ESTIMATION AND INFERENCE FOR MINIMIZER AND MINIMUM OF CONVEX FUNCTIONS: OPTIMALITY, ADAPTIVITY AND UNCERTAINTY PRINCIPLES

BY T. TONY CAI[a], RAN CHEN[b] AND YUANCHENG ZHU[c]

Department of Statistics and Data Science, The Wharton School, University Of Pennsylvania, [a]*tcai@wharton.upenn.edu,* [b]*ran1chen@wharton.upenn.edu,* [c]*yuancheng.zhu@gmail.com*

Optimal estimation and inference for both the minimizer and minimum of a convex regression function under the white noise and nonparametric regression models are studied in a nonasymptotic local minimax framework, where the performance of a procedure is evaluated at individual functions. Fully adaptive and computationally efficient algorithms are proposed and sharp minimax lower bounds are given for both the estimation accuracy and expected length of confidence intervals for the minimizer and minimum.

The nonasymptotic local minimax framework brings out new phenomena in simultaneous estimation and inference for the minimizer and minimum. We establish a novel uncertainty principle that provides a fundamental limit on how well the minimizer and minimum can be estimated simultaneously for any convex regression function. A similar result holds for the expected length of the confidence intervals for the minimizer and minimum.

1. Introduction. Motivated by a range of applications, estimation of and inference for the location and size of the extremum of a nonparametric regression function has been a long-standing problem in statistics; see, for example, Blum (1954), Chen (1988), Kiefer and Wolfowitz (1952). The problem has been investigated in different settings. For fixed design, upper bounds for estimating the minimum over various smoothness classes have been obtained (Müller (1989), Facer and Müller (2003), Shoung and Zhang (2001)). Belitser, Ghosal and van Zanten (2012) and establishes the minimax rate of convergence over a given smoothness class for estimating both the minimizer and minimum. For sequential design, the minimax rate for estimation of the location has been established; see Chen, Huang and Huang (1996), Dippon (2003), Polyak and Tsybakov (1990). Mokkadem and Pelletier (2007) introduce a companion for the Kiefer–Wolfowitz–Blum algorithm in sequential design for estimating both the minimizer and minimum.

Another related line of research is the stochastic continuum-armed bandits, which have been used to model online decision problems under uncertainty. Applications include online auctions, web advertising and adaptive routing. Stochastic continuum-armed bandits can be viewed as aiming to find the maximum of a nonparametric regression function through a sequence of actions. The objective is to minimize the expected total regret, which requires the trade-off between exploration of new information and exploitation of historical information; see, for example, Auer, Ortner and Szepesvári (2007), Kleinberg (2004), Kleinberg, Slivkins and Upfal (2019).

In the present paper, we consider optimal estimation and confidence intervals for the minimizer and minimum of convex functions under both the white noise and nonparametric regression models in a nonasymptotic local minimax framework that evaluates the performance

of any procedure at individual functions. This framework provides a much more precise analysis than the conventional minimax theory, which evaluates the performance of the estimators and confidence intervals in the worst case over a large collection of functions. This framework also brings out new phenomena in simultaneous estimation and inference for the minimizer and minimum.

We first focus on the white noise model, which is given by

$$\text{(1.1)} \qquad dY(t) = f(t)\,dt + \varepsilon\,dW(t), \quad 0 \le t \le 1,$$

where $W(t)$ is a standard Brownian motion, and $\varepsilon > 0$ is the noise level. The drift function f is assumed to be in $\mathcal{F}$, the collection of convex functions defined on $[0, 1]$ with a unique minimizer $Z(f) = \arg\min_{0 \le t \le 1} f(t)$. The minimum value of the function f is denoted by $M(f)$, that is, $M(f) = \min_{0 \le t \le 1} f(t) = f(Z(f))$. The goal is to optimally estimate $Z(f)$ and $M(f)$, as well as construct optimal confidence intervals for $Z(f)$ and $M(f)$. Estimation and inference for the minimizer $Z(f)$ and minimum $M(f)$ under the nonparametric regression model will be discussed later in Section 4.

1.1. *Function-specific benchmarks and uncertainty principle.* As the first step toward evaluating the performance of a procedure at individual convex functions in $\mathcal{F}$, we define the function-specific benchmarks for estimation of the minimizer and minimum, respectively, by

$$\text{(1.2)} \qquad R_z(\varepsilon; f) = \sup_{g \in \mathcal{F}} \inf_{\hat{Z}} \max_{h \in \{f,g\}} \mathbb{E}_h \big| \hat{Z} - Z(h) \big|,$$

$$\text{(1.3)} \qquad R_m(\varepsilon; f) = \sup_{g \in \mathcal{F}} \inf_{\hat{M}} \max_{h \in \{f,g\}} \mathbb{E}_h \big| \hat{M} - M(h) \big|.$$

As in (1.2) and (1.3), we use subscript "z" to denote quantities related to the minimizer and "m" for the minimum throughout the paper. For any given $f \in \mathcal{F}$, the benchmarks $R_z(\varepsilon; f)$ and $R_m(\varepsilon; f)$ quantify the estimation accuracy at f of the minimizer $Z(f)$ and minimum $M(f)$ against the hardest alternative to f within the function class $\mathcal{F}$.

We show that $R_z(\varepsilon; f)$ and $R_m(\varepsilon; f)$ are the right benchmarks for capturing the estimation accuracy at individual functions in $\mathcal{F}$ and will construct adaptive procedures that simultaneously perform within a constant factor of $R_z(\varepsilon; f)$ and $R_m(\varepsilon; f)$ for all $f \in \mathcal{F}$. In addition, it is also shown that any estimator $\hat{Z}$ for the minimizer that is "super-efficient" at some $f_0 \in \mathcal{F}$, that is, it significantly outperforms the benchmark $R_z(\varepsilon; f_0)$, must pay a penalty at another function $f_1 \in \mathcal{F}$, and thus no procedure can uniformly outperform the benchmark. An analogous result holds for the minimum.

More interestingly, the nonasymptotic local minimax framework enables us to establish a novel uncertainty principle for estimating the minimizer and minimum of a convex function. The uncertainty principle reveals an intrinsic tension between the task of estimating the minimizer and that of estimating the minimum. That is, there is a fundamental limit to the estimation accuracy of the minimizer and minimum for all functions in $\mathcal{F}$, and consequently, the minimizer and minimum of a convex function cannot be estimated accurately at the same time. More specifically, it is shown that

$$\text{(1.4)} \qquad R_z(\varepsilon; f) \cdot R_m(\varepsilon; f)^2 \asymp \varepsilon^2$$

for all $f \in \mathcal{F}$. Further, on the lower bound side,

$$\text{(1.5)} \qquad \inf_{f \in \mathcal{F}} R_z(\varepsilon; f) \cdot R_m(\varepsilon; f)^2 \ge \frac{\Phi(-0.5)^3}{2} \varepsilon^2,$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution.

For confidence intervals with a prespecified coverage probability, the hardness of the problem is naturally characterized by the expected length. Let $\mathcal{I}_{z,\alpha}(\mathcal{F})$ and $\mathcal{I}_{m,\alpha}(\mathcal{F})$ be respectively the collection of confidence intervals for the minimizer $Z(f)$ and the minimum $M(f)$ with guaranteed coverage probability $1 - \alpha$ for all $f \in \mathcal{F}$. Let $L(CI)$ be the length of a confidence interval CI. The minimum expected lengths at f of all confidence intervals in $\mathcal{I}_{z,\alpha}(\{f, g\})$ and $\mathcal{I}_{m,\alpha}(\{f, g\})$ with the hardest alternative $g \in \mathcal{F}$ for f are given by

$$(1.6) \qquad L_{z,\alpha}(\varepsilon; f) = \sup_{g \in \mathcal{F}} \inf_{CI \in \mathcal{I}_{z,\alpha}(\{f,g\})} \mathbb{E}_f L(CI),$$

$$(1.7) \qquad L_{m,\alpha}(\varepsilon; f) = \sup_{g \in \mathcal{F}} \inf_{CI \in \mathcal{I}_{m,\alpha}(\{f,g\})} \mathbb{E}_f L(CI).$$

As in the case of estimation, we will first evaluate these benchmarks for the performance of confidence intervals in terms of the local moduli of continuity and then construct data-driven and computationally efficient confidence interval procedures. Furthermore, we also establish the uncertainty principle for the confidence intervals,

$$(1.8) \qquad \inf_{f \in \mathcal{F}} L_{z,\alpha}(\varepsilon; f) \cdot L_{m,\alpha}(\varepsilon; f)^2 \geq C_\alpha \varepsilon^2,$$

where C_α is a positive constant depending on α only. The uncertainty principle (1.8) shows a fundamental limit for the accuracy of simultaneous inference for the minimizer $Z(f)$ and minimum $M(f)$ for any $f \in \mathcal{F}$.

1.2. *Adaptive procedures.* Another major step in our analysis is developing data-driven and computationally efficient algorithms for the construction of adaptive estimators and adaptive confidence intervals as well as establishing the optimality of these procedures at each $f \in \mathcal{F}$.

The key idea behind the construction of the adaptive procedures is to iteratively localize the minimizer by computing the integrals over the relevant subintervals together with a carefully constructed stopping rule. For estimation of the minimum and minimizer, additional estimation procedures are added after the localization steps. For the construction of the confidence intervals, another important idea is to look back a few steps before the stopping time.

The resulting estimators, $\hat{Z}$ for the minimizer $Z(f)$ and $\hat{M}$ for the minimum $M(f)$, are shown to attain within a constant factor of the benchmarks $R_z(\varepsilon; f)$ and $R_m(\varepsilon; f)$ simultaneously for all $f \in \mathcal{F}$,

$$\mathbb{E}_f |\hat{Z} - Z(f)| \leq C_z R_z(\varepsilon; f) \quad \text{and} \quad \mathbb{E}_f |\hat{M} - M(f)| \leq C_m R_m(\varepsilon; f),$$

for some absolute constants C_z and C_m not depending on f. The confidence intervals, $CI_{z,\alpha}$ for the minimizer $Z(f)$ and $CI_{m,\alpha}$ for the minimum $M(f)$, are constructed and shown to be adaptive to individual functions $f \in \mathcal{F}$, while having guaranteed coverage probability $1 - \alpha$. That is, $CI_{z,\alpha} \in \mathcal{I}_{z,\alpha}(\mathcal{F})$ and $CI_{m,\alpha} \in \mathcal{I}_{m,\alpha}(\mathcal{F})$ and for all $f \in \mathcal{F}$,

$$\mathbb{E}_f L(CI_{z,\alpha}) \leq C_z(\alpha) L_{z,\alpha}(\varepsilon; f),$$

$$\mathbb{E}_f L(CI_{m,\alpha}) \leq C_m(\alpha) L_{m,\alpha}(\varepsilon; f),$$

where $C_z(\alpha)$ and $C_m(\alpha)$ are constants depending on α only.

1.3. *Related literature.* In addition to estimation and inference for the location and size of the extremum of a nonparametric regression function mentioned at the beginning of this section, the problems considered in the present paper are also connected to nonparametric estimation and inference under shape constraints, which have also been well studied in the literature.

Nonparametric convex regression has been investigated in various settings, ranging from estimation and confidence bands for the whole function (Birgé (1989), Guntuboyina and Sen (2018), Hengartner and Stark (1995), Dümbgen (1998)), to estimation and inference at a fixed point (Kiefer (1982), Cai, Low and Xia (2013), Cai and Low (2015), Ghosal and Sen (2017)). Deng, Han and Sen (2023) established limiting distributions for some local parameters of a convex regression function, including the minimizer based on the convexity-constrained least squares (CLS) estimator and constructed a confidence interval for the minimizer. As seen in Section B.1 of the Supplementary Material (Cai, Chen and Zhu (2024)), this confidence interval is suboptimal in terms of the expected length under the local minimax framework that we introduce later. It is also much more computationally intensive as it requires solving the CLS problem.

In the context of estimating and inferring the value of a convex function at a fixed point, which is a linear functional, the local minimax framework characterized by the benchmarks (1.2)–(1.3) and (1.6)–(1.7) has been used in Cai, Low and Xia (2013) and Cai and Low (2015). However, the focus of the present paper is on the minimizer and minimum, which are nonlinear functionals. Due to their nonlinear nature, the analysis is much more challenging than it is for the function value at a fixed point.

Another related line of research is stochastic numerical optimization of convex functions. Agarwal et al. (2011) studies stochastic convex optimization with bandit feedback and proposes an algorithm that is shown to be nearly minimax optimal. Chatterjee et al. (2016) uses the framework introduced in Cai and Low (2015) to study the local minimax complexity of stochastic convex optimization based on queries to a first-order oracle that produces an unbiased subgradient in a rather restrictive setting.

1.4. *Organization of the paper.* In Section 2, we analyze individual minimax risks, relating them to appropriate local moduli of continuity and more explicit alternative expressions, and explain the uncertainty principle with a discussion of the connections with the classical minimax framework. Super-efficiency is also considered. In Section 3, we introduce the adaptive procedures for the white noise model and show that they are optimal. In Section 4, we consider the nonparametric regression model. Adaptive procedures are proposed and their optimality is established. Section 5 discusses some future directions. For reasons of space, the numerical results and proofs are given in the Supplementary Material (Cai, Chen and Zhu (2024)).

1.5. *Notation.* We finish this section with some notation that will be used in the rest of the paper. The cdf of the standard normal distribution is denoted by Φ. For $0 < \alpha < 1$, $z_\alpha = \Phi^{-1}(1 - \alpha)$. For two real numbers a and b, $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$. $\|\cdot\|_2$ denotes the L_2 norm (i.e., $\|f\|_2 = \sqrt{\int f(x)^2 \, dx}$). For $f \in L_2[0, 1]$ and $r > 0$, $\mathcal{B}_r(f) = \{g \in L_2[0, 1] : \|g - f\|_2 \leq r\}$ and $\partial \mathcal{B}_r(f) = \{g \in L_2[0, 1] : \|g - f\|_2 = r\}$.

2. Benchmarks and uncertainty principle. In this section, we first introduce the local moduli of continuity and use them to characterize the four benchmarks for estimation and confidence intervals introduced in Section 1.1, which are summarized in Table 1. We provide alternative expressions for the local moduli of continuity that are easier to evaluate. The results are used to establish a novel uncertainty principle, which shows an intrinsic tension between the estimation/inference accuracy for the minimizer and the minimum for all functions in $\mathcal{F}$.

	Estimation	Inference
Minimizer $Z(f)$	$R_z(\varepsilon; f)$	$L_{z,\alpha}(\varepsilon; f)$
Minimum $M(f)$	$R_m(\varepsilon; f)$	$L_{m,\alpha}(\varepsilon; f)$.

2.1. *Local moduli of continuity.* For any given convex function $f \in \mathcal{F}$, we define the following local moduli of continuity, one for the minimizer, and the other for the minimum,

$$(2.1) \qquad \omega_z(\varepsilon; f) = \sup\{|Z(f) - Z(g)| : \|f - g\|_2 \le \varepsilon, g \in \mathcal{F}\},$$

$$(2.2) \qquad \omega_m(\varepsilon; f) = \sup\{|M(f) - M(g)| : \|f - g\|_2 \le \varepsilon, g \in \mathcal{F}\},$$

As in the case of a linear functional, the local moduli $\omega_z(\varepsilon; f)$ and $\omega_m(\varepsilon; f)$ clearly depend on the function f and can be regarded as an analogue of the inverse Fisher information in regular parametric models.

The following theorem characterizes the four benchmarks for estimation and inference in terms of the corresponding local modulus of continuity.

THEOREM 2.1.   *Let* $0 < \alpha < 0.3$. *Then*

$$(2.3) \qquad a_1\omega_z(\varepsilon; f) \le R_z(\varepsilon; f) \le A_1\omega_z(\varepsilon; f),$$

$$(2.4) \qquad a_1\omega_m(\varepsilon; f) \le R_m(\varepsilon; f) \le A_1\omega_m(\varepsilon; f),$$

$$(2.5) \qquad b_\alpha\omega_z(\varepsilon/3; f) \le L_{z,\alpha}(\varepsilon; f) \le B_\alpha\omega_z(\varepsilon; f),$$

$$(2.6) \qquad b_\alpha\omega_m(\varepsilon/3; f) \le L_{m,\alpha}(\varepsilon; f) \le B_\alpha\omega_m(\varepsilon; f),$$

where the constants a_1, A_1, b_α, B_α *can be taken as* $a_1 = \Phi(-0.5) \approx 0.309$, $A_1 = 1.5$, $b_\alpha = 0.6 - 2\alpha$ *and* $B_\alpha = 3(1 - 2\alpha)z_\alpha$.

Theorem 2.1 shows that the four benchmarks can be characterized in terms of the local moduli of continuity. However, these local moduli of continuity are not easy to compute. We now introduce two geometric quantities to facilitate further understanding of these benchmarks. For $f \in \mathcal{F}$, $u \in \mathbb{R}$ and $\varepsilon > 0$, let $f_u(t) = \max\{f(t), u\}$ and define

$$(2.7) \qquad \rho_m(\varepsilon; f) = \sup\{u - M(f) : \|f - f_u\|_2 \le \varepsilon\},$$

$$(2.8) \qquad \rho_z(\varepsilon; f) = \sup\{|t - Z(f)| : f(t) \le \rho_m(\varepsilon; f) + M(f), t \in [0, 1]\}.$$

Obtaining $\rho_m(\varepsilon; f)$ and $\rho_z(\varepsilon; f)$ can be viewed as a *water-filling process*. One adds water into the epigraph defined by the convex function f until the "volume" (measured by $\|\cdot\|_2$) is equal to ε. As illustrated in Figure 1, $\rho_m(\varepsilon; f)$ measures the depth of the water (CD), and $\rho_z(\varepsilon; f)$ captures the width of the water surface (FC). $\rho_m(\varepsilon; f)$ and $\rho_z(\varepsilon; f)$ essentially quantify the flatness of the function f near its minimizer $Z(f)$.

The geometric quantities $\rho_m(\varepsilon; f)$ and $\rho_z(\varepsilon; f)$ defined in (2.7) and (2.8) have the following properties.

PROPOSITION 2.1.   *For* $0 < c < 1$, $f \in \mathcal{F}$,

$$(2.9) \qquad c \le \frac{\rho_m(c\varepsilon; f)}{\rho_m(\varepsilon; f)} \le c^{\frac{2}{3}} \quad and \quad \max\left\{\left(\frac{c}{2}\right)^{\frac{2}{3}}, c\right\} \le \frac{\rho_z(c\varepsilon; f)}{\rho_z(\varepsilon; f)} \le 1.$$
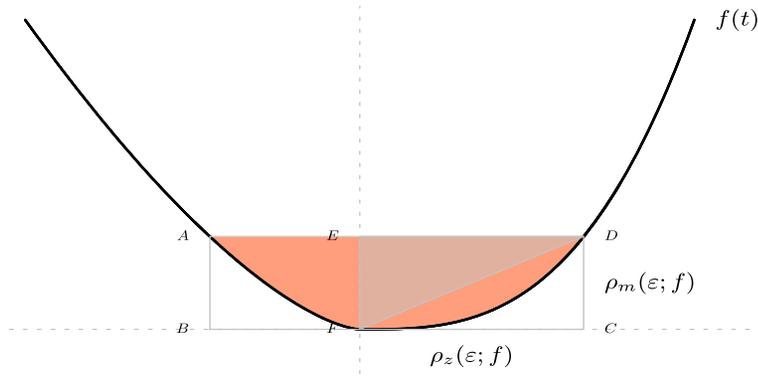
FIG. 1. *Water filling process.*

The following result connects the local moduli of continuity to these two geometric quantities.

PROPOSITION 2.2. *Let $\rho_m(\varepsilon; f)$ and $\rho_z(\varepsilon; f)$ be defined in* (2.7) *and* (2.8), *respectively. Then*

$$(2.10) \qquad \rho_m(\varepsilon; f) \leq \omega_m(\varepsilon; f) \leq 3\rho_m(\varepsilon; f),$$

$$(2.11) \qquad \rho_z(\varepsilon; f) \leq \omega_z(\varepsilon; f) \leq 3\rho_z(\varepsilon; f).$$

Therefore, through the local moduli of continuity, the hardness of the estimation and inference tasks are tied to the geometry of the convex function near its minimizer. Note that as the function gets flatter near its minimizer, $\rho_m(\varepsilon; f)$ decreases while $\rho_z(\varepsilon; f)$ increases. It is useful to calculate $\rho_m(\varepsilon; f)$ and $\rho_z(\varepsilon; f)$ in a concrete example.

EXAMPLE 2.1. Consider the function $f(t) = |t - \frac{1}{2}|^k$ where $k \geq 1$ is a constant. We will calculate $\rho_m(\varepsilon; f)$ and then obtain $\rho_z(\varepsilon; f)$ by first computing $\|f - f_u\|_2^2$ and then setting it to ε^2 to solve for $\rho_m(\varepsilon; f)$.

It is easy to see that in this case $\|f - f_u\|_2^2 = \frac{4k^2}{(2k+1)(k+1)} \cdot u^{\frac{2k+1}{k}}$. Setting $\|f - f_u\|_2^2 = \varepsilon^2$ yields $u = (\frac{(2k+1)(k+1)}{4k^2})^{\frac{k}{2k+1}} \varepsilon^{\frac{2k}{2k+1}}$. Hence,

$$\rho_m(\varepsilon; f) = \left(\frac{(2k+1)(k+1)}{4k^2} \right)^{\frac{k}{2k+1}} \varepsilon^{\frac{2k}{2k+1}}.$$

To compute $\rho_z(\varepsilon; f)$, note that $f^{-1}(u) = \frac{1}{2} \pm u^{\frac{1}{k}} = \frac{1}{2} \pm (\frac{(2k+1)(k+1)}{4k^2})^{\frac{1}{2k+1}} \varepsilon^{\frac{2}{2k+1}}$. Hence,

$$\rho_z(\varepsilon; f) = \min \left\{ \left(\frac{(2k+1)(k+1)}{4k^2} \right)^{\frac{1}{2k+1}} \varepsilon^{\frac{2}{2k+1}}, \frac{1}{2} \right\}.$$

Proposition 2.2 then yields tight bounds for the local moduli of continuity $\omega_m(\varepsilon; f)$ and $\omega_z(\varepsilon; f)$.

REMARK 2.1. Note that the results obtained in Example 2.1 can be extended to a class of convex functions. For $f \in \mathcal{F}$ satisfying

$$0 < \varliminf_{t \to Z(f)} \frac{f(t) - M(f)}{|t - Z(f)|^k} \leq \varlimsup_{t \to Z(f)} \frac{f(t) - M(f)}{|t - Z(f)|^k} < \infty$$

for some $k \geq 1$, it is easy to show that

$$\omega_m(\varepsilon; f) \sim \varepsilon^{\frac{2k}{2k+1}}, \qquad \omega_z(\varepsilon; f) \sim \varepsilon^{\frac{2}{2k+1}}, \quad \text{as } \varepsilon \to 0^+.$$

2.2. *Uncertainty principle.* Section 2.1 provides a precise characterization of the four benchmarks under the nonasymptotic local minimax framework in terms of the local moduli of continuity and the geometric quantities $\rho_m(\varepsilon; f)$ and $\rho_z(\varepsilon; f)$. These results yield a novel uncertainty principle.

THEOREM 2.2 (Uncertainty principle). *Let* $R_z(\varepsilon; f)$, $R_m(\varepsilon; f)$, $L_{z,\alpha}(\varepsilon; f)$ *and* $L_{m,\alpha}(\varepsilon; f)$ *be defined as in* (1.2)–(1.7). *Let* $0 < \alpha < 0.3$. *Then for any* $f \in \mathcal{F}$,

$$(2.12) \qquad 274\varepsilon^2 > R_z(\varepsilon; f) \cdot R_m(\varepsilon; f)^2 \geq \frac{\Phi(-0.5)^3}{2}\varepsilon^2,$$

$$(2.13) \qquad 3^7 \cdot (1 - 2\alpha)^3\varepsilon^2 > L_{z,\alpha}(\varepsilon; f) \cdot L_{m,\alpha}(\varepsilon; f)^2 \geq \frac{(0.6 - 2\alpha)^3}{18}\varepsilon^2.$$

Note that the bounds in (2.12) and (2.13) are universal for all $f \in \mathcal{F}$ and show that there is a fundamental limit to the accuracy of estimation and inference for the minimizer and minimum of a convex function. Our finding here states that the minimizer and the minimum of a convex function cannot be estimated accurately at the same time. This statistical uncertainty principle comes from an intrinsic relationship between the two operators $Z(\cdot)$ and $M(\cdot)$: For any convex function $f \in \mathcal{F}$ and any $r > 0$, there exists $g \in \partial\mathcal{B}_r(f) \cap \mathcal{F}$ such that

$$(2.14) \qquad |Z(g) - Z(f)| \cdot |M(g) - M(f)|^2 \geq \frac{1}{2}\left(\frac{r}{\varepsilon}\right)^2 \cdot \varepsilon^2,$$

where $r/\varepsilon = \|(f - g)/\varepsilon\|_2$ characterizes the probabilistic distance between the two convex functions f and g under the white noise model.

REMARK 2.2. To the best of our knowledge, the uncertainty principles established in this paper are the first of their kind in nonparametric statistics in that they reveal the fundamental tensions between estimation/inference of different quantities. It is shown in the Supplementary Material (Cai, Chen and Zhu ((2024), Section B.3)) that similar uncertainty principles also hold for certain subclasses of convex functions. Note that it is not possible to establish such results using conventional minimax analysis where the performance is measured in a worst-case sense over a large parameter space.

2.3. *Penalty for superefficiency.* We have shown that the estimation benchmarks $R_z(\varepsilon; f)$ and $R_m(\varepsilon; f)$ defined in (1.2) and (1.3) can be characterized by the local moduli of continuity. Before we show in Section 3 that these benchmarks are indeed achievable by adaptive procedures, we first prove that they cannot be essentially outperformed by any estimator uniformly over $\mathcal{F}$. The benchmarks $R_z(\varepsilon; f)$ and $R_m(\varepsilon; f)$ play a role analogous to the information lower bound in classical statistics.

THEOREM 2.3 (Penalty for superefficiency). *For any estimator* $\hat{Z}$, *if* $\mathbb{E}_{f_0}|\hat{Z} - Z(f_0)| \leq \gamma R_z(\varepsilon; f_0)$ *for some* $f_0 \in \mathcal{F}$ *and* $\gamma < 0.1$, *then there exists* $f_1 \in \mathcal{F}$ *such that*

$$(2.15) \qquad \mathbb{E}_{f_1}(|\hat{Z} - Z(f_1)|) \geq \frac{1}{40}\left(\log\frac{1}{\gamma}\right)^{1/3} R_z(\varepsilon; f_1).$$

Similarly, for any estimator $\hat{M}$, *if* $\mathbb{E}_{f_0}|\hat{M} - M(f_0)| \leq \gamma R_m(\varepsilon; f_0)$ *for some* $f_0 \in \mathcal{F}$ *and* $\gamma < 0.1$, *then there exists* $f_1 \in \mathcal{F}$ *such that*

$$(2.16) \qquad \mathbb{E}_{f_1}|\hat{M} - M(f_1)| \geq \frac{1}{8}\left(\log\frac{1}{\gamma}\right)^{1/3} R_m(\varepsilon; f_1).$$
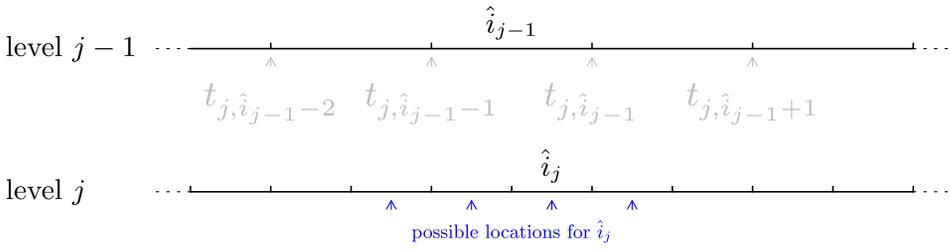
FIG. 2. *Illustration of the localization step. At level j, the middle two intervals are the two subintervals of the selected interval at level $j-1$. One adjacent interval of the same length on each side is added and the interval at level j is selected among these four intervals.*

REMARK 2.3. Theorem 2.3 shows that if an estimator of $Z(f)$ or $M(f)$ is superefficient at some $f_0 \in \mathcal{F}$ in the sense of outperforming the benchmark by a factor of γ for some small $\gamma > 0$, then it must be subefficient at some $f_1 \in \mathcal{F}$ by underperforming the benchmark by at least a factor of $(\log \frac{1}{\gamma})^{\frac{1}{3}}$.

3. Adaptive procedures and optimality. We now turn to the construction of data-driven and computationally efficient algorithms for estimation and confidence intervals for the minimizer $Z(f)$ and minimum $M(f)$ under the white noise model. The procedures are shown to be adaptive to each individual function $f \in \mathcal{F}$ in the sense that they simultaneously achieve, up to a universal constant, the corresponding benchmarks $R_z(\varepsilon; f)$, $R_m(\varepsilon; f)$, $L_{z,\alpha}(\varepsilon; f)$ and $L_{m,\alpha}(\varepsilon; f)$ for all $f \in \mathcal{F}$. These results are much stronger than what can be obtained from a conventional minimax analysis.

3.1. *The construction.* There are three main building blocks in the construction of the estimators and confidence intervals: Localization, stopping and estimation/inference.

In the localization step, we begin with the full interval $[0, 1]$. Then, iteratively, we halve the intervals and select one halved interval among a set of halved intervals depending on the interval selected in the previous iteration. This set of halved intervals include the two resulting subintervals of the previously selected interval and one neighboring halved interval, when such an interval exists, on both sides. The selection rule is to choose the one with the smallest integral of the white noise process over it. See Figure 2 for an illustration of the localization step.

The second step of the construction is the stopping rule. The localization step is iterative, so one needs to determine when there is no further gain and stop the iteration. The integral over each selected interval is a random variable and can be viewed as an estimate of the minimum times the length of the interval. The intuition is that, as the iteration progresses, the bias decreases and the variance increases. As shown in Figure 3, the basic idea is to use the differences of the integrals over the two neighboring intervals 5 blocks away from the current designated interval when such intervals exist on both sides. If either of the differences is smaller than 2 standard deviations, then the iteration stops.

After selecting the final subinterval, the last step in the construction is the estimation/inference for both the minimum and minimizer, which will be described separately later. The detailed construction is given as follows.
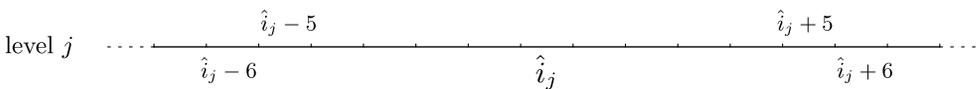

FIG. 3. *Illustration of the stopping rule.*

3.1.1. *Sample splitting.* For technical reasons, we split the data into three independent pieces to ensure independence of the data used in the three steps of the construction. This is done as follows.

Let $B_1(t)$ and $B_2(t)$ be two independent standard Brownian motions, and both be independent of the observed data Y. For $t \in [0, 1]$, let

(3.1)

$$Y_l(t) = Y(t) + \frac{\sqrt{2}}{2}\varepsilon B_1(t) + \frac{\sqrt{6}}{2}\varepsilon B_2(t),$$

$$Y_s(t) = Y(t) + \frac{\sqrt{2}}{2}\varepsilon B_1(t) - \frac{\sqrt{6}}{2}\varepsilon B_2(t),$$

$$Y_e(t) = Y(t) - \sqrt{2}\varepsilon B_1(t).$$

Then $Y_l(\cdot)$, $Y_s(\cdot)$ and $Y_e(\cdot)$ are independent and can be written as

(3.2)

$$dY_l(t) = f(t)\,dt + \sqrt{3}\varepsilon\,dW_1(t),$$

$$dY_s(t) = f(t)\,dt + \sqrt{3}\varepsilon\,dW_2(t),$$

$$dY_e(t) = f(t)\,dt + \sqrt{3}\varepsilon\,dW_3(t),$$

where W_1, W_2 and W_3 are independent standard Brownian motions.

We now have three independent copies: Y_l is used for localization, Y_s for stopping and Y_e for the construction of the final estimator and confidence interval for the minimum.

REMARK 3.1. For mere estimation and inference for the minimizer, the copy Y_e is not needed, and it suffices to split into two independent copies with smaller variance. This leads to slightly better performance. In addition, the noise levels of the three processes Y_l, Y_s and Y_e can be different. For the simplicity and ease of presentation, we split the original sample into three independent and homoskedastic copies for estimation and inference for both the minimizer and minimum.

3.1.2. *Localization.* For $j = 0, 1, \ldots,$ and $i = 0, 1, \ldots, 2^j$, let

(3.3) $\quad m_j = 2^{-j}, \quad t_{j,i} = i \cdot m_j, \quad$ and $\quad i_j^* = \max\{i : Z(f) \in [t_{j,i-1}, t_{j,i}]\}.$

That is, at level j for $j = 0, 1, \ldots,$ the i_j^*th subinterval is the one containing the minimizer $Z(f)$. For $j = 0, 1, \ldots$ and $i = 1, 2, \ldots, 2^j$, define

$$X_{j,i} = \int_{t_{j,i-1}}^{t_{j,i}} dY_l(t),$$

where Y_l is one of the three independent copies constructed above through sample splitting. For convenience, we define $X_{j,i} = +\infty$ for $j = 0, 1, \ldots$ and $i \in \mathbb{Z} \setminus \{1, 2, \ldots, 2^j\}$.

Let $\hat{i}_0 = 1$ and for $j = 1, 2, \ldots,$ let

$$\hat{i}_j = \underset{2\hat{i}_{j-1}-2 \leq i \leq 2\hat{i}_{j-1}+1}{\arg\min} X_{j,i}.$$

Note that given the value of $\hat{i}_{j-1}$ at level $j-1$, in the next iteration the procedure halves the interval $[t_{\hat{i}_{j-1}-1}, t_{\hat{i}_{j-1}}]$ into two subintervals and selects the interval $[t_{\hat{i}_j-1}, t_{\hat{i}_j}]$ at level j from these and their immediate neighboring subintervals. So, i only ranges over 4 possible values at level j; see Figure 2 for an illustration.

3.1.3. *Stopping rule.* It is necessary to have a stopping rule to select a final subinterval constructed in the localization iterations. We use another independent copy Y_s constructed in the sample splitting step to devise a stopping rule. For $j = 0, 1, \ldots$ and $i = 1, 2, \ldots, 2^j$, let

$$\tilde{X}_{j,i} = \int_{t_{j,i-1}}^{t_{j,i}} \, dY_s(t).$$

Again, for convenience, we define $\tilde{X}_{j,i} = +\infty$ for $j = 0, 1, \ldots$ and $i \in \mathbb{Z} \setminus \{1, 2, \ldots, 2^j\}$. Let the statistic T_j be defined as

$$T_j = \min\{\tilde{X}_{j,\hat{i}_j+6} - \tilde{X}_{j,\hat{i}_j+5}, \tilde{X}_{j,\hat{i}_j-6} - \tilde{X}_{j,\hat{i}_j-5}\},$$

where we use the convention $+\infty - x = +\infty$ and $\min\{+\infty, x\} = x$, for any $-\infty \leq x \leq \infty$.

The stopping rule is based on the value of T_j. It is helpful to provide some intuition before formally defining the stopping rule. Intuitively, the algorithm should stop at a place where the signal to noise ratio of T_j is small or where the signal is negative. Let $\sigma_j^2 = 6m_j\varepsilon^2$. It is easy to see that when $\tilde{X}_{j,\hat{i}_j+6} - \tilde{X}_{j,\hat{i}_j+5} < \infty$,

(3.4) $$\tilde{X}_{j,\hat{i}_j+6} - \tilde{X}_{j,\hat{i}_j+5} | \hat{i}_j \sim N\left(\int_{t_{j,\hat{i}_j+5}}^{t_{j,\hat{i}_j+6}} (f(t + m_j) - f(t)) \, dt, \sigma_j^2 \right).$$

Note that the standard deviation σ_j decreases at the rate $\frac{1}{\sqrt{2}}$ as j increases. We now turn to the mean of $\tilde{X}_{j,\hat{i}_j+6} - \tilde{X}_{j,\hat{i}_j+5} | \hat{i}_j$. Recall the notation introduced in (3.3). It is easy to see that the algorithm should stop as soon as $\int_{t_{j,\hat{i}_j+5}}^{t_{j,\hat{i}_j+6}} (f(t + m_j) - f(t)) \, dt$ turns negative, since for any $\hat{i}_j$, if $\int_{t_{j,\hat{i}_j+5}}^{t_{j,\hat{i}_j+6}} (f(t+m_j) - f(t)) \, dt < 0$, then $|\hat{i}_j - i_j^*| \geq 5$, and consequently, $|\hat{i}_{j_1} - i_{j_1}^*| \geq 5$ for any $j_1 \geq j$. When $\int_{t_{j,\hat{i}_j+5}}^{t_{j,\hat{i}_j+6}} (f(t + m_j) - f(t)) \, dt$ is positive, a careful analysis in the proof shows that it shrinks at a rate faster than or equal to $\frac{1}{4}$ as j increases. Analogous results hold for $\tilde{X}_{j,\hat{i}_j-6} - \tilde{X}_{j,\hat{i}_j-5} | \hat{i}_j$.

Finally, the iterations stop at level $\hat{j}$ where

$$\hat{j} = \min\left\{ j : \frac{T_j}{\sigma_j} \leq 2 \right\}.$$

The subinterval containing the minimizer $Z(f)$ is localized to be $[t_{\hat{j},\hat{i}_{\hat{j}}-1}, t_{\hat{j},\hat{i}_{\hat{j}}}]$.

3.1.4. *Estimation and inference.* After the final subinterval $[t_{\hat{j},\hat{i}_{\hat{j}}-1}, t_{\hat{j},\hat{i}_{\hat{j}}}]$ is obtained, we then use it to construct estimators and confidence intervals for $Z(f)$ and $M(f)$. We begin with the minimizer $Z(f)$. The estimator of $Z(f)$ is given by the midpoint of the interval $[t_{\hat{j},\hat{i}_{\hat{j}}-1}, t_{\hat{j},\hat{i}_{\hat{j}}}]$, that is,

(3.5) $$\hat{Z} = \frac{t_{\hat{j},\hat{i}_{\hat{j}}} + t_{\hat{j},\hat{i}_{\hat{j}}-1}}{2}.$$

To construct the confidence interval for $Z(f)$, one needs to take a few steps to the left and to the right at level $\hat{j}$. Let $K_\alpha = \lceil \frac{\log \alpha}{\log \Phi(-2)} \rceil$ and define

$$L = \max\{0, \hat{i}_{\hat{j}} - 12 \times 2^{K_\alpha} + 1\}, \qquad U = \min\{2^{\hat{j}}, \hat{i}_{\hat{j}} + 12 \times 2^{K_\alpha} - 2\}.$$

The $1 - \alpha$ confidence interval for $Z(f)$ is given by

(3.6)
$$CI_{z,\alpha} = [t_{\hat{j},L}, t_{\hat{j},U}].$$

For estimation of and confidence interval for the minimum $M(f)$, define

$$\bar{X}_{j,i} = \int_{t_{j,i-1}}^{t_{j,i}} Y_e(t)\, dt.$$

Let $\tilde{i}_{\hat{j}} = \hat{i}_{\hat{j}} + 2(\mathbb{1}\{\bar{X}_{\hat{j},\hat{i}_{\hat{j}}+6} - \bar{X}_{\hat{j},\hat{i}_{\hat{j}}+5} \leq 2\hat{\sigma}_{\hat{j}}\} - \mathbb{1}\{\bar{X}_{\hat{j},\hat{i}_{\hat{j}}-6} - \bar{X}_{\hat{j},\hat{i}_{\hat{j}}-5} \leq 2\hat{\sigma}_{\hat{j}}\})$ and define the final estimator of the minimum $M(f)$ by

(3.7)
$$\hat{M} = \frac{1}{m_{\hat{j}}} \bar{X}_{\hat{j},\tilde{i}_{\hat{j}}}.$$

We now construct confidence interval for $M(f)$. Recall that $K_\alpha = \lceil \frac{\log \alpha}{\log \Phi(-2)} \rceil$. Compared with the confidence interval for the minimizer, we take four more blocks on each side at the level $(\hat{j} - K_{\frac{\alpha}{4}} - 1)_+$. More specifically, we define

$$t_L = t_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+, \hat{i}_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+}-5}, \qquad t_R = t_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+, \hat{i}_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+}+4}.$$

Set

(3.8)
$$\tilde{K}_\alpha = \max\{4, 2 + \lceil \log_2 (2 + z_{\alpha/3}) \rceil\}.$$

Note that at level $\hat{j} + \tilde{K}_{\frac{\alpha}{4}}$, the indices of the intervals with right endpoints t_L, t_R, respectively, are

$$i_L = t_L \cdot 2^{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}} \quad \text{and} \quad i_R = t_R \cdot 2^{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}.$$

Note also that $i_R - i_L = 9 \times 2^{1+\tilde{K}_{\frac{\alpha}{4}}+K_{\frac{\alpha}{4}}}$, which only depends on α. Define an intermediate estimator of the minimum $M(f)$ by

$$\hat{f}_1 = \frac{1}{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}} \min_{i_L < i \leq i_R} \bar{X}_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}, i}.$$

Let F_n be the cumulative distribution function of $\tilde{v}_n = \max\{v_1, \ldots, v_n\}$, where $v_1, \ldots, v_n \overset{\text{i.i.d.}}{\sim} N(0, 1)$, and define

(3.9)
$$S_{n,\beta} = F_n^{-1}(1 - \beta).$$

In other words, $S_{n,\beta}$ is the $(1 - \beta)$ quantile of the distribution of the maximum of n i.i.d. standard normal variables. Let

$$f_{lo} = \hat{f}_1 - z_{\alpha/4} \frac{\sqrt{3}\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}} - \frac{\sqrt{3}\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}}, \qquad f_{hi} = \hat{f}_1 + S_{i_R-i_L, \frac{\alpha}{4}} \cdot \frac{\sqrt{3}\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}}.$$

Then the $(1 - \alpha)$ level confidence interval for $M(f)$ is defined as

(3.10)
$$CI_{m,\alpha} = [f_{lo}, f_{hi}].$$

3.2. *Statistical optimality.* Now we establish the optimality of the adaptive procedures constructed in Section 3.1. The results show that the data-driven estimators and the confidence intervals attain their corresponding local minimax risks/lengths in the sense of staying within a constant multiplier. We begin with the estimator of the minimizer.

THEOREM 3.1 (Estimation of minimizer). *The estimator $\hat{Z}$ defined in (3.5) satisfies*

$$\mathbb{E}_f|\hat{Z} - Z(f)| < 53\rho_z(\varepsilon; f) \le C_z R_z(\varepsilon; f), \quad \text{for all } f \in \mathcal{F},$$

where $C_z > 0$ is an absolute constant.

The following result holds for the confidence interval $CI_{z,\alpha}$.

THEOREM 3.2 (Confidence interval for the minimizer). *Let $0 < \alpha < 0.3$. The confidence interval $CI_{z,\alpha}$ given in (3.6) is a $(1 - \alpha)$ level confidence interval for the minimizer $Z(f)$ and its expected length satisfies*

$$\mathbb{E}_f L(CI_{z,\alpha}) \le (24 \times 2^{K_\alpha} - 3) \times 17.5 \times \rho_z(\varepsilon; f) \le C_{z,\alpha} L_{z,\alpha}(\varepsilon; f), \quad \text{for all } f \in \mathcal{F},$$

where $K_\alpha = \lceil \frac{\log \alpha}{\log \Phi(-2)} \rceil$ and $C_{z,\alpha}$ is a constant depending on α only.

Similarly, the estimator and confidence interval for the minimum $M(f)$ are within a constant factor of the benchmarks simultaneously for all $f \in \mathcal{F}$.

THEOREM 3.3 (Estimation of minimum). *The estimator $\hat{M}$ defined in (3.7) satisfies*

$$\mathbb{E}_f|\hat{M} - M(f)| < 449\rho_m(\varepsilon; f) \le C_m R_m(\varepsilon; f), \quad \text{for all } f \in \mathcal{F},$$

where $C_m > 0$ is an absolute constant.

THEOREM 3.4 (Confidence interval for the minimum). *The confidence interval $CI_{m,\alpha}$ given in (3.10) is a $(1 - \alpha)$ confidence interval for the minimum $M(f)$ and when $0 < \alpha < 0.3$, its expected length satisfies*

$$\mathbb{E}_f L(CI_{m,\alpha}) \le c_{m,\alpha} \rho_m(\varepsilon; f) \le C_{m,\alpha} L_{m,\alpha}(\varepsilon; f), \quad \text{for all } f \in \mathcal{F},$$

where $c_{m,\alpha}$ and $C_{m,\alpha}$ are constants depending on α only.

4. Nonparametric regression. We have so far focused on the white noise model. The procedures and results presented in the previous sections can be extended to nonparametric regression, where we observe

(4.1) $$y_i = f(x_i) + \sigma z_i, \quad i = 0, 1, 2, \ldots, n,$$

with $x_i = \frac{i}{n}$, and $z_i \overset{\text{i.i.d.}}{\sim} N(0, 1)$. The noise level σ is assumed to be known. The tasks are the same as before: construct optimal estimators and confidence intervals for the minimizer and minimum of $f \in \mathcal{F}$.

4.1. *Benchmarks and discretization errors.* Analogous to the benchmarks for the white noise model defined in equations (1.2), (1.3), (1.6), (1.7), we define similar benchmarks for the nonparametric regression model (4.1) with $n + 1$ equally spaced observations. Denote by $\mathcal{I}_{z,\alpha,n}(\mathfrak{F})$ and $\mathcal{I}_{m,\alpha,n}(\mathfrak{F})$, respectively, the collections of $(1 - \alpha)$ level confidence intervals for $Z(f)$ and $M(f)$ on a function class $\mathfrak{F}$ under the regression model (4.1) and let

$$\tilde{R}_{z,n}(\sigma; f) = \sup_{g \in \mathcal{F}} \inf_{\hat{Z}} \max_{h \in \{f,g\}} \mathbb{E}_h |\hat{Z} - Z(h)|,$$

$$\tilde{R}_{m,n}(\sigma; f) = \sup_{g \in \mathcal{F}} \inf_{\hat{M}} \max_{h \in \{f,g\}} \mathbb{E}_h |\hat{M} - M(h)|,$$

(4.2)

$$\tilde{L}_{z,\alpha,n}(\sigma; f) = \sup_{g \in \mathcal{F}} \inf_{CI \in \mathcal{I}_{z,\alpha,n}(\{f,g\})} \mathbb{E}_f L(CI),$$

$$\tilde{L}_{m,\alpha,n}(\sigma; f) = \sup_{g \in \mathcal{F}} \inf_{CI \in \mathcal{I}_{m,\alpha,n}(\{f,g\})} \mathbb{E}_f L(CI).$$

Compared with the white noise model, estimation and inference for both $Z(f)$ and $M(f)$ incur additional discretization errors, even in the noiseless case; see the Supplementary Material (Cai, Chen and Zhu ((2024), Section C.11)) for further discussion.

4.2. *Data-driven procedures.* Similar to the white noise model, we first split the data into three independent copies and then construct the estimators and confidence intervals for $Z(f)$ and $M(f)$ in three major steps: localization, stopping and estimation/inference.

4.2.1. *Data splitting.* Let $z_{1,0}, z_{1,1}, \ldots, z_{1,n}, z_{2,0}, z_{2,1}, \ldots, z_{2,n}$ be i.i.d. standard normal random variables, and all be independent of the observed data $\{y_1, \ldots, y_n\}$. We construct the following three sequences:

(4.3)

$$y_{l,i} = y_i + \frac{\sqrt{2}}{2}\sigma z_{1,i} + \frac{\sqrt{6}}{2}\sigma z_{2,i},$$

$$y_{s,i} = y_i + \frac{\sqrt{2}}{2}\sigma z_{1,i} - \frac{\sqrt{6}}{2}\sigma z_{2,i},$$

$$y_{e,i} = y_i - \sqrt{2}\sigma z_{1,i},$$

for $i = 0, \ldots, n$. For convenience, let $y_{l,i} = y_{s,i} = y_{e,i} = \infty$ for $i \notin \{0, 1, \ldots, n\}$. It is easy to see that these random variables are all independent with the same variance $3\sigma^2$ for $i \in \{0, 1, \ldots, n\}$. We will use $\{y_{l,i}\}$ for localization, $\{y_{s,i}\}$ for devising the stopping rule, and $\{y_{e,i}\}$ for constructing the final estimation and inference procedures.

Let $J = \lfloor \log_2(n + 1) \rfloor$. For $j = 0, 1, \ldots, J$, $i = 1, 2, \ldots, \lfloor \frac{n+1}{2^{J-j-1}} \rfloor$, the ith block at level j consists of $\{x_{(i-1)2^{J-j}}, x_{(i-1)2^{J-j}+1}, \ldots, x_{i \cdot 2^{J-j}-1}\}$. Denote the sum of the observations in the ith block at level j for the sequence u ($u = l, s, e$) as

$$Y_{j,i,u} = \sum_{k=(i-1)2^{J-j}}^{i \cdot 2^{J-j}-1} y_{u,k}, \quad \text{for } u = l, s, e.$$

Again, let $Y_{j,i,u} = +\infty$ when $i \in \mathbb{Z} \backslash \{1, 2, \ldots, \lfloor \frac{n+1}{2^{J-j-1}} \rfloor\}$, for $u = l, s, e$.

4.2.2. *Localization.* We now use $\{y_{l,i}, i = 0, \ldots, n\}$ to construct a localization procedure. Let $\hat{\imath}_0 = 1$, and for $j = 1, 2, \ldots, J$, let

$$\hat{\imath}_j = \underset{\max\{2\hat{\imath}_{j-1}-2,1\} \leq i \leq \min\{2\hat{\imath}_{j-1}+1, \lfloor \frac{n+1}{2^{J-j}} \rfloor\}}{\arg\min} Y_{j,i,l}.$$

This is similar to the localization step in the white noise model. In each iteration, the blocks at the previous level are split into two subblocks. The ith block at level $j - 1$ is split into two blocks, the $(2i - 1)$st block and $2i$th block, at level j. For a given $\hat{\mathtt{i}}_{j-1}$, $\hat{\mathtt{i}}_j$ is the subblock with the smallest sum among the two subblocks of $(j - 1)$-level-block $\hat{\mathtt{i}}_{j-1}$ and their immediate neighboring subblocks.

4.2.3. *Stopping rule.*
Similar to the stopping rule for the white noise model, define the statistic $\mathtt{T}_j$ as

$$\mathtt{T}_j = \min\{Y_{j,\hat{\mathtt{i}}_j+6,s} - Y_{j,\hat{\mathtt{i}}_j+5,s}, Y_{j,\hat{\mathtt{i}}_j-6,s} - Y_{j,\hat{\mathtt{i}}_j-5,s}\}.$$

Let $\tilde{\sigma}_j^2 = 6 \times 2^{J-j}\sigma^2$. It is easy to see that when $Y_{j,\hat{\mathtt{i}}_j+6,s} - Y_{j,\hat{\mathtt{i}}_j+5,s} < \infty$,

$$(4.4) \qquad Y_{j,\hat{\mathtt{i}}_j+6,s} - Y_{j,\hat{\mathtt{i}}_j+5,s} | \hat{\mathtt{i}}_j \sim N\left(\sum_{k=(\hat{\mathtt{i}}_j+4)2^{J-j}}^{(\hat{\mathtt{i}}_j+5)2^{J-j}-1} f(x_{k+2^{J-j}}) - f(x_k), \tilde{\sigma}_j^2\right).$$

Define

$$\check{j} = \begin{cases} \min\{j : \mathtt{T}_j \leq 2\tilde{\sigma}_j\} & \text{if } \{j : \mathtt{T}_j \leq 2\tilde{\sigma}_j\} \cap \{0, 1, 2, \ldots, J\} \neq \varnothing, \\ \infty & \text{otherwise,} \end{cases}$$

and terminate the algorithm at level $\hat{j} = \min\{J, \check{j}\}$. So, either $\mathtt{T}_j$ triggers the stopping rule for some $0 \leq j \leq J$ or the algorithm reaches the highest possible level J.

With the localization strategy and the stopping rule, the final block, the $\hat{\mathtt{i}}_{\hat{j}}$th block at level $\hat{j}$ is given by $\{x_k : (\hat{\mathtt{i}}_{\hat{j}} - 1)2^{J-\hat{j}} \leq k \leq \hat{\mathtt{i}}_{\hat{j}}2^{J-\hat{j}} - 1\}$.

4.2.4. *Estimation and inference.*
After we have our final block, $\hat{\mathtt{i}}_{\hat{j}}$th block at level $\hat{j}$, we use it to construct estimators and confidence intervals for the minimizer $Z(f)$ and the minimum $M(f)$. We start with the estimation of $Z(f)$. The estimator of $Z(f)$ is given as follows:

$$(4.5) \qquad \hat{Z} = \begin{cases} -\dfrac{1}{2n} + \dfrac{1}{n}(2^{J-\hat{j}}\hat{\mathtt{i}}_{\hat{j}} - 2^{J-\hat{j}-1}) & \check{j} < \infty, \\ \dfrac{1}{n} \arg\min\limits_{\hat{\mathtt{i}}_{\hat{j}}-2 \leq i \leq \hat{\mathtt{i}}_{\hat{j}}+2} y_{e,i-1} - \dfrac{1}{n} & \check{j} = \infty. \end{cases}$$

To construct the confidence interval for $Z(f)$, we take a few adjacent blocks to the left and right of $\hat{\mathtt{i}}_{\hat{j}}$th block at level $\hat{j}$. Let

$$\mathtt{L} = \max\{0, \hat{\mathtt{i}}_{\hat{j}} - 12 \times 2^{K_{\alpha/2}} + 1\} \quad \text{and}$$

$$\mathtt{U} = \min\{\lceil(n+1)2^{\hat{j}-J}\rceil, \hat{\mathtt{i}}_{\hat{j}} + 12 \times 2^{K_{\alpha/2}} - 2\}.$$

When $\check{j} < \infty$, let

$$t_{lo} = \frac{2^{J-\hat{j}}}{n}\mathtt{L} - \frac{1}{2n} \quad \text{and} \quad t_{hi} = \frac{2^{J-\hat{j}}}{n}\mathtt{U} - \frac{1}{2n}.$$

When $\check{j} = \infty$, t_{lo} and t_{hi} are calculated by the following Algorithm 1. Note that $\check{j} = \infty$ means that the procedure is forced to end and the discretization error can be dominant.

Algorithm 1 first iteratively shrinks the original interval $[t_{lo} - \frac{1}{n}, t_{hi} + \frac{1}{n}]$ to find the minimizer $\frac{i_m}{n}$ of the function f among the $n + 1$ sample points with high probability. In each

---

Algorithm 1 Computing t_{lo} and t_{hi} when $\check{j} = \infty$

$L \leftarrow \max\{1, \hat{\mathfrak{i}}_{\check{j}} - 12 \times 2^{K_{\alpha/2}}\} - 1$, $U \leftarrow \min\{n + 1, \hat{\mathfrak{i}}_{\check{j}} + 12 \times 2^{K_{\alpha/2}}\} - 1$, $\alpha_1 \leftarrow \frac{\alpha}{8}$,
$\alpha_2 = \alpha/24$

Generate $z_{3,0}, z_{3,1} \ldots, z_{3,n} \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$

$i_l \leftarrow \min\{\{U\} \cup \{i \in [L, U - 1] : y_{e,i} + \sqrt{3}\sigma z_{3,i} - (y_{e,i+1} + \sqrt{3}\sigma z_{3,i+1}) \leq 2\sqrt{3}\sigma z_{\alpha_1}\}\}$

$i_r \leftarrow \max\{\{L - 1\} \cup \{i \in [L, U - 1] : y_{e,i} + \sqrt{3}\sigma z_{3,i} - (y_{e,i+1} + \sqrt{3}\sigma z_{3,i+1}) \geq -2\sqrt{3}\sigma z_{\alpha_1}\}\}$

if $i_l = U$ **then**

  if $i_l = n$ and $y_{e,n-2} - y_{e,n-1} - \sqrt{3}\sigma(z_{3,n-2} - z_{3,n-1}) + 2\sqrt{6}\sigma z_{\alpha_2} > 0$ **then**

    $t_{hi} \leftarrow 1$

    $t_{lo} \leftarrow ((-\frac{y_{e,n} - y_{e,n-1} - \sqrt{3}\sigma(z_{3,n} - z_{3,n-1}) + 2\sqrt{6}\sigma z_{\alpha_2}}{n(y_{e,n-2} - y_{e,n-1} - \sqrt{3}\sigma(z_{3,n-2} - z_{3,n-1}) + 2\sqrt{6}\sigma z_{\alpha_2})} + \frac{n-1}{n}) \vee \frac{n-1}{n}) \wedge \frac{n}{n}$,

  else

    $t_{lo} = t_{hi} = U/n$

  end if

end if

if $i_r = L - 1$ **then**

  if $i_r = -1$ and $y_{e,2} - y_{e,1} - \sqrt{3}\sigma(z_{3,2} - z_{3,1}) + 2\sqrt{6}\sigma z_{\alpha_2} > 0$ **then**

    $t_{hi} \leftarrow ((\frac{y_{e,0} - y_{e,1} - \sqrt{3}\sigma(z_{3,0} - z_{3,1}) + 2\sqrt{6}\sigma z_{\alpha_2}}{n(y_{e,2} - y_{e,1} - \sqrt{3}\sigma(z_{3,2} - z_{3,1}) + 2\sqrt{6}\sigma z_{\alpha_2})} + \frac{1}{n}) \vee \frac{0}{n}) \wedge \frac{1}{n}$, $t_{lo} = 0$

  else

    $t_{lo} = t_{hi} = 0$

  end if

end if

if $(i_l - U)(i_r - L + 1) \neq 0$ **then**

  $i_{lo} \leftarrow (i_l - 1) \vee L$, $i_{hi} \leftarrow (i_r + 2) \wedge U$

  if $i_{hi} - i_{lo} \geq 3$ or $(i_{hi} - n)i_{lo} = 0$, **then**

    $t_{lo} = i_{lo}/n$, $t_{hi} = i_{hi}/n$

  else if $y_{e,i_{hi}+1} - y_{e,i_{hi}} - \sqrt{3}\sigma(z_{3,i_{hi}+1} - z_{3,i_{hi}}) \leq -2\sqrt{6}\sigma z_{\alpha_2}$ or $y_{e,i_{lo}-1} - y_{e,i_{lo}} - \sqrt{3}\sigma(z_{3,i_{lo}-1} - z_{3,i_{lo}}) \leq -2\sqrt{6}\sigma z_{\alpha_2}$ **then**

    $t_{lo} = t_{hi} = (i_{hi} + i_{lo})/2n$

  else

    $t_{hi} \leftarrow ((\frac{y_{e,i_{hi}-1} - y_{e,i_{hi}} - \sqrt{3}\sigma(z_{3,i_{hi}-1} - z_{3,i_{hi}}) + 2\sqrt{6}\sigma z_{\alpha_2}}{n(y_{e,i_{hi}+1} - y_{e,i_{hi}} - \sqrt{3}\sigma(z_{3,i_{hi}+1} - z_{3,i_{hi}}) + 2\sqrt{6}\sigma z_{\alpha_2})} + \frac{i_{hi}}{n}) \vee \frac{i_{hi}-1}{n}) \wedge \frac{i_{hi}}{n}$

    $t_{lo} \leftarrow ((-\frac{y_{e,i_{lo}+1} - y_{e,i_{lo}} - \sqrt{3}\sigma(z_{3,i_{lo}+1} - z_{3,i_{lo}}) + 2\sqrt{6}\sigma z_{\alpha_2}}{n(y_{e,i_{lo}-1} - y_{e,i_{lo}} - \sqrt{3}\sigma(z_{3,i_{lo}-1} - z_{3,i_{lo}}) + 2\sqrt{6}\sigma z_{\alpha_2})} + \frac{i_{lo}}{n}) \vee \frac{i_{lo}}{n}) \wedge \frac{i_{lo}+1}{n}$

  end if

end if

---

iteration, the algorithm tests whether the slopes of the segments on both ends are positive or negative. It shrinks the left end with negative slope (on the left), or shrinks the right end with positive slope (on the right) or stops if no further shrinking is needed on either side.

Note that the minimizer of any convex function with given values at these $n + 1$ points is smaller than the intersection of the following two lines:

$$(4.6) \qquad y = f\left(\frac{i_m}{n}\right) \quad \text{and} \quad y = \frac{f(\frac{i_m+2}{n}) - f(\frac{i_m+1}{n})}{1/n}\left(t - \frac{i_m + 1}{n}\right) + f\left(\frac{i_m + 1}{n}\right).$$

Note that these two lines are determined by $f(\frac{i_m}{n})$, $f(\frac{i_m+1}{n})$ and $f(\frac{i_m+2}{n})$ only. Given the noisy observations at these three points, $\frac{i_m}{n}$, $\frac{i_m+1}{n}$ and $\frac{i_m+2}{n}$, the range of these two lines and the intersection can be inferred, and the right side of the interval can then be shrunk accordingly.

Similar is done for the left side of the confidence interval. In addition, boundary cases and other complications need to be considered, which are handled in Algorithm 1.

Note that our construction and the theoretical results only rely on convexity. In particular, the existence of second-order derivative is not needed as it is commonly assumed in the literature. This is an important contributing factor to optimality under the nonasymptotic local minimax framework.

The $(1 - \alpha)$-level confidence interval for the minimizer $Z(f)$ is given by

$$(4.7) \qquad \mathtt{CI}_{z,\alpha} = [t_{lo} \wedge t_{hi}, t_{hi}]$$

We now construct the estimator and confidence interval for the minimum $M(f)$. Let

$$\Delta = \mathbb{1}\{Y_{\hat{\jmath},\hat{\imath}_{\hat{\jmath}}+6,s} - Y_{\hat{\jmath},\hat{\imath}_{\hat{\jmath}}+5,s} \le 2\sqrt{6}\sigma\sqrt{2^{J-\hat{\jmath}}}\}$$

$$- \mathbb{1}\{Y_{\hat{\jmath},\hat{\imath}_{\hat{\jmath}}-6,s} - Y_{\hat{\jmath},\hat{\imath}_{\hat{\jmath}}-5,s} \le 2\sqrt{6}\sigma\sqrt{2^{J-\hat{\jmath}}}\}$$

and define

$$(4.8) \qquad \tilde{\imath}_{\hat{\jmath}} = \begin{cases} \hat{\imath}_{\hat{\jmath}} + 2\Delta & \text{if } \breve{\jmath} < \infty, \\ \underset{\hat{\imath}_{\hat{\jmath}}-2 \le i \le \hat{\imath}_{\hat{\jmath}}+2}{\arg\min} \ y_{e,i-1} & \text{if } \breve{\jmath} = \infty. \end{cases}$$

The estimator of $M(f)$ is then given by the average of the observations of the copy for estimation and inference in the $\tilde{\imath}_{\hat{\jmath}}$th block at level $\hat{\jmath}$,

$$(4.9) \qquad \hat{M} = \frac{1}{2^{J-\hat{\jmath}}} Y_{\hat{\jmath},\tilde{\imath}_{\hat{\jmath}},e}.$$

To construct the confidence interval for $M(f)$, we specify two levels j_s and j_l, with

$$j_s = \max\{0, \hat{\jmath} - K_{\frac{\alpha}{4}} - 1\} \quad \text{and} \quad j_l = \min\{J, \hat{\jmath} + \tilde{K}_{\frac{\alpha}{4}}\},$$

where $\tilde{K}_{\frac{\alpha}{4}}$ is defined as in equation (3.8). It will be shown that at level j_s, $Z(f)$ is within four blocks of the chosen block with probability at least $1 - \frac{\alpha}{4}$, and at level j_l, with probability at least $1 - \frac{\alpha}{4}$, the length of the block is no larger than $\rho_z(\frac{\sigma}{\sqrt{n}}; f)$. Define

$$I_{lo} = \max\{1, 2^{j_l-j_s}(\hat{\imath}_{j_s} - 5)\}, \qquad I_{hi} = \min\left\{2^{j_l-j_s}(\hat{\imath}_{j_s} + 4) + 1, \left\lceil \frac{n+1}{2^{J-j_l}} \right\rceil\right\}.$$

It can be shown that $Z(f)$ lies with high probability in the interval $[\frac{2^{J-j_l}(I_{lo}-1)}{n}, \frac{2^{J-j_l}I_{hi}-1}{n}] \cap [0,1]$. Define an intermediate estimator for $M(f)$ by

$$\hat{f}_1 = \min_{I_{lo} \le i \le I_{hi}} \frac{1}{2^{J-j_l}} Y_{j_l,i,e}.$$

Let

$$f_{hi} = \hat{f}_1 + S_{I_{hi}-I_{lo}+1,\frac{\alpha}{4}} \frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}},$$

where $S_{n,\beta}$ is defined in equation (3.9) in Section 3. This is the upper limit of the confidence interval, now we define the lower limit f_{lo}.

Algorithm 2 Computing $\mathtt{f}_{lo}$ when $\hat{\hat{\jmath}} + \tilde{K}_{\frac{\alpha}{4}} > J$

$H \leftarrow S_{I_{hi}-I_{lo}+3, \frac{\alpha}{8}} \sqrt{3}\sigma, k_l \leftarrow I_{lo} - 1, k_r \leftarrow I_{hi} - 2$

if $I_{lo} = 1$ **then**
  $v_{r,0}(t) \leftarrow \frac{y_{e,2} - y_{e,1} + 2H}{1/n}(t - 1/n) + y_{e,1} - H, h(0) \leftarrow \min_{t \in [0,1/n]} v_{r,0}(t), k_l \leftarrow I_{lo}$
end if

if $I_{hi} - 1 = n$ **then**
  $v_{l,n-1}(t) \leftarrow \frac{y_{e,n-1} - y_{e,n-2} - 2H}{1/n}(t - \frac{n-1}{n}) + y_{e,n-1} - H, h(n-1) = \min_{t \in [\frac{n-1}{n}, 1]} v_{l,n-1}(t),$
  $k_r \leftarrow I_{hi} - 3$
end if

for $i = k_l, \ldots, k_r$ **do**
  Define two linear functions:
  $v_{l,i}(t) = \frac{y_{e,i} - y_{e,i-1} - 2H}{1/n}(t - x_i) + y_{e,i} - H, v_{r,i} = \frac{y_{e,i+2} - y_{e,i+1} + 2H}{1/n}(t - x_{i+1}) + y_{e,i+1} - H$
  $h(i) = \min_{t \in [x_i, x_{i+1}]} \max\{v_{l,i}(t), v_{r,i}(t)\}$
end for

$\mathtt{f}_{lo} \leftarrow \min\{h(i) : I_{lo} - 1 \le i \le I_{hi} - 2\} \wedge \mathtt{f}_{hi}$

When $\hat{\hat{\jmath}} + \tilde{K}_{\frac{\alpha}{4}} \le J$, let

$$\mathtt{f}_{lo} = \hat{\mathtt{f}}_1 - (z_{\alpha/4} + 1)\frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}}.$$

When $\hat{\hat{\jmath}} + \tilde{K}_{\frac{\alpha}{4}} > J$, we compute $\mathtt{f}_{lo}$ by Algorithm 2, which is based on the geometric property of the convex function f that for any $1 \le k \le n - 2$,

$$
(4.10) \quad
\begin{aligned}
\inf_{t \in [x_k, x_{k+1}]} f(t) &\ge \inf_{t \in [\frac{k}{n}, \frac{k+1}{n}]} \max\Bigg\{ \frac{f(x_k) - f(x_{k-1})}{1/n}(t - x_k) + f(x_k), \\
&\qquad \frac{f(x_{k+2}) - f(x_{k+1})}{1/n}(t - x_{k+1}) + f(x_{k+1}) \Bigg\}.
\end{aligned}
$$

Note that $h(i)$ in Algorithm 2 is derived from one or two linear functions, so given the relationship of the function values at two end points of the corresponding interval, it has an explicit form. Hence, the procedure is still computationally efficient.

The $(1 - \alpha)$-level confidence interval for the minimum $M(f)$ is given by

$$(4.11) \quad \mathtt{CI}_{m,\alpha} = [\mathtt{f}_{lo}, \mathtt{f}_{hi}].$$

REMARK 4.1. As mentioned in the Introduction, Agarwal et al. (2011) propose an algorithm for stochastic convex optimization with bandit feedback. While both our procedures and the method in Agarwal et al. (2011) include an ingredient trying to localize the minimizer through shrinking intervals by exploiting the convexity of the underlying function, the two methods are essentially different due to the significant differences in both the designs and loss functions. The goal of exploiting convexity in Agarwal et al. (2011) is mainly to determine the direction of shrinking their intervals, while ours is mainly for deciding when to stop and what to do after stopping.

4.3. *Statistical optimality.* Now we establish the optimality of the adaptive procedures constructed in Section 4.2. The regression model is similar to the white noise model, but with additional discretization errors. The results show that our data-driven procedures are simultaneously optimal (up to a constant factor) for all $f \in \mathcal{F}$. We begin with the estimator of the minimizer.

THEOREM 4.1 (Estimation of the minimizer).   *The estimator $\hat{Z}$ of the minimizer $Z(f)$ defined in* (4.5) *satisfies*

$$(4.12) \qquad \mathbb{E}_f|\hat{Z} - Z(f)| \le C_1 \tilde{R}_{z,n}(\sigma; f), \quad for\ all\ f \in \mathcal{F},$$

where $C_1 > 0$ is an absolute constant.

The following result holds for the confidence interval $\mathrm{CI}_{z,\alpha}$ of $Z(f)$.

THEOREM 4.2.   *Let $0 < \alpha < 0.3$. The confidence interval $\mathrm{CI}_{z,\alpha}$ given in* (4.7) *is a* $(1 - \alpha)$*-level confidence interval for the minimizer $Z(f)$ and its expected length satisfies*

$$\mathbb{E}_f L(\mathrm{CI}_{z,\alpha}) \le C_{2,\alpha} \tilde{L}_{z,\alpha,n}(\sigma; f), \quad for\ all\ f \in \mathcal{F},$$

where $C_{2,\alpha}$ is a constant depending on α only.

Similarly, the estimator and confidence interval for the minimum $M(f)$ are within a constant factor of the benchmarks simultaneously for all $f \in \mathcal{F}$.

THEOREM 4.3 (Estimation for the minimum).   *The estimator $\hat{M}$ defined in* (4.9) *satisfies*

$$\mathbb{E}_f|\hat{M} - M(f)| \le C_3 \tilde{R}_{m,n}(\sigma; f), \quad for\ all\ f \in \mathcal{F},$$

where C_3 is an absolute constant.

THEOREM 4.4.   *Let $0 < \alpha < 0.3$. The confidence interval $\mathrm{CI}_{m,\alpha}$ given in* (4.11) *is a* $(1 - \alpha)$*-level confidence interval and its expected length satisfies*

$$\mathbb{E}_f L(\mathrm{CI}_{m,\alpha}) \le C_{4,\alpha} \tilde{L}_{m,\alpha,n}(\sigma; f), \quad for\ all\ f \in \mathcal{F},$$

where $C_{4,\alpha}$ is a constant depending only on α.

In addition to statistical optimality, the proposed algorithms are computationally fast. We conducted numerical experiments and the results are in the Supplementary Material (Cai, Chen and Zhu ((2024), Section A)). A comparison of our algorithms with convexity-constrained least squares based methods (e.g., Deng, Han and Sen (2023)) and is also given in the Supplementary Material.

5. Discussion.   In the present paper, we studied optimal estimation and inference for the minimizer and minimum of a convex function under a nonasymptotic local minimax framework. We show in the Supplementary Material (Cai, Chen and Zhu ((2024), Section B.2)) that the results in this paper can be readily used to establish the optimal rates of convergence over the convex smoothness classes under the classical minimax framework.

A key advantage of the nonasymptotic local minimax framework is its ability to characterize the difficulty of estimating individual functions and to demonstrate novel phenomena that are not observable in the classical minimax theory. The uncertainty principle established in this paper highlights the fundamental tension between the accuracy of estimating the minimizer and that of estimating the minimum of a convex function. Similar results also apply to inference accuracy. It would be of great interest to establish uncertainty principles for other statistical problems, such as stochastic optimization with bandit feedback under shape constraints.

A more conventional approach to assessing the difficulty at individual functions is to consider the minimax risk over a local neighborhood of a given function in the parameter space. However, this entails defining a topology/metric over the parameter space and specifying the

size of the local neighborhood, which can be challenging due to variations across problems. Metrics such as L_1, L_2 or weighted L_2 distances are often employed. In contrast, our framework does not require specifying topology/metric on the parameter space or the size of a local neighborhood. This makes it more convenient and easier to use.

The correct form of the conventional local minimax benchmark for the minimizer under (1.1) is given by $R_{\mathrm{con},z}(\varepsilon; f) = \inf_{\hat{Z}} \sup_{h \in B(f,\varepsilon)} \mathbb{E}_h |\hat{Z} - Z(h)|$, where $B(f,\varepsilon) = \{h \in \mathcal{F} : \|h - f\|_2 \le \varepsilon\}$ is the ε-neighborhood of f in $\mathcal{F}$. The order of $R_{\mathrm{con},z}(\varepsilon; f)$ is clearly no smaller than our local minimax benchmark $R_z(\varepsilon; f)$ because

$$
\begin{aligned}
R_{\mathrm{con},z}(\varepsilon; f) &= \inf_{\hat{Z}} \sup_{h \in B(f,\varepsilon)} \mathbb{E}_h |\hat{Z} - Z(h)| \\
&\ge \inf_{\hat{Z}} \sup_{h \in B(f,\varepsilon)} \sup_{g \in \{f,h\}} \mathbb{E}_g |\hat{Z} - Z(g)| \\
&\ge \Phi(-0.5)\omega(\varepsilon; f) \sim R_z(\varepsilon; f),
\end{aligned}
$$

where the last inequality holds because of $\lim_{\eta \to \varepsilon} \sup_{g:\|g-f\|_2=\eta} |Z(g) - Z(f)| = \omega(\varepsilon; f)$ and the last step is due to Theorem 2.1. Analogous results hold for the other three problems. This means that our benchmarks are at least as stringent as their conventional counterparts. Therefore, our procedures are also adaptive and optimal under the conventional local minimax framework.

The present work can be extended in different directions in addition to the aforementioned ones. For estimation, the results can be easily generalized to the ℓ_q loss for $q > 1$. It is interesting to consider the extremum under more general shape constraints such as s-convexity. In addition, estimation and inference for other nonlinear functionals such as the quadratic functional, entropies and divergences under a nonasymptotic local minimax framework can be studied. We expect the penalty-of-superefficiency property to hold in these problems and our approach to be particularly helpful for the construction of the confidence intervals. Another important direction is to apply our framework to other statistical models such as estimation and inference for the mode and the maximum of a log concave density function based on i.i.d. observations. We expect similar uncertainty principles to hold.

SUPPLEMENTARY MATERIAL

Supplement to "Estimation and inference for minimizer and minimum of convex functions: Optimality, adaptivity and uncertainty principles" (DOI: 10.1214/24-AOS2355SUPP; .pdf). Section A of the supplement presents the simulation results. Section B discusses the comparisons with CLS based methods and the connection with the classical minimax framework. The proofs of the main results and the technical lemmas are given respectively in Sections C and D.

REFERENCES

AGARWAL, A., FOSTER, D. P., HSU, D. J., KAKADE, S. M. and RAKHLIN, A. (2011). Stochastic convex optimization with bandit feedback. In *Advances in Neural Information Processing Systems* (J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K. Q. Weinberger, eds.) **24**. Curran Associates, Red Hook.

AUER, P., ORTNER, R. and SZEPESVÁRI, C. (2007). Improved rates for the stochastic continuum-armed bandit problem. In *Learning Theory. Lecture Notes in Computer Science* **4539** 454–468. Springer, Berlin. MR2397605 https://doi.org/10.1007/978-3-540-72927-3_33

BELITSER, E., GHOSAL, S. and VAN ZANTEN, H. (2012). Optimal two-stage procedures for estimating location and size of the maximum of a multivariate regression function. *Ann. Statist.* **40** 2850–2876. MR3097962 https://doi.org/10.1214/12-AOS1053

BIRGÉ, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann. Statist.* **17** 1532–1549. MR1026298 https://doi.org/10.1214/aos/1176347380

BLUM, J. R. (1954). Multidimensional stochastic approximation methods. *Ann. Math. Stat.* **25** 737–744. MR0065092 https://doi.org/10.1214/aoms/1177728659

CAI, T. T., CHEN, R. and ZHU, Y. (2024). Supplement to "Estimation and Inference for Minimizer and Minimum of Convex Functions: Optimality, Adaptivity and Uncertainty Principles." https://doi.org/10.1214/24-AOS2355SUPP

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

CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068

CHATTERJEE, S., DUCHI, J. C., LAFFERTY, J. and ZHU, Y. (2016). Local minimax complexity of stochastic convex optimization. In *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon and R. Garnett, eds.) **29**. Curran Associates, Red Hook.

CHEN, H. (1988). Lower rate of convergence for locating a maximum of a function. *Ann. Statist.* **16** 1330–1334. MR0959206 https://doi.org/10.1214/aos/1176350965

CHEN, H., HUANG, M.-N. L. and HUANG, W.-J. (1996). Estimation of the location of the maximum of a regression function using extreme order statistics. *J. Multivariate Anal.* **57** 191–214. MR1391168 https://doi.org/10.1006/jmva.1996.0029

DENG, H., HAN, Q. and SEN, B. (2023). Inference for local parameters in convexity constrained models. *J. Amer. Statist. Assoc.* **118** 2721–2735. MR4681616 https://doi.org/10.1080/01621459.2022.2071721

DIPPON, J. (2003). Accelerated randomized stochastic optimization. *Ann. Statist.* **31** 1260–1281. MR2001650 https://doi.org/10.1214/aos/1059655913

DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26** 288–314. MR1611768 https://doi.org/10.1214/aos/1030563987

FACER, M. R. and MÜLLER, H.-G. (2003). Nonparametric estimation of the location of a maximum in a response surface. *J. Multivariate Anal.* **87** 191–217. MR2007268 https://doi.org/10.1016/S0047-259X(03)00030-7

GHOSAL, P. and SEN, B. (2017). On univariate convex regression. *Sankhya A* **79** 215–253. MR3707421 https://doi.org/10.1007/s13171-017-0104-8

GUNTUBOYINA, A. and SEN, B. (2018). Nonparametric shape-restricted regression. *Statist. Sci.* **33** 568–594. MR3881209 https://doi.org/10.1214/18-STS665

HENGARTNER, N. W. and STARK, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* **23** 525–550. MR1332580 https://doi.org/10.1214/aos/1176324534

KIEFER, J. (1982). Optimum rates for non-parametric density and regression estimates under order restrictions. *Stat. Probab. Essays Honor CR Rao* **419** 428.

KIEFER, J. and WOLFOWITZ, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Stat.* **23** 462–466. MR0050243 https://doi.org/10.1214/aoms/1177729392

KLEINBERG, R. (2004). Nearly tight bounds for the continuum-armed bandit problem. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, *NIPS'04* 697–704. MIT Press, Cambridge, MA, USA.

KLEINBERG, R., SLIVKINS, A. and UPFAL, E. (2019). Bandits and experts in metric spaces. *J. ACM* **66** 30. MR3962354 https://doi.org/10.1145/3299873

MOKKADEM, A. and PELLETIER, M. (2007). A companion for the Kiefer–Wolfowitz–Blum stochastic approximation algorithm. *Ann. Statist.* **35** 1749–1772. MR2351104 https://doi.org/10.1214/009053606000001451

MÜLLER, H.-G. (1989). Adaptive nonparametric peak estimation. *Ann. Statist.* **17** 1053–1069. MR1015137 https://doi.org/10.1214/aos/1176347255

POLYAK, B. T. and TSYBAKOV, A. B. (1990). Optimal orders of accuracy for search algorithms of stochastic optimization. *Problemy Peredachi Informatsii* **26** 45–53. MR1074128

SHOUNG, J.-M. and ZHANG, C.-H. (2001). Least squares estimators of the mode of a unimodal regression function. *Ann. Statist.* **29** 648–665. MR1865335 https://doi.org/10.1214/aos/1009210684