be represented in a similar form as in (1.1)–(1.2), namely

$$\mathbb{C}(f_0, x_0) =_d K(f_0, x_0) \cdot \mathbb{D}_{\boldsymbol{\alpha}}.$$

Here, $K(f_0, x_0)$ is a constant depending on the local structure of the regression function f_0 at x_0 to be specified in Section 2, and $\mathbb{D}_{\boldsymbol{\alpha}}$ is the nonstandard limiting distribution playing the similar role as the Chernoff distribution $\mathbb{Z}$ in univariate problems.

One important and canonical setting for (1.5) is the following: Suppose (i) f_0 depends only through its first s coordinates ($0 \le s \le d$), and all nontrivial first-order partial derivatives of f_0 are nonvanishing at x_0: $\partial_k f_0(x_0) > 0$, $1 \le k \le s$, and (ii) the design points $\{X_i\}$ are either of a balanced fixed lattice design (see Section 2 for a precise definition) or a random design with uniform distribution on $[0, 1]^d$. In this setting, (1.5) reduces to

$$\big(n/\sigma^2\big)^{\frac{1}{2+s}}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \left\{\prod_{k=1}^{s}\big(\partial_k f_0(x_0)/2\big)\right\}^{\frac{1}{2+s}} \cdot \mathbb{D}_{(\underbrace{1, \ldots, 1}_{s \text{ many 1's}}, \infty, \ldots, \infty)}.$$

When $s = d = 1$, we recover the familiar limit distribution theory for univariate isotonic least squares estimator.

The limit theory in (1.5), as we will see in Section 2, implies that *the max–min block estimator* (1.4) *automatically adapts to the local smoothness structures and the intrinsic dimension of* f_0. The local adaptation is in similar spirit to [9, 21, 64], who showed that univariate shape-restricted MLEs/LSEs adapt to local smoothness of the truth. It should be emphasized here that local smoothness to which adaptation occurs specifically refers to the number of *vanishing* (partial) derivatives. A distinct feature for the max–min block estimator (1.4) here is that both (i) the local rate of convergence, that is, $n_*^{\frac{1}{2+\sum_{k\in\mathcal{D}_*}\alpha_k^{-1}}}$, and (ii) the dependence on $\{f_0, x_0\}$ whenever explicit, that is, the constant $K(f_0, x_0)$ in the limit distribution, are *optimal in a local asymptotic minimax sense for all possible local smoothness levels*. So in this sense the limit distribution theory for the max–min block estimator (1.4) in the form of (1.5) is the best one can hope for in the problem of multiple isotonic regression.

Another interesting consequence of (1.5) and its local asymptotic minimaxity is that *the optimal local rates of convergence are in general not the same in fixed lattice and random designs*. In fact, the local rate in the fixed lattice design case is *no slower* than that in the random design case, and can be much faster when (a) the local smoothness levels of the isotonic regression function, or (b) the sizes of the lattice, differ substantially along different dimensions. The reason for the discrepancy in the local rates can be attributed to the fact that significant imbalance in (a) or (b) screens out dimensions with "low regularity" that do not contribute to the asymptotics in the fixed lattice design case. Here dimensions with "low regularity," loosely speaking, refer to those with low smoothness levels in (a), and to those with sparsely spaced design points in (b).

The proof of the limit theory (1.5) in general dimensions is significantly more challenging than its univariate counterpart. Indeed, thanks to the "switching relation" put forward by [30], it is now well understood that limiting distributions for various univariate monotonicity shape constrained estimators can be obtained via the argmax continuous mapping theory, upon a proper *one-sided* localization (typically) on the order of a cube-root rate (cf. [63]). In contrast, in the multiple isotonic regression problem we consider here, the key step in the proof is a *two-sided* localization technique, where the stochastic orders of the length of all sides of the rectangle over which the max–min block estimator (1.4) takes average, need to be estimated sharply *both from above and below*. These estimates bring about substantial technical challenges as opposed to univariate problems.

Finally, we mention the work of [22], in which global risk bounds in L_q norms for the max–min block estimator (1.4) are thoroughly studied. Risk bounds in global metrics, as already mentioned in (D1), have a limited scope of structures for adaptation due to the strict global smoothness of the isotonic functions. Our local limit distribution theory (1.5) can therefore also be viewed as a further step in understanding the adaptive behavior of the max–min block estimator (1.4) to a rich class of structures that are exhibited only through local properties of the isotonic regression function.

The rest of the paper is organized as follows. In Section 2, we present the limit distribution theory (1.5) for the max–min block estimator (1.4), and discuss its many implications. In Section 3, we establish a local asymptotic minimax lower bound, showing the information-theoretic optimality of the limit theorem (1.5). Due to the highly technical nature of the proofs, Section 4 is devoted to an outline of the main ideas in the proofs. Section 5 concludes the paper with a brief discussion. All the proof details are presented in Section 6 and the Supplementary Material [44].

1.2. *Notation*. For a real-valued measurable function f defined on $(\mathcal{X}, \mathcal{A}, P)$, $\|f\|_{L_p(P)} \equiv \|f\|_{P,p} \equiv (P|f|^p)^{1/p}$ denotes the usual L_p-norm under P, and $\|f\|_\infty \equiv \sup_{x\in\mathcal{X}}|f(x)|$. Let $(\mathcal{F}, \|\cdot\|)$ be a subset of the normed space of real functions $f : \mathcal{X} \to \mathbb{R}$. For $\varepsilon > 0$, let $\mathcal{N}(\varepsilon, \mathcal{F}, \|\cdot\|)$ be the ε-covering number of $\mathcal{F}$; see page 83 of [63] for more details.

For the regression model (1.3), for any $A \subset [0, 1]^d$, define

$$(1.6)\qquad \bar{Y}|_A \equiv \frac{1}{n_A}\sum_{i:X_i\in A} Y_i,\qquad \bar{f}_0|_A \equiv \frac{1}{n_A}\sum_{i:X_i\in A} f_0(X_i),\qquad \bar{\xi}|_A \equiv \frac{1}{n_A}\sum_{i:X_i\in A}\xi_i$$

where $n_A \equiv |\{i : X_i \in A\}|$.

For two real numbers a, b, $a \vee b \equiv \max\{a, b\}$ and $a \wedge b \equiv \min\{a, b\}$. For $x \in \mathbb{R}^d$, let $\|x\|_p$ denote its p-norm ($0 \le p \le \infty$). For any $x, y \in \mathbb{R}^d$, let $[x, y] \equiv \prod_{k=1}^d [x_k \wedge y_k, x_k \vee y_k]$, $xy \equiv (x_k y_k)_{k=1}^d$, and $x \wedge (\vee) y \equiv (x_k \wedge (\vee) y_k)_{k=1}^d$. For $\ell_1, \ell_2 \in \{1, \ldots, d\}$, we let $\mathbf{1}_{[\ell_1:\ell_2]} \in \mathbb{R}^d$ be such that $(\mathbf{1}_{[\ell_1:\ell_2]})_k = \mathbf{1}_{\ell_1\le k\le\ell_2}$, and $\mathbf{1} \equiv \mathbf{1}_{[1:d]}$ for simplicity. C_x will denote a generic finite constant that depends only on a generic quantity x, whose numeric value may change from line to line unless otherwise specified. $a \lesssim_x b$ and $a \gtrsim_x b$ mean $a \le C_x b$ and $a \ge C_x b$, respectively, and $a \asymp_x b$ means $a \lesssim_x b$ and $a \gtrsim_x b$ [$a \lesssim b$ means $a \le Cb$ for some absolute constant C]. $\mathcal{O}_{\mathbf{P}}$ and $\mathfrak{o}_{\mathbf{P}}$ denote the usual big and small O notation in probability. $\rightsquigarrow$ is reserved for weak convergence. For two integers $k_1 > k_2$, we interpret $\sum_{k=k_1}^{k_2} \equiv 0$, $\prod_{k=k_1}^{k_2} \equiv 1$. We also interpret $(\infty)^{-1} \equiv 0$, $0/0 \equiv 0$.

For $f : \mathbb{R}^d \to \mathbb{R}$, and $k \in \{1, \ldots, d\}$, $\alpha_k \in \mathbb{Z}_{\ge 1}$, let $\partial_k^{\alpha_k} f(x) \equiv \frac{d^{\alpha_k}}{dx_k^{\alpha_k}} f(x)$. For a multi-index $\boldsymbol{j} = (j_1, \ldots, j_d) \in \mathbb{Z}_{\ge 0}^d$, let $\partial^{\boldsymbol{j}} \equiv \partial_1^{j_1} \cdots \partial_d^{j_d}$, and $\boldsymbol{j}! \equiv j_1! \cdots j_d!$ and $x^{\boldsymbol{j}} \equiv x_1^{j_1} \ldots x_d^{j_d}$

for $x \in \mathbb{R}^d$. For $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}_{\geq 1}^d$ in Assumption A below, that is, for some $0 \leq s \leq d$, $1 \leq \alpha_1, \ldots, \alpha_s < \infty = \alpha_{s+1} = \cdots = \alpha_d$, let $J(\boldsymbol{\alpha})$ (resp., $J_*(\boldsymbol{\alpha})$) be the set of all $\boldsymbol{j} = (j_1, \ldots, j_d) \in \mathbb{Z}_{\geq 0}^d$ satisfying $0 < \sum_{k=1}^s j_k/\alpha_k \leq 1$ (resp., $\sum_{k=1}^s j_k/\alpha_k = 1$) and $j_k = 0$ for $s+1 \leq k \leq d$, and let $J_0(\boldsymbol{\alpha}) \equiv J(\boldsymbol{\alpha}) \cup \{\mathbf{0}\}$. We often write $J = J(\boldsymbol{\alpha})$, $J_* = J_*(\boldsymbol{\alpha})$ and $J_0 = J_0(\boldsymbol{\alpha})$ if no confusion arises. The set J, J_* will play a crucial role below in determining $\boldsymbol{j}$'s for which $\partial^{\boldsymbol{j}} f_0(x_0)$ can be nonzero under Assumption A; cf. Lemma 1.

2. Limit distribution theory.

2.1. *Assumptions.* We first state the assumptions on the local smoothness of f_0 at the point of interest $x_0 \in (0,1)^d$ and the intrinsic dimension of f_0.

ASSUMPTION A. f_0 is coordinate-wise nondecreasing (i.e., $f_0 \in \mathcal{F}_d$), and is $\boldsymbol{\alpha}$-smooth at x_0 with intrinsic dimension s, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$ with integers $1 \leq \alpha_1, \ldots, \alpha_s < \infty = \alpha_{s+1} = \cdots = \alpha_d$, $0 \leq s \leq d$, in the sense that $\partial_k^{j_k} f_0(x_0) = 0$ for $1 \leq j_k \leq \alpha_k - 1$ and $\partial_k^{\alpha_k} f_0(x_0) \neq 0$, $1 \leq k \leq s$, and in rectangles of the form $\bigcap_{k=1}^d \{|(x - x_0)_k| \leq L_0 \cdot (r_n)_k\}$, $r_n = (\omega_n^{1/\alpha_1}, \ldots, \omega_n^{1/\alpha_d})$ with $\omega_n > 0$, the Taylor expansion of f_0 satisfies for all $L_0 > 0$,

$$\lim_{\omega_n \searrow 0} \omega_n^{-1} \sup_{\substack{x \in [0,1]^d, \\ |(x-x_0)_k| \leq L_0 \cdot (r_n)_k, \\ 1 \leq k \leq d}} \left| f_0(x) - \sum_{\boldsymbol{j} \in J_0} \frac{\partial^{\boldsymbol{j}} f_0(x_0)}{\boldsymbol{j}!} (x - x_0)^{\boldsymbol{j}} \right| = 0.$$

Assumption A concerns the local smoothness of f_0 at a fixed-point x_0, allowing for potentially different local smoothness levels along different coordinates $\{1, \ldots, s\}$. The Taylor expansion, which includes all terms of order ω_n or larger in a small hyperrectangle, interestingly features different rates ω_n^{1/α_k} in different dimensions in the x-domain. This is quite different from the Taylor expansion in Euclidean balls which includes all terms with $\|\boldsymbol{j}\|_1 \leq \alpha$ for a certain smoothness index α. Our expansion has the prescribed convergence rate if f_0 is locally $C^{\max_{1 \leq k \leq s} \alpha_k}$ at x_0 and depends only through its first s coordinates. Note that Assumption A is interesting mostly from a local perspective. Indeed, if this condition holds for all $x_0 \in (0,1)^d$ with some $1 \leq \alpha_1, \ldots, \alpha_s < \infty$, then we must have $\alpha_1 = \cdots = \alpha_s = 1$.

Now we consider a few examples that satisfy Assumption A with different values of $\boldsymbol{\alpha}$'s. In the following examples, we consider $d = 2$ and $x_0 = (1/2, 1/2)$ unless otherwise specified.

EXAMPLE 1. Let $f_0^{(1)}(x_1, x_2) = x_1 + x_2$. Then $\alpha_1 = \alpha_2 = 1$.

EXAMPLE 2. Let $f_0^{(2)}(x_1, x_2) = x_1$. Then $s = 1$ with $\alpha_1 = 1, \alpha_2 = \infty$.

EXAMPLE 3. Let $f_0^{(3)}(x_1, x_2) = (x_1 + x_2)\mathbf{1}_{0 \leq x_1 \leq 1/4} + 8x_1 \cdot \mathbf{1}_{1/4 < x_1 < 3/4} + 8(x_1 + x_2) \times \mathbf{1}_{3/4 \leq x_1 \leq 1}$. Then $s = 1$ with $\alpha_1 = 1, \alpha_2 = \infty$ for $x_0 \in (1/4, 3/4) \times (0,1)$, and $s = 2$ with $\alpha_1 = \alpha_2 = 1$ for $x_0 \in (0, 1/4) \times (0,1) \cup (3/4, 1) \times (0,1)$.

Example 2 is a canonical example for which the regression function is globally of intrinsic dimension 1, while in Example 3 the function can be locally of intrinsic dimension 1 in the strip $(1/4, 3/4) \times (0,1)$.

EXAMPLE 4. Let $f_0^{(4)}(x_1, x_2) = (x_1 - 1/2)^3 + (x_2 - 1/2)^3$. Then $\alpha_1 = 3, \alpha_2 = 3$.

EXAMPLE 5. Let $f_0^{(5)}(x_1, x_2) = (x_1 - 1/2)^3 + (x_1 - 1/2)^2(x_2 - 1/2) + (x_1 - 1/2) \times (x_2 - 1/2)^2 + (x_2 - 1/2)^3$. Then $\alpha_1 = 3, \alpha_2 = 3$.

Example 4 and Example 5 both share the same local smoothness level $\boldsymbol{\alpha} = (3, 3)$, but are quite different in that for $f_0^{(4)}$ all mixed derivatives vanish, while for $f_0^{(5)}$ certain mixed derivatives do no vanish: $\partial^{\boldsymbol{j}} f_0^{(5)}(x_0) \neq 0$ for $\boldsymbol{j} \in \{(1, 2), (2, 1)\}$.

LEMMA 1. *The following statements hold*:

1. *Suppose Assumption A holds.* α_k *must be odd and* $\partial_k^{\alpha_k} f_0(x_0) > 0$ *for* $1 \le k \le s$.
2. *Suppose Assumption A holds. Any mixed derivative of the form* $\partial^{\boldsymbol{j}} f_0(x_0)$, $\mathbf{0} \neq \boldsymbol{j} \in J \setminus J_*$, *vanishes at* x_0, *and thus for some* $\varepsilon_1 > 0$, $L_1 > 0$ *depending only on* f_0, x_0,

$$\sup_{0<\omega_n\le\varepsilon_1} \sup_{\substack{x\in[0,1]^d,\\ |(x-x_0)_k|\le(r_n)_k,\\ 1\le k\le d}} \omega_n^{-1}\left|f_0(x) - f_0(x_0)\right| \le L_1.$$

3. *Let* $J_1 \equiv \{\boldsymbol{j} \in J_* : \|\boldsymbol{j}\|_0 > 1\}$. *Then* $J_1 = \varnothing$ *if and only if* $|J_*| = s$, *if and only if* $\{\alpha_k\}_{k=1}^s$ *is a set of relative primes, that is, the greatest common divisor of* $\{\alpha_{k_1}, \alpha_{k_2}\}$ *is* 1 *for all* $1 \le k_1 < k_2 \le s$.
4. *When* $\boldsymbol{\alpha}$ *is such that* $J_1 \neq \varnothing$, *there exists some* $f \in \mathcal{F}_d$ *for which* f *satisfies Assumption A with* $\boldsymbol{\alpha}$, *but* $\partial^{\boldsymbol{j}} f(x_0) \neq 0$ *for some* $\boldsymbol{j} \in J_1$.

PROOF. See Appendix B in [44]. □

Lemma 1 reveals an important and unique feature of multiple isotonic functions compared with smooth functions: If f_0 satisfies the "marginal smoothness" Assumption A with $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ at x_0, then the only possible nonzero mixed derivatives $\partial^{\boldsymbol{j}} f_0(x_0)$ in the Taylor expansion must have critical order $\boldsymbol{j} \in J_*$ satisfying $\sum_{k=1}^s j_k/\alpha_k = 1$. Such possible nonzero mixed derivatives cannot be ruled out under Assumption A as soon as certain pair of $\{\alpha_k\}$ has a nontrivial common divisor. The importance of such a feature lies in the fact that these mixed derivatives $\{\partial^{\boldsymbol{j}} f_0(x_0) : \boldsymbol{j} \in J_*\}$ contribute to the convergence rate of the same order as the marginal derivatives $\{\partial_k^{\alpha_k} f_0(x_0)\}$, in rectangles of the form $\bigcap_{k=1}^d \{|(x - x_0)_k| \le L_0 \cdot (r_n)_k\}$, $r_n = (\omega_n^{1/\alpha_1}, \dots, \omega_n^{1/\alpha_d})$ with $\omega_n \searrow 0$. Hence adaptation of the max–min estimator (1.4) to marginal smoothness levels—which only uses marginal information in rectangles—becomes possible.

Next we state the assumptions on the design of the covariates.

ASSUMPTION B. The design points $\{X_i\}_{i=1}^n$ satisfy either of the following:

- (*Fixed design*) $\{X_i\}$'s follow a $\boldsymbol{\beta}$-fixed lattice design: there exist some $\{\beta_1, \dots, \beta_d\} \subset (0, 1)$ with $\sum_{k=1}^d \beta_k = 1$ such that $x_0 \in \{X_i\}_{i=1}^n = \prod_{k=1}^d \{x_{1,k}, \dots, x_{n_k,k}\}$, where $\{x_{1,k}, \dots, x_{n_k,k}\}$ are equally spaced in [0, 1] (i.e., $|x_{j,k} - x_{j+1,k}| = 1/n_k$ for all $j = 1, \dots, n_k - 1$) and $n_k = \lfloor n^{\beta_k} \rfloor$.
- (*Random design*) $\{X_i\}$'s follow i.i.d. random design with law P independent of $\{\xi_i\}$'s. The Lebesgue density π of P is bounded away from 0 and ∞ on $[0, 1]^d$ and is continuous over an open set containing the region $\{((x_0)_1, \dots, (x_0)_s, x_{s+1}, \dots, x_d) : 0 \le x_k \le 1, s + 1 \le k \le d\}$.

In the fixed lattice design case, we use β_k to control the size of the lattice in dimension k. A balanced fixed lattice design refers to the special case with $\beta_k = 1/d$ for all $k = 1, \dots, d$.

In the random design case, the continuity of the density π is imposed over the region where asymptotics take place.

We choose without loss of generality the index $\{1,\ldots,d\}$ such that

$$0 \le \alpha_1\beta_1 \le \cdots \le \alpha_s\beta_s \le \cdots \le \alpha_d\beta_d \le \infty. \tag{2.1}$$

This requirement facilitates the statement of our main Theorem 1 below. Otherwise, we may find some permutation τ of $\{1,\ldots,d\}$ for which $\alpha_{\tau(1)}\beta_{\tau(1)} \le \cdots \le \alpha_{\tau(d)}\beta_{\tau(d)}$, and consider the coordinate-wise nondecreasing function $\tilde{f}_0(x_1,\ldots,\ldots,x_d) \equiv f_0(x_{\tau(1)},\ldots,x_{\tau(d)})$. Such a reparametrization is compatible with Assumption A since $\alpha_k\beta_k = \infty$ if and only if $\alpha_k = \infty$.

2.2. *Limit distribution theory.* Let $x_0 \in (0,1)^d$. Let $\pi(x) \equiv 1$ in the $\boldsymbol{\beta}$-fixed lattice design case and $\pi(x) \equiv \mathrm{d}P/(\mathrm{d}x_1\cdots\mathrm{d}x_d)$ be the Lebesgue density of P in the random design case. For any $0 \le s \le d$, and $h_1, h_2 \in \mathbb{R}^d_{\ge 0}$ such that $(h_1)_k \le (x_0)_k$, $(h_2)_k \le (1-x_0)_k$ for all $s+1 \le k \le d$, let

$$\mathcal{I}_\pi^{[s+1:d]}(h_1,h_2) \equiv \int_{\substack{(x_0-h_1)_k \le x_k \le (x_0+h_2)_k \\ s+1\le k\le d}} \pi\big((x_0)_1,\ldots,(x_0)_s, x_{s+1},\ldots,x_d\big)\,\mathrm{d}x_{s+1}\cdots\mathrm{d}x_d,$$

and $\mathcal{I}_\pi^{[d+1:d]} \equiv \pi(x_0)$. The integration above is carried out over the region $\{(x_0-h_1)_k \le x_k \le (x_0+h_2)_k : s+1 \le k \le d\} \subset [0,1]^{d-s}$ with the integrand given by the Lebesgue density π of the design distribution P. In the $\boldsymbol{\beta}$-fixed lattice design case, $\mathcal{I}_\pi^{[s+1:d]}(h_1,h_2) = \prod_{k=s+1}^d (h_1+h_2)_k$.

Let κ_*, n_* be defined in Table 1 below. Let the limit process $\mathbb{C}(f_0,x_0)$ be defined by

$$\begin{aligned}&\mathbb{C}(f_0,x_0)\\ &\equiv \sup_{\substack{h_1>0,\\ (h_1)_k\le(x_0)_k,\\ s+1\le k\le d}} \inf_{\substack{h_2>0,\\ (h_2)_k\le(1-x_0)_k,\\ s+1\le k\le d}} \left[\frac{\mathbb{G}(h_1,h_2)}{\prod_{k=\kappa_*}^s((h_1)_k+(h_2)_k)\mathcal{I}_\pi^{[s+1:d]}(h_1,h_2)} + \bar{f}_0(h_1,h_2;x_0)\right],\end{aligned} \tag{2.2}$$

where $\mathbb{G}$ is a centered Gaussian process defined on $\mathbb{R}^d_{\ge 0}\times\mathbb{R}^d_{\ge 0}$ with the following covariance structure: for any (h_1,h_2), (h_1',h_2'),

$$\begin{aligned}&\mathrm{Cov}\big(\mathbb{G}(h_1,h_2),\mathbb{G}(h_1',h_2')\big)\\ &\quad = \prod_{k=\kappa_*}^s \big((h_1)_k \wedge (h_1')_k + (h_2)_k \wedge (h_2')_k\big)\cdot \mathcal{I}_\pi^{[s+1:d]}(h_1\wedge h_1', h_2\wedge h_2')\end{aligned}$$

and

$$\bar{f}_0(h_1,h_2;x_0) \equiv \sum_{\substack{\boldsymbol{j}\in J_*,\\ j_k=0,1\le k\le \kappa_*-1}} \frac{\partial^{\boldsymbol{j}} f_0(x_0)}{(\boldsymbol{j}+\boldsymbol{1})!}\prod_{k=\kappa_*}^s \frac{(h_2)_k^{j_k+1} - (-h_1)_k^{j_k+1}}{(h_2)_k+(h_1)_k}.$$

TABLE 1
Definitions of κ_*, n_*

	$\boldsymbol{\beta}$-fixed lattice design	Random design
κ_*	$\arg\max_{1\le\ell\le d} \frac{\sum_{k=\ell}^d \beta_k}{2+\sum_{k=\ell}^s \alpha_k^{-1}}$	1
n_*	$n^{\sum_{k=\kappa_*}^d \beta_k}$	n

Furthermore, let $\mathbb{D}_{\boldsymbol{\alpha}}$ be defined by

$$\mathbb{D}_{\boldsymbol{\alpha}} \equiv \sup_{\substack{h_1>0,\\(h_1)_k\leq(x_0)_k,\\s+1\leq k\leq d}} \inf_{\substack{h_2>0,\\(h_2)_k\leq(1-x_0)_k,\\s+1\leq k\leq d}} \Bigg[\frac{\mathbb{G}(h_1,h_2)}{\prod_{k=\kappa_*}^{s}((h_1)_k+(h_2)_k)\mathcal{I}_\pi^{[s+1:d]}(h_1,h_2)} + \sum_{k=\kappa_*}^{s}\frac{(h_2)_k^{\alpha_k+1}-(h_1)_k^{\alpha_k+1}}{(h_2)_k+(h_1)_k}\Bigg]. \tag{2.3}$$

With these definitions, we are now in position to state the main result of this paper.

THEOREM 1. *Suppose Assumptions A–B hold, and the errors $\{\xi_i\}$ are i.i.d. mean-zero with finite variance $\mathbb{E}\xi_1^2=\sigma^2<\infty$ (and are independent of $\{X_i\}$ in the random design case). With κ_*, n_* defined in Table 1, we have the following local rate of convergence*:

$$(n_*/\sigma^2)^{\frac{1}{2+\sum_{k=\kappa_*}^{s}\alpha_k^{-1}}}\big(\hat{f}_n(x_0)-f_0(x_0)\big)=\mathcal{O}_{\mathbf{P}}(1).$$

If κ_* is uniquely defined, with $\mathbb{C}(f_0,x_0)$ defined in (2.2), the following limit theory holds:

$$(n_*/\sigma^2)^{\frac{1}{2+\sum_{k=\kappa_*}^{s}\alpha_k^{-1}}}\big(\hat{f}_n(x_0)-f_0(x_0)\big)\rightsquigarrow\mathbb{C}(f_0,x_0).$$

Furthermore, if either $\{\alpha_k\}$ is a set of relative primes or all mixed derivatives of f_0 vanish at x_0 in J_*, then

$$\mathbb{C}(f_0,x_0)=_d K(f_0,x_0)\cdot\mathbb{D}_{\boldsymbol{\alpha}},$$

where $K(f_0,x_0)=\{\prod_{k=\kappa_*}^{s}(\partial_k^{\alpha_k}f_0(x_0)/(\alpha_k+1)!)^{1/\alpha_k}\}^{\frac{1}{2+\sum_{k=\kappa_*}^{s}\alpha_k^{-1}}}$ and $\mathbb{D}_{\boldsymbol{\alpha}}$ is defined in (2.3).

PROOF. See Section 6. □

REMARK 1. A few technical remarks:

1. In the $\boldsymbol{\beta}$-fixed lattice design case where $\pi(x)=1$ is used to calculate $\mathcal{I}_\pi^{[s+1:d]}(h_1,h_2)$, the covariance structure of $\mathbb{G}$ is simpler:

$$\operatorname{Cov}\big(\mathbb{G}(h_1,h_2),\mathbb{G}(h_1',h_2')\big)=\prod_{k=\kappa_*}^{d}\big((h_1)_k\wedge(h_1')_k+(h_2)_k\wedge(h_2')_k\big), \tag{2.4}$$

and can be represented as follows: let $d_*\equiv d-\kappa_*+1$, and let $\{\mathbb{B}_i : i\in\{1,2\}^{d_*}\}$ be independent Brownian sheets on $\mathbb{R}_{\geq 0}^{d_*}$. For any $h_1,h_2\in\mathbb{R}_{\geq 0}^d$, let $\bar{\mathbb{G}}(h_1,h_2)\equiv\sum_{i\in\{1,2\}^{d_*}}\mathbb{B}_i((h_{i_1})_{\kappa_*},\ldots,(h_{i_{d_*}})_d)$. Then $\mathbb{G}(\cdot,\cdot)=_d\bar{\mathbb{G}}(\cdot,\cdot)$. A similar representation holds in the random design case for $s=d$.

2. $\mathbb{C}(f_0,x_0)$ has at least a sub-Gaussian tail, and hence admits moments of all orders (cf. Lemma 4).

3. When either $\{\alpha_k\}$ is a set of relative primes or all mixed derivatives of f_0 vanish at x_0 in J_*, the following self-similarity property of the process $\mathbb{G}(\cdot,\cdot)$ is essential for the representation $\mathbb{C}(f_0,x_0)=_d K(f_0,x_0)\cdot\mathbb{D}_{\boldsymbol{\alpha}}$: for $\gamma\in\mathbb{R}_{\geq 0}^d$ with $\gamma_1=\cdots=\gamma_{\kappa_*-1}=0,\gamma_{\kappa_*}\ldots,\gamma_s,\gamma_{s+1}=\cdots=\gamma_d=1$, $\mathbb{G}(\gamma\cdot,\gamma\cdot)=_d(\prod_{k=\kappa_*}^{s}\gamma_k)^{1/2}\cdot\mathbb{G}(\cdot,\cdot)$.

4. $\mathbb{D}_{\boldsymbol{\alpha}}$ (and $\mathbb{C}(f_0,x_0)$) can be represented by sup-inf over the summation of a stochastic term plus a nonrandom drift term, similar to that of the Chernoff distribution; see (2.6) below for an explicit derivation of $\mathbb{D}_1$ being the Chernoff distribution.

5. Although implicit in notation, $\mathbb{D}_{\boldsymbol{\alpha}}$ can depend on x_0 through $\mathcal{I}_\pi^{[s+1:d]}$. However, such dependence disappears under (i) the fixed lattice design and (ii) the random design with uniform distribution. For general distributions P in the random design case, if $s = d$ and $\mathbb{C}(f_0, x_0) =_d K(f_0, x_0) \cdot \mathbb{D}_{\boldsymbol{\alpha}}$, the dependence of x_0 within $\mathbb{D}_{\boldsymbol{\alpha}}$ can be assimilated into the constant. In fact, by taking σ to be $\sigma/\sqrt{\pi(x_0)}$ in (6.17), we have

$$\begin{aligned}&\big(\pi(x_0)n/\sigma^2\big)^{\frac{1}{2+\sum_{k=1}^d \alpha_k^{-1}}}\big(\hat{f}_n(x_0) - f_0(x_0)\big)\\&\quad\rightsquigarrow K(f_0, x_0) \sup_{h_1>0}\inf_{h_2>0}\left[\frac{\mathbb{G}(h_1, h_2)}{\prod_{k=1}^d((h_1)_k + (h_2)_k)} + \prod_{k=1}^d \frac{(h_2)_k^{\alpha_k+1} - (h_1)_k^{\alpha_k+1}}{(h_2)_k + (h_1)_k}\right],\end{aligned}$$

where $\mathbb{G}$ is the Gaussian process with the covariance structure (2.4).

Theorem 1 shows that the max–min block estimator (1.4) adapts to the local smoothness levels $\{\alpha_k\}$ and the intrinsic dimension s of the isotonic regression function f_0, in both the fixed lattice and random design settings. One particularly interesting consequence of the above theorem is that *the adaptive local rates for the fixed lattice and random design cases are in general not the same*. Indeed,

$$\begin{aligned}\omega_n^{\text{fixed}} &\equiv n^{-\frac{\sum_{k=\kappa_*}^d \beta_k}{2+\sum_{k=\kappa_*}^s \alpha_k^{-1}}} = n^{-\max_{1\le\ell\le d}\frac{\sum_{k=\ell}^d \beta_k}{2+\sum_{k=\ell}^s \alpha_k^{-1}}},\\ \omega_n^{\text{random}} &\equiv n^{-\frac{1}{2+\sum_{k=1}^s \alpha_k^{-1}}} = n^{-\frac{\sum_{k=1}^d \beta_k}{2+\sum_{k=1}^s \alpha_k^{-1}}},\end{aligned}$$

so that $\omega_n^{\text{fixed}} \le \omega_n^{\text{random}}$, that is, the local rate in the fixed lattice design case is *no slower* than that in the random design case.

The following proposition gives an equivalent definition of κ_* in the fixed lattice design case in Theorem 1.

PROPOSITION 1. *The following are equivalent under* (2.1):

(1) *The maximizer of* $\ell \mapsto \frac{\sum_{k=\ell}^d \beta_k}{2+\sum_{k=\ell}^s \alpha_k^{-1}}$ *is unique and* $\kappa_* = \arg\max_{1\le\ell\le d}\frac{\sum_{k=\ell}^d \beta_k}{2+\sum_{k=\ell}^s \alpha_k^{-1}}$.

(2) *For any* $1 \le \ell \le d$, $\frac{\alpha_\ell^{-1}}{2+\sum_{k=\ell}^s \alpha_k^{-1}} \ne \frac{\beta_\ell}{\sum_{k=\ell}^d \beta_k}$, *and* $\kappa_* = \min\{1 \le \ell \le d : \frac{\alpha_\ell^{-1}}{2+\sum_{k=\ell}^s \alpha_k^{-1}} < \frac{\beta_\ell}{\sum_{k=\ell}^d \beta_k}\} = \min\{1 \le \ell \le d : (\omega_n^{(\ell)})^{1/\alpha_\ell} n^{\beta_\ell} > 1\}$. *Here*, $\omega_n^{(\ell)} \equiv n^{-\frac{\sum_{k=\ell}^d \beta_k}{2+\sum_{k=\ell}^s \alpha_k^{-1}}}$ *is the unique solution of the fixed-point equation*

$$\omega = \frac{1}{\sqrt{\prod_{k=\ell}^d(\omega^{1/\alpha_k} n^{\beta_k})}}. \tag{2.5}$$

PROOF. By algebra, for any relationship $\sim$ in the set $\{<, \le, >, \ge\}$, we have (i) $\frac{\sum_{k=\ell}^d \beta_k}{2+\sum_{k=\ell}^s \alpha_k^{-1}} \sim \frac{\sum_{k=\ell+1}^d \beta_k}{2+\sum_{k=\ell+1}^s \alpha_k^{-1}}$ if and only if (ii) $\frac{\beta_\ell}{\sum_{k=\ell}^d \beta_k} \sim \frac{\alpha_\ell^{-1}}{2+\sum_{k=\ell}^s \alpha_k^{-1}}$ if and only if (iii) $2 \sim \sum_{k=\ell}^d \beta_k(\frac{1}{\alpha_\ell\beta_\ell} - \frac{1}{\alpha_k\beta_k}) \equiv \psi(\ell)$. Under the ordering (2.1), $\ell \mapsto \psi(\ell)$ is nonincreasing, so the statement (1) holds if and only if $\sim$ is taken as $<$ for all $1 \le \ell \le \kappa_* - 1$ and as $>$ for $\kappa_* \le \ell \le d$ in (i), if and only if the same $\sim$ are taken in (ii), if and only if the statement (2) holds. $\square$

Proposition 1(2) shows that κ_* can be determined by a sequence of bias-variance equations in (2.5). This gives an interesting interpretation of the quantities κ_*, n_* in the $\boldsymbol{\beta}$-fixed lattice design case:

• $\kappa_* \ (\leq (s+1) \wedge d)$ can be viewed as a "critical dimension" in the sense that samples in dimensions $\{1, \ldots \kappa_* - 1\}$ do not contribute in the asymptotics of $\hat{f}_n$. In other words, the limit distribution of $\hat{f}_n$ is fully driven by samples in dimensions $\{\kappa_*, \ldots, d\}$. The uniqueness of the maximizer of $\ell \mapsto \frac{\sum_{k=\ell}^d \beta_k}{2+\sum_{k=\ell}^s \alpha_k^{-1}}$ gives a well-defined κ_* and, therefore, the limiting distribution $\mathbb{C}(f_0, x_0)$.

• $n_* = n^{\sum_{k=\kappa_*}^d \beta_k}$ can be viewed as the "effective sample size" over the effective dimensions $\{\kappa_*, \ldots, d\}$ for the asymptotics of $\hat{f}_n$.

In contrast, in the random design case the "critical dimension" κ_* is always $\kappa_* = 1$, as long as the Lebesgue density of P is suitably regular at x_0. In this setting, all dimensions $\{1, \ldots, d\}$ are effective, and the "effective sample size" is simply $n_* = n$.

The local rate of convergence in Theorem 1 can now be interpreted very naturally: the exponent for the "effective sample size" n_*, namely, $\frac{1}{2+\sum_{k=\kappa_*}^s \alpha_k^{-1}}$ becomes the local smoothness along "effective dimensions" $\{\kappa_*, \ldots, d\}$ (note that $\alpha_k^{-1} = 0$ for $s+1 \leq k \leq d$).

REMARK 2. In the special case that f_0 is globally flat, that is, $f_0 \equiv c$ for some $c \in \mathbb{R}$, we have $\alpha_1 = \cdots = \alpha_d = \infty$ and, therefore, the local rate of convergence for the max–min block estimator (1.4) is parametric $\mathcal{O}_{\mathbf{P}}(n^{-1/2})$. When f_0 is locally flat, it is shown by [18] (see also [29]) that in the closely related univariate monotone density estimation problem, the Grenander estimator converges at a parametric rate, with a limiting distribution involving the maximal interval contained in the flat region. In the multivariate case, the shape for locally flat regions can be quite complicated. For example, for $d = 2$ and any upper set $U \subset [1/2, 1]^2$, consider $f_0 \equiv \mathbf{1}_U$. Then the local rate of convergence for the max–min block estimator (1.4) is still parametric at, say, $(1/4, 1/4)$, but the limiting distribution would depend crucially on the exact shape of U. It is an interesting open problem to characterize all possible locally flat regions and derive the corresponding limiting distributions for the max–min block estimator (1.4).

2.3. *Comparison of local rates.*

In this section, we make comparisons of the local rates in different fixed lattice and random designs. As will be seen below, the discrepancy of the local rates appears when either the local smoothness levels of the isotonic regression function, or the sizes of the lattice differ substantially along different dimensions.

2.3.1. *Difference in local rates due to imbalanced local smoothness levels.*

Consider the case where the local smoothness levels are imbalanced with $\alpha_1 = \cdots = \alpha_s = \alpha, \alpha_{s+1} = \cdots = \alpha_d = \infty$ for some $\alpha \geq 1$, $1 \leq s < d$, while the sizes of the lattice are balanced with $\beta_1 = \cdots = \beta_d = 1/d$. By Theorem 1, we have the following corollary:

COROLLARY 1. *Suppose that the assumptions in Theorem 1 hold with $\alpha_1 = \cdots = \alpha_s = \alpha \geq 1, \alpha_{s+1} = \cdots = \alpha_d = \infty$ for some $1 \leq s < d$ and $\beta_1 = \cdots = \beta_d = 1/d$. Then in the fixed lattice design case*,

$$(n/\sigma^2)^{\frac{1}{2+s/\alpha}}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \mathbb{C}(f_0, x_0), \quad \alpha > (d-s)/2;$$

$$(n^{1-s/d}/\sigma^2)^{1/2}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \mathbb{C}(f_0, x_0), \quad \alpha < (d-s)/2.$$

In the random design case,

$$(n/\sigma^2)^{\frac{1}{2+s/\alpha}}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \mathbb{C}(f_0, x_0).$$

The local rates can be written more compactly:

- (*Fixed design*) $\hat{f}_n(x_0) - f_0(x_0) = \mathcal{O}_{\mathbf{P}}(n^{-\max\{\frac{1}{2+s/\alpha}, \frac{1}{2}\cdot\frac{d-s}{d}\}})$.
- (*Random design*) $\hat{f}_n(x_0) - f_0(x_0) = \mathcal{O}_{\mathbf{P}}(n^{-\frac{1}{2+s/\alpha}})$.

Below we consider two scenarios according to the phase transition boundary $\alpha = (d-s)/2$ given above in the balanced fixed lattice design case.

(Scenario 1: $\alpha > (d-s)/2$.) In this case, $\kappa_* = 1$ in the fixed lattice design case, so $\omega_n^{\mathrm{fixed}} = \omega_n^{\mathrm{random}} = n^{-\frac{1}{2+s/\alpha}}$. This includes the important case of $s = d$. In the special case where $d = 1$ and $\alpha_1 = 1$, Corollary 1 reduces to the limit distribution theory for univariate isotonic regression: Suppose for simplicity we consider the fixed balanced lattice design, or the uniform random design. Then $\mathbb{C}(f_0, x_0) =_d K(f_0, x_0) \cdot \mathbb{D}_1 = (f_0'(x_0)/2)^{1/3} \cdot \mathbb{D}_1$, where $\mathbb{D}_1$ is the well-known (rescaled) Chernoff distribution. To see this, with $\mathbb{B}$ denoting the standard two-sided Brownian motion starting at 0, we have (cf. Section 3.3 of [35])

$$
\begin{aligned}
\mathbb{D}_1 &= \sup_{h_1>0}\inf_{h_2>0}\left[\frac{\mathbb{G}(h_1,h_2)}{h_1+h_2} + (h_2 - h_1)\right] \\
&=_d \sup_{-h_1<0}\inf_{h_2>0}\left[\frac{(\mathbb{B}(h_2)+h_2^2) - (\mathbb{B}(-h_1) + (-h_1)^2)}{h_2 - (-h_1)}\right] \qquad (2.6)\\
&= \text{slope at zero of the greatest convex minorant of } t \mapsto \mathbb{B}(t) + t^2 \\
&=_d \text{slope at zero of the least concave majorant of } t \mapsto \mathbb{B}(t) - t^2.
\end{aligned}
$$

It is also interesting to observe that for the most natural case $\alpha_1 = \cdots = \alpha_d = 1$, the local rate is $\mathcal{O}_{\mathbf{P}}(n^{-1/(2+d)})$. This local rate is, somewhat surprisingly, *faster* than the global minimax rate $\mathcal{O}(n^{-1/2d})$ in L_2 metric for $d \geq 3$, cf. [42]. The reason for this is that the global minimax rate in L_2 metric is dominated by the antichain structure of the multiple isotonic regression functions (cf. [42]), while the smoothness constraint rules out such a structure locally at a fixed point. To put the problem in other words, the global minimax rate in L_2 metric is too conservative in capturing the smoothness structure of the isotonic functions as soon as $d \geq 3$.

(Scenario 2: $\alpha < (d-s)/2$.) In this case, $\kappa_* = s+1 > 1$, so the local rate of convergence in the fixed lattice design case is much faster than that in the random design case: $\omega_n^{\mathrm{fixed}} \ll \omega_n^{\mathrm{random}}$.

Let us consider one concrete situation to better understand this phenomenon: $\alpha = 1, s = 1$ and $d > 3$. The local rate is then $\mathcal{O}_{\mathbf{P}}(n^{-\frac{1}{2}\cdot\frac{d-1}{d}})$ in the fixed lattice design case, and is $\mathcal{O}_{\mathbf{P}}(n^{-1/3})$ in the random design case. Suppose for simplicity the regression function $f_0(x_0) = f_0((x_0)_1, \ldots, (x_0)_d) = g_0((x_0)_1)$ for some one-dimensional nondecreasing function g_0. Consider fixed lattice and random design cases separately:

- In the fixed lattice design case, the oracle estimator first takes sample mean in dimensions 2 to d, and then performs isotonic regression in dimension 1 with reduced variance $\sigma_1^2 \equiv \sigma^2/n^{(d-1)/d}$ and sample size $n_1 = n^{1/d}$. However, as long as $d > 3$, there are no longer large samples within the oracle bandwidth $(\sigma_1^2/n_1)^{1/3} = (\sigma^2/n)^{1/3} \ll n^{-1/d}$ in dimension 1 due to the smoothness. This means that the oracle estimator is simply the sample mean over dimensions 2 to d with a convergence rate $n^{-\frac{1}{2}\cdot\frac{d-1}{d}}$ when $d > 3$. See the left panel of Figure 1.
- In the random design case, since the first coordinates of the design points are distinct with probability one, the oracle estimator is the one-dimensional estimator with a bandwidth on the order of $n^{-1/3}$. This gives the usual convergence rate $n^{-1/3}$. See the right panel of Figure 1.

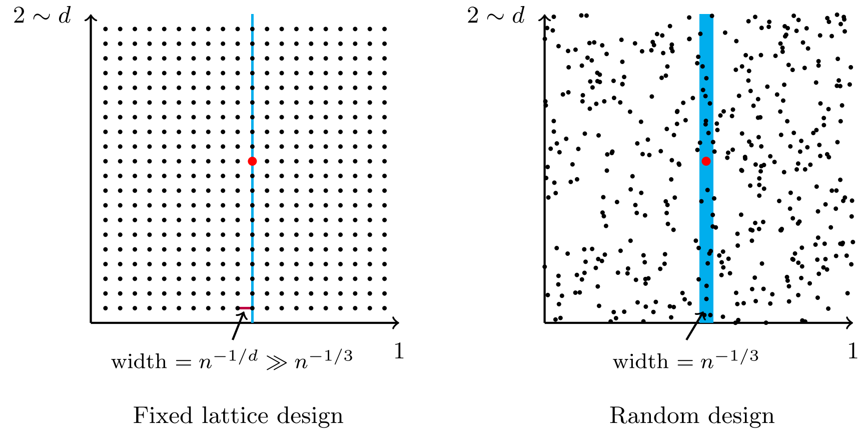

FIG. 1. *Illustration of the oracle estimator in the balanced fixed lattice and random design cases. Horizontal direction = dimension 1. Vertical direction = dimensions 2 to d. Red point = x_0. Blue strip = samples over which the oracle estimator takes the average.*

Corollary 1 with $\alpha = 1, s = 1, d > 3$ can then be understood as saying that the max–min block estimator (1.4) mimics this oracle behavior in terms of the local rate of convergence, in both the fixed lattice and random design cases. In more general settings of Corollary 1, as soon as the local smoothness levels $\alpha < (d - s)/2$, the first s dimensions are screened out in the fixed lattice design case, so the asymptotics only take place over pure noises in dimensions $\{s + 1, \dots, d\}$.

2.3.2. *Difference in local rates due to imbalanced lattice sizes.* Consider the case where the local smoothness levels are balanced with $\alpha_1 = \cdots = \alpha_d = \alpha \geq 1$, while the sizes of the lattice are imbalanced with $\beta_1 \leq \cdots \leq \beta_d$. Using $\sum_{k=1}^d \beta_k = 1$, Theorem 1 and the equivalent definition of κ_* in Proposition 1, we have the following.

COROLLARY 2. *Suppose that the assumptions in Theorem 1 hold with $\alpha_1 = \cdots = \alpha_d = \alpha \geq 1$, and $\beta_1 \leq \cdots \leq \beta_d$.*

In the fixed lattice design case, suppose $\beta_\ell \neq \frac{1-\sum_{k=1}^{\ell-1}\beta_k}{2\alpha+d-\ell+1}$ for all $1 \leq \ell \leq d$. Let $\kappa_* \equiv \min\{1 \leq \ell \leq d : \beta_\ell > \frac{1-\sum_{k=1}^{\ell-1}\beta_k}{2\alpha+d-\ell+1}\}$ and $d_* \equiv d - \kappa_* + 1$. Then

$$\big(n^{\sum_{k=\kappa_*}^d \beta_k}/\sigma^2\big)^{1/(2+d_*/\alpha)}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \mathbb{C}(f_0, x_0).$$

In the random design case,

$$\big(n/\sigma^2\big)^{\frac{1}{2+d/\alpha}}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \mathbb{C}(f_0, x_0).$$

Alternatively, we may write the local rates more compactly:

- (*Fixed design*) $\hat{f}_n(x_0) - f_0(x_0) = \mathcal{O}_{\mathbf{P}}(n^{-\max_{1\leq\ell\leq d}\frac{\sum_{k=\ell}^d \beta_k}{2+(d-\ell+1)/\alpha}})$.
- (*Random design*) $\hat{f}_n(x_0) - f_0(x_0) = \mathcal{O}_{\mathbf{P}}(n^{-\frac{1}{2+d/\alpha}})$.

The basic pattern for the phase transition phenomenon here is similar to the discussion in the previous section. The difference is that now it is the imbalance of the sizes of the lattice in different dimensions, rather than the local smoothness levels, that screens out dimensions with too sparsely spaced design points. See Figure 2 for the concrete phase transition in the line segment $\{\beta_1 + \beta_2 = 1, \beta_1 \leq \beta_2\}$ for $d = 2$ and in the triangle $\{\beta_1 + \beta_2 + \beta_3 = 1, \beta_1 \leq \beta_2 \leq \beta_3\}$ for $d = 3$.

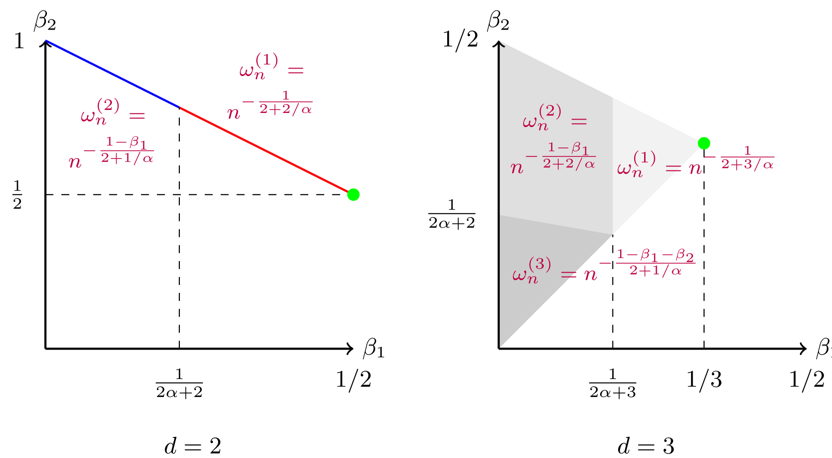

FIG. 2. *Phase transitions of the local rate of convergence for the $\boldsymbol{\beta}$-fixed lattice design in $d=2$ (left panel, in the line segment $\{\beta_1 \le 1/2, \beta_2 = 1-\beta_1\}$) and $d=3$ (right panel, in the triangle $\{\beta_1 \le 1/3, \beta_1 \le \beta_2 \le (1-\beta_1)/2\}$). $\omega_n^{(\ell)}$ = local rate with $\kappa_* = \ell$. Local rates in fixed lattice and random design cases match for $\kappa_* = 1$. Green points = values of (β_1, β_2) for the balanced fixed lattice design.*

REMARK 3. The improvement of the local rate in the fixed lattice design case over the random design case is strongly tied to the lattice structure. It is possible that the local rate in nonlattice fixed design case is slower than that in the random design case. For instance, in the extreme case in $d=2$, suppose that the fixed design points are all located on a straight line $\{(x_1, x_2) \subset [0,1]^2 : x_2 - 1/2 = \theta(x_1 - 1/2), 0 \le x_1 \le 1\}$ with $\theta < 0$. Then we do not have consistent estimation in general as $g(\cdot) \equiv f(\cdot, \theta\cdot)$ may not be monotonically nondecreasing. The situations for general fixed designs will be more complicated. Although we expect that the local rates in “most” fixed design cases will match that in the random design case, it remains an open question to give a complete characterization.

2.4. *An illustrative simulation result.* We present here an illustrative simulation result to assess the accuracy of the distributional approximation in Theorem 1 for the case $d=2$ and $\alpha_1 = \alpha_2 = 1$.

We consider two isotonic regression functions: $f_1(x) = e^{x_1+x_2}$ and $f_2(x) = e(x_1+x_2)$, and $x_0 = (1/2, 1/2)$. Clearly, f_2 is linearization of f_1 at x_0. In particular, the product of the partial derivatives of the two isotonic functions are the same at x_0, that is, $K(f_1, x_0) = K(f_2, x_0)$. Theorem 1 indicates that the limiting distributions for the weighted statistics $n^{1/4}(\hat{f}_n(x_0) - f_i(x_0))$ will be the same for $i = 1, 2$.

We consider the following lattice sizes in the simulation: 15×15, 20×20, 25×25, 30×30. Figure 3 plots the empirical cumulative distribution functions for the weighted statistics $n^{1/4}(\hat{f}_n(x_0) - f_i(x_0))(i = 1, 2)$ based on $B = 300$ repetitions with i.i.d. normal errors $\mathcal{N}(0, 1)$; it shows that even for the quite small lattice of size 15×15, the max–min block estimator (1.4) already achieves reasonable distributional approximation. Not surprisingly, the shapes of the empirical cumulative distribution functions for f_1, f_2 are rather similar. This can be further verified through the QQ-plots for different lattice sizes in Figure 4.

3. Local asymptotic minimax lower bound. We derive in Theorem 1 the precise limiting distribution of the max–min block estimator (1.4) with a local rate of convergence and a limit distribution depending on the unknown smoothness of the regression function at the point of interest. It is natural to wonder if the local rate and the limiting distribution are optimal.

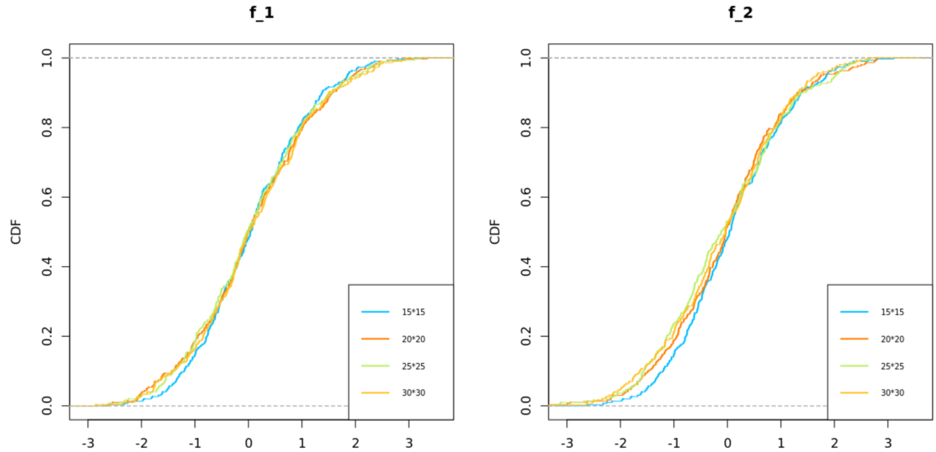

FIG. 3. *Empirical cumulative distribution functions for* $n^{1/4}(\hat{f}_n(x_0) - f_i(x_0))$ $(i = 1, 2)$ *based on* $B = 300$ *repetitions. Here,* $d = 2$, $f_1(x_1, x_2) = e^{x_1 + x_2}$ *in the left panel and* $f_2(x_1, x_2) = e(x_1 + x_2)$ *in the right panel, and* $x_0 = (1/2, 1/2)$. *Lattice sizes*: $15 \times 15, 20 \times 20, 25 \times 25, 30 \times 30$.

THEOREM 2. *Suppose Assumptions A–B hold, and the errors* $\{\xi_i\}$ *are i.i.d.* $\mathcal{N}(0, \sigma^2)$. *Then with* κ_*, n_* *defined in Theorem 1, we have*

$$\sup_{\tau>0} \liminf_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\mathcal{F}_d:\ell_2^2(f,f_0)\le\tau\sigma^2/n} \mathbb{E}_f\big[(n_*/\sigma^2)^{\frac{1}{2+\sum_{k=\kappa_*}^{s}\alpha_k^{-1}}} |\tilde{f}_n(x_0) - f(x_0)|\big]$$
$$\ge L_{d,\|\boldsymbol{\alpha}\|_\infty,P,f_0}.$$

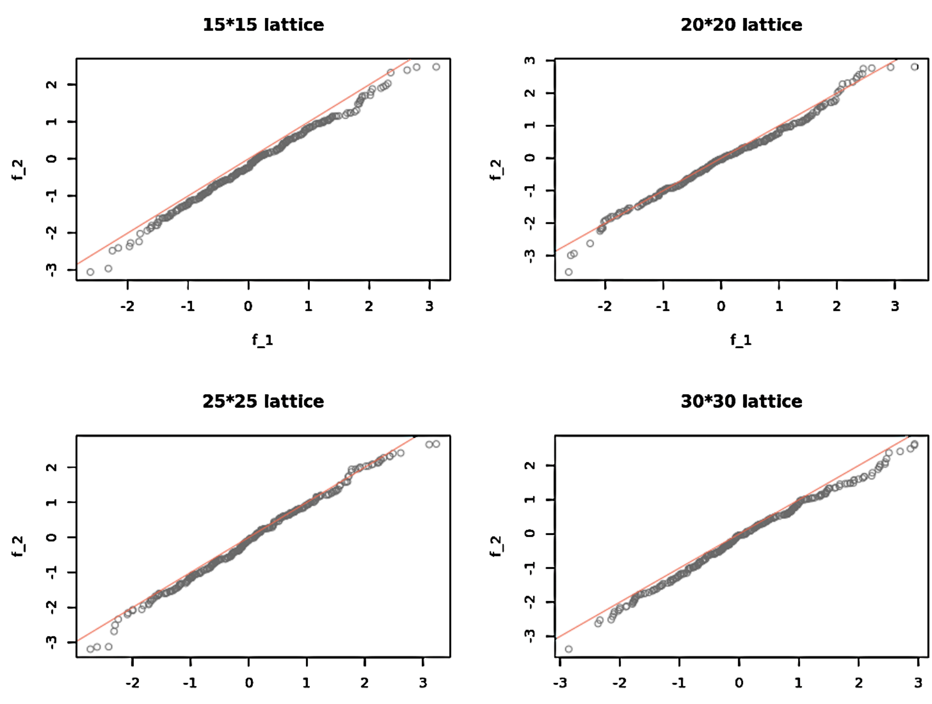

FIG. 4. *QQ-plots for the distributions of* $n^{1/4}(\hat{f}_n(x_0) - f_1(x_0))$ *(horizontal) versus* $n^{1/4}(\hat{f}_n(x_0) - f_2(x_0))$ *(vertical) based on the same simulation as in Figure 3. Lattice sizes*: $15 \times 15, 20 \times 20, 25 \times 25, 30 \times 30$.

Furthermore, if all mixed derivatives of f_0 *vanish at* x_0, *then*

$$\sup_{\tau>0}\liminf_{n\to\infty}\inf_{\tilde f_n}\sup_{f\in\mathcal{F}_d:\ell_2^2(f,f_0)\le\tau\sigma^2/n}\mathbb{E}_f\big[(n_*/\sigma^2)^{\frac{1}{2+\sum_{k=\kappa_*}^s\alpha_k^{-1}}}|\tilde f_n(x_0)-f(x_0)|\big]$$

$$\ge L_{d,\|\boldsymbol{\alpha}\|_\infty,P}\cdot\bigg(\prod_{k=\kappa_*}^s\bigg(\frac{\partial_k^{\alpha_k}f_0(x_0)}{(\alpha_k+1)!}\bigg)^{1/\alpha_k}\bigg)^{\frac{1}{2+\sum_{k=\kappa_*}^s\alpha_k^{-1}}}.$$

Here, $\|\boldsymbol{\alpha}\|_\infty\equiv\max_{\kappa_*\le k\le s}\alpha_k$, $\ell_2^2(f,f_0)\equiv n^{-1}\sum_{i=1}^n(f(X_i)-f_0(X_i))^2$ *for the fixed lattice design case and* $\ell_2^2(f,f_0)\equiv P(f-f_0)^2$ *for the random design case.*

PROOF. See Appendix A in [44]. □

Theorem 2 shows that the max–min block estimator (1.4) enjoys a strong oracle property: Both (i) the adaptive local rate of convergence and (ii) the dependence on the constants, whenever explicit, concerning the unknown regression function f_0 in the limit distribution are optimal in a local asymptotic minimax sense, up to a constant factor depending only on $d,\|\boldsymbol{\alpha}\|_\infty,P$. Note that here the local minimax lower bound is computed over an ℓ_2-ball with radius of order $\mathcal{O}(n^{-1/2})$.

REMARK 4. The dependence of the constant $L_{d,\|\boldsymbol{\alpha}\|_\infty,P}$ on P in the second claim of Theorem 2 can be further improved if $s=d$ in the random design setting: a slight modification of the proof of Theorem 2 shows that $L_{d,\|\boldsymbol{\alpha}\|_\infty,P}=(\pi(x_0))^{-1/(2+\sum_{k=1}^d\alpha_k^{-1})}L_{d,\|\boldsymbol{\alpha}\|_\infty}$. The limit distribution of max–min block estimator (1.4) achieves the optimal dependence on $\pi(x_0)$ in this setting; cf. Remark 1.

REMARK 5. In a related block-decreasing density estimation problem, [55] establishes a local minimax lower bound in the special case of $\alpha_1=\cdots=\alpha_d=1$. Their result and Theorem 2 concern different problems but are similar in spirit. Global minimax lower bounds in L_1 and global risk bounds for histogram-type estimators in the same model are studied in [15].

4. Outline of the proofs.

4.1. *Outline for the proof of Theorem* 1. The proof of Theorem 1 is rather involved, so we highlight the main proof ideas here. Recall that $\kappa_*=1$ in the random design case. Let

$$\omega_n\equiv n_*^{-\frac{1}{2+\sum_{k=\kappa_*}^s\alpha_k^{-1}}},\qquad r_n\equiv(\omega_n^{1/\alpha_1},\ldots,\omega_n^{1/\alpha_d})\mathbf{1}_{[\kappa_*:d]}.\tag{4.1}$$

The components of r_n indicate the localization rate along each dimension. Now we may reparametrize the max–min block estimator (1.4) on the scale of the rate vector r_n. To this end, let $h_1^*,h_2^*\in\mathbb{R}^d_{\ge0}$ be such that

$$\hat f_n(x_0)=\max_{h_1\ge0}\min_{h_2\ge0}\bar Y|_{[x_0-h_1r_n,x_0+h_2r_n]}=\bar Y|_{[x_0-h_1^*r_n,x_0+h_2^*r_n]}.\tag{4.2}$$

We remind the reader that $\bar Y|.$ is the average response over subsets as defined in (1.6). Such a reparametrization relates the problem to its limit Gaussian version. Note that the first κ_*-1 coordinates of r_n is 0, but this is no problem: we will show that such h_1^*,h_2^* exist with high probability.

For any $c > 1$, define the localized max–min block estimator

$$\hat{f}_{n,c}(x_0) \equiv \max_{c^{-\gamma_*}\mathbf{1} \le h_1 \le c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1} \le h_2 \le c\mathbf{1}} \bar{Y}|_{[x_0 - h_1 r_n, x_0 + h_2 r_n]}, \tag{4.3}$$

where $\gamma_* > 0$ is chosen large enough. The difference of (4.3) compared with (4.2) is that the range of max and min in the global estimator (4.2) is restricted to a compact rectangle away from 0 and ∞ in (4.3), so the squared bias and variance in the partial sum process in (4.3) are on the same order. We may therefore expect a nondegenerate limit theory for properly normalized (4.2) by "interchanging" the limit and max–min operations in (4.3) and showing that (4.2) and (4.3) are sufficiently "close" to each other.

Formally, we need the following localization-delocalization result, due to [57].

PROPOSITION 2. *Suppose that three sequences of random variables* $\{W_{n,c}\}$, $\{W_n\}$ *and* $\{W_c\}$ *satisfy the following conditions*:

1. $\lim_{c\to\infty} \limsup_{n\to\infty} \mathbb{P}(W_{n,c} \ne W_n) = 0$.
2. *For every* $c > 0$, $W_{n,c} \rightsquigarrow W_c$ *as* $n \to \infty$.
3. $\lim_{c\to\infty} \mathbb{P}(W_c \ne W) = 0$.

Then $W_n \rightsquigarrow W$ *as* $n \to \infty$.

We choose

$$W_{n,c} \equiv \omega_n^{-1}\big(\hat{f}_{n,c}(x_0) - f_0(x_0)\big), \qquad W_n \equiv \omega_n^{-1}\big(\hat{f}_n(x_0) - f_0(x_0)\big).$$

Now we need to verify the conditions of Proposition 2 for the above defined $W_{n,c}$ and W_n, and find out the limit process W_c.

To illustrate the most important ideas in our proof, we focus on the simplest 2-dimensional balanced fixed lattice design case where $\alpha_1 = \alpha_2 = 1$, $r_n = (n^{-1/4}, n^{-1/4})$ and $\omega_n = n^{-1/4}$ in (4.1); cf. Assumption A.

The first step is to show that the block max–min estimator (4.2) can be localized through its local version (4.3):

$$\lim_{c\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\hat{f}_n(x_0) \ne \hat{f}_{n,c}(x_0)\big) = 0. \tag{4.4}$$

It is easy to see that proving (4.4) reduces to showing that h_1^* and h_2^* are both bounded away from ∞ and 0 in probability, which we refer to as the *large deviation* and *small deviation* problems, respectively. In other words, neither the bias nor the variance of the partial sum process in (4.2) within the block $[x_0 - h_1^* r_n, x_0 + h_2^* r_n]$ will be too large. We accomplish this goal for, for example, h_2^*, in several steps:

(a) First, similarly as in many one-dimensional problems, we establish a local rate of convergence:

$$\hat{f}_n(x_0) - f_0(x_0) = \mathcal{O}_{\mathbf{P}}\big(n^{-1/4}\big).$$

(b) Next, we handle the large deviation problem. In other words, we want to show that $\max\{(h_2^*)_1, (h_2^*)_2\} \le c$ with high probability for large $c > 0$. By the max–min formula,

$$\hat{f}_n(x_0) - f_0(x_0) \ge \underbrace{\big(\bar{f}_0|_{[x_0 - n^{-1/4}, x_0 + h_2^* n^{-1/4}]} - f_0(x_0)\big)}_{\text{bias}} + \underbrace{\bar{\xi}|_{[x_0 - n^{-1/4}, x_0 + h_2^* n^{-1/4}]}}_{\text{noise}}.$$

The bias term can be handled via (localized) Taylor expansion: suppose $\max\{(h_2^*)_1, (h_2^*)_2\} > c \gg 1$ (see, e.g., the thick red and blue lines in the left panel of Figure 5), then

$$\text{bias} \gtrsim n^{-1/4}\big(\max\{(h_2^*)_1, (h_2^*)_2\} - \text{const.}\big) \gtrsim c \cdot n^{-1/4}.$$

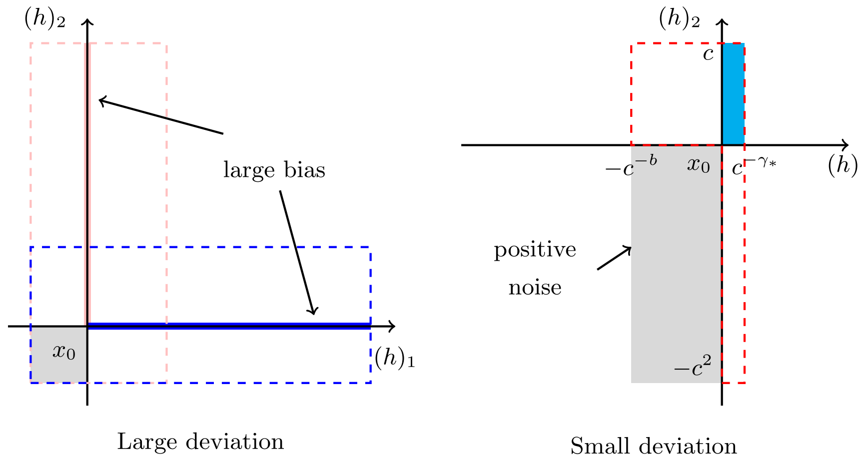

FIG. 5. *Proof outlines*: *Large and small deviation problems*.

The noise term is essentially contributed by the shaded area in the left panel of Figure 5:

$$\text{noise} \gtrsim -|\mathcal{O}_{\mathbf{P}}(1)|n^{-1/4}.$$

Combining the above estimates, if $\max\{(h_2^*)_1, (h_2^*)_2\} > c$, then

$$n^{1/4}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \gtrsim c - |\mathcal{O}_{\mathbf{P}}(1)|,$$

which by (a) should occur with small probability for large $c > 0$.

(c) Finally, we handle the small deviation problem. By (b), we may assume that $\max\{(h_2^*)_1, (h_2^*)_2\} \le c$. Let $0 < b < \gamma_*$ be constants to be determined. Suppose now $(h_2^*)_1 < c^{-\gamma_*}$ (as in the blue strip in the right panel of Figure 5). Using the max–min formula again (but in a different way) and lower bounding the bias yield that

$$\begin{aligned}
\hat{f}_n(x_0) - f_0(x_0) &\ge \max_{\substack{0\le (h_1)_1 \le c^{-b}\\ 0\le (h_1)_2\le c^2}} \bar{\xi}|_{[x_0 - h_1 n^{-1/4}, x_0 + h_2^* n^{-1/4}]} \\
&\quad + \min_{\substack{0\le (h_1)_1 \le c^{-b}\\ 0\le (h_1)_2\le c^2}} \big(\bar{f}_0|_{[x_0 - h_1 n^{-1/4}, x_0 + h_2^* n^{-1/4}]} - f_0(x_0)\big) \\
&\gtrsim n^{-1/4}\bigg[\max_{\substack{0\le (h_1)_1 \le c^{-b}\\ 0\le (h_1)_2\le c^2}} \frac{\mathbb{G}(h_1, h_2^*)}{\prod_{k=1}^2 (h_1 + h_2^*)_k} - C_1 \cdot c^2\bigg].
\end{aligned}$$

The idea here is to choose a larger block for h_1 compared with the localized block for h_2^*. This creates a large positive fluctuation of the noise process $\mathbb{G}(\cdot,\cdot)$ within the shaded region that dominates the relatively small fluctuation within the region in the red dashed line; see the right panel in Figure 5. Indeed, it can be shown that for $\gamma_* < b + 1$, the following small fluctuation holds with high probability: $\mathbb{G}(h_1, h_2^*) - \mathbb{G}(h_1, 0) \ge -C_2 \cdot \sqrt{c^{2-\gamma_*}\log c}$ for large $C_2 > 0$. The scaling $c^{2-\gamma_*}$ is the order of the area of the region within the red dashed line in the right panel of Figure 5. On the other hand, the large positive fluctuation $\max_{0\le (h_1)_1\le c^{-b}, 0\le (h_1)_2\le c^2} \mathbb{G}(h_1, 0) \ge C_3 \cdot \sqrt{c^{2-b}}$ holds for small $C_3 > 0$ with high probability (see the shaded area in the right panel of Figure 5). Therefore, with high probability, for c

large,

$$
\begin{aligned}
n^{1/4}\big(\hat{f}_n(x_0) - f_0(x_0)\big) &\gtrsim \frac{C_3\sqrt{c^{2-b}} - C_2\sqrt{c^{2-\gamma_*}\log c}}{(c^2+c)(c^{-b}+c^{-\gamma_*})} - C_1 c^2 \\
&\geq C_4 c^{(b/2)-1} - C_1 c^2.
\end{aligned}
$$

Now by choosing $b > 6$ and $\gamma_* \in (b, b+1)$, the above display can only occur with small probability for large c by (a), so $(h_2^*)_1 > c^{-\gamma_*}$ with high probability.

Once the first step (4.4) is completed, we may proceed with the second step and conclude that

$$
\begin{aligned}
W_{n,c} &\equiv \max_{c^{-\gamma_*}\mathbf{1}\leq h_1\leq c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1}\leq h_2\leq c\mathbf{1}} \mathbb{U}_n(h_1,h_2) \\
&= \max_{c^{-\gamma_*}\mathbf{1}\leq h_1\leq c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1}\leq h_2\leq c\mathbf{1}} \big[n^{1/4}\bar{\xi}|_{[x_0-h_1 n^{-1/4}, x_0+h_2 n^{-1/4}]} \\
&\quad + n^{1/4}\big(\bar{f}_0|_{[x_0-h_1 n^{-1/4}, x_0+h_2 n^{-1/4}]} - f_0(x_0)\big)\big] \\
&\rightsquigarrow \max_{c^{-\gamma_*}\mathbf{1}\leq h_1\leq c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1}\leq h_2\leq c\mathbf{1}} \Bigg[\frac{\sigma\cdot\mathbb{G}(h_1,h_2)}{\prod_{k=1}^2((h_1)_k+(h_2)_k)} \\
&\quad + \frac{1}{2}\sum_{k=1}^2 \partial_k f_0(x_0)\big((h_2)_k-(h_1)_k\big)\Bigg] \\
&\equiv \max_{c^{-\gamma_*}\mathbf{1}\leq h_1\leq c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1}\leq h_2\leq c\mathbf{1}} \mathbb{U}(h_1,h_2) \equiv W_c.
\end{aligned}
$$

To establish the weak convergence in the above display, we need to establish weak convergence of the process $\mathbb{U}_n$ to $\mathbb{U}$ in $\ell^\infty([c^{-\gamma_*}\mathbf{1}, c\mathbf{1}] \times [c^{-\gamma_*}\mathbf{1}, c\mathbf{1}])$. Finite dimensional convergence follows immediately by the Taylor expansion in Assumption A. Asymptotic equicontinuity will be verified by general tools developed for uniform central limit theorems for partial sum processes (essentially) in [1] and further developed in [63]. Note that such asymptotic equicontinuity is possible as the max–min formula searches over rectangles, rather than upper and lower sets as for the least squares estimator.

The last step attempts at localizing the limit of $W \equiv W_\infty$ by showing that $\tilde{h}_1, \tilde{h}_2$ are bounded away from ∞ and 0 in probability, where $\tilde{h}_1, \tilde{h}_2$ are such that $W = \mathbb{U}(\tilde{h}_1, \tilde{h}_2)$. This problem can be viewed as the limit Gaussian analogue of the first step and, therefore, shares a similar proof strategy as detailed above, but further simplifications are possible due to the exact Gaussian structure in the limit.

Finally, we list the key properties of the partial sum and limit processes that are used in the proofs. For simplicity, we only consider fixed balanced lattice design with $\alpha_1 = \cdots = \alpha_d = 1$. Fix $\rho \in (0,1)$. Let $\omega_{n,\rho} \equiv n^{-\frac{1-\rho}{1+d(1-\rho)}}$, $r_{n,\rho} \equiv \omega_{n,\rho}\mathbf{1}$ and for any $h_1, h_2 \in \mathbb{R}^d_{\geq 0}$, let

$$
\mathbb{G}_{n,\rho}(h_1,h_2) \equiv \omega_{n,\rho}^{-1}\big(n\omega_{n,\rho}^d\big)^{-1} \sum_{i: x_0 - h_1 r_{n,\rho} \leq X_i \leq x_0 + h_2 r_{n,\rho}} \xi_i.
$$

Suppose the following properties hold:

(P1) $\mathbb{G}_{n,\rho} \rightsquigarrow \mathbb{G}_\rho$ in $\ell^\infty([0, c\mathbf{1}] \times [0, c\mathbf{1}])$ for any $c > 0$, and the limit process $\mathbb{G}_\rho(\cdot,\cdot)$ is separable and self-similar with index $\rho \in (0,1)$ in the sense that for any $\gamma \in \mathbb{R}^d_{\geq 0}$, $\mathbb{G}_\rho(\gamma\cdot, \gamma\cdot) =_d (\prod_{k=1}^d \gamma_k)^\rho \cdot \mathbb{G}_\rho(\cdot,\cdot)$.

(P2) It holds that

$$
\sup_{h>0} \frac{|\mathbb{G}_{n,\rho}(h,\mathbf{1})| \vee |\mathbb{G}_{n,\rho}(\mathbf{1},h)|}{\prod_{k=1}^d (h_k+1)} + \sup_{h>0} \frac{|\mathbb{G}_{\rho}(h,\mathbf{1})| \vee |\mathbb{G}_{\rho}(\mathbf{1},h)|}{\prod_{k=1}^d (h_k+1)} = \mathcal{O}_{\mathbf{P}}(1).
$$

(P3) $\mathbb{P}(\sup_{0\le h\le \mathbf{1}} \mathbb{G}_\rho(h,0)\le 0)=\mathbb{P}(\sup_{0\le h\le \mathbf{1}} \mathbb{G}_\rho(0,h)\le 0)=0$.

Then

$$\omega_{n,\rho}^{-1}\big(\hat{f}_n(x_0)-f_0(x_0)\big)\rightsquigarrow K_\rho(f_0,x_0)\cdot \mathbb{D}_{d,\rho},$$

where $K_\rho(f_0,x_0)\equiv\{\prod_{k=1}^d(\partial_k f_0(x_0)/2)\}^{\frac{1-\rho}{1+d(1-\rho)}}$ and

$$\mathbb{D}_{d,\rho}\equiv \sup_{h_1>0}\inf_{h_2>0}\left[\frac{\mathbb{G}_\rho(h_1,h_2)}{\prod_{k=1}^d((h_1)_k+(h_2)_k)}+\sum_{k=1}^d((h_2)_k-(h_1)_k)\right].$$

Some comments on the properties (P1)–(P3):

- (P1) requires a functional limit theory for the process $\mathbb{G}_{n,\rho}$ to its limit $\mathbb{G}_\rho$ over all compact rectangles. The self-similarity index $\rho\in(0,1)$ reflects the dependence structure within the errors $\{\xi_i\}$. In the i.i.d. case as considered in this paper, $\rho=1/2$.
- (P2) requires some uniform control of the moduli of the processes $\mathbb{G}_{n,\rho},\mathbb{G}_\rho$. That the first term in (P2) being stochastically bounded can also be written as

$$\sup_{h>0}\omega_{n,\rho}^{-1}\big(|\bar{\xi}|_{[x_0-hr_{n,\rho},x_0+r_{n,\rho}]}|\vee|\bar{\xi}|_{[x_0-r_{n,\rho},x_0+hr_{n,\rho}]}|\big)=\mathcal{O}_{\mathbf{P}}(1).$$

For i.i.d. errors, this can established by martingale properties of the partial sum process. The stochastic boundedness of the second term in (P2) can be verified by good tail estimates on $\mathbb{G}_\rho$.
- (P3) is a regularity condition on the limit process $\mathbb{G}_\rho$. In the univariate case with i.i.d. errors, this can be easily verified by the reflection principle of Brownian motion.

It is also possible to consider, with substantially increased technicalities, more general assumptions on the local smoothness of f_0 at x_0 and the design points as in Assumptions A–B, but we shall omit these details here.

4.2. *Outline for the proof of Theorem* 2. The basic minimax machinery we use is the following.

PROPOSITION 3. *Suppose that the errors ξ_i's are i.i.d. $\mathcal{N}(0,\sigma^2)$. Let f_n be such that $n\ell_2^2(f_n,f_0)\le\alpha$ and $|f_n(x_0)-f_0(x_0)|\ge\gamma_n$. Then*

$$\inf_{\tilde{f}_n}\sup_{f\in\{f_n,f_0\}}\mathbb{E}_f\big[|\tilde{f}_n(x_0)-f(x_0)|\big]\ge\frac{\gamma_n}{8}\exp\left(-\frac{\alpha}{2\sigma^2}\right).$$

The ℓ_2 metric is defined in the statement of Theorem 2.

Results of this type are well known in the context of density estimation [34, 50], which can also be viewed a special case of minimax reduction scheme with two hypothesis (cf. [61]). We provide a (short) proof in Appendix A in [44] in the context of regression for the convenience of the reader.

Now the problem reduces to that of finding a permissible perturbation f_n. By the second step in the outline for the proof of Theorem 1, the constants concerning the unknown regression function in the limiting distribution essentially come from the local Taylor expansion, so it is tempting to consider the local perturbation function of the form

$$\tilde{f}_n(x)=\begin{cases} f_0(x_0-hr_n) & \text{if } x\in[x_0-hr_n,x_0],\\ f_0(x) & \text{otherwise,}\end{cases}$$

with a good choice of $h\in\mathbb{R}^d$. The complication arises from the fact that $\tilde{f}_n\notin\mathcal{F}_d$ in general, so suitable modifications are needed for a valid construction of f_n. Details can be found in Appendix A in [44].

5. Discussion and final remarks. We developed in this paper pointwise limiting distribution theory for the max–min block estimator (1.4) under local smoothness conditions at a fixed point, and considered both fixed lattice and random designs. One important question is how to use this limit distribution theory for inference. The common bootstrap is known to be inconsistent in a closely related univariate monotone density estimation problem [53, 60], so simple bootstrap procedures are unlikely to succeed here as well. It is an important yet nontrivial task to develop a tuning-free inference method for the multiple isotonic regression model (1.3), even with the limiting distribution theory developed in this paper. A detailed study for the inference problem will therefore be pursued elsewhere.

It should however be mentioned that the limit distribution theory is foundational for further developments of inference methods. In the univariate isotonic regression, the limit distribution theory [16, 64] is essential both in terms of the results and the proof techniques for the popular inference procedure based on likelihood ratio test developed much later (cf. [11, 13]). We expect that the results and proof techniques in this paper will also be useful in this and other directions. However, such developments would not be parallel to the univariate analysis based on the linearity of the monotonicity relationship graph.

From another angle, the limiting distribution theory developed in this paper differs markedly from the univariate cases where usually maximum likelihood/least squares estimators are studied. The common difficulty for these MLE/LSEs in higher dimensions is that the underlying "empirical process" (= partial sum process in (1.4) in our setting) is not tight in the large sample limit, so it is hard to obtain limit distribution theories for these estimators. Our results and techniques can therefore also be viewed as a compliment to the extensive literature on the limit theories for univariate MLE/LSEs.

6. Proof of Theorem 1. Let $\varepsilon_0 > 0$ be such that f_0 is (sufficiently) differentiable on the rectangle $[x_0 - \varepsilon_0 \mathbf{1}, x_0 + \varepsilon_0 \mathbf{1}]$. Recall $\kappa_* = 1$ in the random design case, and $\omega_n \in \mathbb{R}$, $r_n \in \mathbb{R}^d$ are defined in (4.1). Let $d_* \equiv d - \kappa_* + 1$ and $s_* \equiv s - \kappa_* + 1$. We often omit the requirement that $[x_u, x_v] \cap \{X_i\} \neq \varnothing$ in (1.4) for notational simplicity. In the random design setting, we assume that P is the uniform distribution on $[0, 1]^d$ to avoid unnecessary notational digressions; then the covariance structure of $\mathbb{G}$ is given by the simplified expression (2.4).

In the sequel, we consider separately the cases for $1 \le \kappa_* < s + 1$, and $\kappa_* = s + 1$. In the former case, there is at least one nontrivial term in the Taylor expansion of f_0 at x_0, so the problem of limiting distribution is essentially local. In the latter case, since there is only noise present, so the problem is nonlocal.

6.1. *Local rate of convergence.* In this subsection, we establish the local rate of convergence for $\hat{f}_n(x_0)$. This corresponds to step (a) in Section 4.1.

PROPOSITION 4. *Let $1 \le \kappa_* < s + 1$. Assume the same conditions as in Theorem 1. Then $\hat{f}_n(x_0) - f_0(x_0) = \mathcal{O}_{\mathbf{P}}(\omega_n)$.*

We need the following to control the contribution of the noise.

LEMMA 2. *Assume the same conditions as in Theorem 1. Let r_n be as in (4.1). Then for any fixed $\tau > 0$, in both fixed lattice and random design cases, we have $\sup_{h>0} |\bar{\xi}|_{[x_0 - h r_n, x_0 + \tau r_n]}| = \mathcal{O}_{\mathbf{P}}(\omega_n)$.*

PROOF. See Appendix B in [44]. $\square$

PROOF OF PROPOSITION 4. Let x_u^* be such that $\hat{f}_n(x_0) = \min_{x_v \ge x_0} \bar{Y}|_{[x_u^*, x_v]}$. Fix a small enough $\tau > 0$. Since for n large, $[x_0, x_0 + \tau r_n] \cap \{X_i\} \neq \varnothing$ holds in the fixed lattice design case, and with probability tending to one in the random design case, using the max–min formula we have

$$\hat{f}_n(x_0) \le \bar{f}_0|_{[x_u^*, x_0+\tau r_n]} + \bar{\xi}|_{[x_u^*, x_0+\tau r_n]}. \tag{6.1}$$

In both fixed lattice and random design cases, by monotonicity of f_0, we have for n large enough,

$$\begin{aligned} \bar{f}_0|_{[x_u^*, x_0+\tau r_n]} - f_0(x_0) &\le f_0(x_0+\tau r_n) - f_0(x_0) \\ &= \sum_{j \in J_*} \frac{\partial^j f_0(x_0)}{j!}\big(1+\mathfrak{o}(1)\big)(\tau r_n)^j = \mathcal{O}(\omega_n). \end{aligned} \tag{6.2}$$

On the other hand, in both fixed lattice and random design cases, Lemma 2 entails that

$$\sup_{h_1>0} |\bar{\xi}|_{[x_0-h_1 r_n, x_0+\tau r_n]}| = \mathcal{O}_{\mathbf{P}}(\omega_n). \tag{6.3}$$

The one-sided claim follows by combining (6.1)–(6.3). The other direction follows from similar arguments. □

6.2. *Localizing the estimator.* In this subsection, we tackle the large and small deviation problems, that is, steps (b)–(c), as described in Section 4.1.

PROPOSITION 5. *Let $1 \le \kappa_* < s+1$. Assume the same conditions as in Theorem 1 and κ_* is unique. Then*

$$\lim_{c\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\hat{f}_n(x_0) \ne \hat{f}_{n,c}(x_0)\big) = 0.$$

In other words, $\lim_{c\to\infty} \limsup_{n\to\infty} \mathbb{P}(W_{n,c} \ne W_n) = 0$.

PROOF. *(Step 1: the large deviation problem.)* In this step, we handle the large deviation problem. For the proof in this step only, define $r_n \equiv (\omega_n^{1/\alpha_1}, \ldots, \omega_n^{1/\alpha_d})$ for notational convenience. Let h_1^*, h_2^* be defined as in (4.2) but using the current r_n, and

$$H_c \equiv \Big\{h \ge 0 : h_\ell = 0, 1 \le \ell \le \kappa_* - 1, \max_{\kappa_* \le k \le d} h_k \le c\Big\}. \tag{6.4}$$

We will show that

$$\lim_{c\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(h_1^* \notin H_c\big) \vee \mathbb{P}\big(h_2^* \notin H_c\big) = 0. \tag{6.5}$$

We only show this for $\mathbb{P}(h_2^* \notin H_c)$; the situation for h_1^* is similar.

For any h_2^*, let $\bar{h}_2^* \equiv h_2^* \wedge (\varepsilon_0 (r_n)_1^{-1}, \ldots, \varepsilon_0 (r_n)_d^{-1})$. Clearly, $\bar{h}_2^* \le h_2^*$. Then by the max–min formula and the monotonicity of f_0, in the fixed lattice design case it holds for any $h_1 \ge 0$ that

$$\begin{aligned} \hat{f}_n(x_0) &\ge \bar{f}_0|_{[x_0-h_1 r_n, x_0+h_2^* r_n]} + \bar{\xi}|_{[x_0-h_1 r_n, x_0+h_2^* r_n]} \\ &\ge \bar{f}_0|_{[x_0-h_1 r_n, x_0+\bar{h}_2^* r_n]} + \bar{\xi}|_{[x_0-h_1 r_n, x_0+h_2^* r_n]}. \end{aligned} \tag{6.6}$$

(Step 1a: fixed design, effective dimension.) First, we will show that we may take $(h_2^*)_\ell = 0$ for $1 \le \ell \le \kappa_* - 1$ with high probability as $n \to \infty$. Note this is only for the fixed lattice design case. Since we have a lattice design, we only need to show that $(h_2^*)_\ell < n^{-\beta_\ell}(r_n)_\ell^{-1}$ for $1 \le \ell \le \kappa_* - 1$. On the event $\bigcup_{1\le \ell \le \kappa_*-1}\{(h_2^*)_\ell \ge n^{-\beta_\ell}(r_n)_\ell^{-1}\}$, we have by Lemma C.1 that for n large enough,

$$\bar{f}_0|_{[x_0 - r_n \mathbf{1}_{[\kappa_*:d]}, x_0+\bar{h}_2^* r_n]} - f_0(x_0) \gtrsim \omega_n \max_{1\le \ell \le \kappa_*-1} (h_2^*)_\ell^{\alpha_\ell} \ge \omega_n \cdot n^\delta,$$

where the last inequality follows from Proposition 1 that $\min_{1\le\ell\le\kappa_*-1} n^{-\beta_\ell}(r_n)_\ell^{-1} = n^\delta$ for some $\delta > 0$. On the other hand, (a slight modification of) Lemma 2 yields that

$$|\bar{\xi}|_{[x_0-r_n\mathbf{1}_{[\kappa_*:d]},x_0+h_2^* r_n]}| \le \sup_{h_2\ge 0} |\bar{\xi}|_{[x_0-r_n\mathbf{1}_{[\kappa_*:d]},x_0+h_2 r_n]}| = \mathcal{O}_{\mathbf{P}}(\omega_n). \tag{6.7}$$

Combined with (6.6) using $h_1 = (1/2)\mathbf{1}_{[\kappa_*:d]}$, this shows that on the event $\bigcup_{1\le\ell\le\kappa_*-1}\{(h_2^*)_\ell \ge 1\}$,

$$\hat{f}_n(x_0) - f_0(x_0) \gtrsim \omega_n \cdot n^\delta\big(1 - \mathfrak{o}_{\mathbf{P}}(1)\big), \tag{6.8}$$

which can only occur with arbitrarily small probability according to Proposition 4. This means

$$\lim_{n\to\infty} \mathbb{P}\bigg(\bigcup_{1\le\ell\le\kappa_*-1} \{(h_2^*)_\ell \ge 1\}\bigg) = 0. \tag{6.9}$$

Hence we may take $(h_2^*)_\ell = 0$ for $1 \le \ell \le \kappa_* - 1$ from now on. We remind again the reader that this claim is for fixed lattice design only.

(Step 1b: fixed and random designs.) Next, consider the event $\{\max_{\kappa_*\le k\le d}(h_2^*)_k > c\}$. For $c > \max_{s+1\le k\le d}(1-x_0)_k$, we only need to consider the event $\{\max_{\kappa_*\le k\le s}(h_2^*)_k > c\}$. Again using Lemma C.1, for c, n large enough, and with probability at least $1 - \mathcal{O}(n^{-2})$ in the random design case,

$$\bar{f}_0|_{[x_0-r_n\mathbf{1}_{[\kappa_*:d]},x_0+\bar{h}_2^* r_n]} - f_0(x_0) \gtrsim \tilde{\omega}_n \gtrsim \omega_n \cdot c.$$

For the noise term, (6.7) holds in both fixed lattice and random design cases. Hence on the intersection of the event $\{\max_{\kappa_*\le k\le d}(h_2^*)_k > c\}$ and an event with probability tending to 1,

$$\hat{f}_n(x_0) - f_0(x_0) \ge \omega_n\big(c - \mathcal{O}_{\mathbf{P}}(1)\big)$$

holds in both fixed lattice and random design cases. However, in view of Proposition 4, this occurs with arbitrarily small probability for large values of $c > 0, n \in \mathbb{N}$. So the event $\{\max_{\kappa_*\le k\le d}(h_2^*)_k > c\}$ must occur with arbitrarily small probability for c, n large enough. This proves (6.5).

(Step 2: the small deviation problem.) In this step, we handle the small deviation problem. r_n is now defined as in (4.1). Fix $\varepsilon > 0$. By Step 1, we may choose $c > 0, n \in \mathbb{N}$ large enough such that the event $\Omega_{\varepsilon,c}^{(0)} \equiv \{h_2^* \in H_c\}$ holds with probability at least $1 - \varepsilon$. Let $a, b, \gamma_* > 0$ with $a > 1, 0 < b < \gamma_* < b + (a-1)$ be constants to be determined later on, and let $\mathcal{H}_{a,b,\gamma_*}(c) \equiv \{(h_1, h_2) \in \mathbb{R}^d_{\ge 0} \times \mathbb{R}^d_{\ge 0} : 0 \le (h_1)_k \le c^a \mathbf{1}_{\kappa_*\le k\le s} + (x_0)_k \mathbf{1}_{s+1\le k\le d}, 0 \le (h_1)_d \le c^{-b}, 0 \le (h_2)_k \le c\mathbf{1}_{\kappa_*\le k\le s} + (1-x_0)_k\mathbf{1}_{s+1\le k\le d}, 0 \le (h_2)_d \le c^{-\gamma_*}\}$ be defined as in Lemma C.6. Consider the event $\Omega_c^{(1)} \equiv \{(h_2^*)_d < c^{-\gamma_*}\}$. For simplicity of notation, we consider $s < d$; the case $s = d$ follows similarly with slightly different estimates due to Lemma C.6. Let Z_{ni} be defined by

$$Z_{ni}(h_1, h_2) \equiv \omega_n \xi_i \mathbf{1}_{X_i\in[x_0-h_1 r_n, x_0+h_2 r_n]}.$$

It is verified in the proof of Lemma C.6 ahead that for any finite $\tau > 0$,

$$\mathbb{G}_n(\cdot,\cdot) \equiv \sum_{i=1}^n Z_{ni}(\cdot,\cdot) \rightsquigarrow \sigma \cdot \mathbb{G}(\cdot,\cdot) \quad \text{in } \ell^\infty\big([0,\tau\mathbf{1}],[0,\tau\mathbf{1}]\big).$$

Hence by Lemma C.6, as long as $c > 0, n \in \mathbb{N}$ are large enough, there exists a constant $C_1 = C_1(d, \sigma, a)$ such that the event

$$\Omega_\varepsilon^{(2)} \equiv \Big\{\sup_{(h_1,h_2)\in\mathcal{H}_{a,b,\gamma_*}(c)} \big|\mathbb{G}_n(h_1,h_2) - \mathbb{G}_n(h_1, h_2\mathbf{1}_{[s+1:d-1]})\big| \le (C_1/\varepsilon)\sqrt{c^{as_*-\gamma_*}\log c}\Big\}$$

holds with probability at least $1-\varepsilon$. On the other hand, by Lemma C.8, for $u \equiv \sqrt{c^{as_*-b}/(x_0)_d} \cdot \rho_\varepsilon$ where ρ_ε is taken from Lemma C.8, if $a>1$ and $c>1$, it holds for n large enough that

$$\begin{aligned}
&\mathbb{P}\Big(\min_{\substack{0\le (h_2)_k\le c\mathbf{1}_{\kappa_*\le k\le d}\\ 0\le (h_2)_d\le c^{-\gamma_*}}} \max_{\substack{0\le (h_1)_k\le c^a\mathbf{1}_{\kappa_*\le k\le d}\\ 0\le (h_1)_d\le c^{-b}}} \mathbb{G}_n(h_1,h_2\mathbf{1}_{[s+1:d-1]})\le u\Big)\\
&\lesssim \mathbb{P}\Big(\min_{\substack{0\le (h_2)_k\le c\mathbf{1}_{s+1\le k\le d}\\ (h_2)_d=0}} \max_{\substack{0\le (h_1)_k\le c^a\mathbf{1}_{\kappa_*\le k\le d}\\ 0\le (h_1)_d\le c^{-b}}} \mathbb{G}(h_1,h_2)\le u\Big)\\
&\le \mathbb{P}\Big(\min_{\substack{0\le (h_2)_k\le \mathbf{1}_{s+1\le k\le d}\\ (h_2)_d=0}} \max_{\substack{(h_1)_k\le \mathbf{1}_{\kappa_*\le k\le d}\\ (h_1)_d\le (x_0)_d}} \sqrt{c^{as_*-b}/(x_0)_d}\cdot \mathbb{G}(h_1,h_2)\le u\Big)\le\varepsilon.
\end{aligned}$$

Hence there exists some constant $C_2=C_2(x_0,\rho_\varepsilon)>0$ such that the event

$$\begin{aligned}
\Omega_\varepsilon^{(3)} \equiv \big\{&\text{for any } 0\le (h_2)_k\le c\mathbf{1}_{\kappa_*\le k\le d}, 0\le (h_2)_d\le c^{-\gamma_*},\\
&\exists h_1 \text{ with } 0\le (h_1)_k\le c^a\mathbf{1}_{\kappa_*\le k\le d}, 0\le (h_1)_d\le c^{-b}\\
&\text{such that } \mathbb{G}_n(h_1,h_2\mathbf{1}_{[s+1:d-1]})\ge C_2\cdot\sqrt{c^{as_*-b}}\big\}
\end{aligned} \tag{6.10}$$

holds with probability at least $1-\varepsilon$ for n large enough.

(Step 2a: fixed design, noise.) In the fixed lattice design case, on the event $\Omega_{\varepsilon,c}^{(0)}\cap\Omega_c^{(1)}\cap\Omega_\varepsilon^{(2)}\cap\Omega_\varepsilon^{(3)}$, we have for $c>0$, $n\in\mathbb{N}$ large enough,

$$\begin{aligned}
&\omega_n^{-1}\max_{\substack{0\le (h_1)_k\le c^a\mathbf{1}_{\kappa_*\le k\le d}\\ 0\le (h_1)_d\le c^{-b}}} \bar{\xi}|_{[x_0-h_1r_n,x_0+h_2^*r_n]}\\
&=\max_{\substack{0\le (h_1)_k\le c^a\mathbf{1}_{\kappa_*\le k\le d}\\ 0\le (h_1)_d\le c^{-b}}} \frac{\sum_{i=1}^n Z_{ni}(h_1,h_2^*)}{\omega_n^2\prod_{k=\kappa_*}^d(\lfloor (h_1r_n)_k n^{\beta_k}\rfloor+\lfloor (h_2^*r_n)_k n^{\beta_k}\rfloor+1)}\\
&\ge \max_{\substack{0\le (h_1)_k\le c^a\mathbf{1}_{\kappa_*\le k\le d}\\ 0\le (h_1)_d\le c^{-b}}} \frac{\mathbb{G}_n(h_1,h_2^*\mathbf{1}_{[s+1:d-1]})-(C_1/\varepsilon)\sqrt{c^{as_*-\gamma_*}\log c}}{\omega_n^2\prod_{k=\kappa_*}^d(\lfloor (h_1r_n)_k n^{\beta_k}\rfloor+\lfloor (h_2^*r_n)_k n^{\beta_k}\rfloor+1)}\\
&\ge \frac{C_2\sqrt{c^{as_*-b}}-(C_1/\varepsilon)\sqrt{c^{as_*-\gamma_*}\log c}}{(c^a+c)^{s_*}(c^{-b}+c^{-\gamma_*})(1+\mathfrak{o}(1))}\ge C_3\cdot c^{(b-as_*)/2}.
\end{aligned} \tag{6.11}$$

(Step 2b: random design, noise.) In the random design case, let $h_1(c),h_2(c)$ be such that $(h_1(c))_k=c^a\mathbf{1}_{\kappa_*\le k\le s}+(x_0)_k\mathbf{1}_{s+1\le k\le d-1}+c^{-b}\mathbf{1}_{k=d}$, and $(h_2(c))_k=c\mathbf{1}_{\kappa_*\le k\le s}+(1-x_0)_k\mathbf{1}_{s+1\le k\le d-1}+c^{-\gamma_*}\mathbf{1}_{k=d}$. Using Bernstein's inequality,

$$\mathbb{P}\big(\big|(\mathbb{P}_n-P)\mathbf{1}_{X\in[x_0-h_1(c)r_n,x_0+h_2(c)r_n]}\big|\ge\sigma_c^2\big)\le Ce^{-C^{-1}n\sigma_c^2},$$

where $\sigma_c^2\equiv\mathrm{Var}(\mathbf{1}_{X\in[x_0-h_1(c)r_n,x_0+h_2(c)r_n]})\asymp c^{as_*-b}\prod_{k=\kappa_*}^d(r_n)_k$ for c,n large. Hence the event

$$\begin{aligned}
\Omega_c^{(4)}\equiv\big\{&\mathbb{P}_n\mathbf{1}_{X\in[x_0-h_1(c)r_n,x_0+h_2(c)r_n]}\\
&\le P\mathbf{1}_{X\in[x_0-h_1(c)r_n,x_0+h_2(c)r_n]}+\sigma_c^2\big\}
\end{aligned} \tag{6.12}$$

occurs with probability at least $1 - Ce^{-C^{-1}n\sigma_c^2}$. So, in the random design setting, on the event $\Omega_{\varepsilon,c}^{(0)} \cap \Omega_c^{(1)} \cap \Omega_\varepsilon^{(2)} \cap \Omega_\varepsilon^{(3)} \cap \Omega_c^{(4)}$, it holds that

$$\begin{aligned}
&\omega_n^{-1} \max_{\substack{0 \le (h_1)_k \le c^a \mathbf{1}_{\kappa_* \le k \le d} \\ 0 \le (h_1)_d \le c^{-b}}} \bar{\xi}|_{[x_0 - h_1 r_n, x_0 + h_2^* r_n]} \\
&= \max_{\substack{0 \le (h_1)_k \le c^a \mathbf{1}_{\kappa_* \le k \le d} \\ 0 \le (h_1)_d \le c^{-b}}} \frac{\sum_{i=1}^n Z_{ni}(h_1, h_2^*)}{\omega_n^2 (1 \vee n\mathbb{P}_n \mathbf{1}_{X \in [x_0 - h_1 r_n, x_0 + h_2^* r_n]})} \qquad (6.13) \\
&\ge \frac{C_2 \sqrt{c^{as_* - b}} - (C_1/\varepsilon)\sqrt{c^{as_* - \gamma_*} \log c}}{(c^a + c)^{s_*}(c^{-b} + c^{-\gamma_*}) + (1 + \mathfrak{o}(1))c^{as_* - b}} \ge C_3 \cdot c^{(b - as_*)/2}.
\end{aligned}$$

(Step 2c: fixed and random designs, bias.) On the other hand, in both fixed lattice and random design cases,

$$\begin{aligned}
&\min_{\substack{0 \le (h_1)_k \le c^a \mathbf{1}_{\kappa_* \le k \le d} \\ 0 \le (h_1)_d \le c^{-b}}} \bar{f}_0|_{[x_0 - h_1 r_n, x_0 + h_2^* r_n]} - f_0(x_0) \qquad (6.14) \\
&\ge f_0\big(x_0 - c^a \mathbf{1}_{[\kappa_*:d]} r_n\big) - f_0(x_0) \ge -C_4 \cdot c^{a \max_{\kappa_* \le k \le s} \alpha_k} \cdot \omega_n.
\end{aligned}$$

Combining the estimates (6.11), (6.13) and (6.14), we see that for fixed $\varepsilon > 0$, if $c > 0, n \in \mathbb{N}$ are chosen large enough, on the intersection of the event $\{(h_2^*)_d < c^{-\gamma_*}\}$ and an event with probability at least $1 - 4\varepsilon$, it holds that

$$\begin{aligned}
\hat{f}_n(x_0) - f_0(x_0) &\ge \max_{\substack{0 \le (h_1)_k \le c^a \mathbf{1}_{\kappa_* \le k \le d} \\ 0 \le (h_1)_d \le c^{-b}}} \bar{\xi}|_{[x_0 - h_1 r_n, x_0 + h_2^* r_n]} \\
&\quad + \min_{\substack{0 \le (h_1)_k \le c^a \mathbf{1}_{\kappa_* \le k \le d} \\ 0 \le (h_1)_d \le c^{-b}}} \bar{f}_0|_{[x_0 - h_1 r_n, x_0 + h_2^* r_n]} - f_0(x_0) \\
&\ge \omega_n c^{a \max_{\kappa_* \le k \le s} \alpha_k}\big(C_3 \cdot c^{(b - as_*)/2 - a \max_{\kappa_* \le k \le s} \alpha_k} - C_4\big).
\end{aligned}$$

We choose $a \ge 3, b \ge 2(1 + a \max_{\kappa_* \le k \le s} \alpha_k) + as_*$ and $\gamma_* = b + 1$, so that the above display can only occur with arbitrarily small probability by Proposition 4 for large $c > 0, n \in \mathbb{N}$. Hence the event $\{(h_2^*)_d < c^{-\gamma_*}\}$, and thereby $\{\min_{\kappa_* \le k \le d}(h_2^*)_k < c^{-\gamma_*}\}$, must occur with arbitrarily small probability for large $c > 0$. The small deviation for h_1^* can be handled similarly so we omit the details. $\square$

6.3. *Compact convergence.* In this subsection, we establish the weak convergence of the localized process $W_{n,c}$ to the localized limit W_c.

PROPOSITION 6. *Let $1 \le \kappa_* < s + 1$. Assume the same conditions as in Theorem 1. For any $c > 1$,*

$$\begin{aligned}
W_{n,c} &\equiv \omega_n^{-1}\big(\hat{f}_{n,c}(x_0) - f_0(x_0)\big) \\
&\rightsquigarrow \max_{c^{-\gamma_*}\mathbf{1} \le h_1 \le c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1} \le h_2 \le c\mathbf{1}} \Bigg[\frac{\sigma \cdot \mathbb{G}(h_1, h_2)}{\prod_{k=\kappa_*}^d ((h_1)_k + (h_2)_k)} \\
&\quad + \sum_{\substack{\boldsymbol{j} \in J_*, \\ j_k = 0, 1 \le k \le \kappa_* - 1}} \frac{\partial^{\boldsymbol{j}} f_0(x_0)}{(\boldsymbol{j} + \mathbf{1})!} \prod_{k=\kappa_*}^s \frac{(h_2)_k^{j_k+1} - (-h_1)_k^{j_k+1}}{(h_2)_k + (h_1)_k}\Bigg] \\
&\equiv W_c.
\end{aligned}$$

Here, σ and $\mathbb{G}$ are specified in Theorem 1.

We need the following functional central limit theorem.

LEMMA 3. *For any* $h_1, h_2 > 0$, *let*

$$\mathbb{G}_n(h_1,h_2) \equiv \omega_n \sum_{i: x_0 - h_1 r_n \le X_i \le x_0 + h_2 r_n} \xi_i.$$

Then for any $c > 1$, $\mathbb{G}_n \rightsquigarrow \sigma \cdot \mathbb{G}$ *in* $\ell^\infty([c^{-\gamma_*}\mathbf{1}, c\mathbf{1}] \times [c^{-\gamma_*}\mathbf{1}, c\mathbf{1}])$.

PROOF. See Appendix B in [44]. □

PROOF OF PROPOSITION 6. Note that in the fixed lattice design case,

$$\begin{aligned} W_{n,c} &= \max_{c^{-\gamma_*}\mathbf{1} \le h_1 \le c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1} \le h_2 \le c\mathbf{1}} \frac{\omega_n^{-1}}{\prod_{k=\kappa_*}^d (\lfloor (h_1 r_n)_k n^{\beta_k} \rfloor + \lfloor (h_2 r_n)_k n^{\beta_k}\rfloor + 1)} \\ &\quad \times \Big[\sum_{i: x_0 - h_1 r_n \le X_i \le x_0 + h_2 r_n} \xi_i + \sum_{i: x_0 - h_1 r_n \le X_i \le x_0 + h_2 r_n} \big(f_0(X_i) - f_0(x_0)\big) \Big] \\ &= \max_{c^{-\gamma_*}\mathbf{1} \le h_1 \le c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1} \le h_2 \le c\mathbf{1}} \Bigg[\frac{\mathbb{G}_n(h_1,h_2)}{\prod_{k=\kappa_*}^d ((h_1)_k + (h_2)_k)} \cdot \big(1 + \mathfrak{o}(1)\big) \\ &\quad + \sum_{\substack{\boldsymbol{j} \in J_*, \\ j_k = 0, 1 \le k \le \kappa_* - 1}} \frac{\partial^{\boldsymbol{j}} f_0(x_0)}{(\boldsymbol{j}+\mathbf{1})!} \prod_{k=\kappa_*}^s \frac{(h_2)_k^{j_k+1} - (-h_1)_k^{j_k+1}}{(h_2)_k + (h_1)_k} \Bigg] + \mathfrak{o}(1). \end{aligned}$$

Here, the last equality follows from Lemma C.1: for $c^{-\gamma_*}\mathbf{1} \le h_1, h_2 \le c\mathbf{1}$,

$$\begin{aligned} &\frac{\sum_{i: x_0 - h_1 r_n \le X_i \le x_0 + h_2 r_n} (f_0(X_i) - f_0(x_0))}{\prod_{k=\kappa_*}^d (\lfloor (h_1 r_n)_k n^{\beta_k} \rfloor + \lfloor (h_2 r_n)_k n^{\beta_k}\rfloor + 1)} \\ &\quad = \mathfrak{o}(\omega_n) + \omega_n \sum_{\substack{\boldsymbol{j} \in J_*, \\ j_k = 0, 1 \le k \le \kappa_* - 1}} \frac{\partial^{\boldsymbol{j}} f_0(x_0)}{(\boldsymbol{j}+\mathbf{1})!} \prod_{k=\kappa_*}^s \frac{(h_2)_k^{j_k+1} - (-h_1)_k^{j_k+1}}{(h_2)_k + (h_1)_k}. \end{aligned}$$

Since the map $\max_{c^{-\gamma_*}\mathbf{1} \le h_1 \le c\mathbf{1}} \min_{c^{-\gamma_*}\mathbf{1} \le h_2 \le c\mathbf{1}} : \mathbb{R}^{[c^{-\gamma_*}\mathbf{1}, c\mathbf{1}] \times [c^{-\gamma_*}\mathbf{1}, c\mathbf{1}]} \to \mathbb{R}$ is continuous with respect to $\|\cdot\|_\infty$ on $[c^{-\gamma_*}\mathbf{1}, c\mathbf{1}] \times [c^{-\gamma_*}\mathbf{1}, c\mathbf{1}]$, the claim of the proposition for the fixed lattice design case follows by Lemma 3 and the continuous mapping theorem. The random design case follows from similar arguments by using Lemma C.2. □

6.4. *Localizing the limit.* In this subsection, we establish that the limit W can be localized through W_c in the sense described in Section 4.1.

PROPOSITION 7. *Let* $1 \le \kappa_* < s + 1$. *For* $c > 1$, *let* W_c *be as in Proposition* 6, *and* $W \equiv W_\infty$. *Then* $\lim_{c \to \infty} \mathbb{P}(W_c \ne W) = 0$.

To prove Proposition 7, we need the following.

LEMMA 4. *Let* W *be defined as in Proposition* 7, *then* $\|W\|_{\psi_2} < \infty$. *Here*, $\|\cdot\|_{\psi_2}$ *is the sub-Gaussian Orcliz norm* (*definition*; *see*, *e.g.*, [63]).

LEMMA 5. *Let $\mathbb{G}$ be defined as in Theorem 1. Then for any $u \geq 1$,*

$$\mathbb{P}\left(\max_{h_1>0} \frac{|\mathbb{G}(h_1, \mathbf{1})|}{\prod_{k=\kappa_*}^d ((h_1)_k + 1)} > u\right) \leq C_d \exp(-u^2/C_d).$$

Here, $C_d > 0$ is a constant depending only on d.

PROOFS. See Appendix B in [44]. □

PROOF OF PROPOSITION 7. For simplicity of notation, we assume $\sigma = 1$ without loss of generality and set $a_{\boldsymbol{j}} \equiv \partial^{\boldsymbol{j}} f_0(x_0)/(\boldsymbol{j}+\mathbf{1})!$. The strategy of the proof largely follows that of Proposition 5, but with some simplifications. Let

$$\text{(6.15)}\quad \mathbb{U}(h_1, h_2) \equiv \frac{\mathbb{G}(h_1, h_2)}{\prod_{k=\kappa_*}^d ((h_1)_k + (h_2)_k)} + \sum_{\substack{\boldsymbol{j} \in J_*, \\ j_k = 0, 1 \leq k \leq \kappa_*-1}} a_{\boldsymbol{j}} \prod_{k=\kappa_*}^s \frac{(h_2)_k^{j_k+1} - (-h_1)_k^{j_k+1}}{(h_2)_k + (h_1)_k}.$$

Let $h_1^*, h_2^* \in \mathbb{R}^d_{\geq 0}$ be such that $W = \mathbb{U}(h_1^*, h_2^*)$. Since the Gaussian process $\mathbb{G}$ only depends on the last d_* coordinates of its arguments, we may assume that $(h_i^*)_\ell = 0$ for $1 \leq \ell \leq \kappa_* - 1$ and $i = 1, 2$.

(Step 1.) We will first show that

$$\text{(6.16)}\qquad \lim_{c\to\infty} \mathbb{P}(h_1^* \notin H_c^*) \vee \mathbb{P}(h_2^* \notin H_c^*) = 0,$$

where H_c^* is defined in (6.4). We only need to prove that $\{\max_{\kappa_* \leq k \leq s} (h_2^*)_k \leq c\}$ holds with large probability for c large. Using a similar argument for the inequality as in the proof of Lemma C.1, on the event $\{\max_{\kappa_* \leq k \leq s} (h_2^*)_k > c\}$,

$$\begin{aligned} W \geq \mathbb{U}(\mathbf{1}, h_2^*) &= \frac{\mathbb{G}(\mathbf{1}, h_2^*)}{\prod_{k=\kappa_*}^d ((h_2^*)_k + 1)} + \sum_{\substack{\boldsymbol{j} \in J_*, \\ j_k = 0, 1 \leq k \leq \kappa_*-1}} a_{\boldsymbol{j}} \prod_{k=\kappa_*}^s \frac{(h_2^*)_k^{j_k+1} - (-1)^{j_k+1}}{(h_2^*)_k + 1} \\ &\geq - \sup_{h_2 \geq 0} \frac{|\mathbb{G}(\mathbf{1}, h_2)|}{\prod_{k=\kappa_*}^d ((h_2)_k + 1)} + \mathcal{O}\Big(\max_{\kappa_* \leq k \leq s} \big(1 \vee (h_2^*)_k^{\alpha_k}\big)\Big) \\ &\geq -\big|\mathcal{O}_{\mathbf{P}}(1)\big| + \mathcal{O}(c). \end{aligned}$$

The last inequality follows from Lemma 5. On the other hand, by Lemma 4 we know that $W = \mathcal{O}_{\mathbf{P}}(1)$, this means that necessarily $\lim_{c\to\infty} \mathbb{P}(h_2^* \notin H_c^*) = 0$. Similarly, we can show that $\lim_{c\to\infty} \mathbb{P}(h_1^* \notin H_c^*) = 0$, thereby proving the claim (6.16).

(Step 2.) Next, we handle the small deviation problem. Using similar arguments as in the proof of Proposition 5, on the intersection of the event $\{(h_2^*)_d < c^{-\gamma_*}\}$ and an event with probability at least $1 - 4\varepsilon$, it holds that

$$\begin{aligned} W &\geq \max_{\substack{0 \leq (h_1)_k \leq c^a \mathbf{1}_{\kappa_* \leq k \leq d} \\ 0 \leq (h_1)_d \leq c^{-b}}} \min_{\substack{0 \leq (h_2)_k \leq c \mathbf{1}_{\kappa_* \leq k \leq d} \\ 0 \leq (h_2)_d \leq c^{-\gamma_*}}} \frac{\mathbb{G}(h_1, h_2)}{\prod_{k=\kappa_*}^d (h_1 + h_2)_k} - C_4 \cdot c^{a \max_{\kappa_* \leq k \leq s} \alpha_k} \\ &\geq c^{a \max_{\kappa_* \leq k \leq s} \alpha_k} \big(C_3 \cdot c^{(b - as_*)/2 - a \max_{\kappa_* \leq k \leq s} \alpha_k} - C_4\big) \to \infty \end{aligned}$$

as $c \to \infty$ by choosing $a \geq 3$, $b \geq 2(1 + a \max_{\kappa_* \leq k \leq s} \alpha_k) + as_*$ and $\gamma_* = b + 1$. The claim follows from Lemma 4. □

6.5. *Completion of the proof of Theorem 1 for* $1 \le \kappa_* < s+1$. PROOF OF THEOREM 1. By Proposition 2 combined with Propositions 5-7, it follows that $\omega_n^{-1}(\hat{f}_n(x_0) - f_0(x_0))$ converges in distribution to the desired random variable (up to a scaling factor of σ). Hence we only need to verify the distributional equality in the statement of the theorem, when all mixed derivatives of f_0 vanish at x_0 in J_*. To this end, let $\mathbb{U}$ be defined as in the proof of Proposition 7 in (6.15) (with all mixed derivative terms vanishing). Then for $\gamma_0, \gamma_1 = \cdots = \gamma_{\kappa_*-1} = 0, \gamma_{\kappa_*} \ldots, \gamma_s, \gamma_{s+1} = \cdots = \gamma_d = 1$ such that

$$\gamma_0 \left(\prod_{k=\kappa_*}^{d} \gamma_k\right)^{-1/2} = \sigma, \qquad \gamma_0 \gamma_k^{\alpha_k} = \frac{\partial_k^{\alpha_k} f_0(x_0)}{(\alpha_k+1)!}, \quad \kappa_* \le k \le s. \tag{6.17}$$

We have

$$\begin{aligned}
&\gamma_0 \cdot \mathbb{U}\big((\gamma_k (h_1)_k)_{k=1}^d, (\gamma_k (h_2)_k)_{k=1}^d\big) \\
&= \gamma_0 \cdot \frac{\mathbb{G}((\gamma_k (h_1)_k)_{k=1}^d, (\gamma_k (h_2)_k)_{k=1}^d)}{\prod_{k=\kappa_*}^d \gamma_k \cdot \prod_{k=\kappa_*}^d ((h_1)_k + (h_2)_k)} + \gamma_0 \sum_{k=\kappa_*}^{s} \gamma_k^{\alpha_k} \frac{(h_2)_k^{\alpha_k+1} - (h_1)_k^{\alpha_k+1}}{(h_2)_k + (h_1)_k} \\
&=_d \left[\gamma_0 \left(\prod_{k=\kappa_*}^{d} \gamma_k\right)^{-1/2}\right] \cdot \frac{\mathbb{G}(h_1, h_2)}{\prod_{k=\kappa_*}^d ((h_1)_k + (h_2)_k)} + \sum_{k=\kappa_*}^{s} (\gamma_0 \gamma_k^{\alpha_k}) \frac{(h_2)_k^{\alpha_k+1} - (h_1)_k^{\alpha_k+1}}{(h_2)_k + (h_1)_k} \\
&= \frac{\sigma \cdot \mathbb{G}(h_1, h_2)}{\prod_{k=\kappa_*}^d ((h_1)_k + (h_2)_k)} + \sum_{k=\kappa_*}^{s} \frac{\partial_k^{\alpha_k} f_0(x_0)}{(\alpha_k+1)!} \frac{(h_2)_k^{\alpha_k+1} - (h_1)_k^{\alpha_k+1}}{(h_2)_k + (h_1)_k}.
\end{aligned}$$

This shows that under the choice (6.17),

$$\begin{aligned}
&\omega_n^{-1}(\hat{f}_n(x_0) - f_0(x_0)) \\
&\quad \rightsquigarrow \sup_{\substack{h_1>0,\\ (h_1)_k \le (x_0)_k,\\ s+1\le k\le d}} \inf_{\substack{h_2>0,\\ (h_2)_k \le (1-x_0)_k,\\ s+1\le k\le d}} \gamma_0 \cdot \mathbb{U}\big((\gamma_k (h_1)_k)_{k=1}^d, (\gamma_k (h_2)_k)_{k=1}^d\big) \\
&\quad = \sup_{\substack{h_1>0,\\ (h_1)_k \le (x_0)_k,\\ s+1\le k\le d}} \inf_{\substack{h_2>0,\\ (h_2)_k \le (1-x_0)_k,\\ s+1\le k\le d}} \gamma_0 \cdot \mathbb{U}(h_1, h_2).
\end{aligned}$$

Finally, we only need to note that solving (6.17) yields that

$$\gamma_0 = \left(\sigma^2 \prod_{k=\kappa_*}^{s} \left(\frac{\partial_k^{\alpha_k} f_0(x_0)}{(\alpha_k+1)!}\right)^{1/\alpha_k}\right)^{\frac{1}{2+\sum_{k=\kappa_*}^{s} \alpha_k^{-1}}}.$$

This completes the proof. $\square$

6.6. *Proof of Theorem 1 for* $\kappa_* = s+1$. PROOF OF THEOREM 1. The strategy of the proof follows the general principle developed for the case $\kappa_* < s+1$, so we only provide a sketch for the fixed lattice design case.

First, by similar arguments as in the proof of Proposition 4, we can establish a local rate of convergence

$$n_*^{1/2}\big(\hat{f}_n(x_0) - f_0(x_0)\big) = \mathcal{O}_{\mathbf{P}}(1). \tag{6.18}$$

Second, note that $(h_1^*)_k \le (x_0)_k, (h_2^*)_k \le (1-x_0)_k$ for $s+1 \le k \le d$ so there is no large deviation problem. For the small deviation problem, let $2 < b < \gamma_*$ be some fixed constants, and consider the event $\Omega_c^{(1)} \equiv \{(h_2^*)_d < c^{-\gamma_*}\}$. We may show, similar to Lemma C.6, that there exists some $C_1 = C_1(\sigma, d)$ such that for $c > 0, n \in \mathbb{N}$ large enough, the event

$$\Omega_\varepsilon^{(2)} \equiv \Big\{ \sup_{\substack{0\le (h_1)_k \le (x_0)_k \mathbf{1}_{s+1\le k\le d}\\ 0\le (h_1)_d \le c^{-b}\\ 0\le (h_2)_k\le (1-x_0)_k\mathbf{1}_{s+1\le k\le d}\\ (h_2)_d\le c^{-\gamma_*}}} \big|\mathbb{G}_n(h_1,h_2) - \mathbb{G}_n(h_1, h_2\mathbf{1}_{[1:d-1]})\big| \le (C_1/\varepsilon)\sqrt{c^{-\gamma_*}\log c}\Big\}$$

holds with probability at least $1-\varepsilon$ for n large enough. By Lemma C.8, there exists some $C_2 = C_2(\varepsilon) > 0$ such that the event $\Omega_\varepsilon^{(3)} \equiv \{$for any $0 \le (h_2)_k \le (1-x_0)_k\mathbf{1}_{s+1\le k\le d}, (h_2)_d = 0$, there exists $0 \le (h_1)_k \le (x_0)_k\mathbf{1}_{s+1\le k\le d}, (h_1)_d \le c^{-b}$ such that $\mathbb{G}_n(h_1, h_2\mathbf{1}_{[1:d-1]}) \ge C_2\sqrt{c^{-b}}\}$ holds with probability $1-\varepsilon$ for n large enough. On the event $\Omega_c^{(1)} \cap \Omega_\varepsilon^{(2)} \cap \Omega_\varepsilon^{(3)}$, we have

$$\begin{aligned}
&n_*^{1/2}\big(\hat f_n(x_0) - f_0(x_0)\big)\\
&\quad\ge n_*^{1/2} \max_{\substack{0\le (h_1)_k\le (x_0)_k\mathbf{1}_{s+1\le k\le d}\\ 0\le (h_1)_d\le c^{-b}}} \bar\xi|_{[x_0-h_1\mathbf{1}_{[s+1:d]}, x_0+h_2^*\mathbf{1}_{[s+1:d]}]}\\
&\quad\ge \max_{\substack{0\le (h_1)_k\le (x_0)_k\mathbf{1}_{s+1\le k\le d}\\ 0\le (h_1)_d\le c^{-b}}} \frac{\mathbb{G}_n(h_1, h_2^*\mathbf{1}_{[1:d-1]}) - (C_1/\varepsilon)\sqrt{c^{-\gamma_*}\log c}}{n_*^{-1}\cdot \prod_{k=s+1}^d (\lfloor (h_1 r_n)_k n^{\beta_k}\rfloor + \lfloor (h_2^* r_n)_k n^{\beta_k}\rfloor + 1)}\\
&\quad\ge \frac{C_2\sqrt{c^{-b}} - (C_1/\varepsilon)\sqrt{c^{-\gamma_*}\log c}}{(c^{-b}+c^{-\gamma_*})(1+\mathfrak{o}(1))}.
\end{aligned}$$

Hence for $c > 0, n \in \mathbb{N}$ large enough, on the intersection of $\Omega_c^{(1)}$ and an event with probability at least $1-2\varepsilon$,

$$n_*^{1/2}\big(\hat f_n(x_0) - f_0(x_0)\big) \ge C_3 \cdot c^{b/2},$$

where $C_3 = C_3(C_1, C_2, \varepsilon)$. However, by (6.18), this cannot occur with high probability for large $c > 0$. This concludes the small deviation problem. The rest of the proofs parallels that in the case $\kappa_* < s+1$ so we omit details. □

Acknowledgments. Q. Han is supported in part by NSF Grant DMS-1916221. C.-H. Zhang is supported in part by NSF Grants DMS-1513378, DMS-1721495, IIS-1741390 and CCF-1934924. The authors would like to thank four referees for offering a large number of suggestions that significantly improve the presentation of the article.

SUPPLEMENTARY MATERIAL

Supplement: Additional proofs (DOI: 10.1214/19-AOS1928SUPP; .pdf). In the Supplementary Material [44], we provide detailed proofs for Theorem 2 and all supporting lemmas.

REFERENCES

[1] ALEXANDER, K. S. (1987). Central limit theorems for stochastic processes under random entropy conditions. *Probab. Theory Related Fields* **75** 351–378. MR0890284 https://doi.org/10.1007/BF00318707

[2] ANEVSKI, D. and HÖSSJER, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34** 1874–1930. MR2283721 https://doi.org/10.1214/009053606000000443

[3] ANEVSKI, D. and PASTUKHOV, V. (2018). The asymptotic distribution of the isotonic regression estimator over a general countable pre-ordered set. *Electron. J. Stat.* **12** 4180–4208. MR3890765 https://doi.org/10.1214/18-EJS1507

[4] ANEVSKI, D. and SOULIER, P. (2011). Monotone spectral density estimation. *Ann. Statist.* **39** 418–438. MR2797852 https://doi.org/10.1214/10-AOS804

[5] BAGCHI, P., BANERJEE, M. and STOEV, S. A. (2016). Inference for monotone functions under short- and long-range dependence: Confidence intervals and new universal limits. *J. Amer. Statist. Assoc.* **111** 1634–1647. MR3601723 https://doi.org/10.1080/01621459.2015.1100622

[6] BALABDAOUI, F., DUROT, C. and KOLADJO, F. (2017). On asymptotics of the discrete convex LSE of a p.m.f. *Bernoulli* **23** 1449–1480. MR3624867 https://doi.org/10.3150/15-BEJ754

[7] BALABDAOUI, F. and JANKOWSKI, H. (2016). Maximum likelihood estimation of a unimodal probability mass function. *Statist. Sinica* **26** 1061–1086. MR3559943

[8] BALABDAOUI, F., JANKOWSKI, H., RUFIBACH, K. and PAVLIDES, M. (2013). Asymptotics of the discrete log-concave maximum likelihood estimator and related applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 769–790. MR3091658 https://doi.org/10.1111/rssb.12011

[9] BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075 https://doi.org/10.1214/08-AOS609

[10] BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k-monotone density: Limit distribution theory and the spline connection. *Ann. Statist.* **35** 2536–2564. MR2382657 https://doi.org/10.1214/009053607000000262

[11] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693 https://doi.org/10.1214/009053606000001578

[12] BANERJEE, M. and MCKEAGUE, I. W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.* **35** 543–574. MR2336859 https://doi.org/10.1214/009053606000001415

[13] BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743 https://doi.org/10.1214/aos/1015345959

[14] BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. MR3782383 https://doi.org/10.1214/17-AOS1566

[15] BIAU, G. and DEVROYE, L. (2003). On the risk of estimates for block decreasing densities. *J. Multivariate Anal.* **86** 143–165. MR1994726 https://doi.org/10.1016/S0047-259X(02)00028-3

[16] 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

[17] BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927–961. MR1926165 https://doi.org/10.1214/aos/1031689014

[18] CAROLAN, C. and DYKSTRA, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canad. J. Statist.* **27** 557–566. MR1745821 https://doi.org/10.2307/3316111

[19] 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 https://doi.org/10.1214/15-AOS1324

[20] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788 https://doi.org/10.3150/16-BEJ865

[21] CHEN, Y. and WELLNER, J. A. (2016). On convex least squares estimation when the truth is linear. *Electron. J. Stat.* **10** 171–209. MR3466180 https://doi.org/10.1214/15-EJS1098

[22] DENG, H. and ZHANG, C.-H. (2018). Isotonic regression in multi-dimensional spaces and graphs. Preprint. Available at arXiv:1812.08944.

[23] DOSS, C. R. and WELLNER, J. A. (2019). Inference for the mode of a log-concave density. *Ann. Statist.* **47** 2950–2976. MR3988778 https://doi.org/10.1214/18-AOS1770

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

[25] DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699 https://doi.org/10.1214/009053606000001497

[26] DUROT, C., KULIKOV, V. N. and LOPUHAÄ, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Ann. Statist.* **40** 1578–1608. MR3015036 https://doi.org/10.1214/12-AOS1015

[27] DUROT, C. and LOPUHAÄ, H. P. (2018). Limit theory in monotone function estimation. *Statist. Sci.* **33** 547–567. MR3881208 https://doi.org/10.1214/18-STS664

[28] FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2017). On integrated L_1 convergence rate of an isotonic regression estimator for multivariate observations. Preprint. Available at arXiv:1710.04813.

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

[30] 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). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052

[31] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738

[32] GROENEBOOM, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1994). *Lecture Notes in Math*. **1648** 67–164. Springer, Berlin. MR1600884 https://doi.org/10.1007/BFb0095675

[33] GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist*. **27** 1316–1347. MR1740109 https://doi.org/10.1214/aos/1017938928

[34] GROENEBOOM, P. and JONGBLOED, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann. Statist*. **23** 1518–1542. MR1370294 https://doi.org/10.1214/aos/1176324310

[35] 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 https://doi.org/10.1017/CBO9781139020893

[36] GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist*. **43** 2019–2054. MR3375875 https://doi.org/10.1214/15-AOS1335

[37] 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 https://doi.org/10.1214/aos/1015345957

[38] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist*. **29** 1653–1698. MR1891742 https://doi.org/10.1214/aos/1015345958

[39] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5

[40] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621 https://doi.org/10.1007/s00440-014-0595-3

[41] HAN, Q. (2019). Global empirical risk minimizers with "shape constraints" are rate optimal in general dimensions. Preprint. Available at arXiv:1905.12823.

[42] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist*. **47** 2440–2471. MR3988762 https://doi.org/10.1214/18-AOS1753

[43] HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann. Statist*. **44** 1332–1359. MR3485962 https://doi.org/10.1214/15-AOS1408

[44] HAN, Q. and ZHANG, C.-H. (2020). Supplement to "Limit distribution theory for multiple isotonic regression." https://doi.org/10.1214/19-AOS1928SUPP

[45] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat*. **22** 3–33. MR1334065

[46] HUANG, Y. and ZHANG, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist*. **22** 1256–1274. MR1311975 https://doi.org/10.1214/aos/1176325628

[47] JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist*. **42** 625–653. MR3210981 https://doi.org/10.1214/13-AOS1196

[48] JANKOWSKI, H. K. and WELLNER, J. A. (2009). Estimation of a discrete monotone distribution. *Electron. J. Stat*. **3** 1567–1605. MR2578839 https://doi.org/10.1214/09-EJS526

[49] JANKOWSKI, H. K. and WELLNER, J. A. (2009). Nonparametric estimation of a convex bathtub-shaped hazard function. *Bernoulli* **15** 1010–1035. MR2597581 https://doi.org/10.3150/09-BEJ202

[50] JONGBLOED, G. (2000). Minimax lower bounds and moduli of continuity. *Statist. Probab. Lett*. **50** 279–284. MR1792307 https://doi.org/10.1016/S0167-7152(00)00104-8

[51] KIM, A. K. H., GUNTUBOYINA, A. and SAMWORTH, R. J. (2018). Adaptation in log-concave density estimation. *Ann. Statist*. **46** 2279–2306. MR3845018 https://doi.org/10.1214/17-AOS1619

[52] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist*. **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[53] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen*. *Inst. Math. Stat.* (*IMS*) *Collect*. **1** 282–292. IMS, Beachwood, OH. MR2462212 https://doi.org/10.1214/193940307000000202

[54] KULIKOV, V. N. and LOPUHAÄ, H. P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *Ann. Statist*. **33** 2228–2255. MR2211085 https://doi.org/10.1214/009053605000000462

[55] PAVLIDES, M. G. (2012). Local asymptotic minimax theory for block-decreasing densities. *J. Statist. Plann. Inference* **142** 2322–2329. MR2911847 https://doi.org/10.1016/j.jspi.2012.03.003

[56] POLONIK, W. (1998). The silhouette, concentration functions and ML-density estimation under order restrictions. *Ann. Statist*. **26** 1857–1877. MR1673281 https://doi.org/10.1214/aos/1024691360

[57] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā*, *Ser. A* **31** 23–36. MR0267677

[58] PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann*. *Math*. *Stat*. **41** 507–519. MR0260133 https://doi.org/10.1214/aoms/1177697091

[59] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

[60] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann*. *Statist*. **38** 1953–1977. MR2676880 https://doi.org/10.1214/09-AOS777

[61] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[62] VAN EEDEN, C. (1958). *Testing and Estimating Ordered Parameters of Probability Distributions*. Mathematical Centre, Amsterdam. MR0102874

[63] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

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

[65] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann*. *Statist*. **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864