Minimax Lower Bound of k-Monotone Estimation in the Sup-norm

Teresa M. Lebair
Laboratory for Physical Sciences
University of Maryland, College Park
Email: tlebair@lps.umd.edu

Jinglai   Shen
Department of Mathematics and Statistics
University of Maryland, Baltimore County
Email: shenj@umbc.edu

Abstract—Belonging to the framework of shape constrained estimation, k-monotone estimation refers to the nonparametric estimation of univariate k-monotone functions, e.g., monotone and convex functions. This paper develops minimax lower bounds for k-monotone regression problems under the sup-norm for general k by constructing a family of k-monotone piecewise polynomial functions (or hypotheses) belonging to suitable Hölder and Sobolev classes. After establishing that these hypotheses satisfy several properties, we employ results from general minimax lower bound theory to obtain the desired k-monotone regression minimax lower bound. Implications and extensions are also discussed.

I. INTRODUCTION

A univariate function is k-monotone for $k \in \mathbb{N}$ if its $(k-1)$th derivative is increasing; examples of k-monotone functions include monotone ($k = 1$) and convex ($k = 2$) functions. The goal of k-monotone estimation is to develop effective estimators that preserve the k-monotone constraint. An important special case of shape constrained estimation, k-monotone estimation has received considerable interest in statistics and system identification [1], [12]. In particular, monotone and convex estimation have been extensively studied in the literature [2], [4], [9], [10], [13], [14], [16], [17]. Motivated by numerous applications of k-monotone estimation when $k > 2$, e.g., insurance [3], qualitative simulation [5], and attitude control [11], the asymptotic performance of a k-monotone estimator for general k was recently studied in [8]. Specifically, a two stage k-monotone B-spline regression estimator was proposed. Supported by the critical uniform Lipschitz property, it was shown that this estimator achieves the "optimal" convergence rate for general k over a suitable function class under the sup-norm and other norms [8].

A related question, pertaining to k-monotone estimation minimax lower bound theory, asks whether the convergence rate attained in [8] is strict for any k-monotone estimator. To answer this question, we adapt results from general minimax lower bound theory to establish minimax lower bounds for k-monotone regression problems. Particularly, inspired by the recent work on convex estimation minimax lower bounds in the sup-norm [9], we construct a family of k-monotone piecewise polynomial functions (or hypotheses) in suitable Hölder and Sobolev classes. These hypotheses lead to the desired minimax lower bound; implications and extensions are

also discussed. The results presented in this paper originally appear in the first author's Ph.D. thesis [7].

This paper is organized as follows. In Section II, we present the k-monotone regression problem and state the main result (cf. Theorem II.1). Next, in Section III we assemble the hypothesis functions. A proof of the main result is given in Section IV. Finally, several extensions and implications are presented in Section V.

Notation: Given a function $g : [a,b] \to \mathbb{R}$, denote its sup(remum)-norm and L_2-norm by $\|g\|_\infty := \sup_{x \in [a,b]} |g(x)|$ and $\|g\|_{L_2} := (\int_a^b (g(x))^2 \, dx)^{1/2}$, respectively. Define the pth integral of g on $[a,b]$ as

$$\mathcal{I}_{[a,b]}^{(p)}(g) := \tag{1}$$

$$\begin{cases} \int_a^b g(t_1) \, dt_1 & \text{if } p = 1 \\ \int_a^b \int_a^{t_p} \cdots \int_a^{t_2} g(t_1) \, dt_1 \ldots \, dt_{p-1} \, dt_p & \text{otherwise.} \end{cases}$$

Let $f^{(q)}$ denote the qth derivative of the function f. For two sequences of positive numbers (a_n) and (b_n), we write $a_n \asymp b_n$ if there exist constants $c_1, c_2 > 0$, such that $c_1 \leq \liminf_{n \to \infty} a_n/b_n \leq \limsup_{n \to \infty} a_n/b_n \leq c_2$. Finally, for $n, m \in \mathbb{N}$ with $n \leq m$, $[n : m]$ denotes the set of integers $\{n, n+1, \ldots, m\}$.

II. PROBLEM FORMULATION

Fix $k \in \mathbb{N}$, $r \in (k-1, k]$ and $L > 0$. Let $\gamma := r - k + 1$. The family of k-monotone univariate functions on $[0,1]$ is

$$\mathcal{S}_k := \left\{ f : [0,1] \to \mathbb{R} \ \middle| \ \text{the } (k-1)\text{th derivative } f^{(k-1)} \right.$$
$$\text{exists a.e. on } [0,1], \text{ and}$$
$$\left(f^{(k-1)}(x_1) - f^{(k-1)}(x_2) \right) \cdot \left(x_1 - x_2 \right) \geq 0$$
$$\left. \text{when } f^{(k-1)}(x_1), f^{(k-1)}(x_2) \text{ exist} \right\}. \tag{2}$$

When $k = 1, 2$, $\mathcal{S}_k$ represents the set of monotone functions and continuous convex functions, respectively. Additionally, denote the Hölder class of functions as H_L^r, i.e.,

$$H_L^r := \left\{ f : [0,1] \to \mathbb{R} \ \middle| \ |f^{(k-1)}(x) - f^{(k-1)}(y)| \right.$$
$$\left. \leq L|x - y|^\gamma, \ \forall \ x, y \in [0,1] \right\}. \tag{3}$$

Finally, set $\mathcal{S}_{k,H}(r, L) := \mathcal{S}_k \cap H_L^r$.

Consider the k-monotone regression problem

$$y_i = f(x_i) + \sigma \epsilon_i, \qquad (4)$$

where the x_i's are evenly spaced design points on the unit interval, i.e., $x_i = i/n$ for all $i \in [0:n]$, n denotes the sample size, $f : [0,1] \to \mathbb{R}$ is an unknown function in $\mathcal{S}_{k,H}(r,L)$, and the ϵ_i's are iid standard normal errors. Our goal is to establish a lower bound under the sup-norm on the minimax risk associated with the collection of estimators that preserve the k-monotone constraint of $f \in \mathcal{S}_{k,H}(r,L)$, for the nonparametric model (4). Specifically, the main result of this paper is presented in the following theorem.

Theorem II.1. *Fix $k \in \mathbb{N}$, $r \in (k-1,k]$, and consider the regression problem (4). There is a constant $c > 0$ such that*

$$\liminf_{n \to \infty} \inf_{\widehat{f}_n} \sup_{f \in \mathcal{S}_{k,H}(r,L)} \left(\frac{n}{\log n} \right)^{\frac{r}{2r+1}} \mathbb{E}(\|\widehat{f}_n - f\|_\infty) \geq c, \quad (5)$$

where $\inf_{\widehat{f}_n}$ denotes the infimum over all constrained estimators $\widehat{f}_n \in \mathcal{S}_k$ on $[0,1]$.

Our strategy to substantiate Theorem II.1, motivated by [15, Theorem 2.10] and [9], amounts to constructing a class of hypothesis functions that lie in $\mathcal{S}_{k,H}(r,L)$. These functions will maintain a suitable distance from each other in the sup-norm, while staying sufficiently close to each other under the L_2-norm. Specifically, this family of M_n hypotheses $f_{j,n}, j \in [0:M_n]$ must satisfy the following three conditions:

(C1) each $f_{j,n} \in \mathcal{S}_{k,H}(r,L)$, $j \in [0:M_n]$;
(C2) whenever $j \neq \ell$, $\|f_{j,n} - f_{\ell,n}\|_\infty \geq 2s_n > 0$, where $s_n \asymp (\log n/n)^{r/(2r+1)}$;
(C3) there exists a constant $c_0 \in (0,1/8)$ such that for all n sufficiently large,

$$\frac{1}{M_n} \sum_{j=1}^{M_n} K(P_j, P_0) \leq c_0 \log(M_n),$$

where P_j denotes the distribution of $(Y_{j,1}, \ldots, Y_{j,n})$, $Y_{j,i} = f_{j,n}(X_i) + \xi_i$, $j \in [1:M_n]$, $X_i = i/n$, the ξ_i's are iid random variables, and $K(\cdot, \cdot)$ denotes the Kullback divergence between two probability measures [6].

We will specify M_n in the later development. We also assume that there exists a constant $p_* > 0$ (independent of n and $f_{j,n}$) such that $K(P_j, P_0) \leq p_* \sum_{i=1}^{n} (f_{j,n}(X_i) - f_{0,n}(X_i))^2$. This assumption holds if the iid random variables $\xi_i \sim N(0, \sigma^2)$ (cf. [15, (2.36)] or [15, Section 2.5, Assumption B]). Hence, the regression problem (4) satisfies this assumption.

III. CONSTRUCTION OF HYPOTHESIS FUNCTIONS

In this section, we construct a family of suitable functions $f_{j,n}$ satisfying (C1)-(C3). Let (K_n) be the increasing sequence of natural numbers given by

$$K_n := \left\lceil \left(\frac{n}{\log n} \right)^{\frac{1}{2r+1}} \right\rceil, \qquad (6)$$

and fix n large. For that fixed n, define $\kappa_i := \frac{i}{K_n}$ for each $i \in [0:K_n]$. To construct the desired functions $f_{j,n}$,

$j \in [0:M_n]$, we begin by constructing increasing functions $h_{p,n}^{[1]}$, $h_{p,n}^{[2]} : [0, \kappa_{2^p}] \to \mathbb{R}$ inductively for $p \in [1:k]$. Our procedure for constructing the $f_{j,n}$'s involves (i) using $h_{k,n}^{[1]}$ and $h_{k,n}^{[2]}$ to construct the $g_{j,n}$'s (cf. (16)-(17) and (19)-(20)) and (ii) integrating the $g_{j,n}$'s a total of $(k-1)$ times to produce the $f_{j,n}$'s (cf. (21)). We choose $h_{k,n}^{[1]}$ and $h_{k,n}^{[2]}$ for this procedure for the following reasons.

(i) In order to meet (C1), each $g_{j,n} = f_{j,n}^{(k-1)}$ must be increasing. Both $h_{k,n}^{[1]}$ and $h_{k,n}^{[2]}$ (used in the construction of each $g_{j,n}$) meet this requirement.

(ii) In view of (C2), the distance (measured with respect to the sup-norm) between $f_{j,n}$ and $f_{\ell,n}$ for $j \neq \ell$ must be non-small. The $f_{j,n}$'s are constructed such that $f_{j,n}(x) = f_{\ell,n}(x)$ when $j \neq \ell$ for all x outside of two small subintervals of $[0,1]$. On these subintervals, $f_{j,n} - f_{\ell,n}$ will be equal to the $(k-1)$th integral of either $h_{k,n}^{[2]} - h_{k,n}^{[1]}$ or $h_{k,n}^{[1]} - h_{k,n}^{[2]}$ (cf. Lemma III.3). This integral is on the order of K_n^{k-1} times smaller than $h_{k,n}^{[2]} - h_{k,n}^{[1]}$ in the sup-norm. Although this integral is relatively small, it will still be large enough for the $f_{j,n}$'s to meet (C2).

(iii) The $(k-1)$th integral of $h_{k,n}^{[2]} - h_{k,n}^{[1]}$, is on the order of K_n^{k-1} times smaller than $h_{k,n}^{[2]} - h_{k,n}^{[1]}$ in the sup-norm, and even smaller in the L_2-norm. Since the $f_{j,n}$'s are constructed such that $|f_{j,n} - f_{\ell,n}|$ is equal to $|\mathcal{I}_{[0,\cdot]}^{(k-1)}(h_{k,n}^{[2]} - h_{k,n}^{[1]})|$ (cf. (1)) on two small subintervals of $[0,1]$ and zero elsewhere for $j \neq \ell$ (cf. Lemma III.3), the L_2-norm of $f_{j,n} - f_{\ell,n}$ will be sufficiently small and allow the $f_{j,n}$'s to meet (C3).

With these ideas in mind, we proceed to construct the auxiliary functions $h_{p,n}^{[1]}$ and $h_{p,n}^{[2]}$, for $p \in [1:k]$ and establish two technical lemmas in Section III-A. In Section III-B, we use the constructions and results from Section III-A to (i) construct the $g_{j,n}$'s, (ii) construct the hypotheses $f_{j,n}$, $j \in [0:M_n]$ by integrating each $g_{j,n}$ a total of $(k-1)$ times, and (iii) establish Lemma III.3, which will pave the way for the proof of Theorem II.1.

A. Construction of Auxiliary Functions

We will now construct the auxiliary functions $h_{p,n}^{[1]}$ and $h_{p,n}^{[2]}$ to be used in the sequel. Let $c_0 \in (0, \frac{1}{8})$, and p_* be the positive constant indicated after condition (C3) in the previous section. Choose

$$\overline{L} := \min \left\{ \frac{L}{2^k}, \frac{k!}{2^{k^k}} \sqrt{\frac{\gamma c_0}{2^{k+3} p_* (2r+1)}} \right\}. \qquad (7)$$

Define the following functions $h_{p,n}^{[1]}, h_{p,n}^{[2]} : [0, \kappa_{2^p}] \to \mathbb{R}$ recursively for $p \in [1:k]$ as follows. First let

$$h_{1,n}^{[1]}(x) := \begin{cases} 0 & \text{if } x \in [0, \kappa_1] \\ \overline{L} K_n^{1-\gamma} (x - \kappa_1) & \text{if } x \in (\kappa_1, \kappa_2] \end{cases} \qquad (8)$$

and

$$h_{1,n}^{[2]}(x) := \begin{cases} \overline{L} K_n^{1-\gamma} x & \text{if } x \in [0, \kappa_1] \\ \overline{L} K_n^{-\gamma} & \text{if } x \in (\kappa_1, \kappa_2]. \end{cases} \qquad (9)$$

Then for $p \in [2 : k]$, define

$$h_{p,n}^{[1]}(x) := \tag{10}$$
$$\begin{cases} h_{p-1,n}^{[1]}(x) & \text{if } x \in [0, \kappa_{2^{p-1}}] \\ h_{p-1,n}^{[2]}(x - \kappa_{2^{p-1}}) + h_{p-1,n}^{[1]}(\kappa_{2^{p-1}}) & \text{if } x \in (\kappa_{2^{p-1}}, \kappa_{2^p}] \end{cases}$$

and

$$h_{p,n}^{[2]}(x) := \tag{11}$$
$$\begin{cases} h_{p-1,n}^{[2]}(x) & \text{if } x \in [0, \kappa_{2^{p-1}}] \\ h_{p-1,n}^{[1]}(x - \kappa_{2^{p-1}}) + h_{p-1,n}^{[2]}(\kappa_{2^{p-1}}) & \text{if } x \in (\kappa_{2^{p-1}}, \kappa_{2^p}]. \end{cases}$$

Figure 1 contains plots of these functions for $p = 1, 2, 3$. Note that both $h_{1,n}^{[1]}$ and $h_{1,n}^{[2]}$ are continuous. By induction, so are $h_{p,n}^{[1]}$ and $h_{p,n}^{[2]}$ for $p \in [2 : k]$. Additionally, for $p \in [2 : k]$, the first "half" of $h_{p,n}^{[1]}$ (i.e., the part defined on $[0, \kappa_{2^{p-1}}]$) is identical to $h_{p-1,n}^{[1]}$, and the second "half" of $h_{p,n}^{[1]}$ (i.e., the part defined on $(\kappa_{2^{p-1}}, \kappa_{2^p}]$), when shifted appropriately, is identical to $h_{p-1,n}^{[2]}$. An analogous relationship holds for $h_{p,n}^{[2]}$, $h_{p-1,n}^{[2]}$ on $[0, \kappa_{2^{p-1}}]$, and $h_{p-1,n}^{[1]}$ on $(\kappa_{2^{p-1}}, \kappa_{2^p}]$.

For each $p \in [1 : k]$, define $\varphi_p : [0, \kappa_{2^k}] \to \mathbb{R}$ such that for all $x \in [0, \kappa_{2^k}]$,

$$\varphi_1(x) := (h_{k,n}^{[2]} - h_{k,n}^{[1]})(x) \text{ and} \tag{12}$$
$$\varphi_p(x) := \mathcal{I}_{[0,x]}^{(p-1)}(\varphi_1), \ p \in [2 : k].$$

Note that on $[0, \kappa_{2^p}]$,

$$\varphi_1(x) = (h_{k,n}^{[2]} - h_{k,n}^{[1]})(x) = (h_{k-1,n}^{[2]} - h_{k-1,n}^{[1]})(x) \tag{13}$$
$$= \cdots = (h_{p,n}^{[2]} - h_{p,n}^{[1]})(x).$$

Also, on $[0, \kappa_{2^k}]$, the qth derivative

$$\varphi_p^{(q)}(x) = \varphi_{p-q}(x), \text{ for all } q \in [1 : p-1]. \tag{14}$$

The following two lemmas will be useful in Section III-B.

Lemma III.1. *For each* $p \in [1 : k]$, $\varphi_p^{(q)}(\kappa_{2^p}) = 0$ *for* $q \in [0 : p-1]$. *Note that for* $p = k$, *we define* $\varphi_k^{(q)}(\kappa_{2^k})$ *as the* qth *right derivative of* φ_k *at* κ_{2^k}.

Proof. We prove this result by induction on p. For $p = 1$, by virtue of (8)-(9), and (12),

$$\varphi_1(\kappa_2) = (h_{1,n}^{[2]} - h_{1,n}^{[1]})(\kappa_2) = \overline{L}K_n^{-\gamma} - \overline{L}K_n^{-\gamma} = 0.$$

Hence, the result holds for $p = 1$. Fix $p \in [2 : k]$ and assume that the result holds for $(p-1)$. In what follows, we show that the result then holds for p.

Consider $q = 0$ first. Since φ_p is $(p-1)$ times differentiable, by Taylor expansion

$$\varphi_p(\kappa_{2^p}) = \sum_{r=0}^{p-2} \frac{\varphi_p^{(r)}(\kappa_{2^{p-1}}) \left(\frac{2^{p-1}}{K_n}\right)^r}{r!} + \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-1)}(\varphi_p^{(p-1)}).$$

By (14) and the induction hypothesis, we have

$$\varphi_p^{(r)}(\kappa_{2^{p-1}}) = \varphi_{p-r}(\kappa_{2^{p-1}}) = \varphi_{p-1}^{(r-1)}(\kappa_{2^{p-1}}) = 0,$$
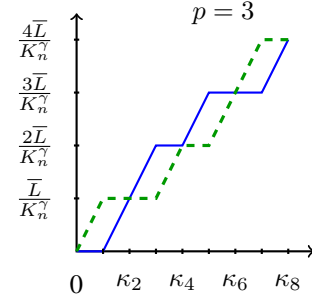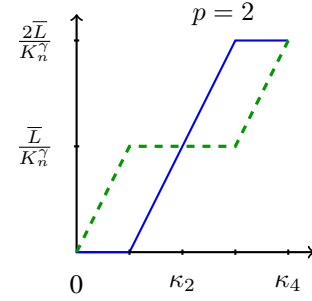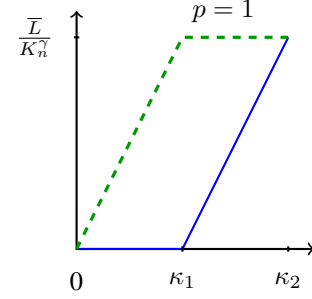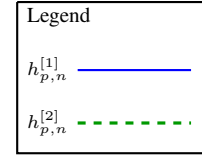

Fig. 1: Plot of $h_{p,n}^{[1]}$ and $h_{p,n}^{[2]}$ for $p = 1, 2, 3$.

for $r \in [1 : p-2]$. Hence, by the above two displays,

$$\varphi_p(\kappa_{2^p})$$
$$= \varphi_p(\kappa_{2^{p-1}}) + \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-1)}(\varphi_p^{(p-1)})$$
$$= \varphi_p(\kappa_{2^{p-1}}) + \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-1)}(h_{p,n}^{[2]} - h_{p,n}^{[1]})$$
$$= \varphi_p(\kappa_{2^{p-1}})$$
$$\quad + \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-1)}\Big((h_{p-1,n}^{[1]} - h_{p-1,n}^{[2]})(\ \cdot\ - \kappa_{2^{p-1}})$$
$$\qquad\qquad\qquad\qquad + \varphi_1(\kappa_{2^{p-1}})\Big)$$
$$= \varphi_p(\kappa_{2^{p-1}}) + \mathcal{I}_{[0, \kappa_{2^{p-1}}]}^{(p-1)}(h_{p-1,n}^{[1]} - h_{p-1,n}^{[2]})$$
$$= \varphi_p(\kappa_{2^{p-1}}) - \varphi_p(\kappa_{2^{p-1}}) = 0,$$

where we use (10)-(14), and $\varphi_1(\kappa_{2^{p-1}}) = \varphi_{p-1}^{(p-2)}(\kappa_{2^{p-1}}) = 0$

via (14) and the induction hypothesis.

Next consider $q \in [1 : p - 2]$. We have that

$$\varphi_p^{(q)}(\kappa_{2^p})$$
$$= \sum_{r=0}^{p-q-2} \frac{\varphi_p^{(q+r)}(\kappa_{2^{p-1}}) \left(\frac{2^{p-1}}{K_n}\right)^r}{r!} + \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-q-1)}(\varphi_p^{(p-1)}).$$

In view of $\varphi_p^{(q+r)}(\kappa_{2^{p-1}}) = \varphi_{p-q-r}(\kappa_{2^{p-1}}) = 0$ for all $r \in [0 : p - q - 2]$ via (14) and the induction hypothesis, we have by the above display, and (10)-(14),

$$\varphi_p^{(q)}(\kappa_{2^p}) = \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-q-1)}(\varphi_1) = \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-q-1)}(h_{p,n}^{[2]} - h_{p,n}^{[1]})$$
$$= \mathcal{I}_{[\kappa_{2^{p-1}}, \kappa_{2^p}]}^{(p-q-1)} \left[(h_{p-1,n}^{[1]} - h_{p-1,n}^{[2]})(\cdot - \kappa_{2^{p-1}}) + \varphi_1(\kappa_{2^{p-1}}) \right]$$
$$= -\mathcal{I}_{[0,\kappa_{2^{p-1}}]}^{(p-q-1)}(\varphi_1) = -\varphi_{p-q}(\kappa_{2^{p-1}}) = 0.$$

Finally, if $q = p - 1$, then by (10)-(14),

$$\varphi_p^{(q)}(\kappa_{2^p}) = \varphi_1(\kappa_{2^p}) = h_{p,n}^{[2]}(\kappa_{2^p}) - h_{p,n}^{[1]}(\kappa_{2^p})$$
$$= \left(h_{p-1,n}^{[1]}(\kappa_{2^{p-1}}) + h_{p-1,n}^{[2]}(\kappa_{2^{p-1}}) \right)$$
$$\quad - \left(h_{p-1,n}^{[2]}(\kappa_{2^{p-1}}) + h_{p-1,n}^{[1]}(\kappa_{2^{p-1}}) \right)$$
$$= 0.$$

Hence the result holds by induction. $\square$

Lemma III.2. *Let* $r \in (k-1, k]$ *and* $\gamma = r - (k-1)$. *For* $\varphi_k : [0, \kappa_{2^k}] \to \mathbb{R}$, $\frac{\overline{L}}{k!} K_n^{-r} \leq \|\varphi_k\|_\infty \leq \frac{\overline{L} \, 2^{k^k}}{k!} K_n^{-r}$.

Proof. Consider $k = 1$. It is easy to see via Figure 1 that

$$\|\varphi_1\|_\infty = \|h_{1,n}^{[2]} - h_{1,n}^{[1]}\|_\infty = \overline{L} K_n^{-\gamma},$$

so the result holds for $k = 1$.

We then consider $k > 1$. On $[0, \kappa_1]$, we have that $\varphi_1(x) = h_{1,n}^{[2]}(x) - h_{1,n}^{[1]}(x) = \overline{L} K_n^{1-\gamma} x$. Hence,

$$\|\varphi_k\|_\infty \geq \varphi_k(\kappa_1) = \mathcal{I}_{[0,\kappa_1]}^{(k-1)}(\varphi_1)$$
$$= \mathcal{I}_{[0,\kappa_1]}^{(k-1)}(\overline{L} K_n^{1-\gamma} \cdot) = \frac{\overline{L}}{k!} K_n^{-r}.$$

We claim that $|\varphi_1(x)| \leq \overline{L} K_n^{-\gamma}$ for all $x \in [0, \kappa_{2^p}]$ for $p \in [1 : k]$, and prove this claim by induction on p. Certainly this claim holds for $p = 1$, via (8)-(9) and the definition of φ_1 (12). If $p \in [2 : k]$, and the result holds for $(p - 1)$, we need only show that $|\varphi_1(x)| \leq \overline{L} K_n^{-\gamma}$ on $(\kappa_{2^{p-1}}, \kappa_{2^p}]$. If $x \in (\kappa_{2^{p-1}}, \kappa_{2^p}]$, then by (10)-(13) and Lemma III.1,

$$|\varphi_1(x)| = |(h_{p,n}^{[2]} - h_{p,n}^{[1]})(x)|$$
$$= |(h_{p-1,n}^{[1]} - h_{p-1,n}^{[2]})(x - \kappa_{2^{p-1}}) + \varphi_1(\kappa_{2^{p-1}})|$$
$$= |(h_{p-1,n}^{[1]} - h_{p-1,n}^{[2]})(x - \kappa_{2^{p-1}})| \leq \overline{L} K_n^{-\gamma},$$

since $(x - \kappa_{2^{p-1}}) \in [0, \kappa_{2^{p-1}}]$. By observing that

$$\|\varphi_k\|_\infty \leq \mathcal{I}_{[0,\kappa_{2^k}]}^{(k-1)}(\|\varphi_1\|_\infty) \leq \mathcal{I}_{[0,\kappa_{2^k}]}^{(k-1)}(\overline{L} K_n^{-\gamma})$$
$$= \frac{\overline{L} \, 2^{k^{k-1}}}{(k-1)!} K_n^{-r} \leq \frac{\overline{L} \, 2^{k^k}}{k!} K_n^{-r}$$

and the established claim, the proof is complete. $\square$

Now that we have established the previous two preliminary results, we are ready to construct the hypothesis functions using $h_{k,n}^{[1]}$ and $h_{k,n}^{[2]}$ in the next section.

B. Construction of the Hypotheses

In the sequel, we construct the $(k-1)$th derivatives of the $f_{j,n}$'s, namely the $g_{j,n}$'s, $j \in [0 : M_n]$. We then integrate the $g_{j,n}$'s to create the $f_{j,n}$'s. In Section IV, we will demonstrate that the $f_{j,n}$'s meet conditions (C1)-(C3) of Section II. To this end, we consider two different cases in constructing the $g_{j,n}$'s.

Case 1: $r \in (k-1, k)$, so that $\gamma = r - k + 1 \in (0, 1)$. In this case, let $M_n := \lfloor K_n^\gamma \rfloor - 1 \in \mathbb{N}$ and define the intervals

$$I_j := [(j-1)K_n^{-\gamma}, (j-1)K_n^{-\gamma} + \kappa_{2^k}) \tag{15}$$

for each $j \in \mathbb{N}$. Then, let $g_{0,n} : [0,1] \to \mathbb{R}$ be such that

$$g_{0,n}(x) := \begin{cases} (j-1)2^{k-1}\overline{L}K_n^{-\gamma} + h_{k,n}^{[1]}(x - (j-1)K_n^{-\gamma}) \\ \qquad\qquad \text{if } x \in I_j \cap [0,1] \\ j \, 2^{k-1}\overline{L}K_n^{-\gamma} \text{ if } x \in [(j-1)K_n^{-\gamma}, j \, K_n^{-\gamma}) \setminus I_j, \end{cases} \tag{16}$$

for all appropriate $j \in \mathbb{N}$, where $h_{k,n}^{[1]}$ is defined in (10). Note that we assume that n is large enough so that $\kappa_{2^k} < K_n^{-\gamma}$, and hence, $g_{0,n}$ is well defined. Next, for each $j \in [1 : M_n]$, we define $g_{j,n} : [0,1] \to \mathbb{R}$ such that

$$g_{j,n}(x) := \begin{cases} (j-1)2^{k-1}\overline{L}K_n^{-\gamma} + h_{k,n}^{[2]}(x - (j-1)K_n^{-\gamma}) \\ \qquad\qquad \text{if } x \in I_j \\ g_{0,n}(x) \qquad\qquad \text{otherwise}, \end{cases} \tag{17}$$

where $h_{k,n}^{[2]}$ is defined in (11). Hence, $g_{j,n}(x) = g_{0,n}(x)$ for all $x \in [0,1] \setminus I_j$. For $x \in I_j$,

$$g_{j,n}(x) - g_{0,n}(x)$$
$$= h_{k,n}^{[2]}(x - (j-1)K_n^{-\gamma}) - h_{k,n}^{[1]}(x - (j-1)K_n^{-\gamma})$$
$$= \varphi_1(x - (j-1)K_n^{-\gamma}).$$

Furthermore, $g_{0,n}$ is continuous, since (i) $h_{k,n}^{[1]}$ is continuous on $[0, \kappa_{2^k}]$, and (ii) it can be shown by induction that $h_{k,n}^{[1]}(\kappa_{2^k}) = 2^{k-1}\overline{L}K_n^{-\gamma}$; thus for each j,

$$(j-1)2^{k-1}\overline{L}K_n^{-\gamma}$$
$$+ h_{k,n}^{[1]}((j-1)K_n^{-\gamma} + \kappa_{2^k} - (j-1)K_n^{-\gamma})$$
$$= j2^{k-1}\overline{L}K_n^{-\gamma}.$$

A similar argument shows that $g_{j,n}$ is continuous, $j \in [1 : M_n]$.

In Figure 2 a plot of several of the $g_{j,n}$'s near the origin is given for $k = 1, 2, 3$.

Case 2: $r = k$, so that $\gamma = 1$. In this case, let $M_n := \lfloor \frac{K_n}{2^k} \rfloor - 1 \in \mathbb{N}$ and define the intervals

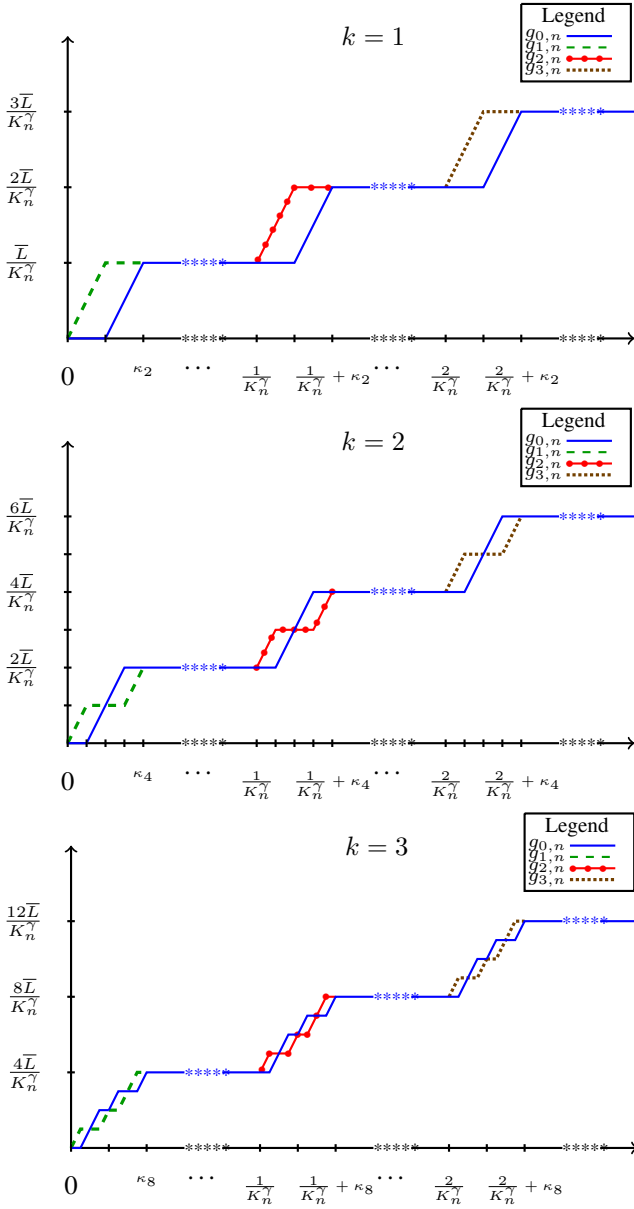$$I_j := [(j-1)2^k K_n^{-1}, j \, 2^k K_n^{-1}) \tag{18}$$

Fig. 2: Plot of the $g_{j,n}$'s near the origin, $\gamma \in (0,1)$.

Fig. 3: Plot of the $g_{j,n}$'s near the origin, $\gamma = 1$.

for each $j \in \mathbb{N}$. Then define the function $g_{0,n} : [0,1] \to \mathbb{R}$ such that

$$g_{0,n}(x) := (j-1)\, 2^{k-1}\overline{L}K_n^{-1} + h_{j,n}^{[1]}(x - (j-1)\, 2^k K_n^{-1})$$
$$\text{if } x \in I_j \cap [0,1] \quad (19)$$

for all appropriate $j \in \mathbb{N}$, where $h_{k,n}^{[1]}$ is defined in (10). Next for each $j \in [1:M_n]$ let $g_{j,n} : [0,1] \to \mathbb{R}$ be such that

$$g_{j,n}(x) := \qquad\qquad\qquad\qquad\qquad\qquad\quad (20)$$
$$\begin{cases} (j-1)2^{k-1}\overline{L}K_n^{-1} + h_{k,n}^{[2]}(x - (j-1)2^k K_n^{-1}) \\ \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } x \in I_j \\ g_{0,n}(x) \qquad\quad \text{otherwise}, \end{cases}$$
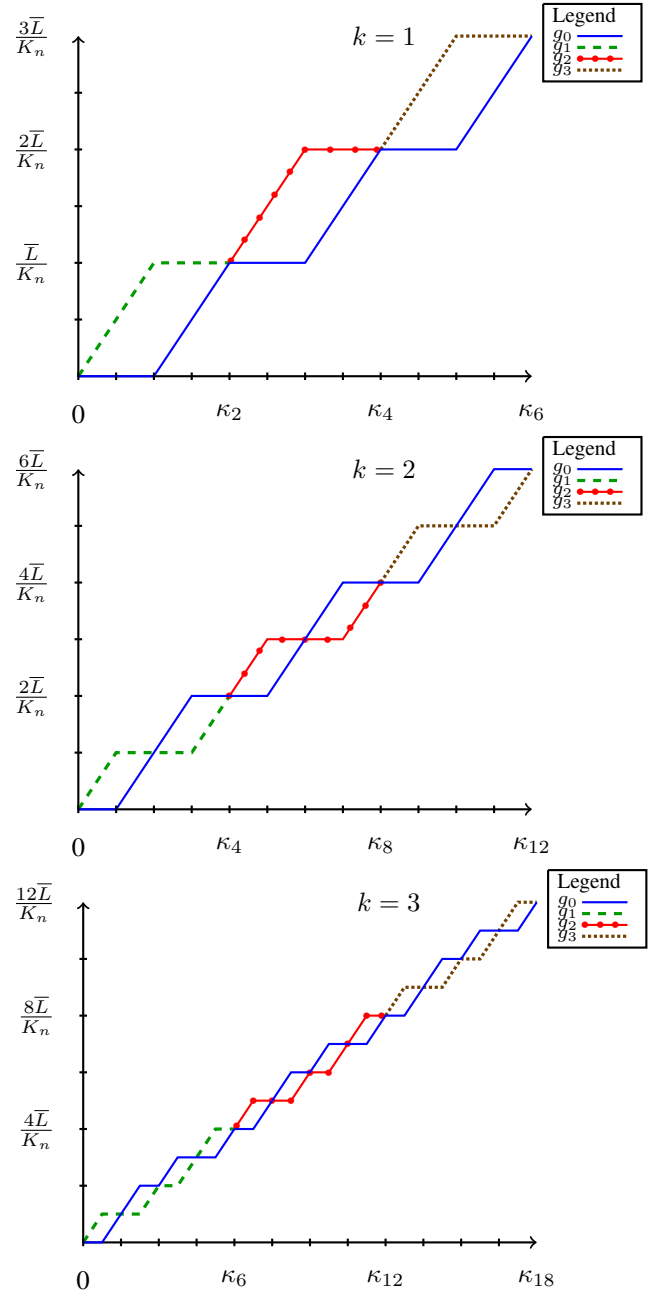
where $h_{k,n}^{[2]}$ is defined in (11). In this case, we again have that, $g_{j,n}(x) = g_{0,n}(x)$ for all $x \in [0,1] \setminus I_j$ and for all $x \in I_j$,

$$g_{j,n}(x) - g_{0,n}(x) = (h_{k,n}^{[2]} - h_{k,n}^{[1]})(x - (j-1)2^k K_n^{-1})$$
$$= \varphi_1(x - (j-1)2^k K_n^{-1}).$$

In Figure 3 a plot of several of the $g_{j,n}$'s near the origin is given for $k = 1, 2, 3$. By an argument similar to that in Case 1, each $g_{j,n}$, $j \in [0 : M_n]$, is continuous on $[0,1]$.

Finally, in either case, define the jth hypothesis function $f_{j,n} : [0,1] \to \mathbb{R}$ such that

$$f_{j,n}(x) := \mathcal{I}_{[0,x]}^{(k-1)}(g_{j,n}), \ \forall \ x \in [0,1]. \qquad (21)$$

Now that we have constructed the hypothesis functions, we will establish one more lemma, before demonstrating that the hypotheses satisfy conditions (C1)-(C3) in Section IV.

Lemma III.3. *If $\gamma \in (0,1)$, then for each $j \in [1:M_n]$, we have*

$$f_{j,n}(x) - f_{0,n}(x) = \begin{cases} \varphi_k(x - (j-1)K_n^{-\gamma}) & \text{if } x \in I_j \\ 0 & \text{otherwise}, \end{cases}$$

where I_j is defined in (15). Alternatively, if $\gamma = 1$, for each $j \in [1:M_n]$, we have

$$f_{j,n}(x) - f_{0,n}(x) = \begin{cases} \varphi_k(x - (j-1)2^k K_n^{-1}) & \text{if } x \in I_j \\ 0 & \text{otherwise}, \end{cases}$$

where I_j is defined in (18).

Proof. Certainly, the result holds for $k = 1$ by (12), (16)-(17), and (19)-(21). In what follows, consider $k > 1$ in two cases.
Case 1: $\gamma \in (0,1)$, so each $f_{j,n}$ is given by (16)-(17), and (21).
Fix $j \in [1:M_n]$. If $x \in [0, (j-1)K_n^{-\gamma})$, then via (17), and (21),

$$(f_{j,n} - f_{0,n})(x) = \mathcal{I}_{[0,x]}^{(k-1)}(g_{j,n} - g_{j,0}) = 0. \quad (22)$$

Suppose that $x \in I_j$ (cf. (15)). By (12), (16)-(17), and (21),

$$\begin{aligned} &(f_{j,n} - f_{0,n})(x) \\ &= \mathcal{I}_{[0,x]}^{(k-1)}(g_{j,n} - g_{j,0}) \\ &= \mathcal{I}_{[(j-1)K_n^{-\gamma},x]}^{(k-1)}\left((h_{k,n}^{[2]} - h_{k,n}^{[1]})(\ \cdot\ - (j-1)K_n^{-\gamma})\right) \\ &= \mathcal{I}_{[0,\ x-(j-1)K_n^{-\gamma}]}^{(k-1)}(\varphi_1) = \varphi_k(x - (j-1)K_n^{-\gamma}). \quad (23) \end{aligned}$$

Finally, consider $x \in [(j-1)K_n^{-\gamma} + \kappa_{2^k}, 1]$. By Lemma III.1, (12)-(14), (16)-(17), and (21),

$$\begin{aligned} 0 &= \varphi_k^{(p)}(\kappa_{2^k}) = \varphi_{k-p}(\kappa_{2^k}) = \mathcal{I}_{[0,\kappa_{2^k}]}^{(k-p-1)}(h_{k,n}^{[2]} - h_{k,n}^{[1]}) \\ &= \mathcal{I}_{[(j-1)K_n^{-\gamma},\ (j-1)K_n^{-\gamma}+\kappa_{2^k}]}^{(k-p-1)} \\ &\qquad\qquad \left((h_{k,n}^{[2]} - h_{k,n}^{[1]})(\ \cdot\ - (j-1)K_n^{-\gamma})\right) \\ &= \mathcal{I}_{[(j-1)K_n^{-\gamma},\ (j-1)K_n^{-\gamma}+\kappa_{2^k}]}^{(k-p-1)}\left(g_{j,n} - g_{0,n}\right) \\ &= \mathcal{I}_{[0,\ (j-1)K_n^{-\gamma}+\kappa_{2^k}]}^{(k-p-1)}\left(g_{j,n} - g_{0,n}\right) \\ &= (f_{j,n} - f_{0,n})^{(p)}((j-1)K_n^{-\gamma} + \kappa_{2^k}), \quad (24) \end{aligned}$$

for $p \in [0:k-1]$. Since $g_{j,n} - g_{0,n}$ is continuous by the discussion below (16)-(17), $f_{j,n} - f_{0,n}$ is $(k-1)$ times continuously differentiable. Hence, via Taylor expansion, (15), (17), (21) and (24),

$$\begin{aligned} (f_{j,n} - f_{0,n})(x) &= (f_{j,n} - f_{0,n})(x) \\ &- \sum_{p=0}^{k-2} \frac{1}{p!}\Big[(f_{j,n} - f_{0,n})^{(p)}\left((j-1)K_n^{-\gamma} + \kappa_{2^k}\right) \\ &\qquad\qquad \times \left(x - [(j-1)K_n^{-\gamma} + \kappa_{2^k}]\right)^p\Big] \\ &= \mathcal{I}_{[(j-1)K_n^{-\gamma}+\kappa_{2^k},\ x]}^{(k-1)}\left((f_{j,n} - f_{0,n})^{(k-1)}\right) \end{aligned}$$

$$\begin{aligned} &= \mathcal{I}_{[(j-1)K_n^{-\gamma}+\kappa_{2^k},\ x]}^{(k-1)}(g_{j,n} - g_{0,n}) \\ &= 0, \quad (25) \end{aligned}$$

and the result holds for Case 1.
Case 2: $\gamma = 1$ so each $f_{j,n}$ is given by (19)-(21). Again, fix $j \in [1:M_n]$. Using arguments similar to those in (22), (23), and (24)-(25), we can also show that Case 2 holds for all $x \in [0, (j-1)2^k K_n^{-1})$, $[(j-1)2^k K_n^{-1}, j\ 2^k K_n^{-1})$, and $[j\ 2^k K_n^{-1}, 1]$, respectively. $\square$

Now that we have constructed the hypotheses and established the previous result, we are ready to show that these functions meet conditions (C1)-(C3) from Section II.

IV. PROOF OF THE MAIN RESULT

We use the previous constructions and results to establish Theorem II.1 in the following proof.

Proof. In each of the following two cases, we demonstrate that the $f_{j,n}$'s of (21) meet conditions (C1)-(C3).
Case 1: $\gamma \in (0,1)$, so that $M_n = \lfloor K_n^\gamma \rfloor - 1$, and the $g_{j,n}$'s are given by (16)-(17). We show as follows that the $f_{j,n}$'s satisfy conditions (C1), (C2), and (C3), respectively.
(1) For all n (and K_n) sufficiently large, the following properties of each $g_{j,n}$ can be easily verified via Figure 2: for $x,y \in [0,1]$, suppose that
(i) $0 < |x-y| \le \kappa_{2^k}$. Then

$$\begin{aligned} \max_j &\frac{|g_{j,n}(x) - g_{j,n}(y)|}{|x-y|} \\ &\le \frac{|g_{0,n}(x) - g_{0,n}(y)|}{|x-y|}\Big|_{x=\kappa_1,\ y=\kappa_2} = \overline{L}K_n^{1-\gamma}. \end{aligned}$$

(ii) $\kappa_{2^k} < |x-y| \le K_n^{-\gamma}$. Then

$$\begin{aligned} \max_j &|g_{j,n}(x) - g_{j,n}(y)| \\ &\le |g_{0,n}(x) - g_{0,n}(y)|\big|_{x=0,\ y=\kappa_{2^k}} = 2^{k-1}\overline{L}K_n^{-\gamma}. \end{aligned}$$

(iii) $K_n^{-\gamma} < |x-y| \le 1$. Without loss of generality, let $x < y$ with $y = qK_n^{-\gamma} + s(x,y)$ for some $q \in \mathbb{N}$ and $0 \le s(x,y) < K_n^{-\gamma}$. It follows from (7) that

$$\begin{aligned} \max_j &\frac{|g_{j,n}(x) - g_{j,n}(y)|}{|x-y|} \\ &\le \max_{i=2^k-1,\ 2^k} \frac{|g_{0,n}(x) - g_{0,n}(y)|}{|x-y|}\Big|_{x=\kappa_1,\ y=qK_n^{-\gamma}+\kappa_i} \\ &\le \frac{2^{k-1}(q+1)\overline{L}K_n^{-\gamma}}{qK_n^{-\gamma} + (2^k-1)K_n^{-1}} \le \frac{2^{k-1}(q+1)\overline{L}K_n^{-\gamma}}{qK_n^{-\gamma}} \\ &\le 2^k\overline{L} \le L. \end{aligned}$$

Since each $g_{j,n}$ is nondecreasing, each $f_{j,n}$ given by (21) belongs to $\mathcal{S}_k$ (cf. (2)). To see that each $f_{j,n}$ is in the Hölder class H_L^r (cf. (3)), we consider the following three scenerios:
(1.1) $0 < |x-y| \le \kappa_{2^k}$. Then, by (i), (7), and (21), we have that

$$\frac{|f_{j,n}^{(k-1)}(x) - f_{j,n}^{(k-1)}(y)|}{|x-y|^\gamma}$$

$$= \frac{|f_{j,n}^{(k-1)}(x) - f_{j,n}^{(k-1)}(y)|}{|x-y|}|x-y|^{1-\gamma}$$
$$\leq \bar{L}K_n^{1-\gamma}\kappa_{2^k}^{1-\gamma} \leq 2^k\bar{L} \leq L.$$

(1.2) $\kappa_{2^k} < |x-y| \leq K_n^{-\gamma}$. Then, by (ii), (7), and (21), we have that
$$\frac{|f_{j,n}^{(k-1)}(x) - f_{j,n}^{(k-1)}(y)|}{|x-y|^\gamma} \leq \frac{2^{k-1}\bar{L}K_n^{-\gamma}}{|x-y|^\gamma}$$
$$\leq 2^{k-1}\bar{L}K_n^{-\gamma}\kappa_{2^k}^{-\gamma} \leq 2^k\bar{L} \leq L.$$

(1.3) $\frac{1}{K_n^\gamma} < |x-y| \leq 1$. By (iii) and (21), we obtain
$$\frac{|f_{j,n}^{(k-1)}(x) - f_{j,n}^{(k-1)}(y)|}{|x-y|^\gamma} \leq \frac{|f_{j,n}^{(k-1)}(x) - f_{j,n}^{(k-1)}(y)|}{|x-y|}$$
$$\leq L.$$

Thus, each $f_{j,n} \in \mathcal{S}_{k,H}(r,L)$, so the $f_{j,n}$'s satisfy (C1).

(2) Suppose that $j \neq \ell$. By Lemma III.3, we have that $f_{j,n}(x) = f_{0,n}(x) = f_{\ell,n}(x)$ for all $x \in [0,1] \setminus (I_j \cup I_\ell)$ (cf. (15)). Hence, $|f_{j,n}(x) - f_{\ell,n}(x)| = 0$ on $[0,1] \setminus (I_j \cup I_\ell)$. Also,
$$|f_{j,n}(x) - f_{\ell,n}(x)| = |f_{j,n}(x) - f_{0,n}(x)|$$
$$= |\varphi_k(x - (j-1)K_n^{-\gamma})| \ \forall \ x \in I_j,$$
and similarly, $|f_{j,n}(x)-f_{\ell,n}(x)| = |\varphi_k(x-(\ell-1)K_n^{-\gamma})| \ \forall \ x \in I_\ell$. Therefore, in view of (6) and Lemma III.2, we see that the $f_{j,n}$'s meet condition (C2) with $s_n := \frac{\bar{L}}{2k!}\left[\left(\frac{n}{\log n}\right)^{\frac{1}{2r+1}}\right]^{-r} \asymp \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}$, as
$$\|f_{j,n} - f_{\ell,n}\|_\infty = \|\varphi_k\|_\infty \geq \frac{\bar{L}}{k!}K_n^{-r} = 2s_n.$$

(3) By the discussion following the statement of (C3) in Section II, there exists $p_* > 0$, independent of n and $f_{j,n}$, such that $K(P_j, P_0) \leq p_*\sum_{i=1}^n (f_{j,n}(X_i) - f_{0,n}(X_i))^2$ where $X_i = \frac{i}{n}$, for all $j \in [1:M_n]$. Also,
$$X_i \in \left[(j-1)K_n^{-\gamma}, (j-1)K_n^{-\gamma} + \kappa_{2^k}\right) \implies$$
$$\lceil n(j-1)K_n^{-\gamma}\rceil \leq i \leq \lfloor n\left((j-1)K_n^{-\gamma} + \kappa_{2^k}\right)\rfloor.$$

Let $a_j := \lceil n(j-1)K_n^{-\gamma}\rceil$ and $b_j := \lfloor n\left((j-1)K_n^{-\gamma} + \kappa_{2^k}\right)\rfloor$. Then by Lemma III.2, Lemma III.3, (6), and (7), for n sufficiently large,
$$K(P_j, P_0) \leq p_*\sum_{i=1}^n \left(f_{j,n}(X_i) - f_{0,n}(X_i)\right)^2$$
$$= p_*\sum_{i=a_j}^{b_j} \left(f_{j,n}(X_i) - f_{0,n}(X_i)\right)^2$$
$$= p_*\sum_{i=a_j}^{b_j} \left(\varphi_k\left(i/n - (j-1)K_n^{-\gamma}\right)\right)^2$$
$$\leq p_*\sum_{i=a_j}^{b_j} \left(\frac{\bar{L}2^{k^k}}{k!}K_n^{-r}\right)^2 \leq p_*2n \ \kappa_{2^k}\left(\frac{\bar{L}2^{k^k}}{k!}K_n^{-r}\right)^2$$

$$= p_*2^{k+1}\left(\frac{\bar{L}2^{k^k}}{k!}\right)^2 nK_n^{-(2r+1)}$$
$$\leq \frac{\gamma c_0}{4(2r+1)}n \ K_n^{-(2r+1)} \leq \frac{\gamma c_0}{4(2r+1)}\log n$$
$$\leq \frac{\gamma c_0}{2(2r+1)}\log\left(\frac{n}{\log n}\right) \leq \frac{\gamma c_0}{2}\log(K_n)$$
$$\leq c_0\log(\lfloor K_n^\gamma\rfloor - 1) = c_0\log M_n$$

for all $j \in [1:M_n]$. Thus, $\frac{1}{M_n}\sum_{j=1}^{M_n} K(P_j, P_0) \leq c_0 \log M_n$, and (C3) holds in Case 1.

Case 2: $\gamma = 1$, so that $M_n = \lfloor \frac{K_n}{2^k}\rfloor - 1$, and the $g_{j,n}$'s are now given by (19)-(20). We now demonstrate that conditions (C1)-(C3) hold for Case 2 in (1)-(3).

(1) It is easy to see that each $g_{j,n}$ is increasing via Figure 3. Hence, each $f_{j,n}$ given by (21) belongs to $\mathcal{S}_k$. Also, it is easy to verify that for any $0 \leq x < y \leq 1$, $\frac{|g_{j,n}(x)-g_{j,n}(y)|}{|x-y|} \leq \bar{L} \leq L$ for each $j \in [0:M_n]$. This thus implies that each $f_{j,n} \in \mathcal{S}_{k,H}(r,L)$. Hence, the $f_{j,n}$'s meet condition (C1).

(2) Suppose $j \neq \ell$. By an argument similar to that in Case 1 and (18),
$$|f_{j,n}(x) - f_{\ell,n}(x)| = \begin{cases} |\varphi_k(x - (j-1)\kappa_{2^k})| & \text{if } x \in I_j \\ |\varphi_k(x - (\ell-1)\kappa_{2^k})| & \text{if } x \in I_\ell \\ 0 & \text{otherwise,} \end{cases}$$
and the $f_{j,n}$'s meet condition (C2) with $s_n := \frac{\bar{L}}{2k!}\left[\left(\frac{n}{\log n}\right)^{\frac{1}{2r+1}}\right]^{-r} \asymp \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}$.

(3) Let $a_j := \lceil n(j-1)\kappa_{2^k}\rceil$ and $b_j := \lfloor n \ j \ \kappa_{2^k}\rfloor$. By an argument similar to that in Case 1, for n sufficiently large, we have that for all $j \in [1:M_n]$,
$$K(P_j, P_0) \leq p_*\sum_{i=a_j}^{b_j} \left(\varphi_k\left(i/n - (j-1)\kappa_{2^k}\right)\right)^2$$
$$\leq p_*\sum_{i=a_j}^{b_j} \left(\frac{\bar{L}2^{k^k}}{k!}K_n^{-r}\right)^2 \leq p_*2n \ \kappa_{2^k}\left(\frac{\bar{L}2^{k^k}}{k!}K_n^{-r}\right)^2$$
$$\leq \frac{c_0}{2}\log(K_n) \leq c_0\log\left(\lfloor\frac{K_n}{2^k}\rfloor - 1\right) = c_0\log(M_n)$$

Hence $\frac{1}{M_n}\sum_{j=1}^{M_n} K(P_j, P_0) \leq c_0 \log M_n$ and condition (C3) holds for Case 2 as well.

We have demonstrated that conditions (C1)-(C3) are satisfied by $f_{j,n}$, $j \in [0:M_n]$. By virtue of the discussion following the statement of Theorem II.1, Theorem II.1 holds. $\square$

V. IMPLICATIONS AND EXTENSIONS

In this section, we state and establish several corollaries. Combining Theorem II.1 and [8, Theorem 4.1], we immediately obtain the following corollary which demonstrates the minimax optimal convergence rate under the sup-norm.

Corollary V.1. *Fix $k \in \mathbb{N}$, $r \in (k-1, k]$, and $L > 0$. Then the regression problem given by (4) satisfies*
$$\inf_{\widehat{f}} \sup_{f \in \mathcal{S}_{k,H}(r,L)} \mathbb{E}\big(\|\widehat{f} - f\|_\infty\big) \asymp \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}, \qquad (26)$$

where $\inf_{\widehat{f}}$ denotes the infimum over all estimators in $\mathcal{S}_k$.

We have studied minimax lower bounds for constrained Hölder clasees, where the ceiling of the Hölder exponent, $\lceil r \rceil$, is equal to k, the order of the k-monotone constraint. In the next corollary, we establish the same minimax lower bounds when $\lceil r \rceil$ is greater than the order of the derivative constraint. Let $\mathcal{S}_{p,H}(r,L) := \mathcal{S}_p \cap H_L^r$ for any $p \in [1:k-1]$. Consider the regression problem given by (4) with $\mathcal{S}_{k,H}(r,L)$ replaced by $\mathcal{S}_{p,H}(r,L)$. We obtain the following corollary.

Corollary V.2. *Fix $k \in \mathbb{N}$, $r \in (k-1,k]$, $L > 0$, and $p \in [1 : k-1]$. There exists a constant $c > 0$ such that*

$$\liminf_{n \to \infty} \inf_{\widehat{f}} \sup_{f \in \mathcal{S}_{p,H}(r,L)} \left(\frac{n}{\log n} \right)^{\frac{r}{2r+1}} \mathbb{E}\big(\|\widehat{f} - f\|_\infty \big) \geq c,$$

where $\inf_{\widehat{f}}$ denotes the infimum over all estimators in $\mathcal{S}_p$.

Proof. In view of the proof of Theorem II.1, it suffices to show that there exists a family of hypothesis functions $f_{j,n}$, $j \in [0 : M_n]$ satisfying (i) $f_{j,n} \in \mathcal{S}_{p,H}(r,L)$ for all j, (ii) (C2) of Section II, and (iii) (C3) of Section II. Because the $g_{j,n}$'s constructed in (16)-(17) and (19)-(20) satisfy $g_{j,n} \geq 0$, we have $f_{j,n}^{(p)}(x) = \mathcal{I}_{[0,x]}^{(k-p-1)}(g_{j,n}) \geq 0$, so each $f_{j,n}$ given by (21) belongs to $\mathcal{S}_p$. Moreover, by the proof of Theorem II.1, each $f_{j,n} \in H_L^r$ (and thus is also in $\mathcal{S}_{p,H}(r,L)$) and satisfies (C2) and (C3). $\square$

We have studied k-monotone estimation over suitable Hölder classes in the sup-norm. In this final corollary, we consider the same problem over a Sobolev class. For fixed $k \in \mathbb{N}$ and $L > 0$, define the Sobolev class of functions

$$\mathcal{W}(k,L) :=$$
$$\left\{ f : [0,1] \to \mathbb{R} \;\middle|\; f \text{ is } (k-1) \text{ times differentiable, with} \right.$$
$$\left. f^{(k-1)} \text{ absolutely continuous, and } \|f^{(k)}\|_{L_2} \leq L \right\}.$$

Let $\mathcal{S}_{p,\mathcal{W}}(k,L) := \mathcal{S}_p \cap \mathcal{W}(k,L)$ with $p \in [1:k]$. Consider the regression problem given by (4) with $\mathcal{S}_{k,H}(r,L)$ replaced by $\mathcal{S}_{p,\mathcal{W}}(k,L)$. We obtain the following corollary.

Corollary V.3. *Fix $k \in \mathbb{N}$, $L > 0$, and $p \in [1:k]$. Then there exists a positive constant c such that*

$$\liminf_{n \to \infty} \inf_{\widehat{f}} \sup_{f \in \mathcal{S}_{p,\mathcal{W}}(k,L)} \left(\frac{n}{\log n} \right)^{\frac{r}{2r+1}} \mathbb{E}\big(\|\widehat{f} - f\|_\infty \big) \geq c,$$

where $\inf_{\widehat{f}}$ denotes the infimum over all estimators in $\mathcal{S}_p$.

Proof. In view of the proof of the previous corollary, it is sufficient show that the $f_{j,n}$'s given by (21) belong to $\mathcal{S}_{p,\mathcal{W}}(k,L)$. Note that each $g_{j,n}$ is Lipschitz continuous, with Lipschitz constant L. Hence, each $f_{j,n}^{(k-1)} = g_{j,n}$ is absolutely continuous. In addition, by the Lipschitz continuity of $g_{j,n}$,

$$\int_0^1 \left(f_{j,n}^{(k)}(x) \right)^2 \, dx = \int_0^1 \left(g_{j,n}'(x) \right)^2 \, dx \leq \int_0^1 L^2 \, dx = L^2.$$

Thus, $\|f_{j,n}^{(k)}\|_{L_2} \leq L$, and each $f_{j,n} \in \mathcal{S}_{p,\mathcal{W}}(k,L)$. $\square$

VI. Conclusion

This paper establishes minimax lower bounds for k-monotone regression problems over a Hölder class in the sup-norm for general k. Combining these results with the minimax upper bounds for k-monotone regression problems established in [8], we obtain optimal minimax convergence rates under the sup-norm. An extension to Sobolev classes is also verified.

References

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

[2] P. Bellec and A. Tsybakov. Sharp oracle bounds for monotone and convex regression through aggregation. *J. Mach. Learn. Res.*, Vol. 16, pp. 1879–1892, 2015.

[3] L. Eeckhoudt and H. Schlesinger. Higher-order risk attitudes. In *Handbook of Insurance*, pp. 41–57. Springer, 2013.

[4] P. Groeneboom, F. Jongbloed, and J.A. Wellner. Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics*, Vol. 29, pp. 1653–1698, 2001.

[5] B.J. Kuipers, C. Chiu, D.T. Dalle Molle, and D.R. Throop. Higher-order derivative constraints in qualitative simulation. *Artificial Intelligence*, Vol. 51, pp. 343–379, 1991.

[6] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. on Inform, Theory*, Vol. 13, pp. 126–127, 1967.

[7] T.M. Lebair. *Constrained Estimation and Approximation using Control, Optimization, and Spline Theory*. Ph.D. Dissertation, University of Maryland, Baltimore County, 2016.

[8] T.M. Lebair and J. Shen. A two stage k-monotone B-spline regression estimator: Uniform Lipschitz property and optimal convergence rate. *Electronic Journal of Statistics*, Vol. 12(1), pp. 1388–1428, 2018.

[9] T.M. Lebair, J. Shen and X. Wang. Minimax lower bound and optimal estimation of convex functions in the sup-norm. *IEEE Trans. Automat. Control*, Vol. 62(7), pp. 3482—3487, 2017.

[10] E. Mammen. Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics*, Vol. 19, pp. 741–759, 1991.

[11] A. K. Sanyal, M. Chellappa, J. Ahmed, J. Shen, and D.S. Bernstein. Globally convergent adaptive tracking of spacecraft angular velocity with inertia identification and adaptive linearization. *Proc. of the 42nd IEEE Conf. on Dec. and Contr.*, pp. 2704–2709, Hawaii, 2003.

[12] J. Shen and T.M. Lebair. Shape restricted smoothing splines via constrained optimal control and nonsmooth Newton's methods. *Automatica*, Vol. 53, pp. 216–224, 2015.

[13] J. Shen and X. Wang. Estimation of monotone functions via P-splines: A constrained dynamical optimization approach. *SIAM J. on Control and Optim.*, Vol. 49(2), pp. 646–671, 2011.

[14] J. Shen and X. Wang. Convex regression via penalized splines: a complementarity approach. *Proc. of 2012 American Control Conference*, pp. 332–337, Montreal, Canada, 2012.

[15] A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2010.

[16] X. Wang and J. Shen. A class of grouped Brunk estimators and penalized spline estimators for monotone regression. *Biometrika*, Vol. 97(3), pp. 585–601, 2010.

[17] X. Wang and J. Shen. Uniform convergence and rate adaptive estimation of convex functions via constrained optimization. *SIAM J. on Control and Optim.*, Vol. 51(4), pp. 2753–2787, 2013.