Submitted to the Annals of Statistics

UNIFORM CONVERGENCE AND RATE ADAPTIVE ESTIMATION OF A CONVEX FUNCTION

By Xiao Wang and Jinglai Shen

Purdue University and University of Maryland Baltimore County

This paper addresses the problem of estimating a convex regression function under both the sup-norm risk and the pointwise risk using B-splines. The presence of the convex constraint complicates various issues in asymptotic analysis, particularly uniform convergence analysis. To overcome this difficulty, we establish the uniform Lipschitz property of optimal spline coefficients in the ℓ_∞-norm by exploiting piecewise linear and polyhedral theory. Based upon this property, it is shown that this estimator attains the optimal rate of convergence over the Hölder class under both the risks. In addition, we construct adaptive estimates under both the sup-norm risk and the pointwise risk. These estimates achieve a maximal risk within a constant factor of the minimax risk over the Hölder class.

1. Introduction. Consider the convex regression problem of the form

$$y_k = f(x_k) + \sigma\epsilon_k, \qquad k = 1, \dots, n, \tag{1.1}$$

where $f : [0,1] \to \mathbb{R}$ is a convex function, the ϵ_i are independent, standard normal errors, $x_i = i/n, i = 1, \dots, n$ are the design points. Let

$$\mathcal{C} = \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, \frac{f(y) - f(x)}{y - x} \le \frac{f(z) - f(y)}{z - y} \text{ if } x < y < z \right\}$$

be the collection of convex functions on $[0,1]$. The goal of this paper is to estimate $f \in \mathcal{C}$ and analyze the performance of the estimate under both the sup-norm risk and the pointwise risk.

The shape restricted inference finds a wide range of applications in numerous important fields, and receives fast growing interest in diverse areas. Examples include reliability (survival functions, hazard functions), medicine (dose-response curve), finance (option price and delivery price), and astronomy (mass functions). Much effort has focused on estimation of a monotone function via the least squares approach (i.e., Brunk's estimator) [1, 5, 28, 32]. For convex or concave regression, the least squares estimator was originally proposed in [15] and its asymptotic properties have been studied by

AMS 2000 subject classifications: Primary 62G08

Keywords and phrases: Adaptive estimate, B-splines, convex regression, minimax risk

[11, 13, 24]. However, the least squares estimators suffer several major deficiencies: (i) they lack smoothness; (ii) they have a non-normal asymptotic distribution [11, 43] with low convergence rates (e.g., of order $n^{1/3}$ for the Brunk's estimator) regardless of the smoothness of the true function; and (iii) they are inconsistent at boundary and have a systematic non-negligible asymptotic bias, as pointed out by [42].

Other estimation procedures have also been developed for shape restricted inference. For instance, Mammen and Thomas-Agnan [25] studied constrained smoothing splines, but their computation is highly complicated; see [36] for a related result via control theoretic splines. A two-step estimator was proposed in [3]: it isotonizes a derivative estimator and then obtains a convex one by integrating the monotone derivative. Meyer [27] developed an algorithm for cubic monotone estimation with an extension to convex constraints and other variants, e.g., increasing-concave constraints. A penalized monotone B-spline estimator was treated in [35]; its asymptotic behaviors were analyzed. Additional results include [12, 28, 30, 41], just to name a few. In spite of the above mentioned progress, many critical questions remain open in convex regression and its asymptotic analysis, especially those related to adaptive estimation over a function class. One of bottle-neck difficulties in adaptive asymptotic analysis is largely due to the lack of uniform convergence properties of an estimator when a shape constraint is imposed.

In this paper, we consider estimation of a convex function in the Hölder class. Let H_L^r denote the Hölder class

$$H_L^r := \Big\{ f : |f^{(\ell)}(x_1) - f^{(\ell)}(x_2)| \le L|x_1 - x_2|^{\gamma}, \quad \forall x_1, x_2 \in [0,1] \Big\},$$

where $\gamma = r - \ell \in (0,1]$. Let $\mathcal{C}_H(r,L) = \mathcal{C} \cap H_L^r$ be the collection of functions in both $\mathcal{C}$ and H_L^r. Since a convex function on $[0,1]$ must be Lipschitz continuous, i.e., $\gamma = 1$ and $\ell = 0$, we have $r \ge 1$ for any $f \in \mathcal{C}(r,L)$. It is well known that, for a fixed r, there exists an estimator, depending on r, which achieves the optimal rate of convergence in H_L^r [38]. For example, the minimax sup-norm risk on H_L^r has an asymptotic order given by

$$\inf_{\hat f} \sup_{f \in H_L^r} \mathbb{E}\big\{ \|\hat f - f\|_\infty \big\} \asymp L^{\frac{1}{2r+1}} \sigma^{\frac{2r}{2r+1}} \Big(\frac{\log n}{n} \Big)^{\frac{r}{2r+1}}, \tag{1.2}$$

where $a \asymp b$ means that a/b is bounded by two positive constants from below and above. However, the existence of an adaptive estimator (independent of r) that achieves the convergence rate in (1.2) uniformly over r is more subtle. When the sup-norm risk is considered, a series of papers, e.g., [2, 9, 18, 22], have shown that the kernel estimator can be used to construct such an adaptive estimator. On the other hand, when the pointwise risk is considered, a full adaptive procedure achieving (1.2) does not exist and a logarithmic penalty term must occur [4, 19]. Specifically, for any $x_0 \in (0,1)$, there exists a positive constant π_1 such that

$$\inf_{\hat{f}} \sup_{f \in H_L^r} \mathbb{E}\{(\hat{f}(x_0) - f(x_0))^2\} \geq \pi_1 L^{\frac{2}{2r+1}} \sigma^{\frac{4r}{2r+1}} \left(\frac{\log n}{n}\right)^{\frac{2r}{2r+1}}. \tag{1.3}$$

Other approaches for pointwise adaptive estimation are reported in [21, 39], where a similar phenomenon occurs. For general discussions of adaptive methods for *unconstrained* functions, see [29, 40] and the references therein.

When a shape constraint is imposed, it was firstly noted in [17] that it does not improve the optimal rate of convergence. Further, it was found in [23] that the extra difference order constraint completely changes the adaptive estimation problem. In particular, Low and Kang [23] proposed a pointwise rate adaptive procedure for monotone estimation in the minimax sense with respect to a Lipschitz parameter. Unfortunately, when this procedure is applied to an interval of fixed points, it does not yield a monotone function as an estimate. Another related adaptive procedure for monotone estimation is given in [6], which studied the least squares estimator and showed that the attained rate of the probabilistic error is uniformly over a shrinking L_2-neighborhood of the true function.

The present paper proposes a B-spline estimator with an arbitrary spline degree for convex regression. The convex shape constraint of an estimator is converted into the similar constraint on spline coefficients. In addition to its conceptual simplicity and numerical efficiency, the obtained B-spline estimator is globally convex, smooth by choosing a suitable spline degree, and attains boundary consistency (as well as at the interior) by selecting a proper number of spline bases. The major part of the paper is devoted to adaptive asymptotic analysis of the B-spline estimator on $\mathcal{C}_H(r, L)$ under both the sup-norm and pointwise risks. Toward this end, it is essential and critical to establish certain uniform convergence properties of the B-spline estimator. However, challenging issues arise due to the presence of constraints. For example, the closed form of optimal spline solutions does not exist in general. Instead, they are characterized by complementarity conditions [10, 34] that give rise to a nonsmooth piecewise linear function of observation data. Due to nonsmooth and combinatorial nature of complementarity problems, a thorough understanding of complementarity conditions and the associated piecewise linear function is far from trivial. In this paper, we exploit optimization techniques, along with adaptive asymptotic statistical tools, to tackle these problems. The major contributions of the paper are:

1. As a key technical contribution of the paper, we establish the uniform Lipschitz property of optimal spline coefficients with respect to the ℓ_∞-norm via piecewise linear and polyhedral theory (cf. Theorem 3.1). Unlike the conventional and generic Lipschitz property in the ℓ_2-norm (which is trivial to show), the attained Lipschitz property in the ℓ_∞-norm requires a nontrivial argument that takes full advantage of the convex shape constraint. It yields a uniform sup-norm bound on variations of spline coefficients regardless of the number of spline bases, leading to more precise and less conservative error estimates in uniform convergence analysis. This property paves the way for asymptotic analysis (e.g., cf. Propositions 7.1–7.3) and construction of adaptive procedures.

2. By exploring the uniform Lipschitz property, we obtain the following results in adaptive asymptotic analysis:

(2.1) For a fixed order r, we show that the proposed B-spline estimator achieves an optimal minimax rate of convergence on $\mathcal{C}_H(r, L)$ under both the sup-norm and pointwise risks (cf. Section 3.1). This result gives rise to an optimal choice of the number of spline bases.

(2.2) Adaptive estimators are constructed under both the sup-norm and pointwise risks over $\mathcal{C}_H(r, L)$ with $r \in [1, 2]$. These estimates achieve a maximum risk within a constant factor of the minimax risk over the Hölder class (cf. Section 3.2). In particular, the pointwise adaptive estimator attains convexity on the interval $[0, 1]$ as well as the minimax risk over an entire range of values of $r \in [1, 2]$ and L.

(2.3) A brief discussion on variance estimation is given in Section 3.3.

The paper is organized as follows. Section 2 formulates the B-spline convex estimator and develops optimality conditions for spline coefficients. The main results of the paper are presented in Section 3, including the uniform Lipschitz property and its implications in adaptive asymptotic analysis. Potential extensions and future research directions are discussed in Section 4. The technical proofs of the main results are given in Sections 5–10.

2. Formulation and Optimality Conditions. Denote the pth degree B-spline basis with knots $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{K_n} = 1$ by $\{B_k^{[p]} : k = 1, \ldots, K_n + p\}$. For simplicity, we consider equally spaced knots, namely, $\kappa_1 = 1/K_n, \kappa_2 = 2/K_n, \ldots, \kappa_{K_n} = 1$. The value of K_n will depend upon n as discussed below. Assume that n/K_n is an integer denoted by M_n. We consider the following convex spline estimator:

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

where the spline coefficients $\hat{b} = \{\hat{b}_k, k = 1, \ldots, K_n + p\}$ minimize

$$\sum_{i=1}^{n} \Big(y_i - \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(x_i) \Big)^2 \tag{2.1}$$

subject to the convex constraint $\Delta^2 b \geq 0$, where Δ is the backward difference operator such that $\Delta b_k = b_k - b_{k-1}$ and $\Delta^2 = \Delta\Delta$.

Let the $n \times (K_n + p)$ design matrix $X = \big[B_k^{[p]}(x_j)\big]_{j,k}$ and denote $\beta_n = \sum_{i=1}^{n} \big(B_k^{[p]}(x_i)\big)^2$ for $k = p+1, \ldots, K_n$. Given a spline degree p, $\big(\beta_n \frac{K_n}{n}\big)$ converges to a positive constant (depending on p only) as $(n/K_n) \to \infty$. Thus there exists a positive constant $C_{\beta,p}$ (depending on p only) such that

$$\beta_n \geq C_{\beta,p} \cdot \frac{n}{K_n}, \qquad \forall\ n, K_n. \tag{2.2}$$

Define the positive definite matrix $\Lambda_p := X^T X/\beta_n \in \mathbb{R}^{(K_n+p)\times(K_n+p)}$ and $\bar{y} := X^T y/\beta_n$, where $y = (y_1, \ldots, y_n)^T$ (we drop the subscript p in Λ_p for notational simplicity). It is easy to verify that for a given spline degree p, Λ is a $(2p+1)$-banded matrix. For instance, when $p = 1$, Λ is tridiagonal. The convex constraint on spline coefficients is characterized by the following polyhedral cone

$$\Omega := \Big\{ b \in \mathbb{R}^{K_n+p} : b_k - 2b_{k+1} + b_{k+2} \geq 0,\ k = 1, \ldots, K_n + p - 2 \Big\}.$$

When the knots are equally spaced, it is easy to verify that if the B-spline coefficient vector $\hat{b}$ is in Ω, then $\hat{f}^{[p]}$ is a convex function. Formulating (2.1) in matrix notation, the underlying optimization problem becomes the following equivalent constrained quadratic program

$$\hat{b} = \arg\min_{b\in\Omega} \frac{1}{2} b^T \Lambda\, b - b^T \bar{y}. \tag{2.3}$$

We first give the characterization of optimality conditions for $\hat{b}$. The conditions are represented by complementarity conditions, which plays a crucial role in addressing analytic and statistical properties of the estimator. We provide a short introduction of the complementarity condition. Two vectors $u = (u_1, \ldots, u_d)^T$ and $v = (v_1, \ldots, v_d)^T$ in $\mathbb{R}^d$ are said to satisfy the *complementarity condition* [7] if $u_i \geq 0$, $v_i \geq 0$, and $u_i\, v_i = 0$ for all $i = 1, \cdots, d$. This condition can be put in a more compact vector form: $0 \leq u \perp v \geq 0$, where $u \perp v$ means that the two vectors are orthogonal, i.e., $u^T v = 0$.

We introduce additional notation. Let

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ & \cdots & & & \cdots & & \\ 1 & 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \in \mathbb{R}^{(K_n+p)\times(K_n+p)},$$

and let $D_2 \in \mathbb{R}^{(K_n+p-2)\times(K_n+p)}$ be the 2nd-order difference matrix such that $D_2 b = [\Delta^2(b_3), \cdots, \Delta^2(b_{K_n+p})]^T$; see (5.3) for the explicit form of D_2.

THEOREM 2.1. *The necessary and sufficient conditions for $\hat{b} \in \Omega$ to minimize (2.3) are*

$$0 \le D_2 \hat{b} \ \perp\ C_{\gamma\bullet}\, C\big(\Lambda \hat{b} - \bar{y}\big) \ge 0, \tag{2.4}$$

$$C_{(K_n+p)\bullet}\,\big(\Lambda \hat{b} - \bar{y}\big) = C_{(K_n+p)\bullet}\, C\big(\Lambda \hat{b} - \bar{y}\big) = 0, \tag{2.5}$$

where the index set $\gamma := \{1, \ldots, K_n + p - 2\}$, and $C_{d\bullet}$ denotes the dth row of C.

2.1. *Piecewise Linear Formulation of Optimal Spline Coefficients.* It follows from Theorem 2.1 that $\hat{b}(\bar{y})$ is characterized by the mixed complementarity conditions. It is known from complementarity and polyhedral theory that $\hat{b}(\bar{y})$ is a continuous piecewise linear function of $\bar{y}$ determined by an index set $\alpha = \{\, i \,|\, (D_2\hat{b})_i = 0\} \subseteq \{1, \ldots, K_n + p - 2\}$ (α may be empty). Indeed, $\hat{b}$ has $2^{(K_n+p-2)}$ linear selection functions, each of which is denoted by $\hat{b}^\alpha$ corresponding to the index set α. Hence, the solution mapping $\bar{y} \mapsto \hat{b}$ is a (continuous) piecewise linear function with $2^{(K_n+p-2)}$ selection functions. The following proposition characterizes each linear selection function; its construction and proof is given in Section 5.2.

PROPOSITION 2.1. *For each index set $\alpha \subseteq \{1, \ldots, K_n + p - 2\}$, let $\ell := K_n + p - |\alpha|$. Then there exists a row independent matrix $F_\alpha \in \mathbb{R}^{\ell\times(K_n+p)}$ such that the linear selection function $\hat{b}^\alpha$ is given by*

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

In view of the above proposition and its construction (cf. Section 5.2), a linear selection function corresponds to an index set α depending on $\bar{y}$ (or y

by somewhat abusing notation). Consequently, the piecewise linear function $\hat{b}$ can be written as

$$\hat{b}(\bar{y}) = F^T_{\alpha(y)}\big(F_{\alpha(y)}\Lambda F^T_{\alpha(y)}\big)^{-1}F_{\alpha(y)}\bar{y}.$$

Let $N(x) := \big[B^{[p]}_1(x), \ldots, B^{[p]}_{K_n+p}(x)\big]^T$. For a given y, the convex B-spline estimator becomes

(2.6)
$$\hat{f}^{[p]}(x) = N^T(x)\hat{b}(\bar{y}) = \frac{1}{n}\sum_{i=1}^{n} N^T(x)F^T_{\alpha(y)}\Big(F_{\alpha(y)}\frac{X^TX}{n}F^T_{\alpha(y)}\Big)^{-1}F_{\alpha(y)}N(x_i)y_i.$$

Denote the weight function in (2.6) by $K_\alpha(s,t)$, i.e.,

$$K_{\alpha(y)}(s,t) = N^T(s)F^T_{\alpha(y)}\Big(F_{\alpha(y)}\frac{X^T(t)X(t)}{n}F^T_{\alpha(y)}\Big)^{-1}F_{\alpha(y)}N(t).$$

Hence, the convex spline estimator is a kernel estimator. However, the kernel depends on the index set α, which in turn relies on the observation y. Therefore, the estimator is *not* a linear but a piecewise linear function in y.

3. Main Results. In this section, we exploit the piecewise linear formulation of $\hat{b}$ to establish the uniform Lipschitz property of $\hat{b}$ in the ℓ_∞-norm. Roughly speaking, this property says that $\hat{b}(\bar{y})$ is a Lipschitz function of $\bar{y}$ with a uniform Lipschitz constant (with respect to the ℓ_∞-norm), regardless of K_n and α. This property is critical in establishing uniform consistency and developing adaptive estimators. Formally, this property is stated in the following theorem whose proof is given in Section 6.

THEOREM 3.1. *Given a spline degree p. There exists a positive constant $c_{\infty,p}$ (dependent on p only) such that*

(1) for any K_n and any index set α,

$$\big\|F^T_\alpha(F_\alpha\Lambda F^T_\alpha)^{-1}F_\alpha\big\|_\infty \le c_{\infty,p};$$

(2) for any K_n,

$$\big\|\hat{b}(u) - \hat{b}(v)\big\|_\infty \le c_{\infty,p}\|u-v\|_\infty, \quad \forall\, u, v \in \mathbb{R}^{K_n+p}.$$

In the next, we apply the uniform Lipschitz property to derive optimal rates of convergence in Section 3.1, construct adaptive estimators under both sup-norm risk and pointwise risk in Section 3.2, and study variance estimation in Section 3.3.

3.1. *Optimal Rate of Convergence.* For any $f \in \mathcal{C}_H(r, L)$ with $r > 1$, we write $\hat{f}_{(r)} := \hat{f}^{[p]}$ when using the spline degree $p = \lceil r - 1 \rceil$ to fit the data. If $r = 1$, then $\hat{f}_{(r)} := \hat{f}^{[p]}$ with $p = 1$, namely, $\hat{f}_{(r)}$ is a piecewise linear spline.

THEOREM 3.2. *Assume $f \in \mathcal{C}_H(r, L)$. Then,*

(1) If K_n is chosen as

$$K_n = \left(\frac{L}{\sigma}\right)^{\frac{2}{2r+1}} \left(\frac{n}{\log n}\right)^{\frac{1}{2r+1}},$$

then there exists a positive constant $\tilde{C}_{1r}$ dependent only on r such that

$$\sup_{f \in \mathcal{C}_H(r,L)} \mathbb{E}\Big(\|\hat{f}_{(r)} - f\|_\infty\Big) \leq \tilde{C}_{1r} L^{\frac{1}{2r+1}} \sigma^{\frac{2r}{2r+1}} \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}. \tag{3.1}$$

(2) For any $x_0 \in [0, 1]$, if K_n is chosen as

$$K_n = \left(\frac{L}{\sigma}\right)^{\frac{2}{2r+1}} n^{\frac{1}{2r+1}},$$

then there exists a positive constant $\tilde{C}_{2r}$ dependent only on r such that

$$\sup_{f \in \mathcal{C}_H(r,L)} \mathbb{E}\Big(|\hat{f}_{(r)}(x_0) - f(x_0)|^2\Big) \leq \tilde{C}_{2r} L^{\frac{2}{2r+1}} \sigma^{\frac{4r}{2r+1}} n^{\frac{2r}{2r+1}}. \tag{3.2}$$

It is known that the maximum likelihood estimate of a convex function is inconsistent at the boundary, which is called the *spiking problem* [42]. In contrast, Theorem 3.2 shows that $\hat{f}_{(r)}$ is uniformly consistent on $[0, 1]$. The optimal choice of K_n is of order $(n/\log n)^{\frac{1}{(2r+1)}}$ and $\|\hat{f}_{(r)} - f\|_\infty$ achieves the optimal rate of convergence, which is of order $(n/\log n)^{\frac{r}{(2r+1)}}$ [29]. Under the pointwise risk, the optimal choice of K_n is of order $n^{\frac{1}{(2r+1)}}$ and the estimator thus achieves the optimal rate of convergence, which is of order $n^{\frac{2r}{(2r+1)}}$ [38].

The next result shows that, for any $f \in \mathcal{C}_H(r, L)$ with $r > 2$, the constrained spline estimator and the unconstrained spline estimator coincide with probability tending to one, provided that $f''(x) > 0$ for all $x \in [0, 1]$.

THEOREM 3.3. *Assume $f \in \mathcal{C}_H(r, L)$ with $r > 2$ and $f''(x) \geq c > 0$ for all $x \in [0, 1]$. Let $\hat{f}^{uc}$ be the unconstrained regression spline estimator. If $n^{-1} K_n^5 \log n \to 0$ and $K_n \to \infty$ as $n \to \infty$, then,*

$$P\big(\hat{f}^{uc}(x) = \hat{f}(x),\ \forall\ x \in [0, 1]\big) \longrightarrow 1.$$

PROOF. Zhou et al. [45] studied the problem of estimating derivatives of a regression function using the corresponding derivatives of regression splines without shape constraint. For any $x \in [\kappa_k, \kappa_{k+1}]$, $k = 0, \dots, K_n - 1$, if $\ell \geq 3$,

$$\mathbb{E}\Big(\frac{d^2}{dx^2}\hat{f}^{uc}(x)\Big) - f''(x) = b(x) + o(K_n^{-r+2}),$$

where

$$b(x) = \frac{K_n^{-(\ell-2)}}{(\ell-1)!}\Big(f^{(\ell)}(\kappa_{k+1}) - f^{(\ell)}(\kappa_k)\Big)B_{\ell-1}\big(K_n(x-\kappa_k)\big)$$

is of order $O(K_n^{-r+2})$, and $B_m(\cdot)$ is the mth Bernoulli polynomial inductively defined as follows:

$$B_0(x) = 1, \qquad B_i(x) = \int_0^x iB_{i-1}(z)dz + b_i,$$

where $b_i = -i\int_0^1\int_0^x B_{i-1}(z)dzdx$ is the ith Bernoulli number. The variance of $\frac{d^2}{dx^2}\hat{f}^{uc}(x)$ is of order $n^{-1}K_n^5$ (cf. [45, Lemma 5.4]). Similar to Lemma 8.1 given in Section 8, it can be shown that

$$\Big\|\frac{d^2}{dx^2}\hat{f}^{uc}(x) - \frac{d^2}{dx^2}\mathbb{E}(\hat{f}^{uc}(x))\Big\|_\infty = O_p\Big(\sqrt{n^{-1}K_n^5\log n}\Big).$$

Therefore, if $n^{-1}K_n^5\log n \to 0$ and $K_n \to \infty$ as $n \to \infty$, $\|\frac{d^2}{dx^2}\hat{f}^{uc} - f''\|_\infty = o_p(1)$. Hence, the unconstrained and constrained estimators are asymptotically equivalent. $\square$

3.2. *Adaptive Estimation.* In this section, we construct adaptive estimators, with respect to both the sup-norm risk and thepointwise risk . These estimates have maximum risks within a constant factor of the minimax risk over $\mathcal{C}_H(r,L)$. We will focus on the function class $\mathcal{C}_H(r,L)$ with $1 \leq r \leq 2$, where the differences between the constrained and unconstrained estimate do no vanish asymptotically.

3.2.1. *Adaptive Estimation under Sup-norm Risks.* It follows from Propositions 7.1 and 7.2 that, for any $r \in [1,2]$, the bounds for the bias and the stochastic term, respectively, are

$$\sup_{f\in\mathcal{C}_H(r,L)} \|\bar{f}_{(r)} - f\|_\infty \leq C_1LK_n^{-r}, \tag{3.3}$$

$$P\Big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \geq u\Big) \leq (K_n+1)\exp\Big\{-\frac{n}{2K_nC_2^2\sigma^2}u^2\Big\}, \tag{3.4}$$

where C_1 and C_2 are two positive constants independent of $r \in [1,2]$. Hence the optimal number of knots is

$$K_{(r)} = \left(\frac{C_2}{C_1 r \sqrt{2(2r+1)}} \right)^{-\frac{2}{2r+1}} \left(\frac{\sigma}{L}\right)^{\frac{-2}{2r+1}} \left(\frac{\log n}{n}\right)^{\frac{-1}{2r+1}}, \tag{3.5}$$

and the optimal rate of convergence is

$$\begin{aligned} \psi_{(r)} &= C_1 L K_{(r)}^{-r} + \sqrt{\frac{2}{2r+1}}\, C_2 \sigma \sqrt{\frac{K_{(r)} \log n}{n}} \\ &= 2 C_1^{\frac{1}{2r+1}} \left(\frac{C_2}{r\sqrt{2(2r+1)}} \right)^{\frac{2r}{2r+1}} L^{\frac{1}{2r+1}} \sigma^{\frac{2r}{2r+1}} \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}. \end{aligned} \tag{3.6}$$

Given n, let $\tau_n := \lceil (\log n)^{1/2} \rceil$, and $r_j := 1 + j/\tau_n$, $j = 0, 1, \ldots, \tau_n$ be the elements in $[1,2]$. We consider the adaptive estimator using the idea of Lepski [20]. Let

$$\hat{k} = \sup \left\{ 0 \le k \le \tau_n : \|\hat{f}_{(r_k)} - \hat{f}_{(r_j)}\|_\infty \le \frac{1+\sqrt{2}}{2}\, \psi_{(r_j)}, \quad \text{for any} \quad j \le k \right\}.$$

Define $\hat{r} := r_{\hat{k}}$. We use $\hat{f}_{(\hat{r})}$ for estimation in the sup-norm distance.

THEOREM 3.4. *The estimator $\hat{f}_{(\hat{r})}$ is a rate adaptive estimator on $\mathcal{C}_H(r,L)$ for the sup-norm distance, i.e., there exists a positive constant π_2 such that*

$$\sup_{r\in[1,2]} \sup_{f\in\mathcal{C}_H(r,L)} \mathbb{E}\big\{ \|\hat{f}_{(\hat{r})} - f\|_\infty \big\} \le \pi_2\, L^{\frac{1}{2r+1}} \sigma^{\frac{2r}{2r+1}} \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}.$$

3.2.2. *Pointwise Adaptive Estimation.* In this section, we construct an estimator which attains the minimax rate of convergence for a whole range of values of $r \in [1,2]$ and L. In the context of convex regression, unlike the earlier work on pointwise adaptive estimation, a fully adaptive procedure can be obtained.

We explore the idea of Low and Kang [23] to construct an adaptive estimate of $f(x_0)$ for any given $x_0 \in (0,1)$. Given the observation data $(y_i)_{i=1}^n$, let

$$\bar{y}_k := \frac{\sum_{i=1}^n y_i I(\kappa_{k-1} < x_i \le \kappa_k)}{\sum_{i=1}^n I(\kappa_{k-1} < x_i \le \kappa_k)}, \quad k = 1, \ldots, K_n. \tag{3.7}$$

Then $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_{K_n})^T$ minimizes $\sum_{k=1}^{K_n} (b_k - \bar{y}_k)^2$ subject to the convex constraint $\Delta^2 b_k \ge 0$, $k = 3, \ldots, K_n$. This indicates that a piecewise constant spline with $p = 0$ is used to fit the data in (2.1). Recall that $M_n = n/K_n$. Let

$$\zeta_k := \frac{1}{n}\Big[(k-1)M_n + \frac{M_n+1}{2}\Big].$$

Hence, ζ_k is the average of the design points on $(\kappa_{k-1}, \kappa_k]$. Let $\widetilde{f}$ denote the piecewise linear function which interpolates $(\zeta_k, \hat{b}_k)$, $k = 1, \dots, K_n$.

Fix $x_0 \in (0, 1)$. For each n, let $d_n \in \mathbb{N}$ satisfy $\zeta_{d_n} < x_0 \leq \zeta_{d_n+1}$. Let $K_{n,j} := 2^j n^{1/5}$, where $K_{n,j}$ depends on j. Further, we let $\bar{y}_{k,j}$ denote the $\bar{y}_k$ defined in (3.7) corresponding to a given $K_{n,j}$, and let $\widetilde{f}_j$ be the estimator $\widetilde{f}$ corresponding to $K_{n,j}$. Fix a real number $\lambda > 0$ such that $P(Z > \lambda) < 1/4$, where Z is a standard normal random variable. Set

$$I_j := I\Big(\Delta\bar{y}_{d_n+4,j} - \Delta\bar{y}_{d_n-2,j} \leq \lambda 2^{\frac{j}{2}+1} n^{-\frac{2}{5}}\sigma\Big) \prod_{i=0}^{j-1} I\Big(\Delta\bar{y}_{d_n+4,i} - \Delta\bar{y}_{d_n-2,i} > \lambda 2^{\frac{i}{2}+1} n^{-\frac{2}{5}}\sigma\Big).$$

Note that exactly one $I_j \neq 0$ and thus the collection $\{I_j\}$ provides a selection procedure for $K_{n,j}$. The adaptive estimator is given by

$$\widetilde{f}(x_0) = \sum_{j=1}^{\infty} \widetilde{f}_j(x_0) I_j. \tag{3.8}$$

THEOREM 3.5. *The estimator in (3.8) is a rate adaptive estimator under the pointwise risk, i.e., for any $x_0 \in (0, 1)$, there exists a positive constant π_3 such that*

$$\sup_{r\in[1,2]} \sup_{f\in\mathcal{C}_H(r,L)} \mathbb{E}|\widetilde{f}(x_0) - f(x_0)|^2 \leq \pi_3 L^{\frac{2}{(2r+1)}} \sigma^{\frac{4r}{(2r+1)}} n^{\frac{-2r}{(2r+1)}}. \tag{3.9}$$

3.3. *Variance Estimation.* In practice, the variance σ^2 is replaced by the estimated variance $\hat{\sigma}^2$ in the above adaptive procedures. We will briefly study the asymptotic properties of the maximum likelihood estimator of σ^2. Given the observation data $y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$ at design points $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, let $\hat{f}_y := \big(\hat{f}^{[p]}(x_1), \dots, \hat{f}^{[p]}(x_n)\big)^T$ with $p = \lceil r-1 \rceil$ and $\vec{f} := \big(f(x_1), \dots, f(x_n)\big)^T$. Let $\alpha(y)$ be an index set corresponding to the optimal coefficient $\hat{b}(X^T y/\beta_n)$ defined in Section 2.1. Then for fixed K_n and p, we have $\hat{f}_y = A_{\alpha(y)} y$, where

$$A_{\alpha(y)} = X F_{\alpha(y)}^T \Big(F_{\alpha(y)} X^T X F_{\alpha(y)}^T\Big)^{-1} F_{\alpha(y)} X^T \in \mathbb{R}^{n\times n}. \tag{3.10}$$

It follows from the similar discussion as in Section 2.1 that $\hat{f}_y : \mathbb{R}^n \to \mathbb{R}^n$ is a continuous piecewise linear function, where each linear selection function is defined by $A_{\alpha(y)}$. The MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\big\|y - A_{\alpha(y)}y\big\|_2^2.$$

THEOREM 3.6. *Assume $f \in \mathcal{C}(r, L)$, $K_n \to \infty$ as $n \to \infty$, and let $p = \lceil r-1 \rceil$. If $K_n = o(n)$, then $\hat{\sigma}^2 \to \sigma^2$ in probability, and if $K_n = o(\sqrt{n})$, then $\sqrt{n}\big(\hat{\sigma}^2 - \sigma^2\big)$ is asymptotically normal with mean zero and variance $2\sigma^4$. Furthermore, if K_n is of order $n^{\frac{1}{2r+1}}$, then $|\mathbb{E}(\hat{\sigma}^2 - \sigma^2)| = O\big(n^{\frac{-2r}{2r+1}}\big)$.*

Variance estimation based on the differences of successive points has been studied originally by Rice [31]. Compared to this estimation and other generalizations, the MLE $\hat{\sigma}^2$ has a smaller asymptotic variance but a slightly larger bias. Meyer and Woodroofe [26] studied the bias reduction variance estimator for a monotone regression model using the least squares method. The bias reduction variance estimator for a convex spline model is nontrivial and shall be addressed in a future paper.

4. Discussion.

We have considered the B-spline estimators for convex regression in this paper. The proposed estimator and asymptotic analysis techniques can be extended to other shape restricted inference problems. For example, it is known that the uniform Lipschitz property (in the ℓ_∞-norm) holds for the monotone constraint [35, 41]. Therefore the minimax optimal convergence rates and adaptive estimators can be established in a similar manner. It is conjectured that the uniform Lipschitz property holds for a higher order difference constraint. However, its development is much more involved and shall be reported in the future.

We have provided the optimal rate adaptive estimators for convex regression under both the sup-norm risk and the pointwise risk. Nonetheless, the question of explicit construction of asymptotically exact adaptive estimation for convex regression over the Hölder classes remains open.

5. Proofs for Section 2.

5.1. *Proof of Theorem 2.1.*

PROOF OF THEOREM 2.1. Write (2.3) as $\min_{b\in\Omega} g(b)$, where the objective function $g(b) := \frac{1}{2}b^T\Lambda b - b^T\bar{y}$. It is clear that g is coercive on $\mathbb{R}^{K_n+p}$ and strictly convex on the closed convex set Ω. This ensures the existence and uniqueness of an optimal solution. Furthermore, since Ω is a polyhedral cone, it is finitely generated by $\{v^1, -v^1, v^2, -v^2, v^3, v^4, \ldots, v^{K_n+p}\}$. Here, for each $k = 3, \ldots, K_n + p$,

$$v^k = \Big(\underbrace{0, \ldots, 0,}_{(k-1)-\text{copies}} \ v^k_k, \ldots, v^k_{K_n+p} \Big)^T = \Big(\underbrace{0, \ldots, 0,}_{(k-1)-\text{copies}} \ 1, 2, \ldots, K_n{+}p{-}k{+}1 \Big)^T,$$

and for $k = 1, 2$,

(5.1)

$$v^1 = \Big(1,\, 0,\, -1,\, -2, \ldots, -(K_n{+}p{-}2)\Big)^T,\ v^2 = \Big(0,\, 1,\, 2,\, 3, \ldots, K_n{+}p{-}1\Big)^T.$$

It is easy to see that $\Delta^2 v^k_j = 0$ for $k = 1, 2$ and all $j > 2$. Hence $\pm v^k \in \Omega$ for $k = 1, 2$, and it can be also verified that $\sum_{k=1}^2 v^k = \mathbf{1}$. Further, any $b = (b_1, \ldots, b_{K_n+p})^T \in \Omega$ can be positively generated as

$$b = \sum_{i=1}^{2} \Big(\max(0, b_i) v^i + \max(0, -b_i)(-v^i) \Big) + \sum_{i=3}^{K_n+p} \Delta^2(b_i) v^i.$$

By using these generators for Ω, we obtain the following necessary and sufficient optimality conditions for an optimizer $\hat{b}$:

$$0 \le D_2 \hat{b} \perp \widetilde{C} \nabla g(\hat{b}) \ge 0 \quad \text{and} \quad \langle v^k, \nabla g(\hat{b}) \rangle = 0, \ \ \forall\, k = 1, 2, \tag{5.2}$$

where $D_2 \in \mathbb{R}^{(K_n-2+p)\times(K_n+p)}$ is given by

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 & 0 \\ & \cdots & & & \cdots & & \cdots & \cdots & \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}, \tag{5.3}$$

and $\widetilde{C} \in \mathbb{R}^{(K_n-2+p)\times(K_n+p)}$ is given by

$$\widetilde{C} = \big[\, v^3 \quad \cdots \quad v^{K_n+p} \,\big]^T$$
$$= \begin{bmatrix} 0 & 0 & 1 & 2 & \cdots & (K_n+p-4) & (K_n+p-3) & (K_n+p-2) \\ 0 & 0 & 0 & 1 & \cdots & \cdots & (K_n+p-4) & (K_n+p-3) \\ & \cdots & & & \cdots & & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

It can be shown via the definitions of v^1 and v^2 in (5.1) that the second optimality condition in (5.2) can be equivalently written as

$$\sum_{i=1}^{K_n+p} \big(\nabla g(\hat{b})\big)_i = 0 \quad \text{and} \quad \sum_{i=1}^{K_n+p} (K_n + p - i + 1)\big(\nabla g(\hat{b})\big)_i = 0,$$

where $\nabla g(b) = \Lambda b - \bar{y}$. This gives rise to the two boundary conditions. Moreover, noting that for any k, the definitions of v^1 and v^2 in (5.1) yield

$$\widetilde{C}_{k\bullet} \nabla g(\hat{b}) = \sum_{i=1}^{k} \sum_{j=1}^{i} \left(\nabla g(\hat{b}) \right)_j = (C^2)_{k\bullet} \nabla g(\hat{b}),$$

we obtain the equivalent condition for the first optimality condition in (5.2):

$$0 \,\leq\, D_2 \, \hat{b} \,\perp\, \left(C^2\right)_{\gamma\bullet} \nabla g(\hat{b}) \,\geq\, 0, \tag{5.4}$$

where $\gamma = \{1, \ldots, K_n + p - 2\}$. By $(C^2)_{\gamma\bullet} = C_{\gamma\bullet} C$, the proof is complete. □

5.2. *Construction and Proof for Proposition 2.1.* We first construct certain equations that yield a linear selection function corresponding to the index set index set $\alpha = \{\, i \,|\, (D_2 \hat{b})_i = 0\} \subseteq \{1, \ldots, K_n + p - 2\}$ (α may be empty). Specifically, for the given $\hat{b}$ and α, we define a vector $\widetilde{b}^{\alpha}$ and an associated family of index sets $\{\beta_i^{\alpha}\}$ in the following steps:

(1) let $\ell_1 := \min_{3 \leq i \leq K_n + p} \{ i \,:\, \Delta^2(\hat{b}_i) = 0 \}$, and $\overline{\ell}_1 := \max_{\ell_1 \leq k \leq K_n + p} \{ k \,:\, \Delta^2(\hat{b}_i) = 0, \ \forall i = \ell_1, \ldots, k \}$. Then inductively define, for $j \geq 1$,

$$\ell_{j+1} := \min_{1 + \overline{\ell}_j \leq i \leq K_n} \{ i \,:\, \Delta^2(\hat{b}_i) = 0 \},$$
$$\overline{\ell}_{j+1} := \max_{\ell_{j+1} \leq k \leq K_n} \{ k \,:\, \Delta^2(\hat{b}_i) = 0, \quad \forall i = \ell_{j+1}, \ldots, k \}.$$

Suppose that we obtain q's such $\ell_i, \overline{\ell}_i$, namely, $\ell_1, \ldots, \ell_q$ and $\overline{\ell}_1, \ldots, \overline{\ell}_q$. Define $\widehat{\beta}^{\alpha}_{\ell_j} := \{ i \,:\, \ell_j - 2 \leq i \leq \overline{\ell}_j \}$ for $j = 1, \ldots, q$. Note that $|\widehat{\beta}^{\alpha}_{\ell_j}| \geq 3$ for each ℓ_j, and for two consecutive index sets, $\ell_{j+1} \geq \overline{\ell}_j + 2$. Thus if the equality holds, then $\widehat{\beta}^{\alpha}_{\ell_j} \cap \widehat{\beta}^{\alpha}_{\ell_{j+1}} = \{ \overline{\ell}_j \}$; otherwise, the two consecutive index sets are disjoint.

(2) let $\widehat{L} := K_n + p + q - | \cup_{i=1}^{q} \widehat{\beta}^{\alpha}_{\ell_i} |$, where $| \cdot |$ denotes the cardinality of an index set. For each $i \in \{1, \ldots, K_n + p\} \setminus \cup_{i=1}^{q} \widehat{\beta}^{\alpha}_{\ell_j}$, define $\widehat{\beta}^{\alpha}_{\ell_s} = \{ i \}$, where $s = (q+1), \ldots, \widehat{L}$.

(3) this step arranges the index sets $\widehat{\beta}^{\alpha}_{\ell_j}$ in a monotone order as follows. For each $\widehat{\beta}^{\alpha}_{\ell_i}$, let $\min(\widehat{\beta}^{\alpha}_{\ell_i})$ denote the least element in $\widehat{\beta}^{\alpha}_{\ell_i}$ (the similar notation will be used for max below). Define $\ell_{s_1} := \arg\min_{\ell_1, \ldots, \ell_{\widehat{L}}} \{ \min(\widehat{\beta}^{\alpha}_{\ell_i}) \}$. Let $\widetilde{\beta}^{\alpha}_1 := \widehat{\beta}^{\alpha}_{\ell_{s_1}}$. Then inductively define for each $j \geq 1$, $\widetilde{\beta}^{\alpha}_{j+1} := \widehat{\beta}^{\alpha}_{\ell_{s_{j+1}}}$, where

$$\ell_{s_{j+1}} := \arg \min_{\{\ell_1, \ldots, \ell_{\widehat{L}}\} \setminus \{\ell_{s_1}, \ldots, \ell_{s_j}\}} \{ \min(\widehat{\beta}^{\alpha}_{\ell_i}) \}.$$

(4) in this step, we regroup the index sets $\widehat{\beta}^{\alpha}_{\ell_j}$ in a way that preserves desired structural properties to be used in the subsequent development. Define $p_0 := 0$ and

$$p_1 := \max\big(1, \quad \max\{k \geq 1 \,:\, \widetilde{\beta}^{\alpha}_j \cap \widetilde{\beta}^{\alpha}_{j+1} \neq \emptyset, \ \forall j = 1, \ldots, k-1\}\big),$$

and $\beta^{\alpha}_1 := \cup_{j=1}^{p_1} \widetilde{\beta}^{\alpha}_j$, the companion index set $\vartheta_1 := \{\min(\widetilde{\beta}^{\alpha}_j), \forall j = 1, \ldots, p_1\} \cup \{\max(\widetilde{\beta}^{\alpha}_{p_1})\}$. Recursively, define, for each $s \geq 1$,

$$p_{s+1} := \max\big(p_s+1, \quad \max\{k \geq p_s+1 \,:\, \widetilde{\beta}^{\alpha}_j \cap \widetilde{\beta}^{\alpha}_{j+1} \neq \emptyset, \ \forall j = p_s+1, \ldots, k-1\}\big),$$

and $\beta^{\alpha}_{s+1} := \cup_{j=p_s+1}^{p_{s+1}} \widetilde{\beta}^{\alpha}_j$, the companion index set $\vartheta_{s+1} := \{\min(\widetilde{\beta}^{\alpha}_j), \forall j = p_s+1, \ldots, p_{s+1}\} \cup \{\max(\widetilde{\beta}^{\alpha}_{p_{s+1}})\}$. Without loss of generality, we assume that the index elements of each ϑ_s are in the strictly increasing order. Hence, any two consecutive index sets in ϑ_s correspond to ℓ_j and $\overline{\ell}_j$ defined in Step (1) with $\ell_{j+1} = \overline{\ell}_j$.

(5) suppose that there are L such the index sets ϑ_s, and let $\vartheta := \cup_{s=1}^{L} \vartheta_s$ whose index elements are in the strictly increasing order. Then $\widetilde{\beta}^{\alpha} := (\overline{\beta}_i)$, where $i \in \vartheta$.

It is clear from the above construction that $\{\beta^{\alpha}_i\}$ forms a finite and disjoint partition of $\{1, \ldots, K_n + p\}$, namely, $\bigcup_{i=1}^{L} \beta^{\alpha}_i = \{1, \ldots, K_n + p\}$ and $\beta^{\alpha}_j \cap \beta^{\alpha}_k = \emptyset$ whenever $j \neq k$.

For a given index set α, we drop the sign restriction (i.e., the inequality $D_2 b \geq 0$) in (2.4) and obtain its corresponding linear selection function from the following (possibly redundant) equations:

$$
\begin{aligned}
&\text{(5.5a)} & (D_2\hat{b})_{\alpha} &= 0, \\
&\text{(5.5b)} & D_2\hat{b} &\perp C_{\gamma\bullet} C\big(\Lambda\hat{b} - \bar{y}\big), \\
&\text{(5.5c)} & C_{\overline{\alpha}\bullet} C\big(\Lambda\hat{b} - \bar{y}\big) &= 0, \\
&\text{(5.5d)} & C_{(K_n+p)\bullet}\left(\Lambda\hat{b} - \bar{y}\right) &= C_{(K_n+p)\bullet}\, C\left(\Lambda\hat{b} - \bar{y}\right) = 0,
\end{aligned}
$$

where $\overline{\alpha} := \{1, \ldots, K_n + p - 2\} \setminus \alpha$. Indeed, we shall use equations (5.5a), (5.5b), and (5.5d) to characterize a linear piece. Let $\widetilde{b}^{\alpha}$ denote the vector constituting the free variables of equation (5.5a). With this construction and notation, we are ready to prove Proposition 2.1 as follows.

Proof of Proposition 2.1. We introduce some notation first. Let $m^{\alpha}_i := |\beta^{\alpha}_i|$ and $h^{\alpha}_i := m^{\alpha}_i - 1$, where $i = 1, \ldots, L$. Note that if $m^{\alpha}_i > 1$, then

$m_i^\alpha \geq 3$ such that $h_i^\alpha \geq 2$ and $|\vartheta_i| \geq 2$. It follows from the definition of β_i^α that $\hat{b}^\alpha = (F_\alpha)^T \widetilde{b}^\alpha$, where the matrix

$$F_\alpha = \begin{bmatrix} F_{\alpha,1} & & & \\ & F_{\alpha,2} & & \\ & & \ddots & \\ & & & F_{\alpha,L} \end{bmatrix} \in \mathbb{R}^{\ell \times (K_n+p)} \tag{5.6}$$

and each matrix block corresponding to β_k^α is given as follows: if $m_k^\alpha = 1$, then $F_{\alpha,k} = 1$; otherwise, assuming that the index elements in ϑ_k in Step (5) above are in the strictly increasing order without loss of generality, and letting $h_{k,j}^\alpha := \vartheta_k(j+1) - \vartheta_k(j) \geq 2$ for each $j = 1, \dots, |\vartheta_k|-1$, we determine $F_{\alpha,k} \in \mathbb{R}^{|\vartheta_k| \times m_k^\alpha}$ from β_k^α constructed in Steps (1)-(5) as
(5.7)

$$F_{\alpha,k} = \begin{bmatrix} 1 & \mathbf{h}^{\alpha,k,1} & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & \widetilde{\mathbf{h}}^{\alpha,k,1} & 1 & \mathbf{h}^{\alpha,k,2} & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \widetilde{\mathbf{h}}^{\alpha,k,2} & 1 & \mathbf{h}^{\alpha,k,3} & \cdots & & & \\ & \cdots & \cdots & & & & & & & \\ & \cdots & \cdots & & & & & & & \\ & & & & & \cdots & \cdots & & & \\ & & & & & \cdots & \cdots & & & \\ & \cdots & & \cdots & & & & \cdots & \mathbf{h}_{\alpha,k,w_k} & 0 \\ & \cdots & & \cdots & & & & \cdots & \widetilde{\mathbf{h}}^{\alpha,k,w_k} & 1 \end{bmatrix},$$

where $w_k := |\vartheta_k| - 1$, and the row vectors

$$\begin{aligned} \mathbf{h}^{\alpha,k,j} &:= \begin{bmatrix} \frac{h_{k,j}^\alpha - 1}{h_{k,j}^\alpha} & \frac{h_{k,j}^\alpha - 2}{h_{k,j}^\alpha} & \cdots & \frac{1}{h_{k,j}^\alpha} \end{bmatrix}, \quad j = 1, \dots, w_k, \\ \widetilde{\mathbf{h}}^{\alpha,k,j} &:= \begin{bmatrix} \frac{1}{h_{k,j}^\alpha} & \frac{2}{h_{k,j}^\alpha} & \cdots & \frac{h_{k,j}^\alpha - 1}{h_{k,j}^\alpha} \end{bmatrix}, \quad j = 1, \dots, w_k. \end{aligned}$$

For notational simplicity, let $v := \Lambda \hat{b} - \bar{y}$. In view of the complementarity condition in (2.4), we have $(D_2 \hat{b})^T C_{\gamma\bullet} C v = 0$. Since $\hat{b} = (F_\alpha)^T \widetilde{b}^\alpha$, $(\widetilde{b}^\alpha)^T F_\alpha (D_2^T C_{\gamma\bullet} C v) = 0$. Moreover, it can be further verified that

$$D_2^T C_{\gamma\bullet} C = \begin{bmatrix} I_{K_n+p-2} & 0_{(K_n+p-2)\times 2} \\ E & 0_{2\times 2} \end{bmatrix} \in \mathbb{R}^{(K_n+p)\times(K_n+p)},$$

where

$$E = \begin{bmatrix} -(K_n+p-1) & -(K_n+p-2) & \cdots & \cdots & -2 \\ K_n+p-2 & K_n+p-3 & \cdots & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{2\times(K_n+p-2)}.$$

It also follows from the boundary conditions $C_{(K_n+p)\bullet}v = C_{(K_n+p)\bullet}Cv = 0$ and elementary row operations that $[-E \quad I_2]v = 0$. Therefore, we obtain $D_2^T C_{\gamma\bullet} Cv = I_{(K_n+p)}\ v = v$. Hence, $(\widetilde{b}^\alpha)^T F_\alpha (D_2^T C_{\gamma\bullet} Cv) = (\widetilde{b}^\alpha)^T F_\alpha v = 0$. Recall that for the given index set α, $\widetilde{b}^\alpha$ corresponds to the free variables of the equation (5.5a). Hence, $\widetilde{b}^\alpha$ is arbitrary such that $F_\alpha v = 0$. This leads to

$$F_\alpha\, \Lambda (F_\alpha)^T\, \widetilde{b}^{\,\alpha} = F_\alpha\, \bar{y}.$$

Letting $\widetilde{\Lambda}^\alpha = F_\alpha\, \Lambda (F_\alpha)^T$ and $\widetilde{y}^{\,\alpha} = F_\alpha\, \bar{y}$, we obtain the linear equation for $\widetilde{b}^{\,\alpha}$. Since F_α is of full row rank and Λ is positive definite, $\widetilde{\Lambda}^\alpha$ is positive definite and hence is invertible. Consequently, we have $\hat{b}^\alpha(\bar{y}) = F_\alpha^T \widetilde{b}^\alpha(\bar{y}) = F_\alpha^T \big(F_\alpha \Lambda F_\alpha^T\big)^{-1} F_\alpha \bar{y}$. □

6. Proof of Theorem 3.1. We divide the proof of Theorem 3.1 into several steps. We first establish a result pertaining to $F_\alpha F_\alpha^T$.

LEMMA 6.1. *For any K_n and α, $F_\alpha F_\alpha^T$ is a strictly diagonally dominant, nonnegative, tridiagonal matrix.*

PROOF. Recall that $\ell := K_n + p - |\alpha|$. For notational simplicity, let $G := F_\alpha F_\alpha^T$. First of all, it is easy to verify via (5.6) and (5.7) that G is the $\ell \times \ell$ tri-diagonal matrix given by

$$\begin{bmatrix} d_{11} & \widetilde{\eta}_1 & 0 & \cdots & \cdots & 0 \\ \widetilde{\eta}_1 & d_{22} & \widetilde{\eta}_2 & & & \\ & \widetilde{\eta}_2 & d_{33} & \widetilde{\eta}_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \widetilde{\eta}_{\ell-2} & d_{(\ell-1)(\ell-1)} & \widetilde{\eta}_{\ell-1} \\ 0 & \cdots & \cdots & 0 & \widetilde{\eta}_{\ell-1} & d_{\ell\ell} \end{bmatrix}.$$

The entries on the three diagonal bands are determined as follows. Consider F_α in (5.6) with L blocks. Fix $k \in \{1, \ldots, L\}$. If $m_k^\alpha = 1$, then $F_{\alpha,k} F_{\alpha,k}^T$ is a real number that appears on the diagonal of G. Denoting this number by d_{ss}, we have $d_{ss} = F_{\alpha,k} F_{\alpha,k}^T = 1$ and $G_{s(s+1)} = G_{(s+1)s} = 0$, $G_{sj} = 0$ for all $j \le s-2$ and $j \ge s+2$. If $m_k^\alpha > 1$, then $F_{\alpha,k} F_{\alpha,k}^T$ is a symmetric, positive definite matrix of order $|\vartheta_k|$ that forms a diagonal block of G. Making use of the structure of $F_{\alpha,k}$ given in the proof of Proposition 2.1 and somewhat lengthy computation, we obtain the following results in two separate cases (recalling $w_k := |\vartheta_k| - 1$):

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

$$\begin{aligned} d_{11} &= 1 + \frac{(h^{\alpha}_{1,1} - 1)(2h^{\alpha}_{1,1} - 1)}{6h^{\alpha}_{1,1}}, \\ \widetilde{\eta}_s &= G_{s(s+1)} = G_{(s+1)s} = \frac{(h^{\alpha}_{1,s})^2 - 1}{6h^{\alpha}_{1,s}}, \quad \forall\ s = 1, \ldots, w_1, \\ d_{ss} &= \frac{2(h^{\alpha}_{1,s-1})^2 + 1}{6h^{\alpha}_{1,s-1}} + \frac{2(h^{\alpha}_{1,s})^2 + 1}{6h^{\alpha}_{1,s}}, \qquad \forall\ s = 2, \ldots, w_1, \\ d_{(w_1+1)(w_1+1)} &= \frac{(h^{\alpha}_{1,w_1} + 1)(2h^{\alpha}_{1,w_1} + 1)}{6h^{\alpha}_{1,w_1}}. \end{aligned}$$

Besides, $G_{(w_1+1)(w_1+2)} = G_{(w_1+2)(w_1+1)} = 0$ and for each $s = 1, \ldots, w_t$, $G_{sj} = 0, \forall j \geq s + 2$ and $j \leq s - 2$. 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 $F_{\alpha,k}F^T_{\alpha,k}$, which is a diagonal entry of G, is denoted by d_{tt}. Then we have

$$\begin{aligned} d_{tt} &= 1 + \frac{(h^{\alpha}_{k,1} - 1)(2h^{\alpha}_{k,1} - 1)}{6h^{\alpha}_{k,1}}, \\ \widetilde{\eta}_{t+s} &= G_{(t+s)(t+s+1)} = \frac{(h^{\alpha}_{k,s})^2 - 1}{6h^{\alpha}_{k,s}}, \qquad \forall\ s = 1, \ldots, w_k, \\ d_{(t+s)(t+s)} &= \frac{2(h^{\alpha}_{k,s+1})^2 + 1}{6h^{\alpha}_{k,s+1}} + \frac{2(h^{\alpha}_{k,s})^2 + 1}{6h^{\alpha}_{k,s}}, \quad \forall\ s = 1, \ldots, w_k - 1, \\ d_{(t+w_k)(t+w_k)} &= \frac{(h^{\alpha}_{k,w_k} + 1)(2h^{\alpha}_{k,w_k} + 1)}{6h^{\alpha}_{k,w_k}}. \end{aligned}$$

In addition, for each $s = t, \ldots, t + w_k + 1$, $G_{sj} = 0$ for all $j \leq s - 2$ and $j \geq s + 2$, and $G_{t(t-1)} = G_{(t+w_k+1)(t+w_k+2)} = 0$.

Due to $G_{t(t-1)} = 0$ and the symmetry of G, we further deduce that if a diagonal entry $d_{tt} = G_{tt}$ with $t \geq 2$ corresponds to a scalar $F_{\alpha,k}F^T_{\alpha,k}$ (i.e., $m^{\alpha}_k = 1$), then $G_{(t-1)t} = 0$. (Recall that $G_{t(t+1)} = 0$ has been shown before.) Similarly, if d_{tt} is the first diagonal entry of $F_{\alpha,k}F^T_{\alpha,k}$, then $G_{(t-1)t} = 0$.

In the next, we show that G is strictly diagonally dominant. For a given $G \in \mathbb{R}^{\ell\times\ell}$, define

$$\xi_1 := d_{11} - |\widetilde{\eta}_1|, \;\; \xi_\ell := d_\ell - |\widetilde{\eta}_{\ell-1}|, \quad \text{and} \;\; \xi_i := d_{ii} - |\widetilde{\eta}_{i-1}| - |\widetilde{\eta}_i|, \;\; \forall\, i \in \{2, \ldots, \ell-1\}.$$

In light of the entries of G obtained above, we have, for each $k \in \{1, \ldots, L\}$, (1.1) if $m_k^\alpha = 1$, then $\xi_i = 1$.

(1.2) if $m_k^\alpha > 1$ with $k = 1$, then (i) the corresponding $\xi_i = d_{11} - |\eta| \geq \frac{1}{2} + \frac{h_{1,1}^\alpha}{6}$; (ii) for $s = 2, \ldots, w_1$, the corresponding $\xi_i = d_{ss} - |G_{s(s-1)}| - |G_{s(s+1)}| \geq \big(h_{1,s-1}^\alpha + h_{1,s}^\alpha\big)/6$; and (iii) the corresponding $\xi_i = d_{(w_1+1)(w_1+1)} - |G_{(w_1+1)w_1}| - |G_{(w_1+1)(w_1+2)}| \geq \frac{1}{2} + \frac{h_{1,w_1}^\alpha}{6}$. The similar results can be obtained for $m_k^\alpha > 1$ with $k = L$ using symmetry.

(1.3) if $m_k^\alpha > 1$ with $k \in \{2, \ldots, L-1\}$, then (i) the corresponding $\xi_i = d_{tt} - |G_{t(t-1)}| - |G_{t(t+1)}| \geq \frac{1}{2} + \frac{h_{k,1}^\alpha}{6}$; (ii) for $s = 1, \ldots, w_k - 1$, the corresponding $\xi_i = d_{(t+s)(t+s)} - |G_{(t+s)(t+s-1)}| - |G_{(t+s)(t+s+1)}| \geq \big(h_{k,s}^\alpha + h_{k,s+1}^\alpha\big)/6$; and (iii) the corresponding $\xi_i = d_{(t+w_k)(t+w_k)} - |G_{(t+w_k)(t+w_k-1)}| - |G_{(t+w_k)(t+w_k+1)}| \geq \frac{1}{2} + \frac{h_{k,w_k}^\alpha}{6}$.

Consequently, $\xi_i > 0$ for all i and G is strictly diagonally dominant. $\square$

For the given F_α, define η_i as the sum of the entries in the ith row of F_α, $i = 1, \ldots, \ell$. We have:

(i) if $m_k^\alpha = 1$, then $\eta_i = 1$.

(ii) if $m_k^\alpha > 1$ with $k = 1$, then (i) for $s = 1$, the corresponding $\eta_i = \frac{1+h_{1,1}^\alpha}{2}$; (ii) for $s = 2, \ldots, w_1$, the corresponding $\eta_i = \big(h_{1,s-1}^\alpha + h_{1,s}^\alpha\big)/2$; and (iii) for $s = w_1 + 1$, the corresponding $\eta_i = \frac{1+h_{1,w_1}^\alpha}{2}$. The similar results can be obtained for $m_k^\alpha > 1$ with $k = L$ using symmetry.

(iii) if $m_k^\alpha > 1$ with $k \in \{2, \ldots, L-1\}$, then (i) the corresponding $\eta_i = \frac{1+h_{k,1}^\alpha}{2}$; (ii) for $s = 1, \ldots, w_k - 1$, the corresponding $\eta_i = \big(h_{k,s}^\alpha + h_{k,s+1}^\alpha\big)/2$; and (iii) the corresponding $\eta_i = \frac{1+h_{k,w_k}^\alpha}{2}$.

Hence, each $\eta_i > 0$. Define the diagonal matrix

$$\Xi_\alpha := \operatorname{diag}\Big(\eta_1^{-1}, \ldots, \eta_\ell^{-1}\Big). \tag{6.1}$$

The next lemma shows the equivalence of the (absolute) row sum of $F_\alpha F_\alpha^T$ and that of F_α.

Lemma 6.2. *For any K_n and α, $\eta_j = \sum_{k=1}^{\ell} (F_\alpha F_\alpha^T)_{jk}$ for each $j = 1, \ldots, \ell$.*

Proof. Let $(F_\alpha)_{j\bullet}$ denote the jth row of F_α. Then

$$\sum_{k=1}^{\ell} |(F_\alpha F_\alpha^T)_{jk}| = \langle (F_\alpha)_{j\bullet}, \sum_{k=1}^{\ell} \big((F_\alpha)_{k\bullet}\big)^T \rangle = \langle (F_\alpha)_{j\bullet}, \mathbf{1} \rangle = \sum_{k=1}^{\ell} (F_\alpha)_{jk} = \eta_j,$$

where we use the fact that the sum of all rows of F_α is $\mathbf{1} := (1, \ldots, 1)$. $\square$

PROPOSITION 6.1. *For any K_n and α, the following statements hold:*

(1) the eigenvalues of $\Xi_\alpha F_\alpha \Lambda F_\alpha^T$ and $\Xi_\alpha F_\alpha F_\alpha^T$ are all positive reals;
(2) $\lambda_{\min}(\Lambda)\lambda_{\min}(\Xi_\alpha F_\alpha F_\alpha^T) \le \lambda_{\min}(\Xi_\alpha F_\alpha \Lambda F_\alpha^T)$ and $\lambda_{\max}(\Xi_\alpha F_\alpha \Lambda F_\alpha^T) \le \lambda_{\max}(\Lambda)\lambda_{\max}(\Xi_\alpha F_\alpha F_\alpha^T)$.

PROOF. (1) For the diagonal matrix Ξ_α, define $\Xi_\alpha^{1/2} := \operatorname{diag}\big(\sqrt{\eta_1^{-1}}, \ldots, \sqrt{\eta_\ell^{-1}}\big)$. Let $\sigma(A)$ denote the spectrum of a square matrix A, i.e., the collection of all eigenvalues of A. We thus have

$$
\begin{aligned}
& \lambda' \in \sigma(\Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2}) \\
\Longleftrightarrow\ & \det\left(\lambda' I - \Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2}\right) = 0 \\
\Longleftrightarrow\ & \det(\Xi_\alpha^{1/2}) \cdot \det\left(\Xi_\alpha^{-1/2} \cdot \lambda' \cdot \Xi_\alpha^{-1/2} - F_\alpha \Lambda F_\alpha^T\right) \cdot \det(\Xi_\alpha^{1/2}) = 0 \\
\Longleftrightarrow\ & \det(\lambda' \cdot \Xi_\alpha^{-1} - F_\alpha \Lambda F_\alpha^T) = 0 \\
\Longleftrightarrow\ & \det(\lambda' I - \Xi_\alpha F_\alpha \Lambda F_\alpha^T) = 0 \\
\Longleftrightarrow\ & \lambda' \in \sigma(\Xi_\alpha F_\alpha \Lambda F_\alpha^T).
\end{aligned}
$$

Since Λ is positive definite and F_α is row linearly independent, $\Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2}$ is positive definite such that all the eigenvalues of $\Xi_\alpha F_\alpha \Lambda F_\alpha^T$ are positive reals. By replacing Λ by the identity matrix, we see that the same holds for the eigenvalues of $\Xi_\alpha F_\alpha F_\alpha^T$.

(2) By Statement (1), $\lambda_{\min}(\Xi_\alpha F_\alpha \Lambda F_\alpha^T) = \lambda_{\min}(\Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2})$ and $\lambda_{\max}(\Xi_\alpha F_\alpha \Lambda F_\alpha^T) = \lambda_{\max}(\Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2})$. Further, for any $x \neq 0$,

$$
\frac{x^T \Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2} x}{x^T x} = \frac{x^T \Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2} x}{x^T \Xi^{1/2} F_\alpha F_\alpha^T \Xi^{1/2} x} \cdot \frac{x^T \Xi_\alpha^{1/2} F_\alpha F_\alpha^T \Xi_\alpha^{1/2} x}{x^T x}.
$$

Therefore, using the fact that all the eigenvalues of Λ are positive, we obtain

$$
\begin{aligned}
\lambda_{\min}(\Lambda)\lambda_{\min}(\Xi^{1/2} F_\alpha F_\alpha^T \Xi^{1/2}) &\le \lambda_{\min}(\Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2}) \le \lambda_{\max}(\Xi_\alpha^{1/2} F_\alpha \Lambda F_\alpha^T \Xi_\alpha^{1/2}) \\
&\le \lambda_{\max}(\Lambda)\lambda_{\max}(\Xi_\alpha^{1/2} F_\alpha F_\alpha^T \Xi_\alpha^{1/2})
\end{aligned}
$$

Since $\lambda_{\min}(\Xi_\alpha^{1/2} F_\alpha F_\alpha^T \Xi_\alpha^{1/2}) = \lambda_{\min}(\Xi_\alpha F_\alpha F_\alpha^T)$ and $\lambda_{\max}(\Xi_\alpha^{1/2} F_\alpha F_\alpha^T \Xi_\alpha^{1/2}) = \lambda_{\max}(\Xi_\alpha F_\alpha F_\alpha^T)$, the desired inequalities follow. $\square$

The following proposition attains uniform upper and lower bounds for the eigenvalues of $\Xi_\alpha F_\alpha F_\alpha^T$, regardless of K_n and α.

PROPOSITION 6.2. *For any K_n and α,*

$$1/3 \leq \lambda_{\min}(\Xi_\alpha F_\alpha F_\alpha^T) \leq \lambda_{\max}(\Xi_\alpha F_\alpha F_\alpha^T) \leq 1.$$

PROOF. (1) Uniform upper bound. By Lemma 6.2, we have $\sum_{k=1}^{\ell} |(F_\alpha F_\alpha^T)_{jk}| = \sum_{k=1}^{\ell} G_{jk} = \eta_j$. Hence, $\sum_{k=1}^{\ell}(\Xi_\alpha F_\alpha F_\alpha^T)_{jk} = 1$ for all $j = 1, \ldots, \ell$. It follows from [16, Corollary 6.1.5] that $\lambda_{\max}(\Xi_\alpha F_\alpha F_\alpha^T) \leq 1$.

(2) Uniform lower bound. To establish this bound, we exploit Gersgorin's Disc Theorem, say [16, Theorem 6.1.1]. Notice that each ξ_i defined in Lemma 6.1 is the difference between the ith diagonal of $G := F_\alpha F_\alpha^T$ and the deleted absolute row sum of the ith row of G. Hence, by Gersgorin's Disc Theorem, we see that $\lambda_{\min}(\Xi_\alpha F_\alpha F_\alpha^T) \geq \min_i(\xi_i/\eta_i)$. Further, using the lower bound of ξ_i given in Lemma 6.1 and the equality for η_i given before Lemma 6.2, it is easy to verify $\xi_i/\eta_i \geq 1/3$ for all i, α and K_n. This yields the desired uniform lower bound. $\square$

The next result establishes a uniform bound on the ℓ_∞-norm of $F_\alpha^T(F_\alpha \Lambda F_\alpha^T)^{-1}F_\alpha$.

PROPOSITION 6.3. *There exists $c_{\infty,p} > 0$ (dependent on p only) such that for any K_n and α, $\|F_\alpha^T(F_\alpha \Lambda F_\alpha^T)^{-1}F_\alpha\|_\infty \leq c_{\infty,p}$.*

PROOF. For any K_n and any α, let Ξ_α be that defined in (6.1). Then

$$\begin{aligned}
\|F_\alpha^T(F_\alpha \Lambda F_\alpha^T)^{-1}F_\alpha\|_\infty &= \|F_\alpha^T \cdot (\Xi_\alpha F_\alpha \Lambda F_\alpha^T)^{-1} \cdot (\Xi_\alpha F_\alpha)\|_\infty \\
&\leq \|F_\alpha^T\|_\infty \cdot \|(\Xi_\alpha F_\alpha \Lambda F_\alpha^T)^{-1}\|_\infty \cdot \|\Xi_\alpha F_\alpha\|_\infty.
\end{aligned}$$

It is easy to verify $\|F_\alpha^T\|_\infty = 1$. Furthermore, due to the definition of Ξ_α, we have $\|\Xi_\alpha F_\alpha\|_\infty = 1$. In what follows, we show that $\|(\Xi_\alpha F_\alpha \Lambda F_\alpha^T)^{-1}\|_\infty$ is uniformly bounded using the banded structure of the matrix and other technical results developed before. This will give rise to a uniform bound.

Let F_α be of ℓ rows. We consider two cases as follows:

(i) $\ell \geq 2(p+1)$. In this case, by the structure of F_α shown in (5.6), we see via straightforward computation that $H := \Xi_\alpha F_\alpha \Lambda F_\alpha^T$ is a banded symmetric matrix with bandwidth p, i.e., $(H)_{ij} = 0$ whenever $|i-j| > p$. It is known from [44, Lemma 6.2] that for a fixed spline degree p, there exist positive constants $\underline{\mu}_p$ and $\overline{\mu}_p$ (dependent on p only) such that $\underline{\mu}_p \leq \lambda_{\min}(\Lambda) \leq \lambda_{\max}(\Lambda) \leq \overline{\mu}_p$ for any K_n. It thus follows from Propositions 6.1 and 6.2 that $\|H\|_2 = \lambda_{\max}(H) \leq \overline{\mu}_p$, where $\overline{\mu}_p$ is independent of K_n and α. Similarly, $\|H^{-1}\|_2 \leq 1/\lambda_{\min}(H) \leq 3/\underline{\mu}_p$. Hence, for F_α with $\ell \geq 2(p+1)$, it follows from [8, Theorem 2.2] that there exists $c' > 0$ (independent of K_n

and α) such that $\|(H^{-1})_{i\bullet}\|_1 \leq c'$ for all $i = 1, \ldots, \ell$, where $(H^{-1})_{i\bullet}$ denotes the ith row of H^{-1}. In other words, $\|H^{-1}\|_\infty \leq c'$.

(ii) $\ell < 2(p+1)$. For any F_α in this case, we introduce the block diagonal matrix $\widetilde{H} := \mathrm{diag}(H, 1, \ldots, 1)$ such that H' has $2(p+1)$ rows. Hence, $\widetilde{H}$ is a banded symmetric matrix with bandwidth p and satisfies $\|\widetilde{H}\|_2 \leq \max(\overline{\mu}_p, 1)$, $\|\widetilde{H}^{-1}\|_2 \leq \max(3/\underline{\mu}_p, 1)$. Thus there exists $c'' > 0$ (independent of K_n and α) such that $\|H^{-1}\|_\infty \leq \|\widetilde{H}^{-1}\|_\infty \leq c''$.

Consequently, $c_{\infty,p} := \max(c', c'')$ is the desired uniform bound with respect to the ℓ_∞-norm. $\square$

Along with the above results, we finally complete the proof of the uniform Lipschitz property below.

Proof of Theorem 3.1. The uniform bound on $\|F_\alpha^T(F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha\|_\infty$ has been established in Proposition 6.3. The second statement follows directly from the continuous and piecewise linear property of $\hat{b}$ and polyhedral theory [10, Proposition 4.2.2]. $\square$

7. Proof of Theorem 3.2. We introduce some notation first. Let $\bar{f}^{[p]}$ be the spline estimator based on noise free data, i.e., $\bar{f}^{[p]}(x) = \sum_{k=1}^{K_n+p} \bar{b}_k B_k^{[p]}(x)$, where

$$\bar{b} := \arg\min_{b \in \Omega} \frac{1}{2} b^T \Lambda b - b^T \mathbb{E}(\bar{y}). \tag{7.1}$$

Propositions 7.1 and 7.2 below give rise to uniform bounds for the bias and stochastic terms of estimation error in the sup-norm, respectively.

Proposition 7.1. *If $1 \leq r' \leq r$, there exists a constant $C_{1r'}$, which depends on r' only, such that*

$$\sup_{f \in \mathcal{C}(r,L)} \|\bar{f}_{(r')} - f\|_\infty \leq C_{1r'} \cdot L \cdot K_n^{-r}. \tag{7.2}$$

In particular, if $r = 2$, then $C_{1r'}$ is independent of r'.

Proof. Consider the case when $1 \leq r \leq 2$ first. Hence $\lceil r' - 1 \rceil = 1$. Let $\tilde{f}$ be a piecewise linear function such that $\tilde{f}(\kappa_k) = f(\kappa_k)$. For any $x \in [\kappa_{k-1}, \kappa_k]$, $k = 1, \ldots, K_n$, there exist $\xi_x, \tilde{\xi}_x \in (\kappa_{k-1}, \kappa_k)$ such that

$$\begin{aligned}
&\tilde{f}(x) - f(x) \\
=&\tilde{f}(\kappa_{k-1}) + K_n\big(\tilde{f}(\kappa_k) - \tilde{f}(\kappa_{k-1})\big)(x - \kappa_{k-1}) - \big[f(\kappa_{k-1}) + f'(\xi_x)(x - \kappa_{k-1})\big] \\
=&\big[f'(\tilde{\xi}_x) - f'(\xi_x)\big](x - \kappa_{k-1}) \leq L|\tilde{\xi}_x - \xi_x|^\ell |x - \kappa_{k-1}| \leq L K_n^{-r}.
\end{aligned}$$

Thus $\|\tilde{f}-f\|_\infty \leq LK_n^{-r}$. Let $\vec{f} := (f(x_1), \dots, f(x_n))^T$, $\vec{\tilde{f}} := (\tilde{f}(x_1), \dots, \tilde{f}(x_n))^T$, and let $\tilde{b}$ be the optimal solution of (7.1) with $\mathbb{E}(\bar{y})$ replaced by $X^T\vec{\tilde{f}}/\beta_n$. Since $\tilde{f}$ is a piecewise linear and convex function, we have $\tilde{b} = (\tilde{f}(\kappa_1), \dots, \tilde{f}(\kappa_{K_n}))^T$. It follows from Theorem 3.1 that

$$\begin{aligned} \|\bar{f}^{[1]} - \tilde{f}\|_\infty &\leq \|\bar{b} - \tilde{b}\|_\infty \leq \frac{c_{\infty,1}}{\beta_n}\|X^T(\vec{f} - \vec{\tilde{f}})\|_\infty \\ &\leq \frac{c_{\infty,1}}{\beta_n}\|X^T\|_\infty \|f - \tilde{f}\|_\infty = c_{\infty,1}\varrho\|f - \tilde{f}\|_\infty, \end{aligned}$$

where $\|X^T\|_\infty = \sum_{i=1}^n B_2^{[1]}(x_i)$ and $\varrho := \sum_{i=1}^n B_2^{[1]}(x_i)/\sum_{i=1}^n B_2^{[1]}(x_i)^2$. Therefore, letting $C_{1r'} := c_{\infty,1}\varrho$ which is independent of r', we have

$$\|\bar{f}^{[1]} - f\|_\infty \leq (1 + c_{\infty,1}\varrho)\|f - \tilde{f}\|_\infty \leq (1 + c_{\infty,1}\varrho)LK_n^{-r} = C_{1r'}LK_n^{-r}.$$

Next, consider the case when $r > 2$. If $r' \leq 2$, a similar argument as above yields (7.2). If $r' > 2$, it is shown in Theorem 3.3 that an unconstrained estimator and the constrained one are asymptotically equivalent. Since (7.2) holds for the unconstrained estimator [44], the proof is complete. □

PROPOSITION 7.2. *There exists a positive constant C_{2r}, which depends on r only, such that for any $u > 0$,*

$$P\Big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \geq u\Big) \leq (K_n + p)\exp\Big\{ -\frac{n}{2K_nC_{2r}^2\sigma^2}u^2\Big\}. \tag{7.3}$$

In particular, if $r \in [1,2]$, then C_{2r} is independent of r.

PROOF. Recall $p = \lceil r-1 \rceil$. By Theorem 3.1 and (2.2), we have

$$\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \leq \frac{\sigma c_{\infty,p}}{\sqrt{\beta_n}} \sup_{k=1,\dots,K_n+p} |\xi_k| = \frac{\sigma c_{\infty,p}}{\sqrt{C_{\beta,p}}}\sqrt{\frac{K_n}{n}} \sup_{k=1,\dots,K_n+p} |\xi_k|,$$

where $\xi_k = \sum_{i=1}^n B_k^{[p]}(x_i)\epsilon_i/\sqrt{\beta_n}$. Letting $C_{2r} := c_{\infty,p}/\sqrt{C_{\beta,p}}$ which is dependent on r only (but independent of r if $r \in [1,2]$), we have

$$\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \ \leq \ \sigma C_{2r}\sqrt{\frac{K_n}{n}}\ \tilde{\Gamma}_r,$$

where $\tilde{\Gamma}_r = \max_{k=1,\dots,K_n+p}|\xi_k|$. Hence, by using the implication: $Z \sim N(0,1) \Longrightarrow P(Z > t) \leq \frac{1}{2}e^{-t^2/2}, \forall\ t \geq 0$, we have

$$\begin{aligned} &P\Big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \geq u\Big) \leq P\Big(\tilde{\Gamma}_r \geq \frac{u}{C_{2r}\sigma}\sqrt{\frac{n}{K_n}}\Big) \\ \leq &(K_n + p)P\Big\{|\xi_k| \geq \frac{u}{C_{2r}\sigma}\sqrt{\frac{n}{K_n}}\Big\} \leq (K_n + p)\exp\Big\{ -\frac{n}{2K_nC_{2r}^2\sigma^2}u^2\Big\}. \end{aligned}$$

□

Let

$$T_n := C_{2r}\sigma\sqrt{\frac{2}{2r+1}}\sqrt{\frac{\log n}{n}}K_n^{\frac{1}{2}}.$$

It follows from Proposition 7.2 that,

$$\begin{aligned}
\mathbb{E}\Big(\|\hat{f}_{(r)}-\bar{f}_{(r)}\|_\infty\Big) &\le T_n + \int_{T_n}^{\infty} P\Big(\|\hat{f}_{(r)}-\bar{f}_{(r)}\|_\infty > t\Big)dt \\
&\le T_n + \int_{T_n}^{\infty} (K_n+p)\exp\Big\{-\frac{n}{2K_nC_{2r}^2\sigma^2}t^2\Big\}dt \\
&\le T_n + \sqrt{\frac{\pi}{2}}\; C_{2r}\sigma\sqrt{n^{-1}K_n}\; (K_n+p)n^{-\frac{1}{2r+1}} \\
&= O(T_n).
\end{aligned}$$

In view of Proposition 7.1 and the above result, we deduce that

$$\mathbb{E}\|\hat{f}_{(r)}-f\|_\infty \le \|\bar{f}_{(r)}-f\|_\infty + \mathbb{E}\|\hat{f}_{(r)}-\bar{f}_{(r)}\|_\infty = O\Big(LK_n^{-r}+\sigma\sqrt{\frac{\log n}{n}}K_n^{\frac{1}{2}}\Big).$$

This shows Statement (1) of Theorem 3.2 by using the optimal choice of K_n.

The next proposition establishes uniform bounds for the stochastic estimation error for a fixed point as well as mean squared error.

PROPOSITION 7.3. *For any given $x_0 \in [0,1]$, there exist two positive constants C_{3r} and C_{4r}, which depend only on r, such that*

$$\mathbb{E}(|\hat{f}_{(r)}(x_0)-\bar{f}_{(r)}(x_0)|^2) \;\le\; C_{3r}\sigma^2n^{-1}K_n, \tag{7.4}$$

$$\mathbb{E}(|\hat{f}_{(r)}(x_0)-\bar{f}_{(r)}(x_0)|^4) \;\le\; C_{4r}\sigma^4n^{-2}K_n^2. \tag{7.5}$$

Furthermore,

$$\mathbb{E}\big(\big\|\hat{f}_{(r)}-\bar{f}_{(r)}\big\|_2^2\big) := \mathbb{E}\big(\int_0^1 \big|\hat{f}_{(r)}(x)-\bar{f}_{(r)}(x)\big|^2dx\big) \;\le\; C_{3r}\sigma^2n^{-1}K_n. \tag{7.6}$$

In particular, if $r\in[1,2]$, then C_{3r} and C_{4r} are independent of r.

PROOF. Recall that $p = \lceil r-1\rceil$, $N(x) = [B_1(x),\dots,B_{K_n+p}(x)]^T \in \mathbb{R}^{K_n+p}$, and $X = [N(x_1),\dots,N(x_n)]^T \in \mathbb{R}^{n\times(K_n+p)}$. Fix $x_0\in[0,1]$. Let $h := N(x_0)\in\mathbb{R}^{K_n+p}$. Note that h has at most p nonzero elements and each of these nonzero elements is positive whose sum is 1. Let G_α be the coefficient matrix of a linear selection function corresponding to an index set α, i.e., $G_\alpha = F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha$. Hence,

$$\begin{aligned} \|G_\alpha h\|_2^2 &= \sum_i \Big(\sum_{h_j>0} (G_\alpha)_{ij} h_j \Big)^2 \le p \cdot \sum_i \sum_{h_j>0} (G_\alpha)_{ij}^2 h_j^2 \\ &\le p \cdot \sum_i \sum_{h_j>0} (G_\alpha)_{ij}^2 h_j \le p \cdot \sum_{h_j>0} h_j \sum_i (G_\alpha)_{ij}^2 \\ &\le p \cdot \sum_{h_j>0} h_j \|G_\alpha\|_\infty^2 \le p \cdot \|G_\alpha\|_\infty^2 = p \cdot c_{\infty,p}^2, \end{aligned}$$

where the first inequality in the third line is due to the symmetry of G_α and the following implication

$$\sum_i \big|(G_\alpha)_{ij}\big| \le \|G_\alpha\|_\infty \implies \sum_i \big|(G_\alpha)_{ij}\big|^2 \le \big(\sum_i \big|(G_\alpha)_{ij}\big| \big)^2 \le \|G_\alpha\|_\infty^2.$$

As a result, in light of Theorem 3.1, we have

$$\max_\alpha \|G_\alpha h\|_2^2 \le p \cdot c_{\infty,p}^2. \tag{7.7}$$

Let $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ be iid random variables with mean zero and variance one, and let $z := (f(x_1), \dots, f(x_n))^T$. Thus $\bar{y} = X^T(z + \sigma\epsilon)/\beta_n$. Hence

$$h^T \hat{b}(\bar{y}) = h^T \hat{b}(X^T(z + \sigma\epsilon)/\beta_n) = \frac{1}{\beta_n} h^T G_{\alpha(\epsilon)} X^T (z + \sigma\epsilon).$$

Furthermore, since $\hat{b}(\cdot)$ is a continuous piecewise linear function on $\mathbb{R}^{K_n+p}$, so is $\hat{b} \circ X^T$ on $\mathbb{R}^n$. It follows from the polyhedral theory that $\hat{b} \circ X^T$ admits a conic subdivision of $\mathbb{R}^n$ [10, 33], i.e., there exist a finite collection of polyhedral cones $\{\mathcal{C}_j\}_{j=1}^q$ and linear functions $\{g^j\}_{j=1}^q$ such that (i) $\bigcup_j \mathcal{C}_j = \mathbb{R}^n$; (ii) each cone $\mathcal{C}_j$ has nonempty interior; (iii) the intersection of any two cones is a common proper face of both cones; and (iv) $\hat{b} \circ X^T$ coincides with g^j on each $\mathcal{C}_j$. For any given $z' \in \mathbb{R}^n$, let $[z, z']$ be a line segment joining z and z'. Starting from z, we assume that the line segment $[z, z']$ intersects some cones in the conic subdivision at $z_1, z_2, \dots, z_{\ell-1} \in \mathbb{R}^n$, and ends at z'. Further, each subsegment of any two consecutive points, such as $[z, z_1], [z_1, z_2], \dots, [z_{\ell-1}, z']$, belongs to a single cone. Hence there exist $\mu_i \in [0, 1], i = 1, \dots, \ell$ with $\ell \le q$, $\sum_{i=1}^\ell \mu_i = 1$ and G_{α_i} such that

$$\begin{aligned} &\hat{b}(X^T z') - \hat{b}(X^T z) \\ &\quad = G_{\alpha_1} X^T (z_1 - z) + G_{\alpha_2} X^T (z_2 - z_1) + \cdots + G_{\alpha_\ell} X^T (z' - z_{\ell-1}) \\ &\quad = \Big(\sum_{i=1}^\ell \mu_i G_{\alpha_i} \Big) X^T (z' - z), \end{aligned}$$

where μ_i and G_{α_i} depend on z' for the fixed z. Since there are q cones in the conic subdivision, we may use the extended tuple $(\mu_i, G_{\alpha_i})_{i=1}^q$ (corresponding to z') to characterize $\hat{b}(X^T z') - \hat{b}(X^T z)$, by setting some $\mu_i = 0$, without loss of generality. Note that if z' is a random variable, so is $(\mu_i, G_{\alpha_i})_{i=1}^q$.

Using $\bar{y} = X^T(z + \sigma\epsilon)/\beta_n$, we have, for the given vector h,

$$\begin{aligned}
&\mathbb{E}\big(|h^T\hat{b}(\bar{y}) - h^T\hat{b}(\mathbb{E}(\bar{y}))|^2\big) \\
&= \frac{1}{\beta_n^2}\mathbb{E}_{(\mu_i, G_{\alpha_i})_{i=1}^q}\Big(\mathbb{E}\big(|h^T\hat{b}(X^T(z+\sigma\epsilon)) - h^T\hat{b}(X^T z)|^2 \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q\big)\Big).
\end{aligned}$$

Moreover, for a fixed tuple $(\mu_i, G_{\alpha_i})_{i=1}^q$,

$$\begin{aligned}
&\mathbb{E}\Big(|h^T\hat{b}(X^T(z+\sigma\epsilon)) - h^T\hat{b}(X^T z)|^2 \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q\Big) \\
&= \sigma^2\mathbb{E}\Big(|\sum_{i=1}^q (\mu_i h^T G_{\alpha_i})X^T\epsilon \cdot \sum_{i=1}^q (\mu_i h^T G_{\alpha_i})X^T\epsilon| \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q\Big) \\
&= \sigma^2\mathbb{E}\Big(|\sum_{i=1}^q (\mu_i h^T G_{\alpha_i} X^T)\epsilon \cdot \epsilon^T(\sum_{j=1}^q \mu_j X G_{\alpha_j} h)| \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q\Big) \\
&= \sigma^2\Big|(\sum_{i=1}^q \mu_i h^T G_{\alpha_i} X^T)(\sum_{j=1}^q \mu_j X G_{\alpha_j} h)\Big| \\
&\leq \sigma^2\Big(\sum_{i=1}^q \mu_i \|X G_{\alpha_i} h\|_2\Big)\cdot\Big(\sum_{j=1}^q \mu_j \|X G_{\alpha_j} h\|_2\Big) \\
&\leq \sigma^2\big(\max_i \|X G_{\alpha_i} h\|_2\big)^2 \leq \sigma^2\|X\|_2^2 \cdot \big(\max_i \|G_{\alpha_i} h\|_2\big)^2 \\
&\leq \sigma^2\beta_n\lambda_{\max}(\Lambda)\cdot p \cdot c_{\infty,p}^2
\end{aligned}$$

where the last inequality is due to $\|X\|_2^2 \leq \beta_n\lambda_{\max}(\Lambda)$ and (7.7). Therefore,

$$\mathbb{E}\big(|h^T\hat{b}(\bar{y}) - h^T\hat{b}(\mathbb{E}(\bar{y}))|^2\big) \leq \frac{1}{\beta_n}\cdot\lambda_{\max}(\Lambda)\cdot p\cdot c_{\infty,p}^2\cdot\sigma^2.$$

Observing that $\lambda_{\max}(\Lambda)$ is uniformly bounded [44] and the uniform bound of β_n in (2.2), we obtain (7.4) for $p = \lceil r-1 \rceil$.

The above argument can be extended to prove (7.6). Indeed, let $h(x) := N(x)$. Thus $\hat{f}_{(r)}(x) - \bar{f}_{(r)}(x) = h^T(x)[\hat{b}(\bar{y}) - \hat{b}(\mathbb{E}(\bar{y}))]$ such that

$$\begin{aligned}
&\mathbb{E}\big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_2^2\big) \\
&= \frac{1}{\beta_n^2}\mathbb{E}_{(\mu_i, G_{\alpha_i})_{i=1}^q}\Big(\mathbb{E}\big(\int_0^1 \big|h^T(x)[\hat{b}(X^T(z+\sigma\epsilon)) - \hat{b}(X^T z)]\big|^2 dx \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q\big)\Big),
\end{aligned}$$

where, for a given tuple $(\mu_i, G_{\alpha_i})_{i=1}^q$,

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^1 \big| h^T(x)[\hat{b}(X^T(z+\sigma\epsilon)) - \hat{b}(X^T z)] \big|^2 dx \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q \Big) \\
&\quad \le \quad \sigma^2 \|X\|_2^2 \int_0^1 \big(\max_i \|G_{\alpha_i} h(x)\|_2 \big)^2 dx \;\le\; \sigma^2 \|X\|_2^2 \cdot (p \cdot c_{\infty,p}^2),
\end{aligned}$$

which yields (7.6).

To show (7.5), we consider

$$\begin{aligned}
&\mathbb{E}(|h^T \hat{b}(\bar{y}) - h^T \hat{b}(\mathbb{E}(\bar{y})|^4) \\
&\quad = \quad \frac{1}{\beta_n^4} \mathbb{E}_{(\mu_i, G_{\alpha_i})_{i=1}^q} \Big(\mathbb{E}\big(|h^T \hat{b}(X^T(z+\sigma\epsilon)) - h^T \hat{b}(X^T z)|^4 \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q \big) \Big).
\end{aligned}$$

For a fixed tuple $(\mu_i, G_{\alpha_i})_{i=1}^q$, let $v := \sum_{i=1}^q \mu_i X G_{\alpha_i} h \in \mathbb{R}^n$ and $\mathbb{E}(\epsilon_i^4) = 3, i = 1, \ldots, n$. We thus have

$$\begin{aligned}
&\mathbb{E}\Big(|h^T \hat{b}(X^T(z+\sigma\epsilon)) - h^T \hat{b}(X^T z)|^4 \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q \Big) \\
&\quad = \quad \mathbb{E}\Big(|(\sum_{i=1}^q \mu_i h^T G_{\alpha_i} X^T)\sigma\epsilon \cdot (\sum_{i=1}^q \mu_i h^T G_{\alpha_i} X^T)\sigma\epsilon|^2 \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q \Big) \\
&\quad = \quad \mathbb{E}\Big(|v^T \sigma\epsilon \cdot \sigma\epsilon^T v|^2 \,\big|\, (\mu_i, G_{\alpha_i})_{i=1}^q \Big) \\
&\quad = \quad \sigma^4 \sum_{i=1}^n v_i^4 \cdot \mathbb{E}(\epsilon_i^4) + \sigma^4 \sum_{i,j=1, i\neq j}^n (v_i v_j)^2 \cdot \mathbb{E}(\epsilon_i^2 \cdot \epsilon_j^2) \\
&\quad = \quad 2\sigma^4 \sum_{i=1}^n v_i^4 + \sigma^4 \sum_{i=1}^n \sum_{j=1}^n v_i^2 v_j^2 \\
&\quad \le \quad 2\sigma^4 \|v\|_2^4 + \sigma^4 (\sum_{i=1}^n v_i^2) \cdot (\sum_{j=1}^n v_j^2) \le 3\sigma^4 \cdot \|v\|_2^4 \\
&\quad \le \quad 3\sigma^4 \cdot (\beta_n \lambda_{\max}(\Lambda) \cdot p \cdot c_{\infty,p}^2)^2,
\end{aligned}$$

where the last inequality is due to $\|v\|_2^2 \le \beta_n \lambda_{\max}(\Lambda) p c_{\infty,p}^2$. This shows that

$$\mathbb{E}\big(|h^T \hat{b}(\bar{y}) - h^T \hat{b}(\mathbb{E}(\bar{y})|^4 \big) \le \frac{2}{\beta_n^2} \cdot \sigma^4 \cdot (\lambda_{\max}(\Lambda) \cdot p \cdot c_{\infty,p}^2)^2.$$

Using the uniform bounds on $\lambda_{\max}(\Lambda)$ and β_n again, we obtain (7.5). □

Propositions 7.1 and 7.3 imply that, for any $x_0 \in [0, 1]$,

$$\sup_{f \in \mathcal{C}_H(r,L)} \mathbb{E}|\hat{f}_{(r)}(x_0) - f(x_0)|^2 \;\le\; C_{1r}^2 L^2 K_n^{-2r} + C_{3r}\sigma^2 K_n n^{-1}. \tag{7.8}$$

This shows Statement (2) of Theorem 3.2 by using the optimal choice of K_n.

8. Proof of Theorem 3.4. Throughout this section, we shall use C_k or c_k with $k \in \mathbb{N}$ to denote positive constants that depend only on L (and σ). We introduce some lemmas. The first lemma, as a complement to (3.4), provides a bound for the stochastic term of estimation error in the sup-norm.

LEMMA 8.1. *There exists a constant $C_2 > 0$ such that*

$$\begin{aligned}(8.1)\qquad &\mathbb{E}\Big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty I\big\{\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \geq u\big\}\Big)\\ &\leq \sqrt{\frac{\pi}{2}} C_2 \sigma \sqrt{n^{-1}K_n}(K_n+1)\exp\Big\{-\frac{nu^2}{2K_nC_2^2\sigma^2}\Big\}.\end{aligned}$$

PROOF. Direct calculation yields that

$$\begin{aligned}&\mathbb{E}\Big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty I\big\{\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \geq u\big\}\Big)\\ &= \int_u^\infty P\Big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \geq t\Big)dt \leq \int_u^\infty (K_n+1)\exp\Big\{-\frac{n}{2K_nC_2^2\sigma^2}t^2\Big\}dt\\ &\leq \sqrt{\frac{\pi}{2}} C_2 \sigma \sqrt{n^{-1}K_n}(K_n+1)\exp\Big\{-\frac{nu^2}{2K_nC_2^2\sigma^2}\Big\}.\end{aligned}$$

This completes the proof. $\square$

It is shown next that it is highly improbable that the estimated $\hat{r}$ is strictly smaller than the true r. We say a few words about notation. Recall that $\tau_n = \lceil (\log n)^{\frac{1}{2}} \rceil$ and define the set $\mathcal{R} := \{r_j \mid r_j = 1 + j/\tau_n,\ j = 0, 1, \ldots, \tau_n\}$.

LEMMA 8.2. *Let $r, d \in [1,2]$ with $d < r$. There exists $C_3 > 0$ such that*

$$\sup_{f \in \mathcal{C}_H(r,L)} P\big(\hat{r} = d\big) \leq C_3 \tau_n n^{-\frac{1}{2d+1}}.$$

PROOF. By the definition of $\hat{r}$ given before Theorem 3.4,

$$\sup_{f \in \mathcal{C}_H(r,L)} P\big(\hat{r} = d\big) \leq \sum_{d \geq r' \in \mathcal{R}} \sup_{f \in \mathcal{C}_H(r,L)} p_\infty(r', d) \leq \tau_n \max_{d \geq r' \in \mathcal{R}} \sup_{f \in \mathcal{C}_H(r,L)} p_\infty(r', d),$$

where

$$p_\infty(r', d) := P\Big(\|\hat{f}_{(r'')} - \hat{f}_{(r')}\|_\infty > \frac{1+\sqrt{2}}{2}\ \psi_{(r')}\Big).$$

Here $\psi_{(r')}$ is defined in (3.6) and $r'' := \min\{r \in \mathcal{R} \mid r > d\}$, i.e., $r'' \in \mathcal{R}$ is closest to d from above. Hence, $r'' > d \geq r'$. In view of (3.3) and (3.4),

$$\begin{aligned}\|\hat{f}_{(r'')} - \hat{f}_{(r')}\|_\infty &\leq \|\hat{f}_{(r'')} - \bar{f}_{(r'')}\|_\infty + \|\hat{f}_{(r')} - \bar{f}_{(r')}\|_\infty + \|\bar{f}_{(r'')} - f\|_\infty + \|\bar{f}_{(r')} - f\|_\infty\\ &\leq \|\hat{f}_{(r')} - \bar{f}_{(r')}\|_\infty + \|\hat{f}_{(r'')} - \bar{f}_{(r'')}\|_\infty + C_1 L K_{(r')}^{-r'} + C_1 L K_{(r'')}^{-r''}\\ &= C_1 L K_{(r')}^{-r'}\big(1 + \omega_{r',r''}\big) + \|\hat{f}_{(r')} - \bar{f}_{(r')}\|_\infty + \|\hat{f}_{(r'')} - \bar{f}_{(r'')}\|_\infty,\end{aligned}$$

where

$$|\omega_{r',r''}| = \frac{K_{(r')}^{r'}}{K_{(r'')}^{r''}} \le c_1\Big(\frac{\log n}{n}\Big)^{\frac{(r''-r')}{(2r''+1)(2r'+1)}} \le c_1\Big(\frac{\log n}{n}\Big)^{\frac{1}{25\tau_n}},$$

for a positive constant c_1 which is bounded away from zero and above. Since $(n^{-1}\log n)^{\sqrt{\log n}} \to 0$ as $n \to \infty$, $\omega_{r',r''}$ converges to zero uniformly for all $r', r'' \in \mathcal{R}$ with the given τ_n. Let $p_\infty(r', d) \le p_{1,\infty} + p_{2,\infty}$, where

$$p_{1,\infty} := P\Big\{\|\hat{f}_{(r')} - \bar{f}_{(r')}\|_\infty \ge \frac{\sqrt{2}}{2}\ \psi_{(r')}(1 - o(1))\Big\},$$
$$p_{2,\infty} := P\Big\{\|\hat{f}_{(r'')} - \bar{f}_{(r'')}\|_\infty \ge \frac{(2r''+1)}{2}\ \psi_{(r'')}\Big\}.$$

By using (3.4), (3.6) and the orders of $K_{(r')}, K_{(r'')}$, we obtain two positive constants c_2 and c_3 such that

$$p_{1,\infty} \le \big(K_{(r')} + 1\big) \cdot n^{-\frac{2}{2r'+1}} \le c_2 n^{-\frac{1}{(2r'+1)}}(\log n)^{-\frac{1}{2r'+1}} \le c_2 n^{-\frac{1}{2d+1}}(\log n)^{-1/5},$$
$$p_{2,\infty} \le \big(K_{(r'')} + 1\big) \cdot n^{-(2r''+1)} \le c_3 n^{-\frac{1}{(2r''+1)}}(\log n)^{-\frac{1}{2r''+1}} \le c_3 n^{-\frac{1}{2d+1}}(\log n)^{-1/5}.$$

Combining the above results, we see that the lemma holds. □

The following lemma develops a uniform bound on the sup-norm risk of $\hat{f}_{(r)}$ for $r \in [1, 2]$.

LEMMA 8.3. *There exist positive constants C_4 and C_5 such that*

$$\sup_{r\in[1,2]}\ \sup_{f\in\mathcal{C}_H(r,L)} P_f\Big\{\psi_{(r)}^{-1}\|\hat{f}_{(r)} - f\|_\infty \ge 1 + \sqrt{2}\ \Big\} \ \le\ C_4 n^{-1},$$
$$\sup_{r\in[1,2]}\ \sup_{f\in\mathcal{C}_H(r,L)} \mathbb{E}\Big(\psi_{(r)}^{-2}\|\hat{f}_{(r)} - f\|_\infty^2\Big) \ \le\ C_5.$$

PROOF. Since

$$\|\hat{f}_{(r)} - f\|_\infty \le \|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty + \|\bar{f}_{(r)} - f\|_\infty \le \|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty + C_1 L K_{(r)}^{-r},$$

we have via (3.6) and $C_1 L K_{(r)}^{-r} = \psi_{(r)}/2$ that, for any $u \ge 1 + \sqrt{2}$,

$$\sup_{f\in\mathcal{C}_H(r,L)} P\Big\{\psi_{(r)}^{-1}\|\hat{f}_{(r)} - f\|_\infty \ge u\Big\} \le P\Big(\|\hat{f}_{(r)} - \bar{f}_{(r)}\|_\infty \ge \big(u - \frac{1}{2}\big)\psi_{(r)}\Big)$$
$$\le (K_{(r)} + 1)e^{-\log n\cdot\frac{(2u-1)^2}{2r+1}}.$$

In view of (3.5), $K_{(r)} = q(r)(n/\log n)^{1/(2r+1)}$ for some function $q(\cdot)$ that is positive and continuous on $[1,2]$. Hence, a constant $q_* > 0$ exists such that $K_{(r)} \le q_*(n/\log n)^{1/(2r+1)}$ for all $r \in [1,2]$. Applying this to $u = 1+\sqrt{2}$, we obtain a positive constant C_4 such that for any $r \in [1,2]$,

$$\sup_{f\in\mathcal{C}_H(r,L)} P\Big\{\psi_{(r)}^{-1}\|\hat{f}_{(r)} - f\|_\infty \ge 1+\sqrt{2}\Big\} \le C_4 n^{-\frac{4\sqrt{2}+8}{2r+1}}(\log n)^{-\frac{1}{2r+1}} \le C_4 n^{-1}.$$

Using the above results, we further get a positive constant c_5 such that

$$\begin{aligned}
\mathbb{E}\Big(\psi_{(r)}^{-2}\|\hat{f}_{(r)} - f\|_\infty^2\Big) &\le (1+\sqrt{2}\ ) + \int_{(1+\sqrt{2})}^{\infty} P\Big(\psi_{(r)}^{-2}\|\hat{f}_{(r)} - f\|_\infty^2 \ge t\Big)dt \\
&\le (1+\sqrt{2}) + \int_{(1+\sqrt{2})}^{\infty} (K_{(r)}+1)\exp\Big\{-\frac{(2t^{1/2}-1)^2}{2r+1}\log n\Big\}dt \\
&\le (1+\sqrt{2}) + c_5 \cdot n^{-\frac{1}{2r+1}} \le (1+\sqrt{2}) + c_5 =: C_5.
\end{aligned}$$

This completes the proof. □

We are ready to prove the theorem. Observe that

$$\sup_{f\in\mathcal{C}_H(r,L)} \mathbb{E}\Big\{\psi_{(r)}^{-1}\|\hat{f}_{(\hat{r})} - f\|_\infty\Big\} \le R^- + R^+,$$

where $R^- := \sup_{f\in\mathcal{C}_H(r,L)} \mathbb{E}\Big\{\psi_{(r)}^{-1}\|\hat{f}_{(\hat{r})} - f\|_\infty\ I\{\hat{r} < r\}\Big\}$, and $R^+ := \sup_{f\in\mathcal{C}_H(r,L)} \mathbb{E}\Big\{\psi_{(r)}^{-1}\|\hat{f}_{(\hat{r})} - f\|_\infty\ I\{\hat{r} \ge r\}\Big\}$. Hence, it suffices to show that

$$\limsup_{n\to\infty} \sup_{r\in[1,2]} R^- = 0, \tag{8.2}$$

$$\limsup_{n\to\infty} \sup_{r\in[1,2]} R^+ < \infty. \tag{8.3}$$

We first prove (8.2). By the definition of $\hat{r}$ given before Theorem 3.4,

$$R^- \le \sum_{r>r'\in\mathcal{R}} \sup_{f\in\mathcal{C}_H(r,L)} \mathbb{E}\Big(\psi_{(r)}^{-1}\|\hat{f}_{(\hat{r})} - f\|_\infty\ I\{\hat{r} = r'\}\Big) \le \rho_1 + \rho_2,$$

where, in view of (3.3) and $C_1 L K_{(r')}^{-r'} = \psi_{(r')}/2$ (for ρ_2 below),

$$\rho_1 := \sum_{r>r'\in\mathcal{R}} \sup_{f\in\mathcal{C}_H(r,L)} P\big(\hat{r} = r'\big)\psi_{(r)}^{-1}\Big(\frac{1}{2}\psi_{(r')} + \frac{\sqrt{2}}{2}\psi_{(r')}\Big),$$

$$\rho_2 := \sum_{r>r'\in\mathcal{R}} \psi_{(r)}^{-1}\mathbb{E}\Big(\Big[\frac{\psi_{(r')}}{2} + \|\hat{f}_{(r')} - \bar{f}_{(r')}\|_\infty\Big] I\{\|\hat{f}_{(r')} - \bar{f}_{(r')}\|_\infty \ge \frac{\sqrt{2}}{2}\psi_{(r')}\}\Big).$$

We will prove $\sup_{r\in[1,2]}\rho_j = o(1)$ as $n\to\infty$, $j=1,2$. Since $r'<r$, it is easy to see that there exists a positive constant c_6 such that

$$\frac{\psi_{(r')}}{\psi_{(r)}} \le c_6\Big(\frac{\log n}{n}\Big)^{\frac{(r-r')}{(2r+1)(2r'+1)}} \le c_6\Big(\frac{\log n}{n}\Big)^{\frac{r-r'}{5(2r+1)}} \le c_6.$$

It thus follows from Lemma 8.2 that, as $n\to\infty$,

$$\rho_1 \le \tau_n\cdot C_3\tau_n n^{-\frac{1}{2r+1}}\cdot\frac{(1+\sqrt{2})c_6}{2} \le C_3\frac{(1+\sqrt{2})c_6}{2}\log n\cdot n^{-\frac{1}{2r+1}} \longrightarrow 0.$$

Further, from (3.4), we deduce the existence of a constant $c_7>0$ such that

$$\begin{aligned}\sum_{r>r'\in\mathcal{R}} &\psi_{(r)}^{-1}\mathbb{E}\Big(\psi_{(r')}I\{\|\hat f_{(r')}-\bar f_{(r')}\|_\infty \ge \frac{\sqrt{2}}{2}\psi_{(r')}\}\Big)\\ &= \sum_{r>r'\in\mathcal{R}}\frac{\psi_{(r')}}{\psi_{(r)}}P\Big\{\|\hat f_{(r')}-\bar f_{(r')}\|_\infty\ge\frac{\sqrt{2}}{2}\psi_{(r')}\Big\} \le \tau_n\cdot c_6\cdot c_7\cdot n^{\frac{-1}{2r'+1}}.\end{aligned}$$

Also, it follows from Lemma 8.1 that a constant $c_8>0$ exists such that for all large n,

$$\begin{aligned}\sum_{r>r'\in\mathcal{R}} &\psi_{(r)}^{-1}\mathbb{E}\Big(\|\hat f_{(r')}-\bar f_{(r')}\|_\infty I\{\|\hat f_{(r')}-\bar f_{(r')}\|_\infty\ge\frac{\sqrt{2}}{2}\psi_{(r')}\}\Big)\\ &= \sum_{r>r'\in\mathcal{R}}\psi_{(r)}^{-1}c_8\big(\log n\big)^{\frac{1}{2(2r'+1)}}\cdot n^{-\frac{r'+3}{2r'+1}} \le \tau_n\psi_{(r)}^{-1}c_8 n^{-\frac{1}{2}}.\end{aligned}$$

By virtue of the above results, we have $\rho_2\to 0$ as $n\to\infty$. This yields (8.2).

We now prove (8.3). Consider the random event $\aleph(r,r') := \big\{\psi_{(r)}^{-1}\|\hat f_{(r')}-f\|_\infty \ge 1+\sqrt{2},\ r'\in\mathcal{R}\big\}$. Then

$$\begin{aligned}R^+ &\le \sup_{f\in\mathcal{C}_H(r,L)}\sum_{r\le r'\in\mathcal{R}}\mathbb{E}\Big(\psi_{(r)}^{-1}\|\hat f_{(r')}-f\|_\infty I\{\hat r=r'\}\Big)\\ &\le (1+\sqrt{2}\,)\sup_{f\in\mathcal{C}_H(r,L)}P\{\hat r\ge r\}\\ &\quad+\sum_{r\le r'\in\mathcal{R}}\sup_{f\in\mathcal{C}_H(r,L)}\mathbb{E}\Big(\psi_{(r)}^{-1}\|\hat f_{(r')}-f\|_\infty I\{\{\hat r=r'\}\cap\aleph(r,r')\}\Big)\\ &\le (1+\sqrt{2}\,)+\sum_{r\le r'\in\mathcal{R}}\left(\sup_{f\in\mathcal{C}_H(r,L)}(\mathbb{E}\big(\psi_{(r)}^{-2}\|\hat f_{(r')}-f\|_\infty^2\big))^{1/2}\sup_{f\in\mathcal{C}_H(r,L)}\rho_f^{1/2}(r,r')\right),\end{aligned}$$

where $\rho_f(r,r') := P(\{\hat r=r'\}\cap\aleph(r,r'))$ for $r'\in\mathcal{R}$. Let $r_* := \min\{r'\in\mathcal{R}:\ r'>r\}$. Hence, $r_*-r\in[0,1/\tau_n)$. In view of (3.6), we have $\psi_{(s)} =$

$p(s)\cdot(n^{-1}\log n)^{s/(2s+1)}$ for a function $p(\cdot)$ that is positive and continuous on $[1,2]$. Let $p_* := \min_{s\in[1,2]} p(s) > 0$. This shows that

$$\psi_{(r_*)} = p(r_*)\Big(\frac{\log n}{n}\Big)^{\frac{r_*}{(2r_*+1)}} \le \Big[1+\frac{p(r_*)-p(r)}{p_*}\Big]p(r)\Big(\frac{\log n}{n}\Big)^{\frac{r}{(2r+1)}} = \big[1+o(1)\big]\psi_{(r)}.$$

Hence, if $\hat r = r' \ge r$, then, by the definition of $\hat r \in \mathcal{R}$, $\|\hat f_{(r')} - \hat f_{(r_*)}\|_\infty \le \frac{1}{2}(1+\sqrt{2})\psi_{r_*} \le \frac{1}{2}(1+\sqrt{2}+o(1))\psi_{(r)}$. Therefore,

$$\begin{aligned}
\psi_{(r)}^{-1}\|\hat f_{(r')} - f\|_\infty &\le \psi_{(r)}^{-1}\cdot\Big(\|\hat f_{(r')} - \hat f_{(r_*)}\|_\infty + \|\hat f_{(r_*)} - f\|_\infty\Big)\\
&\le \frac{1}{2}\big(1+\sqrt{2}+o(1)\big) + \psi_{(r)}^{-1}\psi_{(r_*)}\cdot\big[\psi_{(r_*)}^{-1}\|\hat f_{(r_*)} - f\|_\infty\big]\\
&\le \frac{1}{2}\big(1+\sqrt{2}+o(1)\big) + [1+o(1)]\cdot\psi_{(r_*)}^{-1}\|\hat f_{(r_*)} - f\|_\infty.
\end{aligned}$$

It follows from Lemma 8.3 that for all large n,

$$\mathbb{E}\Big\{\psi_{(r)}^{-2}\|\hat f_{(r')} - f\|_\infty^2\Big\} \le 2\big[4+(1+o(1))\cdot\mathbb{E}\big(\psi_{(r_*)}^{-2}\|\hat f_{(r_*)} - f\|_\infty^2\big)\big] \le 2(4+2C_5).$$

We consider $\rho_f(r,r')$ next. By Lemma 8.3, we have, when $r' = r$,

$$\sup_{f\in\mathcal{C}_H(r,L)} \rho_f(r,r) \le \sup_{f\in\mathcal{C}_H(r,L)} P\{\aleph(r,r)\} \le C_4 n^{-1}.$$

Now consider $\rho_f(r,r')$ with $r' > r$. Since $\bar f_{(s)}(\cdot)$ is a continuous function for any s, there exists a nonrandom point $t_* \in [0,1]$ such that $|\bar f_{(r')}(t_*) - \bar f_{(r_*)}(t_*)| = \|\bar f_{(r')} - \bar f_{(r_*)}\|_\infty$. Let $\xi_* := (\hat f_{(r_*)}(t_*) - \bar f_{(r_*)}(t_*)) - (\hat f_{(r')}(t_*) - \bar f_{(r')}(t_*))$ Therefore, if $\hat r = r' > r$, then

$$\begin{aligned}
\|\bar f_{(r')} - \bar f_{(r_*)}\|_\infty &\le \big|\hat f_{(r')}(t_*) - \hat f_{(r_*)}(t_*)\big| + |\xi_*| \le \|\hat f_{(r')} - \hat f_{(r_*)}\|_\infty + |\xi_*|\\
&\le \frac{1}{2}\big(1+\sqrt{2}+o(1)\big)\psi_{(r)} + |\xi_*|.
\end{aligned}$$

Using this result and $\|\bar f_{(r_*)} - f\|_\infty \le \psi_{(r_*)}/2$, we have

$$\begin{aligned}
\|\hat f_{(r')} - f\|_\infty &\le \|\bar f_{(r')} - \bar f_{(r_*)}\|_\infty + \|\hat f_{(r')} - \bar f_{(r')}\|_\infty + \|\bar f_{(r_*)} - f\|_\infty\\
&\le \frac{1}{2}\big(1+\sqrt{2}+o(1)\big)\psi_{(r)} + |\xi_*| + \|\hat f_{(r')} - \bar f_{(r')}\|_\infty + \frac{1+o(1)}{2}\psi_{(r)}.
\end{aligned}$$

As a result, for $r < r' \in \mathcal{R}$, we further deduce via Markov Inequality,

$$\begin{aligned}
\rho_f(r,r') &\le P\Big(|\xi^*| + \|\hat f_{(r')} - \bar f_{(r')}\|_\infty \ge \frac{\sqrt{2}+o(1)}{2}\psi_{(r)}\Big)\\
&\le P\Big(\|\hat f_{(r')} - \bar f_{(r')}\|_\infty \ge \frac{1.1}{2}\psi_{(r_*)}\Big) + P\Big(|\xi_*| \ge \frac{\sqrt{2}-1.1+o(1)}{2}\psi_{(r_*)}\Big)\\
&\le P\Big(\|\hat f_{(r')} - \bar f_{(r')}\|_\infty \ge \frac{1.1}{2}\psi_{(r_*)}\Big) + 10^2\cdot\psi_{(r_*)}^{-2}\cdot\mathbb{E}(|\xi_*|^2).
\end{aligned}$$

It follows from (3.4) and $r_* \le r' \le 2$ that there exist constants $c_9 > 0$ and $\alpha > 0$ (independent of $r \in [1,2]$) such that $r' > r_* \Rightarrow K_{(r_*)}/K_{(r')} \ge \alpha(n/\log n)^{2/(25\tau_n)} \ge 1$ for all large n and that

$$P\Big(\|\hat f_{(r')} - \bar f_{(r')}\|_\infty \ge \frac{1.1}{2}\psi_{(r_*)}\Big) \le c_9\Big(\frac{n}{\log n}\Big)^{\frac{1}{(2r'+1)}} \cdot \exp\Big\{-\frac{1.21}{2r_*+1}\cdot\frac{K_{(r_*)}}{K_{(r')}}\log n\Big\}$$
$$\le c_9 n^{\frac{1}{(2r'+1)}-\frac{1.21}{(2r_*+1)}} \le c_9 n^{-\frac{0.21}{(2r_*+1)}} \le c_9 n^{-\frac{0.2}{5}}.$$

Using Proposition 7.3, we also obtain a constant $c_{10} > 0$ such that

$$\mathbb{E}(|\xi_*|^2) \le 2\Big[\mathbb{E}(|\hat f_{(r_*)}(t_*) - \bar f_{(r_*)}(t_*)|^2) + \mathbb{E}(|\hat f_{(r')}(t_*) - \bar f_{(r')}(t_*)|^2)\Big]$$
$$\le c_{10}\frac{K_{(r_*)} + K_{(r')}}{n} \le 2\cdot c_{10}\frac{K_{r(*)}}{n}.$$

Noting that $\psi^2_{(r_*)} \ge \widetilde\alpha(K_{(r_*)}/n)\log n$ for a positive constant $\widetilde\alpha$ independent of r_*, we have $\psi^{-2}_{(r_*)}\cdot\mathbb{E}(|\xi_*|^2) \le 2c_{10}\big(\widetilde\alpha\cdot\log n\big)^{-1}$. Hence, there exists a constant $\widetilde c_{10} > 0$ (independent of $r, r' \in [1,2]$) such that for all n sufficiently large,

$$\rho_f^{1/2}(r,r') \le \widetilde c_{10}\cdot\big(\log n\big)^{-1/2} \implies R^+ \le (1+\sqrt2) + \sqrt{2(4+2C_5)}\cdot\widetilde c_{10}.$$

This leads to (8.3), and thus completes the proof of Theorem (3.4).

9. Proof of Theorem 3.5. We establish the following lemma to be used for the analysis of the risk of $\widetilde f(x_0)$.

LEMMA 9.1. *Suppose that f is convex and differentiable on $[0,1]$. Then there exists a positive constant C_6 independent of f such that for each j,*

$$\mathbb{E}\big(|\widetilde f_j(x_0) - f(x_0)|^2 I_j\big) \le C_6 2^j n^{-\frac45}\sigma^2. \tag{9.1}$$

PROOF. Recall that $\zeta_k = \frac1n((k-1)M_n + \frac{M_n+1}{2}), \forall\, k = 1,\dots,K_{n,j}$ Let $\grave f$ and $\check f$ be two piecewise linear functions such that $\grave f(\zeta_k) = \mathbb{E}(\bar y_{k,j})$ and $\check f(\zeta_k) = f(\zeta_k)$, respectively. Note that if M_n is odd, then ζ_k is a design point so that $\mathbb{E}(\bar y_k) - f(\zeta_k) = 0$. Otherwise, direct calculation yields that

$$\mathbb{E}(\bar y_k) - f(\zeta_k) = \frac{1}{M_n}\sum_{j=1}^{M_n}\Big[f(x_{(k-1)M_n+j}) - f(\zeta_k)\Big]$$
$$= \frac{1}{M_n}\sum_{j=1}^{M_n/2}\Big[\big(f(x_{(k-1)M_n+M_n-j+1}) - f(\zeta_k)\big) - \big(f(\zeta_k) - f(x_{(k-1)M_n+j})\big)\Big]$$
$$= \frac{1}{M_n}\sum_{j=1}^{M_n/2}\frac{\frac{M_n}{2}-j}{n}\Bigg[\frac{f(x_{(k-1)M_n+M_n-j+1}) - f(\zeta_k)}{(\frac{M_n}{2}-j)/n} - \frac{f(\zeta_k) - f(x_{(k-1)M_n+j})}{(\frac{M_n}{2}-j)/n}\Bigg].$$

Since f is a convex function, we have $\mathbb{E}(\bar{y}_k) - f(\zeta_k) \geq 0$ and

$$\begin{aligned}\mathbb{E}(\bar{y}_k) - f(\zeta_k) &\leq \frac{1}{M_n}\sum_{j=1}^{M_n/2}\frac{M_n/2}{n}\left[\frac{f(\zeta_{k+1})-f(\zeta_k)}{1/K_{n,j}} - \frac{f(\zeta_k)-f(\zeta_{k-1})}{1/K_{n,j}}\right]\\ &= \frac{1}{4}\Delta^2 f(\zeta_{k+1}).\end{aligned}$$

Hence, for any $x_0 \in (\zeta_{d_n}, \zeta_{d_n+1}]$,

$$\begin{aligned}0 \leq \grave{f}(x_0) - \check{f}(x_0) &\leq \frac{1}{4}\max\left\{\Delta^2 f(\zeta_{d_n+1}), \Delta^2 f(\zeta_{d_n+2})\right\} \leq \Delta^2 f(\zeta_{d_n+1}) + \Delta^2 f(\zeta_{d_n+2})\\ &\leq \Delta f(\zeta_{d_n+2}) - \Delta f(\zeta_{d_n}).\end{aligned}$$

Further, since f is convex, $f(x_0) - f(\zeta_{d_n}) \geq f'(\zeta_{d_n})(x_0 - \zeta_{d_n})$ such that

$$\begin{aligned}0 \leq \check{f}(x_0) - f(x_0) &= f(\zeta_{d_n}) + K_n(f(\zeta_{d_n+1}) - f(\zeta_{d_n}))(x_0 - \zeta_{d_n}) - f(x_0)\\ &\leq [f(\zeta_{d_n+1}) - f(\zeta_{d_n}) - K_n^{-1}f'(\zeta_{d_n})]K_n(x_0 - \zeta_{d_n})\\ &\leq [f(\zeta_{d_n+1}) - f(\zeta_{d_n})] - [f(\zeta_{d_n}) - f(\zeta_{d_n-1})]\\ &\leq \Delta f(\zeta_{d_n+2}) - \Delta f(\zeta_{d_n}).\end{aligned}$$

Let $\tau_j := \mathbb{E}(\Delta\bar{y}_{d_n+4,j} - \Delta\bar{y}_{d_n-2,j})$. Hence, $0 \leq \grave{f}(x_0) - f(x_0) \leq 2\tau_j$.

Notice that for $x_0 \in (\zeta_{d_n}, \zeta_{d_n+1}]$, there exists $\mu \in (0,1]$ such that $\widetilde{f}_j(x_0) = \mu\hat{f}(\zeta_{d_n}) + (1-\mu)\hat{f}(\zeta_{d_n+1})$ and $\grave{f}(x_0) = \mu\bar{f}(\zeta_{d_n}) + (1-\mu)\bar{f}(\zeta_{d_n+1})$, where $\hat{f}$ and $\bar{f}$ are the piecewise constant splines (i.e., $p = 0$) corresponding to the convex constrained least squares for $(\bar{y}_{k,j})$ and $(\mathbb{E}(\bar{y}_{k,j}))$, respectively. It follows from Proposition 7.3 that a positive constant c_{11} exists such that

(9.2)

$$\begin{aligned}&\mathbb{E}[(\widetilde{f}_j(x_0) - \grave{f}(x_0))^2 I_j] \;\leq\; 2\big(\mu^2\mathbb{E}|\hat{f}(\zeta_{d_n}) - \bar{f}(\zeta_{d_n})|^2 + (1-\mu)^2\mathbb{E}|\hat{f}(\zeta_{d_n+1}) - \bar{f}(\zeta_{d_n+1})|^2\big)\\ &\qquad\leq\; 2[\mu + (1-\mu)]C_3\sigma^2 K_{n,j}\cdot n^{-1} \;\leq\; c_{11}\sigma^2 2^j n^{-4/5}.\end{aligned}$$

This implies

$$\begin{aligned}&\mathbb{E}[(\widetilde{f}_j(x_0) - f(x_0))^2 I_j] \;\leq\; 2\Big(\mathbb{E}[(\grave{f}_j(x_0) - f(x_0))^2 I_j] + \mathbb{E}[(\widetilde{f}_j(x_0) - \grave{f}(x_0))^2 I_j]\Big)\\ &\qquad\leq\; 8\tau_j^2\mathbb{E}(I_j) + 2\mathbb{E}[(\widetilde{f}_j(x_0) - \grave{f}(x_0))^2] \leq 8\tau_j^2\mathbb{E}(I_j) + c_{11}\sigma^2 2^{j+1}n^{-4/5}.\end{aligned}$$

If $\tau_j \leq 2\lambda 2^{j/2+1}n^{-2/5}\sigma$, then (9.1) holds. We thus only consider the case when $\tau_j > 2\lambda 2^{j/2+1}n^{-2/5}\sigma$. In this case, note that $\Delta\bar{y}_{d_n+4,j} - \Delta\bar{y}_{d_n-2,j}$ has a normal distribution with mean τ_j and variance $2^{j+2}n^{-4/5}\sigma^2$. Consequently,

$$\begin{aligned}\mathbb{E}(I_j) &\le \mathbb{E}I\Big(\Delta\bar{y}_{d_n+4,j} - \Delta\bar{y}_{d_n-2,j} \le \lambda 2^{\frac{j}{2}+1} n^{-\frac{2}{5}}\sigma\Big)\\ &\le P\left(Z \le \lambda - \frac{\tau_j n^{\frac{2}{5}}}{2^{\frac{j}{2}+1}\sigma}\right) \le P\left(Z \le -\frac{\tau_j n^{\frac{2}{5}}}{2^{\frac{j}{2}+2}\sigma}\right) \le \exp\left(-\frac{\tau_j^2 n^{\frac{4}{5}}}{2^{(j+5)}\sigma^2}\right),\end{aligned}$$

where Z is a standard normal random variable. In view of $\sup_{z>0} z\exp(-\frac{\rho^2 z}{2}) = 2\rho^{-2}e^{-1}$, we obtain $\tau_j^2\mathbb{E}(I_j) \le 2e^{-1}2^{j+4}n^{-4/5}\sigma^2$. Hence (9.1) holds. $\square$

With this lemma, we show as follows:

Proof of Theorem 3.5. Recall that $K_{n,j} := 2^j n^{1/5}$. If $f \in \mathcal{C}_H(r, L)$ with $r \in [1, 2]$, then $\tau_j = \Delta f(\zeta_{d_n+2}) - \Delta f(\zeta_{d_n}) \le 2LK_{n,j}^{-r} = 2L(2^{-j}n^{-1/5})^r$. Let J be the smallest natural number (dependent on r) such that

$$2^J n^{-\frac{4}{5}} \ge L^{\frac{2}{2r+1}}\sigma^{-\frac{2}{2r+1}}n^{-\frac{2r}{2r+1}}. \tag{9.3}$$

If $j \ge J$, then $2^{jr} \ge L^{2r/(2r+1)}\sigma^{-2r/(2r+1)}n^{(4r-2r^2)/(10r+5)}$. Hence, this shows via (9.3) that, for $j \ge J$,

$$\tau_j \le 2L \cdot 2^{-jr} \cdot n^{-\frac{r}{5}} \le 2L^{\frac{1}{2r+1}}\sigma^{\frac{2r}{2r+1}}n^{-\frac{r}{2r+1}} \le 2 \cdot 2^{\frac{J}{2}}n^{-\frac{2}{5}}\sigma. \tag{9.4}$$

Recall that $P(Z > \lambda) < 1/4$ for the positive λ. Then there exists a $\delta > 0$ such that $\gamma := P(Z \ge \lambda - \delta) < 1/4$. Note that $4\gamma < 1$. Given this δ, choose $K \in \mathbb{N}$ (independent of r) such that $1 < \delta \cdot 2^{K/2}$. Since only one $I_j \ne 0$ and $I_j^2 = I_j$, $|\widetilde{f}(x_0) - f(x_0)|^2 = \sum_{j=0}^{\infty}(\widetilde{f}_j(x_0) - f(x_0))^2 I_j$ such that the risk of $\widetilde{f}$ is decomposed into two terms:

$$\mathbb{E}|\widetilde{f}(x_0)-f(x_0)|^2 = \sum_{j=0}^{J+K}\mathbb{E}[(\widetilde{f}_j(x_0)-f(x_0))^2 I_j] + \sum_{j=J+K+1}^{\infty}\mathbb{E}[(\widetilde{f}_j(x_0)-f(x_0))^2 I_j]. \tag{9.5}$$

We consider the first sum in (9.5). It follows from Lemma 9.1 that

$$\sum_{j=0}^{J+K}\mathbb{E}[(\widetilde{f}(x_0) - f(x_0))^2 I_j] \le \sum_{j=0}^{J+K} C_6 2^j n^{-\frac{4}{5}}\sigma^2 \le C_6 2^{K+1}2^J n^{-\frac{4}{5}}\sigma^2.$$

Since J is the smallest integer satisfying (9.3), we have

$$2^J n^{-\frac{4}{5}} \le 2 \cdot L^{\frac{2}{2r+1}}\sigma^{-\frac{2}{2r+1}}n^{-\frac{2r}{2r+1}}. \tag{9.6}$$

Therefore,

$$\sum_{j=0}^{J+K} \mathbb{E}[(\widetilde{f}(x_0) - f(x_0))^2 I_j] \;\le\; C_6 2^{K+2} L^{\frac{2}{2r+1}} \sigma^{\frac{4r}{2r+1}} n^{-\frac{2r}{2r+1}}. \tag{9.7}$$

Consider the second sum in (9.5). We show two technical results. Firstly, since the design points corresponding to $\Delta\bar{y}_{d_n+4,j} - \Delta\bar{y}_{d_n-2,j}$ are disjoint for different j, $\Delta\bar{y}_{d_n+4,j} - \Delta\bar{y}_{d_n-2,j}$ are independent. Hence, for $j > J+K$,

$$\begin{aligned}\mathbb{E}(I_j) &\le \prod_{i=J+K}^{j-1} P\Big(\Delta\bar{y}_{d_n+4,i} - \Delta\bar{y}_{d_n-2,i} > \lambda 2^{\frac{i}{2}+1} n^{-\frac{2}{5}} \sigma\Big) \\ &\le \Big\{P\Big(Z > \lambda - \frac{1}{2^{K/2}}\Big)\Big\}^{j-J-K} \;\le\; \gamma^{j-J-K}.\end{aligned}$$

Secondly, in view of the argument for (9.2) and Proposition 7.3, we have $\mu \in [0,1]$ and a constant $c_{12} > 0$ such that

$$\begin{aligned}&\mathbb{E}|\widetilde{f}_j(x_0) - \grave{f}(x_0)|^4 \\ &\quad\le\; 8\big(\mu^4 \mathbb{E}|\hat{f}(\zeta_{d_n}) - \bar{f}(\zeta_{d_n})|^4 + (1-\mu)^4 \mathbb{E}|\hat{f}(\zeta_{d_n+1}) - \bar{f}(\zeta_{d_n+1})|^4\big) \\ &\quad\le\; 8C_4 \sigma^4 K_{n,j}^2 \cdot n^{-2} \;\le\; c_{12}^2 \sigma^4 \cdot \big(2^j \cdot n^{-\frac{4}{5}}\big)^2.\end{aligned}$$

Using these results and $\gamma \in (0, 1/4)$, we obtain a constant $c_{13} > 0$ such that

$$\begin{aligned}&\sum_{j=J+K+1}^{\infty} \mathbb{E}\Big((\widetilde{f}(x_0) - \grave{f}(x_0))^2 I_j\Big) \\ &\le \sum_{j=J+K+1}^{\infty} \Big(\mathbb{E}|\widetilde{f}(x_0) - \grave{f}(x_0)|^4 \cdot \mathbb{E}(I_j^2)\Big)^{1/2} \;\le\; \sum_{j=J+K+1}^{\infty} c_{12} 2^j n^{-\frac{4}{5}} \sigma^2 \cdot \gamma^{\frac{(j-J-K)}{2}} \\ &\le \sum_{j=J+K+1}^{\infty} c_{12} (4\gamma)^{(j-J)/2} \gamma^{-K/2} 2^J n^{-\frac{4}{5}} \sigma^2 \;\le\; c_{13} 2^J n^{-\frac{4}{5}} \sigma^2.\end{aligned}$$

Furthermore, since f is convex, $\tau_j = \mathbb{E}(\Delta\bar{y}_{d_n+4,j} - \Delta\bar{y}_{d_n-2,j})$ is a decreasing function of j. Therefore, in view of $0 \le \grave{f}(x_0) - f(x_0) \le 2\tau_j$ and (9.4),

$$\sum_{j=J+K+1}^{\infty} \mathbb{E}\Big((\grave{f}(x_0) - f(x_0))^2 I_j\Big) \;\le\; \sum_{j=J+K+1}^{\infty} 4\tau_J^2 \cdot \mathbb{E}(I_j) \;\le\; 4\tau_J^2 \le 16 \cdot 2^J n^{-\frac{4}{5}} \sigma^2.$$

By virtue of the above results and (9.6), we obtain $c_{14} > 0$ such that

$$\sum_{j=J+K+1}^{\infty} \mathbb{E}[(\widetilde{f}(x_0) - f(x_0))^2 I_j] \;\le\; c_{14} L^{\frac{2}{2r+1}} \sigma^{\frac{4r}{2r+1}} n^{-\frac{2r}{2r+1}}. \tag{9.8}$$

Hence, the theorem follows by combining (9.7) and (9.8). □

10. Proof of Theorem 3.6. Recall that $\hat{f}_y := \big(\hat{f}^{[p]}(x_1), \dots, \hat{f}^{[p]}(x_n)\big)^T$ with coefficient matrices A_α defined in (3.10) and $\vec{f} := \big(f(x_1), \dots, f(x_n)\big)^T$.

LEMMA 10.1. *Fix a spline degree p. The following hold:*

(1) For any index set α, $0 \le \operatorname{trace}(A_\alpha) \le c_{\infty,p}(K_n + p)$;
(2) $\mathbb{E}\big[\langle y - \vec{f}, \hat{f}_y - \vec{f}\,\rangle\big] = \sigma^2 \mathbb{E}\big[\operatorname{trace}\big(A_{\alpha(y)}\big)\big] \le c_{\infty,p}\,(K_n + p)\,\sigma^2$.

PROOF. (1) Since A_α is symmetric positive semidefinite, its trace is nonnegative. For a given α, recall that $G_\alpha := F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha$. Hence $A_\alpha = X G_\alpha X^T / \beta_n$. By virtue of $\|G_\alpha\|_\infty \le c_{\infty,p}$ from Theorem 3.1 and $|(\Lambda)_{ij}| \le 1$ for all i, j (by the definition of β_n), we have, for any α, $0 \le \operatorname{trace}\big(A_\alpha\big) = \operatorname{trace}\Big(G_\alpha \cdot \frac{X^T X}{\beta_n}\Big) = \operatorname{trace}(G_\alpha \Lambda) \le c_{\infty,p}(K_n + p)$.

(2) Since $\hat{f}_y : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and piecewise linear, it admits a conic subdivision of $\mathbb{R}^n$ [10, 33], i.e., there exist a finite collection of polyhedral cones $\{\mathcal{C}_j\}_{j=1}^{\ell}$ and linear functions $\{g^j\}_{j=1}^{\ell}$ satisfying the similar conditions as specified in the proof of Proposition 7.3. In particular, each cone $\mathcal{C}_j$ has nonempty interior and $\hat{f}_y$ coincides with g^j on each $\mathcal{C}_j$. Clearly, $g^j(y) = A_\alpha y$ for some index set α. In this case, we write the cone $\mathcal{C}_j$ as $\mathcal{C}_\alpha$. Let $\operatorname{int}(\mathcal{C}_\alpha)$ denote the interior of $\mathcal{C}_\alpha$. Obviously, $\hat{f}_y$ is differentiable on $\operatorname{int}(\mathcal{C}_\alpha)$. Indeed, the (Fréchet-)derivative of $\hat{f}_y$ is A_α for any $y \in \operatorname{int}(\mathcal{C}_\alpha)$. Let $h(y) := \hat{f}_y - \vec{f}$. Since $\mathbb{R}^n \setminus \big(\bigcup_j \operatorname{int}(\mathcal{C}_j)\big)$ has zero measure, h is almost differentiable on $\mathbb{R}^n$ in the sense of [37, Definition 1]. Let $\phi(\mathbf{z})$ be the standard normal density on $\mathbb{R}^n$ with variance σ^2. We have

$$
\begin{aligned}
\mathbb{E}\Big(\Big\|\frac{\partial h}{\partial y}(y)\Big\|\Big) &= \int_{\bigcup_j \operatorname{int}(\mathcal{C}_j)} \Big\|\frac{\partial h}{\partial y}(\mathbf{z} + \vec{f})\Big\| \phi(\mathbf{z}) d\mathbf{z} \\
&= \sum_\alpha \int_{\mathbf{z} + \vec{f} \in \operatorname{int}(\mathcal{C}_\alpha)} \|A_\alpha\| \phi(\mathbf{z}) d\mathbf{z} \le \max_\alpha \|A_\alpha\| < \infty.
\end{aligned}
$$

Letting $Z := \sigma(\epsilon_1, \dots, \epsilon_n)^T$, we have $\mathbb{E}[\langle y - \vec{f}, \hat{f}_y - \vec{f}\rangle] = \mathbb{E}[\langle Z, h(y)\rangle] = \operatorname{trace}\big(\mathbb{E}[Z \cdot h^T(y)]\big)$. By the above results and Stein's Lemma [37, Lemma 2],

$$
\begin{aligned}
&\operatorname{trace}\big(\mathbb{E}[Z \cdot h^T(y)]\big) \\
&= \sigma^2 \operatorname{trace}\Big(\mathbb{E}\Big[\frac{\partial h}{\partial y}(y)\Big]\Big) = \sigma^2 \int_{\bigcup_j \operatorname{int}(\mathcal{C}_j)} \operatorname{trace}\Big(\frac{\partial h}{\partial y}(\mathbf{z} + \vec{f})\Big) \phi(\mathbf{z}) d\mathbf{z} \\
&= \sigma^2 \sum_\alpha \int \operatorname{trace}\Big(\frac{\partial h}{\partial y}(\mathbf{z} + \vec{f})\Big) \phi(\mathbf{z}) \cdot I_{\{\mathbf{z} | \mathbf{z} + \vec{f} \in \operatorname{int}(\mathcal{C}_\alpha)\}} d\mathbf{z} \\
&= \sigma^2 \sum_\alpha \int \operatorname{trace}\big(A_\alpha\big) \phi(\mathbf{z}) \cdot I_{\{\mathbf{z} | \mathbf{z} + \vec{f} \in \operatorname{int}(\mathcal{C}_\alpha)\}} d\mathbf{z} \;=\; \sigma^2 \mathbb{E}\big(\operatorname{trace}(A_{\alpha(y)})\big).
\end{aligned}
$$

Statement (2) thus follows from (1). □

Equipped with the above lemma, we have the proof of Theorem 3.6 below.

Proof of Theorem 3.6. The MLE of σ^2 is $\hat{\sigma}^2 = \|y - \hat{f}_y\|_2^2/n$. Let $R_n := \|y - \vec{f}\|_2^2 - \|y - \hat{f}_y\|_2^2$ such that $\hat{\sigma}^2 = \|y - \vec{f}\|_2^2/n - R_n/n$. Since

$$R_n \,=\, \|y - \vec{f}\|_2^2 - \|y - \hat{f}_y\|_2^2 \,=\, 2\langle y - \vec{f}, \hat{f}_y - \vec{f}\,\rangle - \|\hat{f}_y - \vec{f}\|_2^2,$$

we have, by Lemma 10.1 and (7.8),

$$\begin{aligned} \big|\mathbb{E}(R_n)\big| &\le\, 2\sigma^2 \mathbb{E}(\mathrm{trace}(A_{\alpha(y)})) + \mathbb{E}\|\hat{f}_y - \vec{f}\|_2^2 \\ &\le 2\sigma^2 c_{\infty,p}(K_n + p) + n\big(C_{1r}^2 L^2 K_n^{-2r} + C_{3r}\sigma^2 K_n\, n^{-1}\big), \end{aligned}$$

where $p = \lceil r-1 \rceil$. This shows that $|\mathbb{E}(R_n)| = O(K_n + nK_n^{-2r})$. Hence, we deduce that (i) if $K_n = o(n)$ with $K_n \to \infty$ as $n \to \infty$, then $\hat{\sigma}^2 \to \sigma^2$ in probability; (ii) if $K_n = o(\sqrt{n})$ with $K_n \to \infty$ as $n \to \infty$, then $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ is asymptotically normal with mean zero and variance $2\sigma^4$; and (iii) if K_n is of order $n^{\frac{1}{2r+1}}$, then $\big|\mathbb{E}(\hat{\sigma}^2 - \sigma^2)\big|$ is of order $n^{\frac{-2r}{2r+1}}$. □

References.

[1] Barlow, R.E., Bartholomew, D. J. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression.* Wiley, New York.

[2] Bertin, K. (2004). Minimax exact constant in sup-norm for nonparametric regression with random design. *Journal of Statistical Planning and Inference*, **123**, 225-242.

[3] Birke, M. and Dette, H. (2006). Estimating a convex function in nonparametric regression. *Scandinavian Journal of Statistics*, **34**, 384-404.

[4] Brown, L. and Low, M. (1996). A constrained risk inequality with applications to nonparametric estimation. *Annals of Statistics*, **24**, 2524-2535.

[5] Brunk, H.D. (1958). On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics*, **29**, 437-454.

[6] Cator, E. (2011). Adaptivity and optimality of the monotone least squares estimator. *Bernoulli*, **17**, 714-735.

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

[8] Demko, S. (1977). Inverses of band matrices and local convergence of spline projections. *SIAM Journal on Numerical Analysis*, **14**, 616–619.

[9] Donoho, D. L. (1994). Asymptotic minimax risk for sup-norm loss: solution via optimal recovery. *Probability Theory Related Fields*, **99**, 145-170.

[10] Facchinei, F. and Pang, J.S. (2003). *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer-Verlag, New York.

[11] GROENEBOOM, P., JONGBLOED, G. AND WELLNER, J. A. (2001) Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.*, **29**, 1653-1698.

[12] HALL, P. AND HUANG, L. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, **29**, 625-647.

[13] HANSON, P.L. AND PLEDGER, G. (1976). Consistency in concave regression. *Annals of Statistics*, **4**, 1038-1050.

[14] HASTIE, T. AND TIBSHIRANI, R. (1990). *Generalized Additive Models.* Chapman and Hall, London.

[15] Hildreth, C. (1954). Point estimates of ordinates of concave functions. *Journal of American Statistical Association*, **49**, 598-619.

[16] HORN, R.A. AND JOHNSON, C.R. (1985). *Matrix Analysis.* Cambridge University Press.

[17] KIEFER, J. (1982). 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, 419-428.

[18] KOROSTELEV, A.P. (1993). Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory Probab. Appl.*, **38**, 737-743.

[19] LEMSKII, O.V. (1990). On problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, **35**, 454-466.

[20] LEMSKII, O.V. (1992). On problems of adaptive estimation in white Gaussian noise. In R.Z. Khasminskij (ed.), *Topics in Nonparametric Estimation*, Adv. Soviet Math. **12**, 87-106. Providence, RI: American Mathematical society.

[21] LEPSKI, O. V. AND SPOKOINY, V.G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, **25**, 2512-2546.

[22] LEPSKI, O. V. AND TSYBAKOV, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory Related Fields*, **117**, 17-48.

[23] LOW, M. AND KANG, Y. (2002). Estimating monotone functions. *Statistics & Probability Letters*, **56**, 361-367.

[24] MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.*, **19**, 741-759.

[25] MAMMEN, E. AND THOMAS-AGNAN, C. (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics*, **26**, 239-252.

[26] MEYER, M. AND WOODROOFE, M. (2000). On the degrees of freedom of shape-restricted regression. *Annals of Statistics*, **28**, 1083-1104.

[27] MEYER, M. (2008). Inference using shape-restricted regression splines. *Annals of Applied Statistics*, **2**, 1013-1033.

[28] MUKERJEE, H. (1988). Monotone nonparametric regression. *Annals of Statistics*, **16**, 741-750.

[29] NEMIROVSKI, A.. *Topics in Non-Parametric Statistics.* Available at `http://www2.isye.gatech.edu/~nemirovs/`.

[30] RAMSAY, J.O. (1988). Estimating smooth monotone functions. *J. Roy. Statist. c. Ser. B*, **60**, 365-375.

[31] Rice, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics*, **12**, 1215-1230.

[32] Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley Series in Probability and Mathematical Statistics.

[33] Scholtes, S. (1994). *Introduction to Piecewise Differentiable Equations*. Habilitation thesis, Institut für Statistik und Mathematische Wirtschaftstheorie, Universität Karlsruhe.

[34] Shen, J. and Pang, J.S. (2005). Linear complementarity systems: Zeno states. *SIAM Journal on Control and Optimization*, **44(3)**, 1040–1066.

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

[36] Shen, J. and Wang, X. (2011). A constrained optimal control approach to smoothing splines. *Proc. of the 50th IEEE Conf. on Decision and Control*, 1792–1734, Orlando, Florida.

[37] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, **9**, 1135-1151.

[38] Stone, C.J. (1982). Optimal rate of convergence for nonparametric regression. *Annals of Statistics*, **10**, 1040-1053.

[39] Tsybakov, A.B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Annals of Statistics*, **26**, 2420-2469.

[40] Tsybakov, A.B. (2010). *Introduction to Nonparametric Estimation*. Springer.

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

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

[43] Wright, F.T. (1981). The asymptotic behavior of monotone regression estimates. *Annals of Statistics*, **9**, 443-448.

[44] Zhou, S., Shen, X. and Wolfe, D.A. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics*, **26**, 1760-1782.

[45] Zhou, S. and Wolfe, D. (2000). On derivative estimation in spline regression. *Statistica Sinica*, **10**, 93-108.

Dept. of Statistics
Purdue University
250 N. University Street
West Lafayette, IN 47907
USA

Dept. of Mathematics and Statistics
University of Maryland Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
USA