Minimax Optimal Estimation of Convex Functions in the Supreme Norm

Teresa M. Lebair,* Jinglai Shen,† and Xiao Wang‡

May 7, 2014

Abstract

Estimation of convex functions finds broad applications in engineering and science, while convex shape constraint gives rise to numerous challenges in asymptotic performance analysis. This paper is devoted to minimax optimal estimation of univariate convex functions from the Hölder class in the framework of shape constrained nonparametric estimation. Particularly, the paper establishes the optimal rate of convergence in two steps for the minimax sup-norm risk of convex functions with the Hölder order between one and two. In the first step, by applying information theoretical results on probability measure distance, we establish the minimax lower bound under the supreme norm by constructing a novel family of piecewise quadratic convex functions in the Hölder class. In the second step, we develop a penalized convex spline estimator and establish the minimax upper bound under the supreme norm. Due to the convex shape constraint, the optimality conditions of penalized convex splines are characterized by nonsmooth complementarity conditions. By exploiting complementarity methods, a critical uniform Lipschitz property of optimal spline coefficients in the infinity norm is established. This property, along with asymptotic estimation techniques, leads to uniform bounds for bias and stochastic errors on the entire interval of interest. This further yields the optimal rate of convergence by choosing the suitable number of knots and penalty value. The present paper provides the first rigorous justification of the optimal minimax risk for convex estimation under the supreme norm.

Key Words: shape constrained estimation, convex regression, minimax estimation theory, sup-norm risk, penalized splines, asymptotic analysis, complementarity conditions.

1 Introduction

Nonparametric estimation of shape constrained functions (e.g., monotone/convex functions) receives increasing attention [7, 22, 25, 27, 33], driven by a wide range of applications in science and engineering. Examples include reliability engineering, biomedical research, finance, and astronomy. The goal of shape constrained estimation is to develop an estimator that preserves a pre-specified shape property of a true function, e.g., the monotone or convex property. A challenge in shape constrained estimation is that an estimator is subject to inequality shape constraints, e.g., the monotone or convex constraint. These constraints lead to nonsmooth conditions in estimator characterizations and complicate asymptotic performance analysis.

Considerable progress has been made toward developing and analyzing shape constrained estimators in the framework of nonparametric estimation. For example, estimators that preserve the...
monotone property have been extensively studied in the literature, e.g., [18, 22, 25, 26, 31, 33, 36]. In the realm of convex (or concave) estimation, earlier research is focused on the least squares approach: the least squares convex estimator is first studied in [12] and is shown to be consistent in the interior of the interval of interest [11]. The pointwise rate of convergence for the least squares convex estimator is developed in [17] and pointwise asymptotic distributions are characterized in [10]. A pointwise convex estimation approach is recently introduced in [2]: this approach is non-asymptotic and focuses on the performance of each individual function. To deal with unknown information of a function class, adaptive convex estimation has been proposed in [3, 6], which, however, can only handle a compact sub-interval without including the boundary. To overcome this problem, adaptive convex regression splines are proposed in [34]. Other results include [20, 27, 28].

Given a function class Σ, several critical questions arise when evaluating asymptotic performance of estimators over the function class Σ:

(1) What is the “best” rate of convergence of estimators uniformly on Σ?

(2) How can one construct an estimator that achieves the “best” rate of convergence on Σ?

(3) Is the “best” rate of convergence in (1) strict on Σ for any permissible estimator?

These questions form central research issues in minimax theory of nonparametric estimation [21, 30]. In particular, the first and second questions pertain to the minimax upper bound on Σ and its estimator construction, and the third question is closely related to the minimax lower bound on Σ. For unconstrained estimation, the above questions have been satisfactorily addressed for both the Sobolev and Hölder classes under the $L_2$ or supreme norm; see [21, 30] and references therein for details. This has led to well known optimal rates of convergence over unconstrained function classes (cf. (2) of Section 2). However, if shape constraints are imposed, then minimax asymptotic analysis becomes highly complicated and much fewer results have been reported; some exceptions include [16] for monotone estimation and [2, 5] for convex estimation. It is known that a shape constraint does not improve the unconstrained optimal rate of convergence [13], and it is believed that the same optimal rate holds on a constrained function class but no rigorous justification has been given for general constraints. Furthermore, it is worth pointing out that when the supreme norm (or simply sup-norm) is considered, shape constrained minimax analysis becomes even more challenging due to the following reasons:

(a) Constructive minimax upper bound. Most shape constrained estimators in the literature lack uniform convergence on the entire interval of interest, and they demonstrate poor performance on the boundary of the interval. For example, the shape constrained least squares estimator is inconsistent at the boundary points [35], and the convex estimators developed in [3, 6] deal with a compact sub-interval without including the boundary. Other pointwise estimators, such as [2, 16], are applicable only to a small interval of a fixed interior point, and they do not yield a convex or monotone estimate on the entire interval. This hinders construction of an estimator that achieves the optimal minimax risk under the sup-norm. On the other hand, in order to establish the uniform convergence requested by the sup-norm risk, one inevitably faces many nontrivial issues arising from an underlying nonsmooth optimization formulation of a shape constrained estimator, which call for new tools to handle them.

(b) Minimax lower bound. By minimax lower bound theory, which is based upon information theoretical results on distance between probability measures, it is known that establishing a minimax lower bound amounts to constructing a family of functions (or hypotheses) from a function class satisfying a suitable sup-norm separation order and a small total $L_2$-distance order [30 Section 2]. While it is conceived that there exist many such families, a shape constraint considerably limits choice of a feasible family under the sup-norm, especially when
a higher order shape constraint is imposed (recalling that roughly speaking, the convex constraint places a second order constraint on a function). Therefore, great care needs to be taken in order to meet both the order conditions and shape constraints.

The present paper is devoted to minimax optimal estimation of univariate convex functions from the Hölder class with Hölder order \( r \in (1,2) \) under the sup-norm. Specifically, we develop a two-step procedure to establish the optimal rate of convergence of the sup-norm risk. In the first step, we construct a family of convex functions from the Hölder class that yields the minimax lower bound in the sup-norm. In the second step, a penalized spline based convex estimator is developed and is shown to achieve the minimax upper bound in the sup-norm. It should be mentioned that even though the obtained optimal rate of convergence coincides with the optimal rate for the unconstrained Hölder class as expected, its proof is much more involved than that of the unconstrained case. In order to overcome shape constraint induced difficulties, several new techniques from asymptotic estimation and complementarity theory in constrained optimization are invoked. These new techniques and major contributions of the paper are summarized below.

1. Minimax lower bound. Toward this end, we construct a family of piecewise quadratic convex functions (whose derivatives are increasing and piecewise linear); see Section 3. These functions overlap on most of the interval \([0, 1]\), except on certain small sub-intervals. By careful selection of the slopes of the derivatives of these functions and the length of non-overlapping sub-intervals, we show that the constructed convex functions satisfy the desired order conditions, thus leading to the minimax lower bound. To the best of our knowledge, this construction is the first of its kind for minimax convex estimation. The proposed construction process also sheds light on minimax lower bounds for monotone or higher order derivative constraints.

2. Constructive minimax upper bound. We consider a convex penalized spline (or simply P-spline) estimator subject to the second order difference penalty. The convex shape constraint is converted to the second order difference constraint on spline coefficients. In spite of numerical advantages as well as conceptual simplicity and flexibility of P-splines \([15, 19]\), optimal spline coefficients of the convex P-spline estimator are characterized by nonsmooth complementarity conditions, thanks to the convex shape constraint. The present paper distinguishes the study of P-spline estimators from that in the literature by establishing a critical uniform Lipschitz property of optimal spline coefficients in the infinity norm \([24, 26]\), inspired by uniform convergence required by the sup-norm risk analysis. The proof of uniform Lipschitz property makes extensive use of combinatorial arguments in complementarity theory and the properties of an underlying constrained optimization problem, e.g., the second order difference penalty and piecewise linear formulation of optimal spline coefficients; see Section 4.2. By exploiting the uniform Lipschitz property and asymptotic estimation techniques, we develop uniform bounds for bias and stochastic errors on the entire interval \([0, 1]\) over the Hölder class. These results pave the way for the desired minimax upper bound under the sup-norm.

The paper is organized as follows. In Section 2, we present main results of the paper. Section 3 establishes the minimax lower bound for the rate of convergence under the sup-norm. In Section 4 we develop a convex P-spline estimator and show that this estimator attains the optimal rate of convergence. The paper ends with concluding remarks in Section 5.

2 Problem Formulation and Main Results

Consider the convex regression problem:

\[
y_i = f(x_i) + \sigma \epsilon_i, \quad i = 1, \ldots, n,
\]

where \( f : [0, 1] \to \mathbb{R} \) is an unknown convex function, the constant \( \sigma > 0 \), \( \epsilon_i \) are independent, standard normal errors, \( x_i = i/n, \ i = 1, \ldots, n \) are the equally spaced design points. Let \( f'(\cdot) \)
denote the derivative of $f$. Let
\[ C := \left\{ f : [0, 1] \to \mathbb{R} \mid (f'(x_1) - f'(x_2)) \cdot (x_1 - x_2) \geq 0, \quad \forall x_1, x_2 \in [0, 1] \right\} \]
be the collection of continuous convex functions which are differentiable (almost everywhere) on $[0, 1]$, and $H^r_L$ be the Hölder class with the Hölder exponent (or order) $r \in (1, 2]$ and the Hölder constant $L > 0$, namely,
\[ H^r_L := \left\{ f : [0, 1] \to \mathbb{R} \mid |f'(x_1) - f'(x_2)| \leq L|x_1 - x_2|^\gamma, \quad \forall x_1, x_2 \in [0, 1] \right\}, \]
where $\gamma := r - 1 \in (0, 1]$. Furthermore, let $C_H(r, L) := C \cap H^r_L$ be the collection of functions in both $C$ and $H^r_L$.

For estimation of unconstrained functions over the Hölder class $H^r_L$, it is well known in minimax theory of nonparametric estimation that for a fixed $r$, there exists an estimator (depending on $r$) which achieves the optimal rate of convergence over $H^r_L$ in the sup-norm \[30, 30\]. In fact, the minimax sup-norm risk on $H^r_L$ has an asymptotic order given by
\[ \inf \sup_{{f \in H^r_L}} \mathbb{E}(\|\hat{f} - f\|_\infty) \asymp L \frac{1}{r + 1} \sigma \frac{2^r}{n^{\frac{r}{r+1}}}, \]
where $\hat{f}$ denotes an estimator of a true function $f$, $\mathbb{E}(\cdot)$ is the expectation operator, and $a \asymp b$ means that $a/b$ is bounded by two positive constants from below and above for all $n$ sufficiently large. The goal of this paper is to establish the same asymptotic minimax rate on $C_H(r, L)$ with $r \in (1, 2]$. Specifically, the main result of this paper is presented in the following theorem.

**Theorem 2.1.** Let $r \in (1, 2]$ and $C_H(r, L)$ denote the Hölder class of convex functions defined above. Then there exists a positive constant $C_0$ such that
\[ \inf \sup_{{f \in C_H(r, L)}} \mathbb{E}(\|\hat{f} - f\|_\infty) \asymp C_0 \left(\frac{\log n}{n}\right)^{\frac{r}{r+1}}. \]

The proof of Theorem 2.1 is divided into two sections. Section 3 establishes a lower bound of the optimal rate of convergence via a construction procedure based on multiple hypotheses originating from information theory [30, Theorem 2.5]. Section 4 develops a penalized B-spline based convex estimator that achieves the optimal rate in (3); this gives rise to an upper bound of the optimal convergence rate, and, along with the lower bound, yields the desired optimal rate in (3).

### 3 Minimax Lower Bound of Convex Estimators

In this section, we establish the minimax lower bound of convex estimation of functions in $C_H(r, L)$ with $r \in (1, 2]$ in the sup-norm. The key idea of developing such a lower bound for nonparametric estimators relies on tools for distance of multiple probability measures or hypotheses from information theory [11, 3, 9, 23]; see [30, Section 2] for detailed discussions. It follows from minimax theory (e.g., [30, Theorem 2.5]) that establishing a minimax lower bound over the function class $C_H(r, L)$ in the sup-norm boils down to the construction of a family of functions (or hypotheses) $f_{j,n}, j = 0, 1, \ldots, M_n$ satisfying the following three conditions:

(C1) each $f_{j,n} \in C_H(r, L)$, $j = 0, 1, \ldots, M_n$;

(C2) whenever $j \neq k$, $\|f_{j,n} - f_{k,n}\|_\infty \geq 2s_n > 0$, where $s_n \asymp (\log n/n)^r/(2r+1)$;
Consider the function class \( f \). Suitable functions \( \xi \), all convex estimators on \([0, 1] \) for some constant \( c > 0 \) are constructed, then the following minimax lower bound over \( H \) holds:

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

where \( P_j \) denotes the distribution of \((Y_{j,1}, \ldots, Y_{j,n})\), where \( Y_{j,i} = f_{j,n}(X_i) + \xi_i, i = 1, \ldots, n \) with \( X_i = i/n \) and \( \xi_i \) being iid random variables, and \( K(P, Q) \) denotes the Kullback divergence between the two probability measures \( P \) and \( Q \) [13], i.e.,

\[
K(P, Q) := \begin{cases} 
\int \log \frac{dP}{dQ} dP, & \text{if } P \ll Q, \\
+\infty, & \text{otherwise}
\end{cases}
\]

In addition, we assume that there exists a constant \( p_* > 0 \) (independent of \( n \) and \( f_{j,n} \)) such that

\[
K(P_j, P_0) \leq p_* \sum_{i=1}^{n} \left( f_{j,n}(X_i) - f_{0,n}(X_i) \right)^2. 
\]

This assumption holds true if the iid random variables \( \xi_i \sim N(0, \sigma^2) \) (cf. \[30\] (2.36) or \[30\] Section 2.5, Assumption B). Hence, the convex regression problem defined in (1) satisfies this assumption.

In other words, once a family of functions \( \{f_{j,n}\} \) satisfying the above three conditions is constructed, then the following minimax lower bound over \( C_H(r, L) \) will hold:

\[
\liminf_{n \to \infty} \inf_{f_n, f \in C_H(r, L)} \sup_{f_n} \left( \frac{n}{\log n} \right)^{2r-1} \mathbb{E}(\|\hat{f}_n - f\|_\infty) \geq c
\]

for some constant \( c > 0 \) depending on \( r, L \), and \( p_* \) only, where \( \inf_{\hat{f}_n} \) denotes the infimum over all convex estimators on \([0, 1] \). In view of this, the goal of this section is to construct a family of suitable functions \( f_{j,n} \) satisfying (C1)-(C3).

### 3.1 Construction of the Desired Functions \( f_{j,n} \)

Consider the function class \( C_H(r, L) \) with \( r \in (1, 2] \) and \( L > 0 \), and fix \( c_0 \in (0, \frac{1}{8}) \). Given a sample size \( n \), let \( K_n \) be a positive number depending on \( n \), whose order of \( n \) will be specified below. We construct the desired functions \( f_{j,n} \) in two separate cases:

**Case 1:** \( \gamma := r - 1 \in (0, 1) \).

Let

\[
\tilde{L} := \min \left( \frac{L}{4} \sqrt{\frac{c_0 \gamma}{12p_*}}, \right)
\]

where \( p_* > 0 \) is defined above. We shall define the functions \( f_{j,n}, j = 0, 1, 2, \ldots, [K_n^\gamma] \) as follows. First we define the auxiliary functions \( \tilde{g}_{j,n} \) for \( j = 0, 1, 2, \ldots, [K_n^\gamma] \) as follows. For \( i = 0, 1, 2, \ldots \), let

\[
\tilde{g}_{0,n}(x) := \begin{cases}
2i\tilde{L}K_{n}^{-\gamma} + \tilde{L}K_{n}^{-1-\gamma}(x - \frac{i}{n}), & \text{if } x \in \left[ \frac{K_n^\gamma}{K_n^{\gamma+1}}, \frac{K_n^{\gamma+1}}{n} + \frac{3}{K_n} \right] \\
(2i+1)\tilde{L}K_{n}^{-\gamma}, & \text{if } x \in \left[ \frac{K_n^{\gamma+1}}{n} + \frac{3}{K_n}, \frac{K_n^{\gamma+1}}{n} + \frac{3}{K_n} \right] \\
(2i+1)\tilde{L}K_{n}^{-\gamma} + \tilde{L}K_{n}^{-1-\gamma}[x - \left( \frac{3}{K_n} + \frac{3}{K_n} \right)], & \text{if } x \in \left[ \frac{K_n^{\gamma+1}}{n} + \frac{3}{K_n}, \frac{K_n^{\gamma+1}}{n} + \frac{3}{K_n} \right] \\
(2i+1)\tilde{L}K_{n}^{-\gamma}, & \text{if } x \in \left[ \frac{K_n^{\gamma+1}}{n} + \frac{3}{K_n}, \frac{K_n^{\gamma+1}}{n} + \frac{3}{K_n} \right]
\end{cases}
\]

For each \( j = 1, 2, \ldots, [K_n^\gamma] \), let \( \tilde{g}_{j,n} = \tilde{g}_{0,n} \) everywhere except on \( \left( \frac{j-1}{K_n^{\gamma+1}}, \frac{j}{K_n^{\gamma+1}} \right] \) on which \( \tilde{g}_{j,n} \) is defined as follows:

\[
\tilde{g}_{j,n}(x) := \begin{cases}
2(j-1)\tilde{L}K_{n}^{-\gamma}, & \text{if } x \in \left[ \frac{j-1}{K_n^{\gamma+1}}, \frac{j}{K_n^{\gamma+1}} \right] \\
2(j-1)\tilde{L}K_{n}^{-\gamma} + \tilde{L}K_{n}^{-1-\gamma}[x - \left( \frac{j-1}{K_n^{\gamma+1}} + \frac{1}{K_n} \right)], & \text{if } x \in \left[ \frac{j-1}{K_n^{\gamma+1}} + \frac{1}{K_n}, \frac{j}{K_n^{\gamma+1}} + \frac{1}{K_n} \right] \\
2j\tilde{L}K_{n}^{-\gamma}, & \text{if } x \in \left[ \frac{j}{K_n^{\gamma+1}} + \frac{1}{K_n}, \frac{j}{K_n^{\gamma+1}} + \frac{1}{K_n} \right]
\end{cases}
\]
For each $j = 0, 1, 2, \ldots, \lfloor Kn \rfloor$, let $g_{j,n}$ denote the restriction of $\tilde{g}_{j,n}$ to $[0, 1]$.

**Case 2:** $\gamma = 1$. In this case, choose

$$\tilde{L} := \min \left( L, \sqrt{\frac{c_0}{12p_*}} \right),$$

and define for $i = 0, 1, 2, \ldots$

$$\tilde{g}_{0,n}(x) := \begin{cases} 
L \frac{2j}{K_n} + L(x - \frac{4i}{K_n}) & \text{if } x \in \left[\frac{4i}{K_n}, \frac{4i+1}{K_n}\right] \\
L \frac{2i+1}{K_n} + L(x - \frac{4i+1}{K_n}) & \text{if } x \in \left[\frac{4i+1}{K_n}, \frac{4i+3}{K_n}\right] \\
L \frac{2i+3}{K_n} + L(x - \frac{4i+3}{K_n}) & \text{if } x \in \left[\frac{4i+3}{K_n}, \frac{4i+4}{K_n}\right]
\end{cases}$$

Also define $\tilde{g}_{j,n} = \tilde{g}_{0,n}$ everywhere except on $\left[\frac{4(j-1)}{K_n}, \frac{4j}{K_n}\right)$, on which

$$\tilde{g}_{j,n}(x) := \begin{cases} 
\tilde{L} \frac{2(j-1)}{K_n} + \tilde{L}(x - \frac{4j}{K_n}) & \text{if } x \in \left[\frac{4(j-1)}{K_n}, \frac{4j-3}{K_n}\right] \\
\tilde{L} \frac{2j-3}{K_n} + \tilde{L}(x - \frac{4j-3}{K_n}) & \text{if } x \in \left[\frac{4j-3}{K_n}, \frac{4j}{K_n}\right]
\end{cases}$$

for $j = 1, 2, \ldots, \lfloor Kn \rfloor$. Again, for each $j$, we let $g_{j,n}$ denote the restriction of $\tilde{g}_{j,n}$ to $[0, 1]$.

The plots of the functions $g_{j,n}$, $j = 0, 1, 2, \ldots, \lfloor Kn \rfloor$ near the origin constructed above are displayed in Figures 1 and 2 for Case 1 and Case 2 respectively. Note that in these plots, $g_{0,n}$ often obstructs the view of other $g_{j,n}$’s, but if $j \geq 1$, then $g_{j,n}$ never obstructs the view of any other function.

Finally, in both the cases, for each $j = 0, 1, 2, \ldots, \lfloor Kn \rfloor$, define

$$f_{j,n}(x) := \int_0^x g_{j,n}(t) \, dt, \quad x \in [0, 1].$$

(5)

We present the following theorem for the above construction, whose proof is given in Section 3.2.
Figure 2: Plot of $g_{j,n}$’s near the origin when $\gamma = 1.$

**Theorem 3.1.** Consider the function class $C_H(r, L)$ with $r \in (1, 2], L > 0,$ and $\gamma := r - 1$. Let $K_n = \left(\frac{n}{\log n}\right)^{\frac{1}{r+1}}$ and $M_n := \lfloor K_n^\gamma \rfloor$. Then the functions $f_{j,n}$, $j = 0, 1, \ldots, M_n$ constructed in (5) satisfy conditions (C1)–(C3). Specifically, for all $n$ sufficiently large,

1. Each $f_{j,n} \in C_H(r, L);$ 
2. For all $j, k \in \{0, 1, \ldots, M_n\}$ with $j \neq k$, $\|f_{j,n} - f_{k,n}\|_\infty = 2s_n$, where $s_n \simeq \left(\frac{\log n}{n}\right)^{\frac{r}{r+1}}$; 
3. $\frac{1}{M_n} \sum_{j=1}^{M_n} K(P_j, P_0) \leq c_0 \log(M_n).$

This theorem, together with the similar argument in [30, Theorem 2.5], leads to the lower bound of minimax risk of convex estimation in [4].

**Remark 3.1.** The proposed construction procedure for convex functions on the Hölder class can be extended to other shape constraints, e.g., monotone constraint or shape constraints in terms of higher order derivatives.

### 3.2 Proof of Theorem 3.1

**Proof.** We consider the two cases in the proof: $\gamma \in (0, 1)$, and $\gamma = 1$.

**Case 1:** $\gamma \in (0, 1)$. For all $n$ (and $K_n$) sufficiently large, the following properties of $g_{j,n}$’s can be easily verified with the help of Figure 1 for any $x, y \in [0, 1]$, suppose that
(i) $0 < |x - y| \leq \frac{1}{K_n}$. Then
\[
\max_j \frac{|g_{j,n}(x) - g_{j,n}(y)|}{|x - y|} \leq \frac{|g_{1,n}(x) - g_{1,n}(y)|}{|x - y|} \bigg|_{x=K_n^{-1}, y=2K_n^{-1}} \leq LK_n^{1-\gamma}.
\]

(ii) $\frac{1}{K_n} < |x - y| \leq \frac{1}{K_n^2}$. Then
\[
\max_j |g_{j,n}(x) - g_{j,n}(y)| \leq |g_{0,n}(x) - g_{0,n}(y)| \big|_{x=0, y=4K_n^{-1}} \leq 2LK_n^{-\gamma}.
\]

(iii) $\frac{1}{K_n^2} < |x - y| \leq 1$. Without loss of generality, let $x < y$ with $y = rK_n^{-\gamma} + s(x, y)$ for some $r \in \mathbb{N}$ and $0 \leq s(x, y) < K_n^{-\gamma}$. It can be shown that
\[
\max_j \frac{|g_{j,n}(x) - g_{j,n}(y)|}{|x - y|} \leq \frac{|g_{1,n}(x) - g_{1,n}(y)|}{|x - y|} \bigg|_{x=K_n^{-1}, y=rK_n^{-\gamma} + 3K_n^{-1}} \leq \frac{2(r + 1)LK_n^{-\gamma}}{rK_n^{-\gamma} + 2K_n^{-1}} \leq \frac{2(r + 1)LK_n^{-\gamma}}{rK_n^{-\gamma}} \leq 4L \leq L.
\]

Along with these properties, we show the three conditions as follows:

(1) Obviously, each $g_{j,n}$ is nondecreasing, and hence each $f_{j,n}$ is convex. Furthermore, to show that each function $f_{j,n}$ is in the Hölder class $H^r_L$, we consider the following three cases:

(1.1) $0 < |x - y| \leq \frac{1}{K_n}$. Then, by (i), we have
\[
\frac{|f'_{j,n}(x) - f'_{j,n}(y)|}{|x - y|^\gamma} = \frac{|f'_{j,n}(x) - f'_{j,n}(y)|}{|x - y|} |x - y|^{1-\gamma} \leq LK_n^{1-\gamma} \left(\frac{4}{K_n}\right)^{1-\gamma} \leq 4L \leq L.
\]

(1.2) $\frac{1}{K_n} < |x - y| \leq \frac{1}{K_n^2}$. Then, by (ii), we have
\[
\frac{|f'_{j,n}(x) - f'_{j,n}(y)|}{|x - y|^\gamma} \leq \frac{2LK_n^{-\gamma}}{|x - y|^\gamma} \leq 2LK_n^{1-\gamma} \left(\frac{4}{K_n}\right)^{1-\gamma} \leq 2L \leq L.
\]

(1.3) $\frac{1}{K_n^2} < |x - y| \leq 1$. By (iii), we obtain
\[
\frac{|f'_{j,n}(x) - f'_{j,n}(y)|}{|x - y|^\gamma} \leq \frac{|f'_{j,n}(x) - f'_{j,n}(y)|}{|x - y|} \leq L.
\]

This shows that condition (1) holds.

(2) Let $j, k \in \{0, 1, \ldots, M_n\}$ with $j < k$ without loss of generality. It follows from the definitions of $g_{j,n}$ and $f_{j,n}$ and Figure II that

(2.1) if $j = 0$, then $f'_{j,n}(x) = f'_{k,n}(x)$ for all $x \in [0, 1]$ except on the set $S_{0k} := ((k - 1)K_n^{-\gamma}, (k - 1)K_n^{-\gamma} + 2/K_n) \cup ((k - 1)K_n^{-\gamma} + 2/K_n, (k - 1)K_n^{-\gamma} + 4/K_n)$;

(2.2) if $j \geq 1$, then $f'_{j,n}(x) = f'_{k,n}(x)$ for all $x \in [0, 1]$ except on the set $S_{jk} := ((j - 1)K_n^{-\gamma}, (j - 1)K_n^{-\gamma} + 2/K_n) \cup ((j - 1)K_n^{-\gamma} + 2/K_n, (j - 1)K_n^{-\gamma} + 4/K_n) \cup ((j - 1)K_n^{-\gamma}, (j - 1)K_n^{-\gamma} + 4/K_n) \cup ((k - 1)K_n^{-\gamma} + 2/K_n, (k - 1)K_n^{-\gamma} + 4/K_n)$.

Hence, the set of critical points of $f_{j,n} - f_{k,n}$ is $[0, 1] \setminus S_{jk}$. Furthermore, in view of piecewise linearity of $g_{j,n}$'s, it is easy to see that
(a) \( f_{j,n}(x) - f'_{k,n}(x) = g_{j,n}(x) - g_{k,n}(x) > 0 \) for all \( x \in ((k-1)K^{-\gamma} + 1/K_n, (k-1)K^{-\gamma} + 2/K_n) \), and \( f_{j,n}(x) - f'_{k,n}(x) < 0 \) for all \( x \in ((k-1)K^{-\gamma} + 2/K_n, (k-1)K^{-\gamma} + 3/K_n) \);

(b) for case (2.2), \( f_{j,n}(x) - f'_{k,n}(x) < 0 \) for all \( x \in ((j-1)K^{-\gamma} + 1/K_n, (j-1)K^{-\gamma} + 2/K_n) \), and \( f_{j,n}(x) - f'_{k,n}(x) > 0 \) for all \( x \in ((j-1)K^{-\gamma} + 2/K_n, (j-1)K^{-\gamma} + 3/K_n) \);

(c) \( g_{j,n}(x) = g_{k,n}(x) = 0 \) for all \( x \in [0,1] \setminus S_{jk} \) except \( x = (k-1)K^{-\gamma} + 2/K_n \) and \( x = (j-1)K^{-\gamma} + 2/K_n \) (if \( j \geq 1 \)).

Moreover, \( f_{j,n}(x) = f_{k,n}(x) \) for \( x = 0, 1 \). This shows that \( |f_{j,n}(x) - f_{k,n}(x)| \) achieves a local maximum at \( x^* = (k-1)K^{-\gamma} + 2/K_n \) and/or \( z^* := (j-1)K^{-\gamma} + 2/K_n \) (the latter holds only if \( j \geq 1 \)). Due to the symmetry of non-overlapping regions of \( g_{j,n} \) and \( g_{k,n} \), we have \( \|f_{j,n} - f_{k,n}\|_\infty = |f_{j,n}(x^*) - f_{k,n}(x^*)| = |f_{j,n}(z^*) - f_{k,n}(z^*)| \). Furthermore, it can be verified that \( f_{j,n}(x) = f_{k,n}(x) \) at \( x = (k-1)K^{-\gamma} \). Therefore,

\[
\|f_{j,n} - f_{k,n}\|_\infty = \left| f_{j,n}\left((k-1)K^{-\gamma} + \frac{2}{K_n}\right) - f_{k,n}\left((k-1)K^{-\gamma} + \frac{2}{K_n}\right) \right|
\]

\[
= \left| (f_{j,n} - f_{k,n})\left((k-1)K^{-\gamma} + \frac{2}{K_n}\right) - (f_{j,n} - f_{k,n})\left((k-1)K^{-\gamma}\right) \right|
\]

\[
= \int_{(k-1)K^{-\gamma}}^{(k-1)K^{-\gamma} + \frac{2}{K_n}} \left( g_{j,n}(t) - g_{k,n}(t) \right) dt
\]

\[
= 2 \int_{(k-1)K^{-\gamma}}^{(k-1)K^{-\gamma} + \frac{1}{K_n}} LK_n^{1-\gamma}\left(t - (k-1)K^{-\gamma}\right) dt
\]

\[
= 2 \int_{0}^{\frac{1}{K_n}} LK_n^{1-\gamma}t dt = \bar{L}K_n^{-(1+\gamma)}
\]

\[
= \bar{L}K_n^{-\tau} = 2s_n,
\]

where \( s_n := \bar{L}K_n^{-\tau}/2 = \bar{L}/2 \cdot \left( \frac{\log n}{n} \right)^{\frac{\tau}{\gamma + 1}} \), and thus condition (2) holds.

(3) To show this condition, we first collect a few results about \( f_{j,n} \)'s to be used later:

(3.1) For each \( j = 1, 2, \ldots, M_n \),

\[
\int_0^1 \left( f_{j,n}(x) - f_{0,n}(x) \right)^2 dx = \int_0^1 \left( f_{1,n}(x) - f_{0,n}(x) \right)^2 dx.
\]

(3.2) Let \( h_j := (f_{j,n} - f_{0,n})^2 \) for \( j = 1, \ldots, M_n \). Since \( X_i = i/n \), it follows from analysis of numerical integration and condition (2) that for each \( j \),

\[
\left| \int_0^1 h_j(x) dx - \frac{\sum_{i=1}^n b_j(X_i)}{n} \right| \leq \max_{x \in [0,1]} |h'_j(x)| \cdot \frac{\|g_{j,n} - g_{0,n}\|_\infty}{n} \cdot \|f_{j,n} - f_{0,n}\|_\infty \leq \frac{2\bar{L}^2}{nK_n^{1+2\gamma}}.
\]

(3.3) The following holds:

\[
\int_0^1 \left( f_{1,n}(x) - f_{0,n}(x) \right)^2 dx = 2 \int_{0}^{\frac{1}{K_n}} \left( f_{1,n}(x) - f_{0,n}(x) \right)^2 dx = 2 \int_{0}^{\frac{1}{K_n}} \left( \frac{\bar{L}K_n^{1-\gamma}x^2}{2} \right)^2 dx
\]

\[
+ 2 \int_{0}^{\frac{2}{K_n}} \left( \frac{\bar{L}K_n^{1-\gamma}(x - 1/K_n)}{2} \right)^2 dx = 2\bar{L}^2K_n^{-2\gamma-3} \left( \frac{1}{20} \right) = \frac{2\bar{L}^2}{K_n^{2\gamma+1}} \cdot \frac{46}{60}.
\]
In light of the above results, we have for each $j = 1, \ldots, M_n$,

$$K(P_j, P_0) \leq p_* \sum_{i=1}^{n} (f_{j,n}(X_i) - f_{0,n}(X_i))^2 \leq p_* \left( n \int_0^1 (f_{j,n}(x) - f_{0,n}(x))^2 \, dx + \frac{2L^2}{K_{n+2}} \right)$$

$$\leq p_* \left( n \int_0^1 (f_{1,n}(x) - f_{0,n}(x))^2 \, dx + \frac{2L^2}{K_{n+2}} \right)$$

$$\leq p_* \left( \frac{2\gamma}{K_{n+2}} + \frac{60 + 2\bar{L}}{K_{n+2}} \right)$$

$$\leq \frac{c_0 \gamma}{6} \log(n)$$

for all $n$ sufficiently large, where the last inequality follows from the definition of $\bar{L}$ and the order of $K_n$. Consequently,

$$\frac{1}{M_n} \sum_{j=1}^{M_n} K(P_j, P_0) \leq \frac{c_0 \gamma}{6} \log(n).$$

Finally, since $\gamma \in (0, 1)$, we have, for all $n$ sufficiently large,

$$\log M_n \geq 0.9 \gamma \log K_n = 0.9 \gamma \log \left( \frac{n}{\log n} \right)^{\frac{1}{2(\gamma+1)}} = \left( \frac{0.9 \gamma}{2 \gamma + 3} \right) \log \left( \frac{n}{\log n} \right) \geq \frac{\gamma}{6} \log n.$$

This establishes condition (3) and hence completes the proof for Case 1.

**Case 2:** $\gamma = 1$. We show the three conditions in a similar manner as in Case 1:

1. Clearly, each $f_{j,n}$ is convex on $[0,1]$. Further, it is easy to show via the definition of $g_{j,n}$ and Figure [2] 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 $j = 0, 1, \ldots, [K_n]$. This thus implies that each $f_{j,n} \in C_H(r, L)$, leading to condition (1).

2. Let $0 \leq j < k \leq M_n := [K_n]$. It follows from a similar argument as in (2) of Case 1 that $\|f_{j,n} - f_{k,n}\|_\infty$ is achieved at $x^* = \frac{4(k-1)+2}{K_n}$ and $f_{j,n}(x) = f_{k,n}(x)$ at $x = \frac{4(k-1)}{K_n}$. Therefore, we have

$$\|f_{j,n} - f_{k,n}\|_\infty = \left| f_{j,n} \left( \frac{4(k-1)+2}{K_n} \right) - f_{k,n} \left( \frac{4(k-1)+2}{K_n} \right) \right|$$

$$= \left| (f_{j,n} - f_{k,n}) \left( \frac{4(k-1)+2}{K_n} \right) - (f_{j,n} - f_{k,n}) \left( \frac{4(k-1)}{K_n} \right) \right|$$

$$= \int_{\frac{4(k-1)+2}{K_n}}^{\frac{4(k-1)+1}{K_n}} (g_{j,n}(t) - g_{k,n}(t)) \, dt$$

$$= 2 \int_{\frac{4(k-1)+1}{K_n}}^{\frac{4(k-1)}{K_n}} \bar{L} \left( t - \frac{4(k-1)}{K_n} \right) \, dt$$

$$= 2 \int_{0}^{\frac{1}{K_n}} \bar{L} t \, dt = \bar{L} K_n^{-2} = 2s_n,$$

where $s_n := \bar{L} K_n^{-2}/2 = \bar{L}/2 \cdot \left( \frac{\log n}{n} \right)^{\frac{3}{2}}$ (for $r = 2$), and thus condition (2) holds for $\gamma = 1$. 

10
(3) First of all, it is easy to see that the conditions in (3.1) and (3.2) in Case 1 remain valid for \( \gamma = 1 \). To show the condition in (3.3) for \( \gamma = 1 \), we have

\[
\int_0^1 (f_{1,n}(x) - f_{0,n}(x))^2 \, dx = 2 \int_0^{\pi/n} (f_{1,n}(x) - f_{0,n}(x))^2 \, dx = 2 \int_0^{\pi/n} \left( \frac{Lx^2}{2} \right)^2 \, dx \\
+ 2 \int_0^{\pi/n} \frac{\bar{L}K_n^{-2}}{2} + \bar{L}K_n^{-1} \left( x - \frac{1}{K_n} \right) - \frac{\bar{L}}{2} \left( x - \frac{1}{K_n} \right)^2 \, dx \\
= 2\bar{L}^2 K_n^{-5} \left( \frac{1}{20} + \frac{43}{60} \right).
\]

By using these results and a similar argument as in (3) of Case 1, we have for all \( n \) sufficiently large, \( K(P_j, P_0) \leq \frac{c_0}{6} \log(n) \) for each \( j = 1, \ldots, M_n \) such that

\[
\frac{1}{M_n} \sum_{j=1}^{M_n} K(P_j, P_0) \leq \frac{c_0}{6} \log(n).
\]

Again, in view of

\[
\log M_n \geq 0.9 \log K_n = 0.9 \log \left( \left( \frac{n}{\log n} \right)^{1/2} \right) = \left( \frac{0.9}{5} \right) \log \left( \frac{n}{\log n} \right) \geq \frac{1}{6} \log n
\]

for all \( n \) sufficiently large, we obtain condition (3). This completes the proof for Case 2.

\[\square\]

4 Convex P-spline Estimator and Minimax Upper Bound of Convex Estimators

In this section, we develop a convex P-spline estimator subject to the second order difference penalty that achieves the optimal rate of convergence. We first formulate the convex P-spline estimator as a constrained quadratic optimization problem, and establish optimality conditions for spline coefficients (cf. Section 4.1). We then develop a critical uniform Lipschitz property in the infinity norm for the optimal spline coefficients (cf. Theorem 4.4). Equipped with these results, we establish uniform convergence of the convex P-spline estimator, and this leads to the minimax upper bound under the sup-norm in Section 4.3.

4.1 Convex P-spline Estimator and Optimal Spline Coefficients

Consider a P-spline estimator for the convex estimation problem defined in (1). Specifically, let \( \{ B_k[p] : k = 1, \ldots, K_n + p \} \) be the \( p \)th degree B-spline basis with knots \( 0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{K_n} = 1 \) and extension to \( \kappa_{-p} < \kappa_{-p+1} < \cdots < \kappa_0 \) and \( \kappa K_n < \kappa_{K_n+1} < \cdots < \kappa_{K_n+p} \) on the boundary. For simplicity, consider equally spaced knots, i.e., \( \kappa_k = k/K_n \) with \( k = -p, \ldots, K_n + p \), where the support of each basis function is \( [(k-p-1)/K_n, k/K_n] \cap [0,1], k = 1, \ldots, K_n + p \). The value of \( K_n \) will depend upon \( n \) as shown below, and we also assume \( n/K_n \) to be an integer denoted by \( M_n \). In what follows, we consider the penalized linear B-spline based convex estimator, namely, \( p = 1 \).

Let \( \Delta \) be the backward difference operator, i.e., \( \Delta(b_k) := b_k - b_{k-1} \) and \( \Delta^m(b_k) = \Delta(\Delta^{m-1}(b_k)) \) with \( m \in \mathbb{N} \), and consider the polyhedral cone

\[
\Omega = \{ b \in \mathbb{R}^{K_n+1} | \Delta^2(\Delta(b_k)) = b_k - 2b_{k+1} + b_{k+2} \geq 0, k = 1, \ldots, K_n - 1 \}.
\]
The constrained optimization problem for \( P \)-spline coefficients with the penalty on the \( m \)th order difference is

\[
\hat{b}^{[m]} = \arg \min_{b \in \Omega} \sum_{k=1}^{K_n+1} \left( y_i - \sum_{k=1}^{K_n+1} b_k B_k^{[1]}(x_i) \right)^2 + \lambda^* \sum_{k=m+1}^{K_n+1} (\Delta^m b_k)^2,
\]

where \( \hat{b}^{[m]} \in \mathbb{R}^{K_n+1} \), and \( \lambda^* > 0 \) is the penalty parameter dependent on \( n \) whose order will be determined later. Then the convex \( P \)-spline estimator with \( p = 1 \) is given by:

\[
\hat{f}^{[m]}(x) = \sum_{k=1}^{K_n+1} \hat{b}_k^{[m]} B_k^{[1]}(x).
\]

Since the knots are equally spaced, it is easy to see that if the B-spline coefficient vector \( \hat{b}^{[m]} \) is in \( \Omega \), then \( \hat{f}^{[m]} \) is convex. We consider the second order difference penalty in this paper, i.e., \( m = 2 \).

In order to establish the optimality conditions for the optimal spline coefficient \( \hat{b}^{[m]} \), we introduce more notation. Denote the \( n \times (K_n + 1) \) design matrix by \( X = [B_k^{[1]}(x)]_{j,k} \) and \( \beta_n := \sum_{i=1}^{n} (B_k^{[1]}(x_i))^2 \) for \( k = 2, \ldots, K_n \). Since the knots are equally spaced, \( \beta_n \) is independent of \( k = 2, \ldots, K_n \). For the linear B-spline, \( (\beta_n \Delta^m) \) converges to a positive constant as \( (n/K_n) \to \infty \). Hence, there exists a positive constant \( C_\beta \) such that

\[
\beta_n \geq C_\beta \frac{n}{K_n}, \quad \forall \ n, K_n.
\]

Moreover, define \( \Gamma := X^T X/\beta_n \in \mathbb{R}^{(K_n+1) \times (K_n+1)} \). It is easy to show that \( \Gamma \) is positive definite and tridiagonal. Let \( y = (y_1, \ldots, y_n)^T \) and define the weighted response vector \( \bar{y} := X^T y/\beta_n \in \mathbb{R}^{K_n+1} \). Furthermore, let \( D_2 \in \mathbb{R}^{(K_n-1) \times (K_n+1)} \) be the second-order difference matrix, i.e.,

\[
D_2 = \begin{bmatrix}
1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\
0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 & 0 \\
& & & \ddots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \\
\end{bmatrix}.
\]

Then the convex constraint on spline coefficients is defined by the polyhedral cone \( \Omega := \{ b \in \mathbb{R}^{K_n+1} : D_2 b \geq 0 \} \). Define \( \Lambda := \Gamma + \lambda D_2^T D_2 \), where \( \lambda := \lambda^*/\beta_n > 0 \). Therefore, the underlying optimization problem (6) with \( p = 1 \) and \( m = 2 \) becomes the following quadratic program

\[
\hat{b} = \arg \min_{b \in \Omega} \frac{1}{2} b^T \Lambda b - \bar{y}^T b,
\]

where we drop the superscript in \( \hat{b}^{[m]} \) for notational convenience. In what follows, we treat the optimal spline coefficient vector \( \hat{b} : \mathbb{R}^{K_n+1} \to \mathbb{R}^{K_n+1} \) as a function of \( \bar{y} \). It is known that the function \( \hat{b} \) is piecewise linear and Lipschitz continuous [8]. To obtain a piecewise linear formulation of \( \hat{b} \), consider the optimality conditions in the KKT form:

\[
\Lambda \hat{b} - \bar{y} - D_2^T \chi = 0, \quad 0 \leq \chi \perp D_2 \hat{b} \geq 0,
\]

where \( \chi \in \mathbb{R}^{K_n-1} \) is the Lagrange multiplier, and \( u \perp v \) means that two vectors \( u, v \) are orthogonal, i.e., \( u^T v = 0 \). Here the latter condition in (9) is known as the complementarity condition [4] in constrained optimization. Therefore, linear selection functions (or linear pieces) of \( \hat{b} \) can be
determined by index sets \( \alpha = \{ i \mid (D_2 \hat{b})_i = 0 \} \subseteq \{ 1, \ldots, K_n - 1 \} \), where \( \alpha \) may be empty. For a given index set \( \alpha \), the optimality conditions in (9) yield the equations

\[(D_2)_{\alpha \bullet} \hat{b} = 0, \quad \chi \pi = 0, \quad \Lambda \hat{b} - \bar{y} - ((D_2)_{\alpha \bullet})^T \chi \alpha = 0, \quad (10)\]

where \((D_2)_{\alpha \bullet}\) denotes the rows of \(D_2\) indexed by the index set \(\alpha\), and \(\bar{\alpha}\) is the complement of \(\alpha\), namely, \(\bar{\alpha} := \{ 1, \ldots, K_n - 1 \} \setminus \alpha\). Let \(\hat{b}^\alpha\) denote a linear selection function obtained from (10) corresponding to the index set \(\alpha\), and let \((F_\alpha)^T\) be a matrix formed by the basis of the null space of \((D_2)_{\alpha \bullet}\). (When \(\alpha\) is the empty set, \((F_\alpha)^T\) will be the identity matrix.) In view of the first equation in (10), \(\hat{b}^\alpha\) takes the form \((F_\alpha)^T \tilde{b}^\alpha\), where the vector \(\tilde{b}^\alpha\) consists of the free variables of the equation \((D_2)_{\alpha \bullet} \hat{b} = 0\). It is easy to show via (10) and a standard argument that \((F_\alpha)^T \tilde{b}^\alpha = F_\alpha \bar{y}\). Since \(F_\alpha\) is row linearly independent and \(\Lambda\) is positive definite, we obtain the linear selection function corresponding to \(\alpha\) as

\[
\tilde{b}^\alpha(\bar{y}) = F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha \bar{y}.
\]

Hence, for a fixed \(K_n\), \(\hat{b}\) has \(2^{K_n-1}\) linear selection functions. Note that there are multiple ways to construct (equivalent) piecewise linear formulation of \(\hat{b}\), but the above approach offers flexibility to select suitable \(F_\alpha\) for analytic properties of \(\hat{b}\) in the next section.

4.2 Uniform Lipschitz Property of Optimal Spline Coefficients

The constrained optimization problem (8) does not admit a closed form solution for the optimal spline coefficients, and this poses great difficulties in asymptotical statistical analysis. In this section, we characterize a critical property for the optimal spline solution \(\hat{b}\) (with \(m = 2\) pertaining to the uniform Lipschitz constant, regardless of \(K_n\), \(\alpha\), and \(\lambda\), when the \(\ell_\infty\)-norm is used (cf. Theorem 4.1). Toward this end, we discuss more about the piecewise linear formulation of \(\hat{b}\).

For the given design matrix \(X\), define \(\alpha_n := \sum_{i=1}^{n} B_k^1(x_i) \) for \(k = 1, K_n + 1\), and \(\gamma_n := \sum_{i=1}^{n} B_k^1(x_i)B_{k+1}^1(x_i)\) for \(k = 2, \ldots, K_n\), where \(\gamma_n\) is independent of \(k = 2, \ldots, K_n\) due to the equally spaced knots. Moreover, let \(\theta_n := \alpha_n/\beta_n\), and \(\eta_n := \gamma_n/\beta_n\), where \(\beta_n\) is defined before (7). It is easily verified via the B-spline properties that \(\theta_n \to 1/2\) and \(\eta_n \to 1/4\) as \(n/K_n \to \infty\). Note that the tridiagonal matrix \(\Gamma\) is positive definite and is given by

\[
\Gamma = X^T X / \beta_n = \begin{bmatrix}
\theta_n & \eta_n & 0 & 0 & \cdots & 0 \\
\eta_n & 1 & \eta_n & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\eta_n & 1 & \eta_n & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & \eta_n & \theta_n
\end{bmatrix} \in \mathbb{R}^{(K_n+1) \times (K_n+1)}.
\]

such that the positive definite matrix \(\Lambda = \Gamma + \lambda D_2^T D_2 \in \mathbb{R}^{(K_n+1) \times (K_n+1)}\).

It follows from complementarity theory [4, 8, 26] that the optimal solution \(\hat{b}\) is a piecewise linear and Lipchitz continuous function of \(\bar{y}\) determined by an index set \(\alpha = \{ i \mid (D_2 \hat{b})_i = 0 \} \subseteq \{ 1, \ldots, K_n - 1 \}\) (\(\alpha\) may be empty). Specifically, for given \(\bar{y}\) and \(\alpha\), we define a vector \(\tilde{b}^\alpha\) and an associated family of index sets \(\{ \beta^\alpha_i \}\) as follows. Let the index set \(\vartheta := \{ j - 1 \mid (\Delta^2 \tilde{b})_j = 0, j \in \{ 3, 4, \ldots, K_n + 1 \} \}\). Since \(\tilde{b}_i = (\tilde{b}_{i-1} + \tilde{b}_{i+1})/2\) for each \(i \in \vartheta\), it is easy to see that \(\tilde{b}_i, i \in \vartheta\) are the basic variables of the equation \((D_2)_{\alpha \vartheta} \tilde{b} = 0\), while \(\tilde{b}_i, i \notin \vartheta\) are the free variables. Let \(1 = i_1 < i_2 < \cdots < i_{\ell} = K_n + 1\) be the elements of \(\vartheta\), i.e., the complement of \(\vartheta\), where \(\ell := K_n + 1 - |\alpha|\). Let \(\tilde{b}^\alpha := (\tilde{b}_{i_1}, \ldots, \tilde{b}_{i_\ell})^T\) be the vector of free variables.

A specific basis for the null space of \((D_2)_{\alpha \vartheta}\) is constructed via the linear B-splines \(g_1, g_2, \ldots, g_\ell\) on the interval \([1, K_n + 1]\) with nodes \(1 = i_1 < i_2 < \cdots < i_\ell = K_n + 1\) recently introduced in [34].
Proposition 4.1. If $\beta$ is in the null space of $(D_2)_{\alpha\bullet}$, then the sets $\beta_k := \{1, 2, \ldots, \ell_1 - 1\}$. Next let $\ell_2$ be the smallest element $i_2 > \ell_1$ in $\overline{\vartheta}$ such that $i_2 - i_1 = 1$, and $\beta_2^\alpha := \{\ell_1 + 1, \ell_1 + 2, \ldots, \ell_2\}$. Continuing this process, we obtain $\beta_k^\alpha$'s until we reach $i_k = \ell_1 + \ell_2 + 1, \ldots, L$. If $i_k = K_n$, we also choose the last $\beta_K^\alpha$ as $\{K_n + 1\}$. Suppose that the above process leads to the sets $\beta_1^\alpha, \ldots, \beta_L^\alpha$. It is clear that $\{\beta_i^\alpha\}_{i=1}^L$ forms a disjoint partition of $\{1, \ldots, K_n + 1\}$, namely, $\bigcup_{i=1}^L \beta_i^\alpha = \{1, \ldots, K_n + 1\}$ and $\beta_i^\alpha \cap \beta_j^\alpha = \emptyset$ whenever $j \neq k$. Let $m_i^\alpha := |\beta_i^\alpha|$, where $i = 1, \ldots, L$. Note that if $m_i^\alpha > 1$, then $m_i^\alpha \geq 3$. Recall $\ell := K_n + 1 - |\alpha|$. It follows from the definition of $\beta_i^\alpha$ and the construction of $F_\alpha$ that

$$F_\alpha = \begin{bmatrix} F_{\alpha,1} & F_{\alpha,2} & \cdots & F_{\alpha,L} \end{bmatrix} \in \mathbb{R}^{\ell \times (K_n + 1)},$$

where each matrix block corresponding to $\beta_k^\alpha$ is shown as follows: if $m_k^\alpha = 1$, then $F_{\alpha,k} = 1$; otherwise, assume that the nodes in $\overline{\vartheta}$ on $[\ell_{k-1}, \ell_k]$ are $i_s = \ell_{k-1} < i_{s'+1} < \cdots < i_{s'+w_k} = \ell_k$ for some $w_k$. Let $h_{k,j}^\alpha := i_{s'+j} - i_{s'+j-1} \geq 2$ for each $j = 1, \ldots, w_k$. Then $F_{\alpha,k} \in \mathbb{R}^{(w_k + 1) \times m_k^\alpha}$ constructed from the linear splines (11)-(12) corresponding to $\beta_k^\alpha$ is given by

$$F_{\alpha,k} = \begin{bmatrix} 1 & h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & h_{k,w_k}^\alpha & 0 & 0 & \cdots & 0 \\ 0 & 1 & h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & h_{k,w_k}^\alpha & 0 & \cdots & 0 \\ 0 & 0 & 1 & h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & h_{k,w_k}^\alpha & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & h_{k,w_k}^\alpha & 0 & 0 & \cdots & 0 \\ h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & h_{k,w_k}^\alpha & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & h_{k,w_k}^\alpha & 0 & 0 & \cdots & 0 \\ h_{k,1}^\alpha & h_{k,2}^\alpha & \cdots & h_{k,w_k}^\alpha & h_{k,w_k}^\alpha & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

In view of the above construction of $F_\alpha$ and the linear selection function $\hat{b}^\alpha$, we obtain the following proposition that characterizes the piecewise linear formulation of $\hat{b}$.

**Proposition 4.1.** For each index set $\alpha \subseteq \{1, \ldots, K_n - 1\}$, its corresponding linear selection function $\hat{b}^\alpha$ is given by

$$\hat{b}^\alpha(\bar{y}) = F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha \bar{y},$$

where $F_\alpha \in \mathbb{R}^{\ell \times (K_n + 1)}$ is a row independent matrix given in (13).
Let \( F_\alpha \Gamma F_\alpha^T := G + \lambda H \in \mathbb{R}^{\ell \times \ell} \), where \( G := F_\alpha \Gamma F_\alpha^T \) and \( H := (F_\alpha D_2^T)(F_\alpha D_2^T)^T \). The following result establishes the properties of the matrix \( H \).

**Proposition 4.2.** Let \( m = 2 \). For any \( K_n \) and \( \alpha \), the matrix \( H \) is a symmetric, banded matrix with the bandwidth \( m \). Furthermore, (i) \( 0 \leq H_{ss} \leq 6, \forall s = 1, \ldots, \ell \); (ii) \( |H_{s(s+1)}| = |H_{(s+1)s}| \leq 4, \forall s = 1, \ldots, \ell - 1 \); and (iii) \( |H_{s(s+2)}| = |H_{(s+2)s}| \leq 1, \forall s = 1, \ldots, \ell - 2 \).

**Proof.** Consider \( F_\alpha \) in Eq. (13). Recall that if \( m_0^k = 1 \), then \( F_{\alpha,k} = 1 \) and we define \( w_k := 1 \) and \( h_{k,1}^\alpha := 1 \); otherwise, \( h_{k,j}^\alpha \geq 2 \) for each \( j = 1, \ldots, w_k \). It follows from the structure of \( F_\alpha \) in Eq. (13) and (14) that

\[
F_\alpha D_2^T = 
\begin{bmatrix}
0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & -\frac{\lambda_{1,1}}{h_{1,1}^\alpha} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & -\frac{\lambda_{1,2}}{h_{1,2}^\alpha} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & -\frac{\lambda_{1,3}}{h_{1,3}^\alpha} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & 0 & -\frac{\lambda_{1,4}}{h_{1,4}^\alpha} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\end{bmatrix},
\]  

where the bold face \( 0 \) in the above matrix denotes the zero row of \( (h_{k,j}^\alpha - 1) \) elements if the first nonzero term \( 1/h_{k,j}^\alpha \) (from the top) appears in the column immediately to its right. (By convention, if \( h_{k,j}^\alpha = 1 \), then the zero row vanishes.) For example, the zero row \( 0 \) in the first column has \( (h_{1,1}^\alpha - 1) \) elements, and the next \( 0 \) in the third column has \( (h_{1,2}^\alpha - 1) \) elements.

Clearly, \( H = (F_\alpha D_2^T)(F_\alpha D_2^T)^T \) is symmetric. In light of Eq. (15), we see that \( H_{ij} = 0 \) whenever \( |i - j| \geq 3 \). Furthermore, in view of \( h_{k,j}^\alpha \geq 1 \), we have for each suitable \( s \),

\[
0 \leq H_{ss} \leq \max_{(k,j)} \left(1 + \left(1 + \frac{1}{h_{k,j}^\alpha}\right)^2 + \left(\frac{1}{h_{k,j}^\alpha}\right)^2\right) \leq 6,
\]

\[
|H_{s(s+1)}| \leq 2\max_{(k,j)} \left(\frac{1}{h_{k,j}^\alpha} + 1\right) \leq 4,
\]

\[
|H_{s(s+2)}| \leq \max_{(k,j)} \left(\frac{1}{h_{k,j}^\alpha}, \frac{1}{h_{k,j+1}^\alpha}, \frac{1}{h_{k,j}^\alpha}\right) \leq 1.
\]

This yields the proposition. \( \square \)

The next proposition further establishes important properties of the matrix \( F_\alpha \Gamma F_\alpha^T \) that pave the way for the uniform Lipschitz property of \( b \).

**Proposition 4.3.** There exists \( P \in \mathbb{N} \) such that for any \( K_n \) with \( n/K_n \geq P \) and any index set \( \alpha \), the matrix \( G := F_\alpha \Gamma F_\alpha^T \) is a symmetric, strictly diagonally dominant, and tridiagonal matrix.

**Proof.** Clearly, \( G \) is symmetric. For a given index subset \( \beta_k^\alpha \) corresponding to \( \alpha \) constructed above, let \( \Gamma_{\beta_k^\alpha,\beta_k^\alpha} \) denote the principal submatrix of \( \Gamma \) defined by \( \beta_k^\alpha \). In view of the structure of \( F_\alpha \) in Eq. (13), we have

\[
G = \text{diag} \left( F_{\alpha,1} \Gamma_{\beta_1^\alpha,\beta_1^\alpha} F_{\alpha,1}^T, F_{\alpha,2} \Gamma_{\beta_2^\alpha,\beta_2^\alpha} F_{\alpha,2}^T, \ldots, F_{\alpha,L} \Gamma_{\beta_L^\alpha,\beta_L^\alpha} F_{\alpha,L}^T \right).
\]
In what follows, we drop the subscript \( n \) in \( \theta_n \) and \( q_n \) for notational simplicity. To determine the entries of \( G = (g_{ij}) \), we consider a fixed \( k \in \{1, \ldots, L\} \). If \( m_k^\alpha = 1 \), then \( F_{\alpha,k} \Gamma_{\beta_k}^{\alpha} \beta_k F_{\alpha,k}^T \) is a scalar that appears on the diagonal of \( G \). Denoting this number by \( g_{ss} \), we have

\[
g_{ss} = F_{\alpha,k} \Gamma_{\beta_k}^{\alpha} \beta_k F_{\alpha,k}^T = \begin{cases} \theta, & \text{if } k \in \{1, L\} \\ 1, & \text{otherwise} \end{cases}
\]

and \( g_{s(s+1)} = g_{(s+1)s} = \eta \), \( g_{sj} = 0 \) for all \( j \) with \( |s-j| \geq 2 \).

Now consider \( m_k^\alpha > 1 \). In this case, \( F_{\alpha,k} \Gamma_{\beta_k}^{\alpha} \beta_k F_{\alpha,k}^T \) is a symmetric and positive definite matrix of order \((w_k + 1)\) that forms a diagonal block of \( G \). Making use of the structure of \( F_{\alpha,k} \) given in (14) and somewhat lengthy computation, we obtain the following results in two separate cases:

1. \( k = 1 \) or \( k = L \). For \( k = 1 \),

\[
g_{11} = \theta + \eta - \frac{\eta}{h_{1,1}^\alpha} + (1 + 2\eta) \frac{(h_{1,1}^\alpha - 1)(2h_{1,1}^\alpha - 1)}{6h_{1,1}^\alpha},
\]

\[
g_{s(s+1)} = g_{(s+1)s} = \eta \frac{h_{s,s}^\alpha}{h_{s,s}^\alpha} + (1 + 2\eta) \frac{(h_{s,s}^\alpha - 1)^2 - 1}{6h_{s,s}^\alpha}, \quad \forall \ s = 1, \ldots, w_1,
\]

\[
g_{ss} = (1 + 2\eta) \left[ \frac{2(h_{s,s}^\alpha - 1)^2 + 1}{6h_{s,s}^\alpha} + \frac{2(h_{s,s}^\alpha + 1)^2 + 1}{6h_{s,s}^\alpha} \right] - \left( \frac{1}{h_{s,s}^\alpha - 1} + \frac{1}{h_{s,s}^\alpha} \right) \eta, \quad \forall \ s = 2, \ldots, w_1,
\]

\[
g_{(w_1+1)(w_1+1)} = (1 + 2\eta) \frac{(h_{1,w_1}^\alpha + 1)(2h_{1,w_1}^\alpha + 1)}{6h_{1,w_1}^\alpha} - \left( \frac{1}{h_{1,w_1}^\alpha} \right) \eta.
\]

Besides, \( g_{(w_1+1)(w_1+2)} = g_{(w_1+2)(w_1+1)} = \eta \), and for each \( s = 1, \ldots, w_1 \), \( g_{sj} = 0 \) once \( |s-j| \geq 2 \). For \( k = L \), the similar results can be established by using the symmetry of the rows of \( F_{\alpha,L} \).

2. \( k \in \{2, \ldots, L - 1\} \). In this case, suppose that the \((1, 1)\)-element of \( F_{\alpha,k} \Gamma_{\beta_k}^{\alpha} \beta_k F_{\alpha,k}^T \) corresponds to the diagonal entry \( g_{tt} \) of \( G \), where \( 2 \leq t \leq \ell - 1 \). Then we have

\[
g_{tt} = 1 + \eta - \frac{\eta}{h_{k,1}^\alpha} + (1 + 2\eta) \frac{(h_{k,1}^\alpha - 1)(2h_{k,1}^\alpha - 1)}{6h_{k,1}^\alpha},
\]

\[
g_{(t+s-1)(t+s)} = g_{(t+s)(t+s-1)} = \frac{\eta}{h_{k,s}^\alpha} + (1 + 2\eta) \frac{(h_{k,s}^\alpha - 1)^2 - 1}{6h_{k,s}^\alpha}, \quad \forall \ s = 1, \ldots, w_k,
\]

\[
g_{(t+s)(t+s)} = (1 + 2\eta) \left[ \frac{2(h_{k,s}^\alpha - 1)^2 + 1}{6h_{k,s}^\alpha} + \frac{2(h_{k,s}^\alpha + 1)^2 + 1}{6h_{k,s}^\alpha} \right] - \left( \frac{1}{h_{k,s}^\alpha - 1} + \frac{1}{h_{k,s}^\alpha} \right) \eta,
\]

\[
\forall \ s = 1, \ldots, w_k - 1,
\]

\[
g_{(t+w_k)(t+w_k)} = (1 + 2\eta) \frac{(h_{k,w_k}^\alpha + 1)(2h_{k,w_k}^\alpha + 1)}{6h_{k,w_k}^\alpha} - \left( \frac{1}{h_{k,w_k}^\alpha} \right) \eta.
\]

In addition, \( g_{t(t-1)} = g_{(t+w_k)(t+w_k+1)} = \eta \), and for each \( s = t, \ldots, t + w_k + 1 \), \( g_{sj} = 0 \) once \( |s-j| \geq 2 \).

These results show that \( G \) is tridiagonal.

Next we show that \( G \in \mathbb{R}^{\ell \times \ell} \) is strictly diagonally dominant. Define

\[\xi_t := g_{tt} - |g_{12}|, \quad \xi_t := g_{tt} - |g_{t(t-1)}|, \quad \text{and} \quad \xi_t := g_{tt} - |g_{(t-1)(t)}| - |g_{(t+1)(t+1)}|, \forall \ t \in \{2, \ldots, \ell - 1\}.\]

Recall that \( \theta \to 1/2 \) and \( \eta \to 1/4 \) as \( n/K_n \to \infty \). Hence, there exists \( P \in \mathbb{N} \) such that for any \( K_n \) with \( n/K_n \geq P \) and any \( \alpha \) such that
(3.1) if $m_k^\alpha = 1$, then (i) if $k \in \{1, L\}$, the corresponding $\xi_i = \theta - |\eta| > 1/5$; and (ii) otherwise, the corresponding $\xi_i = 1 - 2|\eta| > 1/3$.

(3.2) if $m_k^\alpha > 1$ with $k = 1$, then

(i) for $s = 1$, in view of $h_{1,1}^\alpha \geq 2$, the corresponding

\[
\xi_i = g_{11} - |g_{12}| = \theta + \eta - \frac{\eta}{h_{1,1}^\alpha} + (1 + 2\eta) \frac{(h_{1,1}^\alpha - 1)(2h_{1,1}^\alpha - 1)}{6h_{1,1}^\alpha} - \frac{\eta}{h_{1,1}^\alpha} + (1 + 2\eta) \frac{(h_{1,1}^\alpha)^2 - 1}{6h_{1,1}^\alpha} \\
= \theta + (1 - \frac{2}{h_{1,1}^\alpha}) \eta + (1 + 2\eta) \left[ \frac{(h_{1,1}^\alpha - 1)(2h_{1,1}^\alpha - 1)}{6h_{1,1}^\alpha} - \frac{(h_{1,1}^\alpha)^2 - 1}{6h_{1,1}^\alpha} \right] \\
\geq \theta + (1 + 2\eta) \left( \frac{h_{1,1}^\alpha}{6} - \frac{1}{2} + \frac{1}{3h_{1,1}^\alpha} \right) \geq \theta + (1 + 2\eta) \left( \frac{h_{1,1}^\alpha}{6} - \frac{1}{2} \right) \\
\geq \theta + (1 + 2\eta) \left( \frac{h_{1,1}^\alpha}{42} + \frac{2}{7} - \frac{1}{2} \right) \\
\geq \frac{1}{14} + (1 + 2\eta) \frac{h_{1,1}^\alpha}{42} > 0.
\]

(ii) for $s = 2, \ldots, w_1$, the corresponding

\[
\xi_i = g_{ss} - |g_{s(s-1)}| - |g_{s(s+1)}| = (1 + 2\eta) \left[ \frac{2(h_{1,s-1}^\alpha)^2 + 1}{6h_{1,s-1}^\alpha} + \frac{2(h_{1,1}^\alpha)^2 + 1}{6h_{1,1}^\alpha} - \frac{(h_{1,s}^\alpha)^2 - 1}{6h_{1,s}^\alpha} - \frac{(h_{1,s-1}^\alpha)^2 - 1}{6h_{1,s-1}^\alpha} \right] \\
- 2 \left( \frac{1}{h_{1,s-1}^\alpha} + \frac{1}{h_{1,s}^\alpha} \right) \eta \\
= (1 + 2\eta) \left( h_{1,s-1}^\alpha + h_{1,s}^\alpha \right) + \left( \frac{1}{h_{1,s-1}^\alpha} + \frac{1}{h_{1,s}^\alpha} \right) \left( \frac{1}{3} - \frac{2\eta}{1 + 2\eta} \right) \\
\geq (1 + 2\eta) \left( h_{1,s-1}^\alpha + h_{1,s}^\alpha \right) > 0,
\]

where the last inequality follows from the facts that $\frac{2\eta}{1 + 2\eta} \to 1/3$ as $n/K_n \to \infty$, and that for any $h_1, h_2 \geq 2$,

\[
\frac{1}{h_1} + \frac{1}{h_2} \leq \frac{h_1 + h_2}{4}.
\]

and (iii) for $s = w_1 + 1$, the corresponding

\[
\xi_i = g_{(w_1+1)(w_1+1)} - |g_{(w_1+1)w_1}| - |g_{(w_1+1)(w_1+2)}| \\
= (1 + 2\eta) \left( \frac{h_{1,w_1}^\alpha + 1}{6h_{1,w_1}^\alpha} + \frac{1}{h_{1,w_1}^\alpha} \right) \eta - \frac{\eta}{h_{1,w_1}^\alpha} + (1 + 2\eta) \frac{(h_{1,w_1}^\alpha)^2 - 1}{6h_{1,w_1}^\alpha} - |\eta| \\
= (1 + 2\eta) \left( \frac{(h_{1,w_1}^\alpha + 1)(2h_{1,w_1}^\alpha + 1)}{6h_{1,w_1}^\alpha} - \frac{(h_{1,w_1}^\alpha)^2 - 1}{6h_{1,w_1}^\alpha} \right) - 2 \left( \frac{1}{h_{1,w_1}^\alpha} \right) \eta \\
= (1 + 2\eta) \left( \frac{h_{1,w_1}^\alpha + 1}{6} + \frac{1}{3h_{1,w_1}^\alpha} + \frac{1}{2} \left( 1 + \frac{1}{h_{1,w_1}^\alpha} \right) \frac{2\eta}{1 + 2\eta} \right) \\
\geq (1 + 2\eta) \left( \frac{h_{1,w_1}^\alpha}{8} + \frac{1}{8} \right) > 0.
\]

The similar results can be obtained for $m_k^\alpha > 1$ with $k = L$ using symmetry.
if $m_k^n > 1$ with $k \in \{2, \ldots, L - 1\}$, then
(i) for the $t$th row, the corresponding
\[
\xi_i = g_{nt} - |g_{(t-1)}| - |g_{(t+1)}| \\
= 1 + \eta - \frac{\eta}{h_{k,1}} + (1 + 2\eta) \left(\frac{h_{k,1}^\alpha - 1}{6h_{k,1}^\alpha} \right) - |\eta| - \frac{\eta}{h_{k,1}^\alpha} + (1 + 2\eta) \left(\frac{h_{k,1}^\alpha - 1}{6h_{k,1}^\alpha} \right)
\]
\[
= (1 - \frac{2\eta}{h_{k,1}^\alpha}) + (1 + 2\eta) \left(\frac{h_{k,1}^\alpha - 1}{6h_{k,1}^\alpha} \right) - \frac{\eta}{h_{k,1}^\alpha} + (1 + 2\eta) \left(\frac{h_{k,1}^\alpha - 1}{6h_{k,1}^\alpha} \right)
\]
\[
\geq (1 - \eta) + (1 + 2\eta) \left(\frac{h_{k,1}^\alpha}{6} + \frac{1}{3h_{k,1}^\alpha} - \frac{1}{2} \right) \geq (1 + 2\eta) \left(\frac{h_{k,1}^\alpha}{7} + \frac{1}{2} \right) \geq (1 + 2\eta) \left(\frac{h_{k,1}^\alpha}{7} + \frac{1}{42} \right) > 0
\]
for all $n/K_n$ sufficiently large.

(ii) for $s = 1, \ldots, w_k - 1$, the corresponding
\[
\xi_i = g_{(s+t)} - |g_{(s+t-1)}| - |g_{(s+t+1)}| \\
= (1 + 2\eta) \left(\frac{2(h_{k,s+1}^\alpha)^2 + 1}{6h_{k,s+1}^\alpha} \right) + (1 + 2\eta) \left(\frac{2(h_{k,s}^\alpha)^2 + 1}{6h_{k,s}^\alpha} \right) - \frac{\eta}{h_{k,s}^\alpha} + (1 + 2\eta) \left(\frac{h_{k,s+1}^\alpha - 1}{6h_{k,s+1}^\alpha} \right)
\]
\[
= (1 + 2\eta) \left(\frac{2(h_{k,s+1}^\alpha)^2 + 1}{6h_{k,s+1}^\alpha} \right) + (1 + 2\eta) \left(\frac{2(h_{k,s}^\alpha)^2 + 1}{6h_{k,s}^\alpha} \right) - \frac{\eta}{h_{k,s}^\alpha} + (1 + 2\eta) \left(\frac{h_{k,s+1}^\alpha - 1}{6h_{k,s+1}^\alpha} \right)
\]
\[
\geq (1 + 2\eta) \left(\frac{h_{k,s+1}^\alpha + h_{k,s}^\alpha}{8} \right) > 0.
\]

and (iii) for $s = w_k$, the corresponding
\[
\xi_i = g_{(s+t+w_k)} - |g_{(s+t+w_k-1)}| - |g_{(s+t+w_k+1)}| \\
= (1 + 2\eta) \left(\frac{h_{k,w_k}^\alpha + 1}{6h_{k,w_k}^\alpha} \right) - \frac{\eta}{h_{k,w_k}^\alpha} + (1 + 2\eta) \left(\frac{h_{k,w_k}^\alpha - 1}{6h_{k,w_k}^\alpha} \right)
\]
\[
= (1 + 2\eta) \left(\frac{h_{k,w_k}^\alpha + 1}{6h_{k,w_k}^\alpha} \right) - \frac{\eta}{h_{k,w_k}^\alpha} + (1 + 2\eta) \left(\frac{h_{k,w_k}^\alpha - 1}{6h_{k,w_k}^\alpha} \right)
\]
\[
\geq (1 + 2\eta) \left(\frac{h_{k,w_k}^\alpha}{8} + \frac{1}{8} \right) > 0.
\]

Consequently, all $\xi_i > 0$ so that $G$ is strictly diagonally dominant.
With the help of the above results, we are ready to prove the uniform Lipschitz property for the optimal spline coefficient vector \( \hat{b} \).

**Theorem 4.1.** There exist positive constants \( \Xi, P \in \mathbb{N} \), and \( \kappa_\infty \) (which is independent of \( K_n \) and \( \lambda \)) such that for any \( \lambda \in [0, \bar{\lambda}] \) and \( K_n \) with \( n/K_n \geq P \),

\[
\| \hat{b}(y^1) - \hat{b}(y^2) \|_\infty \leq \kappa_\infty \| y^1 - y^2 \|_\infty, \quad \forall \ y^1, y^2 \in \mathbb{R}^{K_n+1}.
\]

**Proof.** By Proposition 4.1, it suffices to show the uniform bound of \( \| F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha \|_\infty \). Given any \( K_n \) and any \( \alpha \), recall that \( F_\alpha \Lambda F_\alpha^T = G + \lambda H \), where \( G \) is tridiagonal and \( H \) is banded with bandwidth \( m = 2 \). Hence, \( F_\alpha \Lambda F_\alpha^T \) is banded with bandwidth \( m = 2 \). Define

\[
\tilde{\xi}_i := \| (F_\alpha \Lambda F_\alpha^T)_{ii} \| - \sum_{j \neq i} \| (F_\alpha \Lambda F_\alpha^T)_{ij} \|, \quad \forall \ i = 1, \ldots, \ell.
\]

Let \( P \in \mathbb{N} \) be given in Proposition 4.3. It follows from Propositions 4.2 and 4.3 that for any \( K_n \) with \( n/K_n \geq P \), \( \tilde{\xi}_i \geq \xi_i - 16 \lambda \) for each \( i \), where \( \xi_i \) is defined in the proof of Proposition 4.3. Moreover, Proposition 4.3 shows that \( \xi_i \)'s are uniformly bounded below by a positive constant for any \( K_n \) and \( \alpha \). Therefore, there exists \( \lambda > 0 \) (independent of \( K_n \) and \( \alpha \)) such that for any \( \lambda \in [0, \bar{\lambda}] \), \( \tilde{\xi}_i \geq \xi_i/2 > 0, \forall \ i = 1, \ldots, \ell \) for any \( K_n \) and \( \alpha \).

Let the diagonal matrix \( \Xi := \text{diag}(\tilde{\xi}_1^{-1}, \ldots, \tilde{\xi}_\ell^{-1}) \in \mathbb{R}^{\ell \times \ell} \), which is clearly invertible. We have

\[
\| F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha \|_\infty = \| F_\alpha^T \cdot (\Xi F_\alpha \Lambda F_\alpha^T)^{-1} \cdot (\Xi F_\alpha) \|_\infty \leq \| F_\alpha^T \|_\infty \cdot \| (\Xi F_\alpha \Lambda F_\alpha^T)^{-1} \|_\infty \cdot \| \Xi F_\alpha \|_\infty,
\]

where it is easy to verify \( \| F_\alpha^T \|_\infty = 1 \). Since \( E := \Xi F_\alpha \Lambda F_\alpha^T \) is strictly diagonally dominant with \( E_{ii} - \sum_{j=1,j \neq i}^\ell |E_{ij}| = 1 \) for each \( i \), it follows from the Ahlberg-Nilsen-Varah bound [32] that

\[
\| E^{-1} \|_\infty = \| (\Xi F_\alpha \Lambda F_\alpha^T)^{-1} \|_\infty \leq 1.
\]

Next we establish the bound on \( \| \Xi F_\alpha \|_\infty \) as follows. By using the results for \( \xi_i \) developed in Proposition 4.3 and \( \tilde{\xi}_i \geq \xi_i/2 \), we have for any \( \alpha \), any \( \lambda \in [0, \bar{\lambda}] \), and any \( K_n \) with \( n/K_n \geq P \),

1. if \( m_\alpha^\ell = 1 \), then the absolute sum of the entries in the corresponding row in \( \Xi F_\alpha \) is given by \( 1/\xi_i \leq 10 \).
2. if \( m_\alpha^\ell > 1 \) with \( k = 1 \), then
   2.1 in view of \( h_{1,1} \geq 2 \), the absolute sum of the entries in the row in \( \Xi F_\alpha \) corresponding to \( g_{11} \) is given by
   \[
   \frac{1 + h_{1,1}^\alpha}{2 \xi_i} \leq \frac{1 + h_{1,1}^\alpha}{1 + (1 + 2\eta) h_{1,1}^\alpha} \leq 40.
   \]
   2.2 for \( s = 2, \ldots, w_1 \), the absolute sum of the entries in the row in \( \Xi F_\alpha \) corresponding to \( g_{ss} \) is given by
   \[
   \frac{h_{1,s}^\alpha + h_{1,1}^\alpha}{2 \xi_i} \leq \frac{h_{1,s}^\alpha + h_{1,1}^\alpha}{(1 + 2\eta) \left( \frac{h_{1,s}^\alpha + h_{1,1}^\alpha}{8} \right)} \leq 12.
   \]
3. if \( m_\alpha^\ell > 1 \) with \( k = L \), the absolute sum of the entries in the row in \( \Xi F_\alpha \) corresponding to \( g_{(w_1 + 1)(w_1 + 1)} \) is given by
   \[
   \frac{1 + h_{1,w_1}^\alpha}{2 \xi_i} \leq \frac{1 + h_{1,w_1}^\alpha}{(1 + 2\eta) \left( \frac{h_{1,w_1}^\alpha}{8} + \frac{1}{8} \right)} \leq 12.
   \]

The same results can be obtained for \( m_\alpha^\ell > 1 \) with \( k = L \).
if $m^2_{k,n} > 1$ with $k \in \{2, \ldots, L - 1\}$, then

(3.1) the absolute sum of the entries in the row in $\Xi F_\alpha$ corresponding to $g_{tt}$ is given by

$$\frac{1 + h^\alpha_{k,1}}{2\xi_i} \leq \frac{1 + h^\alpha_{k,1}}{(1 + 2\eta)\left(\frac{h^\alpha_{k,1}}{\kappa} + \frac{1}{2\eta}\right)} \leq 40.$$ 

(3.2) for $s = 1, \ldots, w_k - 1$, the absolute sum of the entries in the row in $\Xi F_\alpha$ corresponding to $g_{(t+s)(t+s)}$ is given by

$$\frac{h^\alpha_{k,s} + h^\alpha_{k,s+1}}{2\xi_i} \leq \frac{h^\alpha_{k,s} + h^\alpha_{k,s+1}}{(1 + 2\eta)\left(\frac{h^\alpha_{k,s+1} + h^\alpha_{k,s}}{\kappa}\right)} \leq 12.$$ 

(3.3) for $s = w_k$, the absolute sum of the entries in the row in $\Xi F_\alpha$ corresponding to $g_{(t+w_k)(t+w_k)}$ is given by

$$\frac{1 + h^\alpha_{k,w_k}}{2\xi_i} \leq \frac{1 + h^\alpha_{k,w_k}}{(1 + 2\eta)\left(\frac{h^\alpha_{k,w_k}}{\kappa} + \frac{1}{2\eta}\right)} \leq 12.$$ 

In view of the above results, we deduce that the existence of a positive constant $\kappa_\infty$, which is independent of $K_n$, $\alpha$, and $\lambda$, such that for all $\lambda \in [0, \bar{\lambda}]$ and $n/K_n \geq P$, $\|\Xi F_\alpha\|_\infty \leq \kappa_\infty$. This in turn implies that $\|F_\alpha^T(F_\alpha \Lambda F_\alpha^T)^{-1}F_\alpha\|_\infty \leq \kappa_\infty$, regardless of $\alpha$, $\lambda$, and $K_n$. Hence, $\|\hat{b}(\bar{y})\|_\infty \leq \kappa_\infty\|\bar{y}\|_\infty$ for any $\bar{y} \in \mathbb{R}^{K_{n+1}}$. Finally, the uniform Lipschitz property of $\hat{b}$ thus follows from the piecewise linear property of $\tilde{b}$ [8, Proposition 4.2.2].

**Remark 4.1.** The uniform Lipschitz property established in Theorem 4.1 for $m = 2$ can be extended to other difference penalties. In fact, it follows from the similar argument as in Proposition 4.2 that for $m \geq 3$, the matrix $H$ is a banded matrix with the bandwidth $m$ and the absolute row sum of $H$ is uniformly bounded. Hence, by choosing a suitable $\bar{\lambda} > 0$ sufficiently small, the uniform Lipschitz property holds. More involved computations show that the uniform Lipschitz property also holds for the first order difference penalty. Nevertheless, it is the second order difference penalty that allows us to obtain the optimal rate of convergence uniformly on $C_H(r, L)$ as shown in the next section.

### 4.3 Convex P-spline Estimator: Optimal Rate of Convergence

In this section, we show that the proposed convex P-spline estimator $\hat{f}^{[1]}$ achieves the optimal convergence rate in the sup-norm, via the uniform Lipschitz property (cf. Theorem 4.1) and asymptotic estimation techniques. To this end, we introduce the following convex P-spline function with $p = 1$ based on noise free data $\bar{f} := (f(x_1), \ldots, f(x_n))^T \in \mathbb{R}^n$, i.e.,

$$\hat{f}^{[1]}(x) = \sum_{k=1}^{K_{n+1}} \hat{b}_k B_k^{[1]}(x),$$

where the linear spline coefficient vector $\tilde{b} = (\tilde{b}_1, \ldots, \tilde{b}_{K_{n+1}})^T \in \mathbb{R}^{K_{n+1}}$ is given by

$$\tilde{b} = \arg\min_{b \in \Omega} \frac{1}{2} b^T \Lambda b - b^T (X^T \bar{f}/\beta_n).$$

Here $\Omega$, $\Lambda$, and $\beta_n$ are defined before [8]. In other words, $\tilde{b} = \tilde{b}(\mathbb{E}(\bar{y}))$. 

20
The following two propositions establish the uniform bounds for bias and stochastic errors of the proposed convex estimator, respectively. The proof of Proposition \[4.4\] \[\text{Proposition 4.2}\] recently established by the authors, and we present its proof to be self-contained and complete. In what follows, let the sup-norm \[\|g\|_\infty := \sup_{x \in [0,1]} |g(x)|\] for a function \(g \in C([0,1])\).

**Proposition 4.4.** Let \(r \in (1,2]\). Then there exist two positive constants \(C_1 \) and \(C_2\) such that for all sufficiently large \(K_n\) with \(n/K_n \geq P\) and for each \(\lambda \in [0,\overline{\lambda}]\) (uniformly in \(f\)),

\[
\sup_{f \in \mathcal{C}_H(r,L)} \|\bar{f}[1] - f\|_\infty \leq C_1LK_n^{-r} + C_2\sqrt{\lambda} \cdot K_n \cdot LK_n^{-r}.
\]

**Proof.** Given a function \(f \in \mathcal{C}_H(r,L)\), we introduce the following functions, in addition to \(\bar{f}[1]\) defined above:

(i) \(\bar{f} : [0,1] \to \mathbb{R}\) is defined by the linear interpolation of \((\kappa_k, f(\kappa_k))\) with \(k = 0,1,\ldots,K_n\). Hence, \(\bar{f}(\kappa_k) = f(\kappa_k)\) for all \(k\).

(ii) Let \(\tilde{f} := (\tilde{f}(x_1),\ldots,\tilde{f}(x_n))^T \in \mathbb{R}^n\) and define \(\hat{f} : [0,1] \to \mathbb{R}\) as \(\hat{f}(x) := \sum_{k=1}^{K_n+1} b_k B_k(x)\),

where \(\hat{b} := (\hat{b}_1,\ldots,\hat{b}_{K_n+1})\) is given by \(\hat{b} = \arg\min_{b \in \Omega} \frac{1}{2} b^T \Lambda b - \tilde{b}^T (X^T \tilde{f}/\beta_n)\).

Clearly, \(\|\tilde{f} - f\|_\infty \leq \|\tilde{f}[1] - \tilde{f}\|_\infty + \|\hat{f} - \tilde{f}\|_\infty + \|\hat{f} - f\|_\infty\), and we obtain the uniform bounds for each term on the right-hand side as follows:

1. \(\|\tilde{f} - f\|_\infty\). For each \(x \in [\kappa_{k-1},\kappa_k]\), \(k = 1,\ldots,K_n\), there exist \(\xi_x, \tilde{\xi}_x \in (\kappa_{k-1},\kappa_k)\) such that

\[
\tilde{f}(x) - f(x) = \hat{f}(\kappa_{k-1}) + \kappa_n(\hat{f}(\kappa_k) - \hat{f}(\kappa_{k-1}))(x - \kappa_{k-1}) - \hat{f}(\kappa_{k}) + \hat{f}'(\tilde{\xi}_x)(x - \kappa_{k-1})
\]

Thus by the Hölder condition, we have, for any \(x \in [\kappa_{k-1},\kappa_k]\), \(\|\tilde{f}(x) - f(x)\| \leq L|\tilde{\xi}_x - \xi_x|^\gamma|x - \kappa_{k-1}| \leq LK_n^{-r}\). This shows \(\|\tilde{f} - f\|_\infty \leq LK_n^{-r}\).

2. \(\|\tilde{f}[1] - \hat{f}\|_\infty\). It follows from the linear B-spline property that there exists a positive constant \(\rho\) (independent of \(n\)) such that \(\|X^T\|_\infty/\beta_n \leq \rho\) for all \(n\). Since both \(\tilde{f}[1]\) and \(\hat{f}\) are piecewise linear functions, we deduce, via the uniform Lipschitz property of the optimal spline coefficient (cf. Theorem \[4.1\]), that for all \(n/K_n\) and \(\lambda > 0\) sufficiently large,

\[
\|\tilde{f}[1] - \hat{f}\|_\infty \leq \|\hat{b} - \tilde{b}\|_\infty \leq \kappa_n \|X^T(\tilde{f} - \hat{f})/\beta_n\|_\infty \leq \kappa_n \|X^T\|_\infty \|f - \tilde{f}\|_\infty \leq \kappa_n \rho LK_n^{-r}.
\]

3. \(\|\hat{f} - \tilde{f}\|_\infty\). Let the vector \(\tilde{b} := (\hat{f}(\kappa_0),\hat{f}(\kappa_1),\ldots,\hat{f}(\kappa_{K_n}))\) \(\in \mathbb{R}^{K_n+1}\). Since \(\hat{f}(\kappa_k) = f(\kappa_k)\) for all \(k\) and \(f\) is convex, we have \(D_2 \hat{b} \geq 0\). Furthermore, due to the piecewise linear property of \(\hat{f}\), we deduce that \(\hat{f}(x) = \sum_{k=1}^{K_n+1} \hat{b}_k B_k(x)\) and \(\tilde{f} = X\hat{b}\). Along with the definition of the optimal spline coefficient vector \(\hat{b}\), this yields

\[
\lambda^* \|D_2 \hat{b}\|_2^2 \leq \lambda^* \|D_2 \hat{b}\|_2^2.
\]

By the virtue of \(\tilde{f} = X\hat{b}\) and \(\Gamma = X^T X/\beta_n \in \mathbb{R}^{(K_n+1) \times (K_n+1)}\), we further have

\[
\frac{1}{2} \|X(\tilde{b} - \hat{b})\|_2^2 \leq \lambda^* \|D_2 \hat{b}\|_2^2 \iff \langle \hat{b} - \tilde{b}, \Gamma(\hat{b} - \tilde{b}) \rangle \leq 2\lambda \|D_2 \hat{b}\|_2^2.
\]
Since $\Gamma$ is symmetric and positive definite, it follows from \cite{37} Lemma 6.2 that there exists a positive constant $\mu$ (independent of $K_n$) such that the smallest real eigenvalue of $\Gamma$ is bounded below by $\mu$ for any $K_n$. Therefore,

$$
\mu\|\hat{b} - \hat{b}\|_2^2 \leq \mu(\hat{b} - \hat{b})^T \Gamma(\hat{b} - \hat{b}) \leq 2\lambda\|D_2\hat{b}\|_2^2.
$$

Moreover, using $\hat{b}_k = f(\kappa_{k-1})$ and the Hölder condition for the true $f$, it is easy to show that

$$
|\Delta^2(\hat{b}_k)| \leq \frac{1}{K_n} L(2K_n)^{-\gamma} \leq 2^{-\gamma} L K_n^{-\gamma}, \quad \forall \ k = 2, \ldots, K_n + 1.
$$

Hence, $\|D_2\hat{b}\|_2 \leq (K_n)^{1/2} L K_n^{-\gamma}$. Letting $C_2 := (2/\mu)^{1/2}$, we have

$$
\|\hat{b} - \tilde{b}\|_2 \leq \sqrt{\frac{2\lambda}{\mu}} \|D_2\hat{b}\|_2 \leq \sqrt{\frac{2\lambda K_n}{\mu}} \cdot L K_n^{-\gamma} \leq C_2 \sqrt{\lambda K_n} \cdot L K_n^{-\gamma}.
$$

Consequently, $\|\hat{f} - \tilde{f}\|_\infty \leq \|\hat{b} - \tilde{b}\|_\infty \leq C_2 \sqrt{\lambda K_n} \cdot L K_n^{-\gamma}$.

Finally, putting the above uniform bounds together, we obtain the desired uniform bound of $\|\hat{f}^{[1]} - f\|_\infty$ for all $f \in \mathcal{H}(r, L)$.

**Proposition 4.5.** Let $r \in (1, 2]$, $\lambda \in (0, \lambda]$, and $K_n$ satisfy $n/K_n \to \infty$ and $K_n/(n^{1/2}/\sqrt{\log n}) \to 0$ as $n \to \infty$. Then there exists a positive constant $C_3$ such that for all $n$ sufficiently large (uniformly in $f$),

$$
\sup_{f \in \mathcal{H}(r, L)} \mathbb{E}(\|\hat{f}^{[1]} - f\|_\infty) \leq C_3 \left(\frac{K_n \log n}{n}\right)^{1/2}.
$$

**Proof.** Define $\xi_k := \sum_{i=1}^{n} B_k^{[1]}(x_i) e_i / \sqrt{\beta_n}$, where $k = 1, \ldots, K_n + 1$. Hence $\xi_k \sim N(0, 1)$ for each $k = 2, \ldots, K_n$. Besides, in view of $\sum_{i=1}^{n} (B_k^{[1]}(x_i))^2 \leq \beta_n$, we see that each $\xi_k$ with $k \in \{1, K_n + 1\}$ has the normal distribution with mean zero and variance not greater than one. Hence, for any $t \geq 0$, $P(|\xi_1| \geq t) \geq P(|\xi_k| \geq t)$ for each $k \in \{1, K_n + 1\}$. Moreover, it follows from the uniform Lipschitz property of $\hat{b}$ (cf. Theorem 4.1) and (7) that

$$
\|\hat{f}^{[1]} - \tilde{f}^{[1]}\|_\infty \leq \frac{\sigma K_n}{\sqrt{\beta_n}} \sup_{k=1,\ldots,K_n+1} |\xi_k| \leq \frac{\sigma K_n}{\sqrt{\beta_n}} \sqrt{\frac{K_n}{n}} \sup_{k=1,\ldots,K_n+1} |\xi_k|.
$$

Defining $C_4 := \kappa_\infty / \sqrt{\sigma}$ and $\bar{\xi} := \max_{k=1,\ldots,K_n+1} |\xi_k|$, we obtain that

$$
\|\hat{f}^{[1]} - \tilde{f}^{[1]}\|_\infty \leq \sigma C_4 \sqrt{\frac{K_n}{n}} \bar{\xi},
$$

and that for any $u \geq 0$, $P(|\xi_2| \geq u) \leq \frac{u}{C_4\sigma \sqrt{\frac{K_n}{n}}} \geq \frac{u}{C_3\sigma} \sqrt{\frac{K_n}{n}} \geq \frac{u}{C_4\sigma} \sqrt{\frac{K_n}{n}} \geq \frac{u}{C_3\sigma} \sqrt{\frac{K_n}{n}}$, for each $k \in \{1, K_n + 1\}$. In light of all these results and the implication: $Z \sim N(0, 1) \Longrightarrow P(Z \geq t) \leq \frac{1}{2} e^{-t^2/2}, \forall \ t \geq 0$, we deduce that for a given $u \geq 0$,

\[
\begin{align*}
P\left(\|\hat{f}^{[1]} - \tilde{f}^{[1]}\|_\infty \geq u\right) & \leq P\left(\xi_k \geq \frac{u}{C_4\sigma \sqrt{\frac{K_n}{n}}} \right) \leq (K_n + 1) P\left(\xi_2 \geq \frac{u}{C_4\sigma \sqrt{\frac{K_n}{n}}} \right) \\
& \leq (K_n + 1) \exp\left\{-\frac{n}{2K_n C_4^2 \sigma^2 u^2}\right\}.
\end{align*}
\]
Let \( T_n := C_4 \sigma \sqrt{\frac{2}{2t+1}} \sqrt{\frac{K_n \log n}{n}} \). It follows from the above result and \( \int_T^\infty e^{-t^2/(2\sigma^2)} dt \leq \sigma e^{-T^2/(2\sigma^2)} \sqrt{\pi/2}, \forall T \geq 0 \) that for any \( f \in \mathcal{C}_H(r, L) \),

\[
\mathbb{E}(\| \hat{f}^{[1]} - \bar{f}^{[1]} \|_\infty) \leq T_n + \int_{T_n}^\infty P(\| \hat{f}^{[1]} - \bar{f}^{[1]} \|_\infty > t) dt \\
\leq T_n + \int_{T_n}^\infty (K_n + 1) \exp \left\{ -\frac{n}{2KnC_3^2 \sigma^2} t^2 \right\} dt \\
\leq T_n + \sqrt{\frac{n}{2}} C_4 \sigma \sqrt{n-1} K_n (K_n + 1)n^{-\frac{1}{2r+1}}.
\]

Hence, if \( K_n \) satisfies \( \lim_{n \to \infty} K_n/(n^{\frac{1}{2r+1}} \sqrt{\log n}) = 0 \), then \( \mathbb{E}(\| \hat{f}^{[1]} - \bar{f}^{[1]} \|_\infty) = O(T_n) \) for all large \( n \). This implies that there exists a positive constant \( C_3 \) independent of \( f \) such that for all \( n \) sufficiently large, \( \mathbb{E}(\| \hat{f}^{[1]} - \bar{f}^{[1]} \|_\infty) \leq C_3 \sqrt{\frac{K_n \log n}{n}} \) for any \( f \in \mathcal{C}_H(r, L) \).

The next theorem shows that by choosing suitable \( K_n \) and \( \lambda^* \), the proposed convex \( P \)-spline estimator achieves the optimal rate of convergence in the sup-norm uniformly on the function class \( \mathcal{C}_H(r, L) \). This thus yields the desired minimax upper bound.

**Theorem 4.2.** If \( K_n \) and \( \lambda^* \) are chosen as

\[
K_n = \left\lceil \left( \frac{n}{\log n} \right)^{\frac{1}{2r+1}} \right\rceil, \quad \lambda^* = \beta_n K_n^{-1},
\]

then there exists a positive constant \( C_0 \) (dependent on \( \sigma, L, r \) only) such that

\[
\sup_{f \in \mathcal{C}_H(r, L)} \mathbb{E}(\| \hat{f}^{[1]} - f \|_\infty) \leq C_0 \left( \frac{\log n}{n} \right)^{\frac{1}{2r+1}}, \quad \forall \text{ large } n.
\]

**Proof.** Clearly, the selected \( K_n \) satisfies \( K_n \to \infty, n/K_n \to \infty \) and \( K_n/(n^{\frac{1}{2r+1}} \sqrt{\log n}) \to 0 \) as \( n \to \infty \). It also follows from the choice of \( \lambda^* \) that \( \lambda = \lambda^*/\beta_n = K_n^{-1} \). Note that \( \lambda \in (0, \lambda^*] \) as \( K_n \to \infty \). By Proposition 4.3, we see that the uniform bound of the bias is given by

\[
\sup_{f \in \mathcal{C}_H(r, L)} \| \hat{f}^{[1]} - f \|_\infty \leq C'_1 L K_n^{-r},
\]

where \( C'_1 \) is a positive constant. Furthermore, in light of Proposition 4.5 we have for all \( n \) sufficiently large,

\[
\mathbb{E}(\| \hat{f}^{[1]} - f \|_\infty) \leq \| \hat{f}^{[1]} - f \|_\infty + \mathbb{E}(\| \hat{f}^{[1]} - \bar{f}^{[1]} \|_\infty) \leq C'_1 L K_n^{-r} + C_3 \sqrt{\frac{K_n \log n}{n}}
\]

for any \( f \in \mathcal{C}_H(r, L) \), where the positive constants \( C'_1 \) and \( C_3 \) are independent of \( f \). It is easy to show that for any fixed \( n \), \( \min_{K_n} (C'_1 L K_n^{-r} + C_3 (\frac{K_n \log n}{n})^{1/2}) \) is achieved when \( C'_1 L K_n^{-r} = C_3 (\frac{K_n \log n}{n})^{1/2} \). Simple calculation further shows that the choice of \( K_n \) and \( \lambda^* \) gives rise to the desired optimal convergence rate in the sup-norm.

**Remark 4.2.** Since the convex \( P \)-spline estimator achieves the uniform convergence on the entire interval \([0, 1]\), it is consistent not only in the interior of \([0, 1]\) but also on the boundary, which is a critical property that many other convex estimators (e.g., the least squares estimator) do not have. Roughly speaking, this is because the convex \( P \)-spline estimator takes advantage of binned data between neighboring knots near a boundary point to yield better estimates under suitable \( \lambda^* \) and \( K_n \), while other estimators do not do so.
5 Concluding Remarks

In this paper, we have established the optimal rate of convergence for the minimax risk of convex estimators under the sup-norm. The results developed in this paper shed light on further research on shape constrained minimax theory. For example, the minimax lower bound is obtained via construction of a family of piecewise quadratic convex functions, and this approach can be extended to other derivative constraints. The minimax upper bound is developed by a convex $P$-spline estimator subject to the second order difference penalty. A key step in the upper bound analysis is the uniform Lipschitz property for optimal spline coefficients. This important property is known to hold for monotone $P$-splines [26], and it is conjectured that the similar property also holds for other shape constrained $P$-splines, but its proof is much more involved and shall be reported in the future. Other related topics include confidence band construction for shape constrained estimators [5].

References

[1] L. Birgé. A new lower bound for multiple hypothesis testing. *IEEE Trans. on Information Theory*, Vol. 51, pp. 1611–1615, 2005.

[2] T. Cai and M. Low. A framework for estimation of convex functions. Technical report, 2012.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2005.

[4] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. Academic Press Inc., (Cambridge 1992).

[5] L. Dümbgen. Optimal confidence bands for shape-restricted curves. *Bernoulli*, Vol. 9, pp. 423–449, 2003.

[6] L. Dümbgen, S. Freitag, and G. Jongbloed. Consistency of concave regression with an application to current-status data. *Mathematical Methods of Statistics*, Vol. 13, pp. 69–81, 2004.

[7] M. Egerstedt and C. Martin. *Control Theoretic Splines*. Princeton University Press, 2010.

[8] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, 2003.

[9] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[10] 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.

[11] P. L. Hanson and G. Pledger. Consistency in concave regression. *Annals of Statistics*, Vol. 4, pp.1038–1050, 1976.

[12] C. Hildreth. Point estimates of ordinates of concave functions. *Journal of American Statistical Association*, Vol. 49, pp. 598–619, 1954.

[13] J. Kiefer. Optimum rates for non-parametric density and regression estimates under order restrictions. In: Kallianpur, G., Krishnaiah, P. R., Ghosh, J.K. (Eds.), Statistics and Probability: Essays in honor of C. R. Rao. North-Holland, Amsterdam, pp. 419–428, 1982.

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

[15] Y. Li and D. Ruppert. On the asymptotics of penalized splines. *Biometrika*, Vol. 95, pp. 415–436, 2008.
[16] M. Low and Y. Kang. Estimating monotone functions. *Statistics & Probability Letters*, Vol. 56, pp. 361–367, 2002.

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

[18] E. Mammen and C. Thomas-Agnan. Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics*, Vol. 26, pp. 239-252, 1999.

[19] B. Marx and P. Eilers. Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science*, Vol. 11, pp. 89–121, 1996.

[20] M. Meyer. Inference using shape-restricted regression splines. *Annals of Applied Statistics*, Vol. 2, pp. 1013–1033, 2008.

[21] A. Nemirovski. *Topics in Non-parametric Statistics*. Lecture on Probability Theory and Statistics, Berlin, Germany: Springer-Verlag, Vol. 1738, Lecture Notes in Mathematics, 2000.

[22] J. Pal and M. Woodroofe. Large sample properties of shape restricted regression estimators with smoothness adjustments. *Statistica Sinica*, Vol. 17, pp. 1601–1616, 2007.

[23] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, 1964.

[24] J. O. Ramsay. Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B*, Vol. 60, pp. 365–375, 1988.

[25] J. Shen and X. Wang. Estimation of shape constrained functions in dynamical systems and its application to genetic networks. *Proc. of American Control Conference*, pp. 5948–5953, Baltimore, 2010.

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

[27] J. Shen and X. Wang. A constrained optimal control approach to smoothing splines. *Proc. of the 50th IEEE Conf. Decision and Control*, pp. 1729–1734, Orlando, FL, 2011.

[28] 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.

[29] C. J. Stone. Optimal rate of convergence for nonparametric regression. *Annals of Statistics*, Vol. 10, pp. 1040–1053, 1982.

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

[31] F. Utreras. Smoothing noisy data under monotonicity constraints: existence, characterization and convergence rates. *Numerische Mathematik*, Vol. 47, pp. 611–625, 1985.

[32] J. M. Varah. A lower bound for the smallest singular value of a matrix. *Linear Algebra and Its Applications*, Vol. 11, pp. 3–5, 1975.

[33] 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.

[34] X. Wang and J. Shen. Uniform convergence and rate adaptive estimation of convex functions via constrained optimization. In press, 2012.

[35] M. B. Woodroofe and J. Sun. A penalized maximum likelihood estimate of $f(0_+)$ when $f$ is nonincreasing. *Statistica Sinica*, Vol. 3, pp. 501–515, 1993.

[36] F. T. Wright. The asymptotic behavior of monotone regression estimates. *Annals of Statistics*, Vol. 9, pp. 443–448, 1981.

[37] S. Zhou, X. Shen, and D. A. Wolfe. Local asymptotics for regression splines and confidence regions. *Annals of Statistics*, Vol. 26, pp. 1760–1782, 1998.