Estimation of Monotone Functions via P-Splines: A Constrained Dynamical Optimization Approach

Jinglai Shen* and Xiao Wang†‡

Original: October, 2009; Revised: April, 2010; Second Revision: November, 2010

Abstract

Estimation of monotone functions has broad applications in statistics, engineering, and science. This paper addresses asymptotic behaviors of monotone penalized spline estimators using constrained dynamical optimization techniques. The underlying regression function is approximated by a B-spline of an arbitrary degree subject to the first-order difference penalty. The optimality conditions for spline coefficients give rise to a size-dependent complementarity problem. As a key technical result of the paper, the uniform Lipschitz property of optimal spline coefficients is established by exploiting piecewise linear and polyhedral theory. This property forms a cornerstone for stochastic boundedness, uniform convergence, and boundary consistency of the monotone estimator. The estimator is then approximated by a solution of a differential equation subject to boundary conditions. This allows the estimator to be represented by a kernel regression estimator defined by a related Green's function of an ODE. The asymptotic normality is established at interior points via the Green's function. The convergence rate is shown to be independent of spline degrees, and the number of knots does not affect asymptotic distribution, provided that it tends to infinity fast enough.

1 Introduction

Various static or dynamic models of biologic, engineering and economic systems contain shape constrained functions; a typical example is monotone functions. Since the exact knowledge of these functions is usually unavailable and measurements pertaining to these functions are contaminated by random noise and disturbances, estimation of these functions becomes a critical problem across many fields [6, 24, 30]. In this paper, we consider the following monotone regression problem: estimate an unknown, nondecreasing function $f : [0, 1] \to \mathbb{R}$ using the sample $\{y_i\}$, where $y_i = f(x_i) + \sigma z_i$, $i = 1, \dots, n$, x_i is the ith design point, y_i is the ith sample, σ is the noise level, and z_i's are independent standard normal variables. In particular, we are interested in asymptotic behaviors of a monotone estimator, i.e., how an estimator behaves for a large n as it is expected that the estimator will be close to the true function f when $n \to \infty$. Focused issues in asymptotic analysis include consistency, asymptotic distribution, and convergence rates as $n \to \infty$.

*Dept. of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250, U.S.A. Email: shenj@umbc.edu.

†Dept. of Statistics, Purdue University, West Lafayette, IN 47909, U.S.A. Email: wangxiao@purdue.edu.

‡The preliminary version of this paper focusing on applications is given in the conference publication [24] without proofs.

The monotone regression problem has received considerable attention in statistics. For example, Brunk's estimator [1]

$$f_n(x_i) = \max_{s\le i} \min_{t\ge i} \frac{y_s + \cdots + y_t}{t-s+1} \tag{1}$$

is a well-known nonparametric maximum likelihood estimator [17]. This estimator has a non-normal asymptotic distribution and its convergence rate is of order $n^{1/3}$ [32]. Further, Brunk's estimator is not satisfactory when the regression function f is smooth. Another relevant approach is monotone smoothing (spline) estimation formulated as a constrained optimal control problem, i.e., to find a nondecreasing function f which minimizes

$$\sum_{i=1}^{n} \Big(y_i - f(x_i)\Big)^2 + \lambda \int_0^1 \Big(f^{(m)}(t)\Big)^2 dt \tag{2}$$

where $\lambda > 0$ is the penalty parameter. Asymptotic properties of this estimator have been developed for $m = 1$, and the attained estimator turns out to be a piecewise linear function [16]. However, a piecewise constant or linear function may not yield a satisfactory approximation in some applications. An example is estimation of the mass of a galaxy in astrophysics [30]. It is known that the mass behaves as a cubic function in certain range, where a piecewise constant or linear fit exhibits poor asymptotic behaviors. In order to obtain a smoother estimator, the second or higher order derivative of the regression function is needed in the penalty term. But this usually leads to tremendous difficulties in both asymptotic analysis and numerical computation due to the presence of the monotone constraint; see [8, 9, 28] for more discussions. Also see [27] for a recent application to (unconstrained) optimal control via smoothing splines. An alternative approach is the polynomial spline technique that has been extensively studied in approximation theory and statistics. Especially the penalized spline regression (simply P-splines or P-spline smoothing) [10, 29] has become popular over the last decade thanks to its highly tractable computation using low rank bases. The methodology and applications of P-spline estimators for *unconstrained* regression problems are discussed in [20], and their theoretical properties can be found in [2, 5, 7].

In this paper, we consider a monotone estimator via P-splines and investigate its asymptotic properties using optimization and ODE techniques, along with statistical theory. The first-order difference penalty is imposed but a B-spline can be of an arbitrary degree. This offers great flexibility for various applications, e.g., the galaxy estimation problem discussed above. The major difficulties of analyzing the monotone P-spline estimator lie in two aspects. First, due to the monotone constraint, the optimal spline coefficients are the solution of a size-dependent complementarity problem and are piecewise linear in $y = (y_1, \ldots, y_n)^T$. Unlike the unconstrained case, however, the explicit forms of these piecewise linear functions are generally unavailable and the number of linear pieces grows exponentially with respect to the number of knots. This complexity hinders a further investigation of analytic properties of the estimator. Second, the monotone condition hampers one from establishing an equivalent kernel of the estimator via matrix techniques widely adopted for the unconstrained counterpart. To overcome these difficulties, new optimization and ODE techniques are proposed that constitute the following contributions of the paper:

• The uniform Lipschitz property of optimal spline coefficients is established via piecewise linear and polyhedral theory (cf. Theorem 3.1). To the best of our knowledge, this critical property is the first of this kind established for a general class of P-spline estimators. Using this property, it is proved that the estimator is stochastically bounded and converges to the regression function in probability uniformly. These results lay a foundation for the subsequent asymptotic analysis.

For example, they lead to consistency of the estimator at boundary.

• Inspired by smoothing spline estimators [11, 16, 25], the monotone P-spline estimator is shown to be approximated by the solution of a dynamical complementarity system subject to boundary conditions. The estimator can be characterized by a kernel estimator, using a Green's function obtained from a related boundary value problem for an ODE. The asymptotic normality of the estimator and the convergence rates are established for different choices of spline degrees via the Green's function.

• The convergence rates of the estimator are shown to be independent of spline degrees and the number of knots, as long as the latter tends to infinity fast enough; see Theorems 4.1 and 4.2. Whereas this observation is pointed out in [7] for certain unconstrained cases, no rigorous justification has been given for the monotone P-spline estimator before. Furthermore, it is shown that if the number of knots grows sufficiently fast, then the modeling bias due to spline approximation is negligible compared to the shrinkage bias due to estimation by a penalized rather than ordinary least squares.

The rest of the paper is organized as follows. The monotone P-spline estimator is formulated and its optimality conditions are characterized in Section 2. Section 3 establishes the uniform Lipschitz property of optimal spline coefficients and indicates its critical implications. In Section 4, the monotone estimator is treated as an approximate solution of an ODE subject to boundary conditions, and the estimator is represented by a kernel regression estimator defined by a related Green's function; its asymptotic behaviors and convergence rates are obtained. Simulations and discussions are given in Section 5, where the proposed monotone estimator is compared with other estimators and various extensions are discussed. Section 6 concludes the paper with summary and remarks on future work.

2 Problem Formulation and Optimality Conditions

The regression function f is approximated by $f^{[p]}(x) = \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(x)$, where $\{B_k^{[p]} : k = 1, \ldots, K_n + p\}$ is the pth degree B-spline basis with knots $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{K_n} = 1$. The value of K_n depends upon n as discussed below. The spline coefficients $\hat{b} = \{\hat{b}_k, k = 1, \ldots, K_n + p\}$ subject to the first-order difference penalty are chosen to minimize

$$\sum_{i=1}^{n} \big[y_i - \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(x_i)\big]^2 + \lambda^* \sum_{k=2}^{K_n+p} \big[\Delta(b_k)\big]^2, \tag{3}$$

where $\lambda^* > 0$ and Δ is the backward difference operator, i.e., $\Delta b_k \equiv b_k - b_{k-1}$. The P-spline estimator is $\hat{f}^{[p]}(x) = \sum_{k=1}^{K_n+p} \hat{b}_k B_k^{[p]}(x)$. We consider the case where both the design points and the knots are equally spaced on the interval $[0, 1]$. We also assume that n/K_n is an integer denoted by M_n. Hence every M_nth design point is a knot, i.e., $\kappa_j = x_{jM_n}$ for $j = 1, \ldots, K_n$. A more general, unequally spaced case is discussed in Subsection 5.2.

When the knots are equally spaced, it is easy to verify that if the B-spline coefficient sequence $\{b_k\}$ is nondecreasing, then $f^{[p]}(x)$ is nondecreasing. Let the polyhedral cone $\Omega \equiv \{b \in \mathbb{R}^{K_n+p} : b_1 \le b_2 \le \cdots \le b_{K_n+p}\}$. Therefore the monotone P-spline coefficients are the unique solution of the following constrained quadratic programming problem

$$\hat{b} = \arg\min_{b \in \Omega} \frac{1}{2}\, b^T(\Gamma_n + \lambda D^T D)b - b^T \bar{y}, \tag{4}$$

where D is the $(K_n+p-1)\times(K_n+p)$ difference matrix with $Db = [\Delta(b_2),\ldots,\Delta(b_{K_n+p})]^T$ and

$$D = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ & \cdots & & & \cdots & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}, \tag{5}$$

and

$$\lambda = \frac{\lambda^*}{\beta_n}, \qquad \Gamma_n = \frac{1}{\beta_n}X^T X, \qquad \text{and} \qquad \bar{y} = \frac{1}{\beta_n}X^T y.$$

Here $\beta_n \equiv \sum_{i=1}^n \big(B_k^{[p]}(x_i)\big)^2$ for $k = p+1,\ldots,K_n$ and the $n\times(K_n+p)$ design matrix $X = \big[B_k^{[p]}(x_j)\big]_{j,k}$. Due to the equally spaced design points, all β_n's are equal for all $k = p+1,\ldots,K_n$.

To characterize the optimality conditions, we introduce more notation. Let C be the $(K_n+p-1)\times(K_n+p)$ matrix given by

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ & \cdots & & & \cdots & & \\ 1 & 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix}.$$

Given two vectors a and b, we write $a \geq 0$ (resp. $b \geq 0$) if each component of a (resp. b) is nonnegative and write $a \perp b$ if a and b are orthogonal, i.e., $a^T b = 0$. Hence, $0 \leq a \perp b \geq 0$ means $a \geq 0$, $b \geq 0$ and $a^T b = 0$. This condition is known as the complementarity condition [3, 4]. With this notation, we obtain the following lemmas for optimality conditions that show $\hat{b}$ is the solution of a mixed linear complementarity problem [3, 4].

Lemma 2.1. *The vector $\hat{b}$ is the (unique) optimal solution of (4) if and only if*

$$0 \leq D\hat{b} \perp -C\Gamma_n\hat{b} + \lambda D\hat{b} + C\bar{y} \geq 0, \tag{6}$$

and

$$\sum_{i=1}^{K_n+p} \big((\Gamma_n\hat{b})_i - \bar{y}_i \big) = 0. \tag{7}$$

Proof. Let $h(b) \equiv \frac{1}{2} b^T(\Gamma_n + \lambda D^T D)b - b^T\bar{y}$ be the objective function for a given $\bar{y}$. Since h is strictly convex and Ω is convex, $\hat{b}$ is the unique (global) minimizer of h on Ω if and only if $\langle \nabla h(\hat{b}), \hat{b}\rangle = 0$ and $\langle \nabla h(\hat{b}), b\rangle \geq 0, \forall\, b \in \Omega$, where the gradient $\nabla h(\hat{b}) = (\Gamma_n + \lambda D^T D)\,\hat{b} - \bar{y}$. Define $v^i \in \mathbb{R}^{K_n+p}$ as $v^i \equiv \big(\underbrace{0,\cdots,0,}_{(i-1)-\text{copies}}\ 1,\cdots,1\big)^T$ for $i = 1,\cdots,K_n+p$. Therefore, Ω is (positively) finitely generated by $\{v^1, -v^1, v^2, \cdots, v^{K_n+p}\}$. Indeed each $b \equiv (b_1,\cdots,b_{K_n+p})^T \in \Omega$ is positively generated as $b = \max(0,b_1)v^1 + \max(0,-b_1)(-v^1) + \sum_{i=2}^{K_n+p}(b_i - b_{i-1})\,v^i$. Let the matrix $V \equiv \big[v^1\ (-v^1)\ v^2\ \cdots\ v^{K_n+p}\big]$ such that $\Omega = \{Vu \,|\, u \geq 0\}$. Hence, letting the vector $u^* \geq 0$ be $u_1^* = \max(0,\hat{b}_1)$, $u_2^* = \max(0,-\hat{b}_1)$ and $u_i^* = \hat{b}_{i-1} - \hat{b}_{i-2}, \forall\, i = 3,\cdots,K_n+p+1$, the optimality conditions are equivalent to $0 \leq u^* \perp w \geq 0$, where $w \in \mathbb{R}^{K_n+p+1}$ with $w_1 = -w_2 = \langle \nabla h(\hat{b}), v^1\rangle$ and $w_{i+1} = \langle \nabla h(\hat{b}), v^i\rangle$, $i = 2,\cdots,K_n+p$. Thus the equivalent optimality conditions become

$$0 \leq D\hat{b} \perp -C\big[(\Gamma_n + \lambda D^T D)\,\hat{b} - \bar{y}\big] \geq 0, \quad \text{and} \quad \sum_{i=1}^{K_n+p}\big[\big((\Gamma_n + \lambda D^T D)\,\hat{b}\big)_i - \bar{y}_i\big] = 0.$$

Since $CD^T = -I$ and $\sum_{i=1}^{K_n+p}(D^T D\,\hat{b})_i = 0$, the lemma follows. $\square$

Lemma 2.2. *The optimality conditions (6)-(7) are, respectively, equivalent to*

$$\lambda^*(\hat{b}_{j+1} - \hat{b}_j) = \Big[\sum_{k=1}^{j}\sum_{i=1}^{n} B_k^{[p]}(x_i)\hat{f}^{[p]}(x_i) - \sum_{k=1}^{j}\sum_{i=1}^{n} B_k^{[p]}(x_i)y_i\Big]_+, \tag{8}$$

for $j = 1, \ldots, K_n + p - 1$, and

$$\sum_{i=1}^{n} \hat{f}^{[p]}(x_i) = \sum_{i=1}^{n} y_i. \tag{9}$$

Proof. Given $\lambda^* > 0$, the optimality condition (6) is equivalent to

$$\lambda^* D\,\hat{b} \;=\; \big[\beta_n C\Gamma_n\,\hat{b} - \lambda^* D\,\hat{b} - \beta_n C\bar{y} + \lambda^* D\,\hat{b}\big]_+ \;=\; \big[CX^T X\,\hat{b} - CX^T y\big]_+,$$

which is further equivalent to (8). It is also clear that (7) is equivalent to (9). $\square$

It shall be shown in Section 4 that the optimality conditions (8)-(9) can be approximated by an ODE with a constrained right-hand side subject to suitable boundary conditions for all large K_n. Indeed, such an ODE gives rise to a dynamic complementarity system [22, 23].

3 Uniform Lipschitz Property of Optimal Spline Coefficients

Since $(\Gamma_n + \lambda D^T D)$ in (4) is positive definite for each $\lambda > 0$, $\hat{b} = (\hat{b}_1, \cdots, \hat{b}_{K_n+p})^T$ is a (vector-valued) continuous piecewise linear function of $\bar{y}$ [4, 23]. However, the closed form of $\hat{b}$ is hard to obtain due to the combinatorial nature of the problem, and this poses a major technical difficulty for asymptotic analysis. In this section it is shown that $\hat{b}(\bar{y})$ satisfies the uniform Lipschitz property in the sense of the ℓ_∞-norm, regardless of K_n and λ, for all sufficiently large n (see Theorem 3.1 below). This property plays a crucial role in establishing stochastic boundedness and uniform consistency of $\hat{f}^{[p]}$ discussed at the end of this section. It should be emphasized that this property is different from the conventional Lipschitz property of a linear complementarity problem of *fixed size* [3] since the Lipschitz constant attained here is invariant to size variation.

Theorem 3.1. *The following hold:*

(a) *Let $p = 0$. For any $K_n \geq 2$ and $\lambda > 0$, $\|\hat{b}(\bar{y}^1) - \hat{b}(\bar{y}^2)\|_\infty \leq \kappa_0\,\|\bar{y}^1 - \bar{y}^2\|_\infty$ with $\kappa_0 = 1$ for all $\bar{y}^1, \bar{y}^2 \in \mathbb{R}^{K_n}$;*

(b) *Let $p = 1$. There exists $\kappa_1 > 0$ such that for all sufficiently large $\lambda > 0$ and n/K_n with $K_n \geq 2$, $\|\,\hat{b}(\bar{y}^1) - \hat{b}(\bar{y}^2)\,\|_\infty \leq \kappa_1\,\|\,\bar{y}^1 - \bar{y}^2\,\|_\infty$ for all $\bar{y}^1, \bar{y}^2 \in \mathbb{R}^{K_n+1}$;*

(c) *Let $p \geq 2$, $\varrho \in (0, 1)$, and $\gamma \in (\varrho, 1)$. Suppose that $K_n \sim n^\gamma$ and $\lambda \sim n^{2(\gamma-\varrho)}$. Then for all n sufficiently large, there exists $\kappa_p > 0$, dependent on p only, such that $\|\hat{b}(\bar{y}^1) - \hat{b}(\bar{y}^2)\|_\infty \leq \kappa_p\,\|\bar{y}^1 - \bar{y}^2\|_\infty$ for all $\bar{y}^1, \bar{y}^2 \in \mathbb{R}^{K_n+p}$.*

To prove this theorem, we establish a piecewise linear formulation of $\hat{b}$ first. Let $\Lambda_n \equiv (\Gamma_n + \lambda D^T D)/(1 + 2\lambda)$, $\overline{b} \equiv \hat{b}/(1 + 2\lambda)$, and $z \equiv \bar{y}/(1 + 2\lambda)$. The optimality conditions (6)-(7) become

$$0 \leq D\,\overline{b} \;\perp \widetilde{C}(\Lambda_n \overline{b} - z) \geq 0, \quad \text{ and } \quad \sum_{i=1}^{K_n+p} \big[\,(\Lambda_n \overline{b})_i - z_i\,\big] = 0, \tag{10}$$

where the $(K_n+p-1)\times(K_n+p)$ matrix $\widetilde{C}$ is given by

$$\widetilde{C}=\begin{bmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}. \tag{11}$$

It is observed from (10) that for each z, the corresponding optimal solution $\bar{b}$ is characterized by an index set $\alpha = \{\, i \mid \big(\widetilde{C}(\Lambda_n \bar{b} - z)\big)_i = 0\} \subseteq \{1,\cdots,K_n+p-1\}$ (α may be empty). For the given $\bar{b}$ and α, define a vector $\widetilde{b}^{\alpha}$ as follows: $\widetilde{b}^{\alpha}_1 \equiv \bar{b}_1$ and $\widetilde{b}^{\alpha}_{i+1} \equiv \bar{b}_{\ell_{i+1}}$ for $i \geq 1$, where $\ell_{i+1} \equiv \min_{1\leq j\leq K_n+p}\{j \mid \bar{b}_j > \widetilde{b}^{\alpha}_i\}$. Hence, the elements of $\widetilde{b}^{\alpha}$ strictly increase as their indices increase. Moreover, for each $\widetilde{b}^{\alpha}_i$, define the index set $\beta^{\alpha}_i = \{j \in \{1,\cdots,K_n+p\} \mid \bar{b}_j = \widetilde{b}^{\alpha}_i\}$. This gives rise to a (finite and disjoint) partition of $\{1,\cdots,K_n+p\}$, namely, $\bigcup_i \beta^{\alpha}_i = \{1,\cdots,K_n+p\}$ and $\beta^{\alpha}_j \cap \beta^{\alpha}_k = \emptyset$ whenever $j \neq k$. It can be shown that $\widetilde{b}^{\alpha}$, and thus $\bar{b}^{\alpha}(z)$ which denotes $\bar{b}(z)$ corresponding to the index set α, is a linear function of z. Hence, for any $z \in \mathbb{R}^{K_n+p}$, $\bar{b}(z) \in \{\bar{b}^{\alpha}(z)\}_{\alpha}$, where $\bar{b}^{\alpha}(z)$ is called a selection function of $\bar{b}(z)$. Therefore, the solution mapping $z \mapsto \bar{b}$ is a (continuous) piecewise linear function with $2^{(K_n+p)}$ selection functions. The same holds true for the mapping $\bar{y} \mapsto \hat{b}$, i.e., $\hat{b}^{\alpha}_i(\bar{y}) \equiv \sum_{j=1}^{K_n+p} \bar{a}^{\alpha}_{ij}\, \bar{y}_j$, where the coefficients $\bar{a}^{\alpha}_{ij}$ pertain to each index set α.

Since the proof of Theorem 3.1 is technical, we sketch its main ideas and outline the key steps as follows. To motivate the main ideas, consider $p \geq 2$ first. In this case, it can be seen from the above construction and Lemma 3.1 that each selection function $\bar{b}^{\alpha}(z)$ is a linear function whose coefficients are defined by the inverse of a $(2p+1)$-diagonal matrix, denoted by $\widetilde{\Lambda}^{\alpha}$. Each $\widetilde{\Lambda}^{\alpha}$ can be decomposed into the sum of a tridiagonal matrix similar to that for $p=1$ and a perturbation matrix that consists of "small" terms (of order λ^{-1} indeed; see Lemma 3.4). Hence, a suitable uniform bound of $(\widetilde{\Lambda}^{\alpha})^{-1}$ in case of $p=0$ or 1 will not only establish the uniform Lipschitz property for $p=0,1$ but also for all $p\geq 2$. In order to apply this perturbation technique for the latter case, a tight bound is expected for $p=1$. A major difficulty of finding such a tight bound for $p=1$ is that the size and elements of $(\widetilde{\Lambda}^{\alpha})^{-1}$ vary, and the number of $\widetilde{\Lambda}^{\alpha}$'s grows exponentially with respect to K_n. To handle this complexity, we exploit the tridiagonal structure of $\widetilde{\Lambda}^{\alpha}$ and show that all $(\widetilde{\Lambda}^{\alpha})^{-1}$ are completely determined by certain sequences with similar properties uniform in α. By fully making use of these properties, it is shown in Proposition 3.1 that for any α, each element of $(\widetilde{\Lambda}^{\alpha})^{-1}$ is positive and bounded above by the $(1,1)$-element of Λ_n^{-1} for all large λ. Based on this, it is then proven in Proposition 3.3 that the sum of (positive) coefficients of a selection function defined by $(\widetilde{\Lambda}^{\alpha})^{-1}$ is bounded above by the infinity norm of some Λ_n^{-1} with the same size as that of $(\widetilde{\Lambda}^{\alpha})^{-1}$. An upper bound of $\|\Lambda_n^{-1}\|_\infty$ uniform in n is further obtained in Proposition 3.4. These results yield the desired bounds for $p=0$ in Proposition 3.2, $p=1$ in Proposition 3.5, and all $p\geq 2$ in Proposition 3.6, respectively. Finally, the polyhedral theory leads to the uniform Lipschitz property for $\hat{b}$ with an arbitrary $p \in \mathbb{Z}_+$ in Subsection 3.3.

3.1 The Case of $p=0$ and $p=1$

For $p=0$ or $p=1$, define $\alpha_n = \sum_{i=1}^n B_k(x_i)^2$ for $k=1, K_n+p$, $\beta_n = \sum_{i=1}^n B_k(x_i)^2$ for $k=2,\ldots,K_n+p-1$, and $\gamma_n = \sum_{i=1}^n B_k(x_i)B_{k+1}(x_i)$ for $k=1,\ldots,K_n+p-1$. Moreover, let $\widetilde{\theta}_n = \alpha_n/\beta_n$, $\widetilde{\eta}_n = \gamma_n/\beta_n$. It is easy to see that (i) for $p=0$, $\widetilde{\theta}_n = 1$ and $\widetilde{\eta}_n = 0$ for all n; and (ii) for $p = 1$, $\widetilde{\theta}_n \to \widetilde{\theta}_*$ and $\widetilde{\eta}_n \to \widetilde{\eta}_*$ as $n/K_n \to \infty$, where $\widetilde{\theta}_* = 1/2$ and $\widetilde{\eta}_* > 0$. We thus have

$$X^T X = \begin{bmatrix} \alpha_n & \gamma_n & 0 & 0 & \cdots & 0 \\ \gamma_n & \beta_n & \gamma_n & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \gamma_n & \beta_n & \gamma_n & 0 \\ & & & \gamma_n & \beta_n & \gamma_n \\ 0 & 0 & \cdots & 0 & \gamma_n & \alpha_n \end{bmatrix}, \quad \Gamma_n = \frac{X^T X}{\beta_n} = \begin{bmatrix} \widetilde{\theta}_n & \widetilde{\eta}_n & 0 & 0 & \cdots & 0 \\ \widetilde{\eta}_n & 1 & \widetilde{\eta}_n & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & & \\ & & \widetilde{\eta}_n & 1 & \widetilde{\eta}_n & 0 \\ & & & \widetilde{\eta}_n & 1 & \widetilde{\eta}_n \\ 0 & 0 & \cdots & 0 & \widetilde{\eta}_n & \widetilde{\theta}_n \end{bmatrix},$$

and

$$\Lambda = \frac{\Gamma_n + \lambda D^T D}{1 + 2\lambda} = \begin{bmatrix} \theta & \eta & 0 & 0 & \cdots & 0 \\ \eta & 1 & \eta & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \eta & 1 & \eta & 0 \\ & & & \eta & 1 & \eta \\ 0 & 0 & \cdots & 0 & \eta & \theta \end{bmatrix}, \tag{12}$$

where the subscript n in Λ_n is dropped for simplicity, $\theta \equiv (\widetilde{\theta}_n + \lambda)/(1+2\lambda)$, and $\eta \equiv (\widetilde{\eta}_n - \lambda)/(1+2\lambda)$ with $\lambda > 0$. Note that $-1/2 < \eta < 0$ for all large λ. Using the notation introduced below (11), we show as follows that each $\widetilde{b}^\alpha$, or equivalently the selection function $\overline{b}^\alpha$, is linear in z.

Lemma 3.1. *For each index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$, $\widetilde{b}^\alpha$ is the (unique) solution of the linear equation $\widetilde{\Lambda}^\alpha \widetilde{b}^\alpha = \widetilde{z}^\alpha$, where the $\ell \times \ell$ matrix $\widetilde{\Lambda}^\alpha$ and the ℓ-vector $\widetilde{z}^\alpha$ are given by*

$$\widetilde{\Lambda}^\alpha = \begin{bmatrix} d_{11} & \eta & 0 & & \cdots & 0 \\ \eta & d_{22} & \eta & & & \\ & \ddots & \ddots & & \ddots & \\ & & & \eta & d_{(\ell-1)(\ell-1)} & \eta \\ 0 & \cdots & 0 & & \eta & d_{\ell\ell} \end{bmatrix}, \quad d_{jj} = \sum_{p,\, q \in \beta_j^\alpha} \Lambda_{pq}, \quad \widetilde{z}_j^\alpha = \sum_{k \in \beta_j^\alpha} z_k$$

Moreover, $\widetilde{\Lambda}^\alpha$ is invertible and $\overline{b}_i^\alpha(z) = \sum_{j=1}^{K_n+p} a_{ij}^\alpha z_j$, $\forall\ i$, where $a_{ij}^\alpha = [(\widetilde{\Lambda}^\alpha)^{-1}]_{ik}$ if $j \in \beta_k^\alpha$.

Proof. For the given index set α, define the matrix $\bar{C}_\alpha = \begin{bmatrix} \mathbf{1}^T \\ \widetilde{C}_{\alpha\bullet} \end{bmatrix}$, where $\mathbf{1} = (1, \cdots, 1)^T$ and $\widetilde{C}_{\alpha\bullet}$ denotes the rows in $\widetilde{C}$ indexed by α. Hence, $\bar{C}_\alpha \Lambda \overline{b} = \bar{C}_\alpha z$. It can be shown via elementary row operations and induction that $\bar{C}_\alpha$ is row equivalent to the matrix $\widehat{C}_\alpha$ whose ith row is given by: $(\widehat{C}_\alpha)_{ij} = \begin{cases} 1, & \text{if } j \in \beta_i^\alpha \\ 0, & \text{otherwise} \end{cases}$. Let $\bar{\alpha}$ be the complement of α in $\{1, \cdots, K_n + p - 1\}$. In view of the complementarity condition in (10), $\big(\widetilde{C}(\Lambda \overline{b} - z)\big)_{\bar{\alpha}} > 0$ implies $(D\hat{b})_{\bar{\alpha}} = 0$ and, in turn, $b_j = b_{j+1}$ for each $j \in \bar{\alpha}$. This shows $\overline{b} = (\widehat{C}_\alpha)^T \widetilde{b}^\alpha$. Therefore, we have $\widehat{C}_\alpha \Lambda (\widehat{C}_\alpha)^T \widetilde{b}^\alpha = \widehat{C}_\alpha z$. Letting $\widetilde{\Lambda}^\alpha = \widehat{C}_\alpha \Lambda (\widehat{C}_\alpha)^T$ and $\widetilde{z}^\alpha = \widehat{C}_\alpha z$, we obtain the desired linear equation for $\widetilde{b}^\alpha$. Since $\widehat{C}_\alpha$ is of full row rank, $\widetilde{\Lambda}^\alpha$ is positive definite and hence is invertible. Finally, the expression for the corresponding $\overline{b}$ follows from the structure of $\widehat{C}_\alpha$ and the definition of β_i^α. □

In the following, let $m_i^\alpha \equiv |\beta_i^\alpha|$ be the cardinality of the index set β_i^α. Let $K_n \geq 2$. Defining $h_i^\alpha \equiv m_i^\alpha - 1$, it can be shown that for a given index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$,

$$d_{ii} = \begin{cases} \theta + h_i^\alpha(1+2\eta) & \text{if } \ell \geq 2 \text{ and } i \in \{1, \ell\} \\ 1 + h_i^\alpha(1+2\eta) & \text{if } \ell \geq 2 \text{ and } i \in \{2, \cdots, \ell-1\} \\ 2(\theta+\eta) + (K_n + p - 2)(1+2\eta) & \text{if } \ell = 1 \end{cases} \tag{13}$$

Note that $1 + 2\eta = (1 + 2\widetilde{\eta}_n)(1+2\lambda)^{-1} > 0$ for a large n and all $\lambda > 0$. Furthermore, for a given index set α with $\ell \geq 2$, let $h^\alpha \equiv (h_1^\alpha, \cdots, h_\ell^\alpha)$ and its corresponding $\widetilde{\Lambda}^\alpha$ can be written as

$$\widetilde{\Lambda}^\alpha(h^\alpha) = \begin{bmatrix} d_{11}(h_1^\alpha) & \eta & 0 & \cdots & 0 \\ \eta & d_{22}(h_2^\alpha) & \eta & & \\ & \ddots & \ddots & \ddots & \\ & & \eta & d_{(\ell-1)(\ell-1)}(h_{\ell-1}^\alpha) & \eta \\ 0 & \cdots & 0 & \eta & d_{\ell\ell}(h_\ell^\alpha) \end{bmatrix}, \tag{14}$$

where each $d_{ii}(h_i^\alpha)$ is defined in (13). To estimate the inverse of each $\widetilde{\Lambda}^\alpha$, we present two lemmas regarding the sequences that characterize $(\widetilde{\Lambda}^\alpha)^{-1}$ in Proposition 3.1 as follows.

Lemma 3.2. *Let $\lambda > 0$ be sufficiently large. Consider the sequence $\{p_i\}_{i=1}^\infty$ defined by*

$$p_1 = \frac{\eta}{\theta}, \qquad p_{i+1} = \frac{\eta}{1 - \eta\, p_i}, \quad \forall\, i \in \mathbb{N}. \tag{15}$$

Then the following hold:

(a) *$\{p_i\}_{i=1}^\infty$ is a strictly increasing sequence with $-1 < p_i < \dfrac{2\eta}{1+\sqrt{1-4\eta^2}} < 0$;*

(b) *for a given $\ell \geq 3$, $0 < p_{\ell-1} - p_i - p_{\ell-1-i} \leq (1-\theta)/(-\eta)$ for all $i = 1, \cdots, \ell-2$.*

Proof. Statement (a) can be proved via mathematical induction. In fact, the sequence $\{p_i\}_{i=1}^\infty$ monotonically converges to the negative root $2\eta(1+\sqrt{1-4\eta^2})^{-1}$ of the equation $\eta x^2 - x + \eta = 0$ from the initial term $p_1 = (\widetilde{\eta}_n - \lambda)/(\widetilde{\theta}_n + \lambda) > -1$. In what follows, we prove (b). The positivity of $p_{\ell-1} - p_i - p_{\ell-1-i}$ follows from (a). To establish its upper bound, define the function $r(x) \equiv (\eta - x + \eta x^2)/(1 - \eta x)$. It is easily verified that $r(p_i) = p_{i+1} - p_i$ and $r'(x) < 0$ for all $x \in [-1, 0]$, i.e., $r(x)$ is decreasing on $[-1, 0]$. We consider $i = 1$ first. In this case,

$$p_{\ell-1} - p_1 - p_{\ell-2} \;=\; (p_{\ell-1} - p_{\ell-2}) - p_1 \;=\; r(p_{\ell-2}) - p_1 \;\leq\; r(p_1) - p_1,$$

where $p_1 \leq p_{\ell-2}$ (cf. (a)) is used in the last step. By making use of the definitions of θ and η, it can be shown via direct but somewhat tedious calculations that for all large $\lambda > 0$,

$$r(p_1) - p_1 \;=\; \frac{\eta}{1 - \frac{\eta^2}{\theta}} - \frac{2\eta}{\theta} \;=\; \frac{\eta(\theta^2 - 2\theta + 2\eta^2)}{\theta(\theta - \eta^2)} \leq \frac{1-\theta}{-\eta}.$$

As a result, (b) holds for $i = 1$. Furthermore, since $\{p_i\}_{i=1}^\infty$ is increasing and $r(x)$ is decreasing, we have $r(p_i) \geq r(p_{\ell-i-2})$ for all $1 \leq i \leq (\lfloor \ell/2 \rfloor - 1)$, where $\lfloor x \rfloor \equiv \max\{n \in \mathbb{N} : n \leq x\}$. This shows that $p_{i+1} - p_i \geq p_{\ell-i-1} - p_{\ell-i-2}$ or equivalently $p_{i+1} + p_{\ell-i-2} \geq p_i + p_{\ell-i-1}$. Hence $p_{\ell-1} - p_{i+1} - p_{\ell-i-2} \leq p_{\ell-1} - p_i - p_{\ell-i-1}$ for all $i = 1, \cdots, (\lfloor \ell/2 \rfloor - 1)$. This result, along with that for $i = 1$, yields the desired upper bound for all $i = 2, \cdots, \ell-2$. □

Lemma 3.3. *Let $\lambda > 0$ be sufficiently large. Consider the sequence $\{q_j\}_{j=1}^{\infty}$ defined by*

$$q_1 = \frac{\eta}{d_{11}(h_1)}, \qquad q_{j+1} = \frac{\eta}{d_{(j+1)(j+1)}(h_{j+1}) - \eta\, q_j}, \quad \forall\, j \in \mathbb{N}, \tag{16}$$

where $d_{11}(h_1) \equiv \theta + h_1(1+2\eta)$ and $d_{jj}(h_j) \equiv 1 + h_j(1+2\eta), \forall\, j > 1$ with real $h_j \geq 0$ for all j. Let $\{p_j\}$ be the sequence defined in (15). Then the following hold:

(a) $-1 < q_j < 0$ *for all j;*

(b) *For any $\ell \in \mathbb{N}$, if $h_j \in \mathbb{Z}_+$ for each $j = 1, \cdots, \ell$, then $p_{\sum_{j=1}^{\ell}(1+h_j)} \leq q_\ell$.*

Proof. (a). We show by induction that $-1 < q_j < 0$ for all j. Since $1+2\eta > 0$, $d_{11} \geq \theta > 0$. Hence $-1 < \eta/\theta \leq q_1 < 0$. Now suppose $-1 < q_j < 0$ for all $j = 1, \cdots, k$ with $k \geq 1$, and consider q_{k+1}. It is easy to verify via the induction hypothesis that $d_{(k+1)(k+1)} - \eta\, q_k > d_{(k+1)(k+1)} + \eta \geq 1+\eta > 0$. In view of $-1/2 < \eta < 0$, we have $-1 < q_{k+1} < 0$. Consequently, $-1 < q_j < 0$ for all j.

(b). We prove this result via induction on ℓ. Consider $\ell = 1$. It is obvious that the desired inequality holds when $h_1 = 0$. Now assume that it also holds for $h_1 = 0, 1, \cdots, k$, where $k \in \mathbb{Z}_+$. Let $h_1 = k+1$. Using the fact that $\eta < 0$, we obtain

$$\frac{\eta^2}{\theta + k(1+2\eta)} + \theta + (k+1)(1+2\eta) = \Big(\frac{\eta^2}{\theta + k(1+2\eta)} + [\theta + k(1+2\eta)]\Big) + (1+2\eta) \geq 2|\eta| + 1 + 2\eta = 1,$$

and thus $1 - \eta^2/[\theta + k(1+2\eta)] \leq \theta + (k+1)(1+2\eta)$. This result, along with the induction hypothesis $p_{1+k} \leq \eta/[\theta + k(1+2\eta)]$, implies $1 - \eta\, p_{1+k} \leq \theta + (k+1)(1+2\eta)$. Therefore $p_{1+k} \leq q_1$, and this completes the proof for $\ell = 1$.

To carry out an induction for a general $\ell \in \mathbb{N}$, we show the following inequality first:

$$p_{i+(h+1)} \leq \frac{\eta}{1 + h(1+2\eta) - \eta\, p_i}, \qquad \forall \;\; h \in \mathbb{Z}_+, \tag{17}$$

where $p_i \in (-1, 0)$ is the ith term of the sequence $\{p_j\}$. Clearly, this inequality holds when $h = 0$. Suppose that it holds for $h = 0, 1, \cdots, k$, where $k \in \mathbb{Z}_+$. Let $h = k+1$. By the definition of $p_{i+(h+1)}$, it is sufficient to show $1 - \eta p_{i+(k+1)} \leq 1 + (k+1)(1+2\eta) - \eta p_i$, which is equivalent to

$$(k+1)(1+2\eta) \geq \eta\big(p_i - p_{i+(k+1)}\big) \tag{18}$$

To prove (18), define the real number $a \equiv 1 + k(1+2\eta) > 1$ and the function $g(x) \equiv x - \eta/(a - \eta x)$. Notice that $g'(x) = [a + \eta(1-x)][a - \eta(1+x)]/(a-\eta x)^2 > 0$ for all $x \in [-1, 0]$, where we use $-1/2 < \eta < 0$ for all λ sufficiently large. Hence $\eta g(p_i) \leq \eta g(-1) = (-\eta)(a+2\eta)/(a+\eta)$. Since $a + 2\eta = (k+1)(1+2\eta) > 0$ and $0 < -\eta/(a+\eta) < 1$, we then have $\eta g(p_i) \leq \eta g(-1) \leq (k+1)(1+2\eta)$. Moreover, using the induction hypothesis, we further deduce

$$\eta\big(p_i - p_{i+(k+1)}\big) \leq \eta\Big(p_i - \frac{\eta}{1 + k(1+2\eta) - \eta p_i}\Big) = \eta\, g(p_i) \leq (k+1)(1+2\eta).$$

Consequently, the inequality (18), as well as (17), holds.

Returning to the proof for a general $\ell \in \mathbb{N}$, we assume that the lemma holds for $1, \cdots, \ell$. Consider $\ell + 1$. Using the inequality (17) and the induction hypothesis, we have

$$\begin{aligned} p_{\sum_{j=1}^{\ell+1}(1+h_j)} &= p_{\sum_{j=1}^{\ell}(1+h_j) + (1+h_{\ell+1})} \leq \frac{\eta}{1 + h_{\ell+1}(1+2\eta) - \eta\, p_{\sum_{j=1}^{\ell}(1+h_j)}} \\ &\leq \frac{\eta}{1 + h_{\ell+1}(1+2\eta) - \eta\, q_\ell} = q_{\ell+1} \end{aligned}$$

It follows from the induction principle that (b) holds true. □

In the following, let $\ell = |\alpha|$ and $\mathbb{R}^\ell_+$ denote the nonnegative orthant of $\mathbb{R}^\ell$. The next proposition shows that each element of $(\widetilde{\Lambda}^\alpha)^{-1}$ is positive and is bounded above by the $(1,1)$-element of Λ_n^{-1} under suitable order conditions. The latter result shall be used in Proposition 3.6 of Subsection 3.2.

Proposition 3.1. *The following hold:*

(a) *Let $p = 0$ or 1, $\lambda > 0$ be sufficiently large, and $K_n \geq 2$. For a given index set α with $\ell \geq 1$ and any $h \in \mathbb{R}^\ell_+$, each element of $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}$ is positive.*

(b) *Let $p = 1$. Suppose that $n/K_n \to \infty$, $\lambda(n) \to \infty$, and $K_n/\sqrt{\lambda(n)} \to \infty$ as $n \to \infty$. Then for all n sufficiently large, each element of $\big(\widetilde{\Lambda}^\alpha\big)^{-1}$ is not greater than $(\Lambda^{-1})_{11}$ for any index set $\alpha \subseteq \{1, \cdots, K_n\}$.*

Proof. (a). The statement holds trivially when $\ell = 1$; we address $\ell \geq 2$ as follows. Given an index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$, let $\mathbf{e}_i^\alpha \equiv (\underbrace{0, \cdots, 0,}_{(i-1)\text{ copies}}\ 1, 0, \cdots, 0)^T \in \mathbb{R}^\ell$, and $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_i^\alpha \equiv (c_{i1}^\alpha, \cdots, c_{i\ell}^\alpha)^T$. To ease the presentation, we consider $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_1^\alpha$ first. In this case,

$$\begin{bmatrix} d_{11}(h_1) & \eta & 0 & \cdots & 0 \\ \eta & d_{22}(h_2) & \eta & & \\ & \ddots & \ddots & \ddots & \\ & & \eta & d_{(\ell-1)(\ell-1)}(h_{\ell-1}) & \eta \\ 0 & \cdots & 0 & \eta & d_{\ell\ell}(h_\ell) \end{bmatrix} \begin{bmatrix} c_{11}^\alpha \\ c_{12}^\alpha \\ \vdots \\ c_{1(\ell-1)}^\alpha \\ c_{1\ell}^\alpha \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

The above linear equation is row equivalent to

$$\begin{bmatrix} d_{11}(h_1) - q_{\ell-1}\,\eta & 0 & 0 & \cdots & 0 \\ q_{\ell-1} & 1 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & q_2 & 1 & 0 \\ 0 & \cdots & 0 & q_1 & 1 \end{bmatrix} \begin{bmatrix} c_{11}^\alpha \\ c_{12}^\alpha \\ \vdots \\ c_{1(\ell-1)}^\alpha \\ c_{1\ell}^\alpha \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{19}$$

where $q_1 = \eta/d_{\ell\ell}(h_\ell)$, and $q_{i+1} = \dfrac{\eta}{d_{(\ell-i)(\ell-i)}(h_{\ell-i}) - \eta\, q_i}$ for $i = 1, \cdots, \ell-2$. It follows from (a) of Lemma 3.3 that $-1 < q_i < 0$ for all i. By the definition of d_{11} (cf. (13)), we have $d_{11} - q_{\ell-1}\,\eta \geq d_{11} + \eta = (\widetilde{\theta}_n + h_1(1 + 2\widetilde{\eta}_n) + \widetilde{\eta}_n)/(1 + 2\lambda) > 0$. Hence, $c_{11}^\alpha > 0$. Since each $-1 < q_k < 0$, this further shows $0 < c_{1j}^\alpha < c_{11}^\alpha, \forall\, j = 2, \cdots, \ell$ and thus completes the case of $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_1^\alpha$. Moreover, note that $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_\ell^\alpha = \big(c_{1\ell}^\alpha, c_{1(\ell-1)}^\alpha, \cdots, c_{12}^\alpha, c_{11}^\alpha\big)^T$ due to symmetry. Hence each element of $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_\ell^\alpha$ is also positive.

Next we consider $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_i^\alpha$ with $1 < i < \ell$. It can be verified that $\big(c_{i1}^\alpha, \cdots, c_{i\ell}^\alpha\big)^T$ satisfies the following linear equation

$$\begin{bmatrix} 1 & w_1 & 0 & 0 & \cdots & \cdots & 0 \\ & \ddots & \ddots & \ddots & & & \\ & & 1 & w_{i-1} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & d_{ii}(h_i) - \eta(w_{i-1} + q_{\ell-i}) & 0 & \cdots & 0 \\ & & & q_{\ell-i} & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ 0 & \cdots & \cdots & 0 & 0 & q_1 & 1 \end{bmatrix} \begin{bmatrix} c_{i1}^\alpha \\ \vdots \\ c_{i(i-1)}^\alpha \\ c_{ii}^\alpha \\ c_{i(i+1)}^\alpha \\ \vdots \\ c_{i\ell}^\alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{20}$$

where

$$w_1 = \frac{\eta}{d_{11}(h_1)}, \qquad w_{j+1} = \frac{\eta}{d_{(j+1)(j+1)}(h_{j+1}) - \eta\, w_j}, \ \forall\, j = 1, \cdots, i-2,$$

and

$$q_1 = \frac{\eta}{d_{\ell\ell}(h_\ell)}, \qquad q_{k+1} = \frac{\eta}{d_{(\ell-k)(\ell-k)}(h_{\ell-k}) - \eta\, q_k}, \ \forall\, k = 1, \cdots, \ell-i-1.$$

Following (a) of Lemma 3.3, we have $-1 < w_j, q_k < 0$ for all j, k. Since $d_{ii}(h_i) = 1 + h_i(1+2\eta)$, where $h_i \ge 0$ and $-1/2 < \eta < 0$ for all sufficiently large λ, we have $d_{ii}(h_i) - \eta(w_{i-1} + q_{\ell-i}) \ge d_{ii}(h_i) + 2\eta = (h_i+1)(1+2\eta) > 0$. Hence, $c^\alpha_{ii} > 0$, which further implies $0 < c^\alpha_{ij} < c^\alpha_{ii}, \forall\, j \neq i$.

To show (b) for $p = 1$, we address two cases as follows:

(b.1) $\alpha = \{1, \cdots, K_n\}$. Here $\widetilde{\Lambda}^\alpha = \Lambda$ is of order $\ell \equiv (K_n + 1) \ge 3$. Note that $(\Lambda^{-1})_{11} = (\Lambda^{-1})_{\ell\ell} = (1 - \eta\, p_{\ell-1})^{-1}$ and $(\Lambda^{-1})_{ii} = (1 - \eta(p_{\ell-i} + p_{i-1}))^{-1}$ for all $2 \le i \le \ell - 1$, where $\{p_i\}$ is defined in (15). It follows from (b) of Lemma 3.2 that $\theta - \eta p_{\ell-1} \le 1 - \eta(p_{\ell-i} + p_{i-1})$ once $2 \le i \le \ell-1$. Further, (a) of Lemma 3.3 shows that $p_{\ell-1} > -1$ (corresponding to $h_i = 0$) such that $\theta - \eta p_{\ell-1} \ge \theta + \eta > 0$. Therefore $(\Lambda^{-1})_{11} \ge (\Lambda^{-1})_{ii}$ for all $2 \le i \le \ell - 1$. By observing from the proof of (a) that $(\Lambda^{-1})_{ii}$ is greater than off-diagonal entries in the ith column, (b) follows.

(b.2) α is a proper subset of $\{1, \cdots, K_n\}$. Consider two subcases: (i) $\ell = 1$; (ii) $\ell \ge 2$. For subcase (i), note that for all large n, $\widetilde{\Lambda}^\alpha = 2(\theta + \eta) + (K_n - 1)(1 + 2\eta) \sim K_n/(1+2\lambda)$ and

$$1/(\Lambda^{-1})_{11} = \theta - \eta\, p_{K_n} \le \theta - \eta \frac{2\eta}{1 + \sqrt{1 - 4\eta^2}} = \theta + \frac{\sqrt{1-4\eta^2} - 1}{2} \sim \frac{\sqrt{\lambda}}{1 + 2\lambda}.$$

By the given asymptotic behaviors of K_n and λ, we have $(\widetilde{\Lambda}^\alpha)^{-1} \le (\Lambda^{-1})_{11}$ for all large n. For subcase (ii), it suffices to show that $c^\alpha_{ii} \le (\Lambda^{-1})_{11}$ for all $i = 1, \cdots, \ell$. Consider $i = 1$ first. It is seen from (19) that the sequence $\{q_j\}_{j=1}^{\ell-1}$ is defined by some $h_j \in \mathbb{Z}_+, j = 1, \cdots, \ell$ satisfying $\sum_{j=2}^{\ell}(1 + h_j) = K_n - h_1$. Moreover, by noticing that each h_j here corresponds to $h_{\ell+1-j}$ in (16), we deduce from (b) of Lemma 3.3 that $p_{K_n - h_1} = p_{\sum_{j=2}^{\ell}(1+h_j)} \le q_{\ell-1}$. Therefore,

$$d_{11}(h_1) - \eta q_{\ell-1} \ \ge\ \theta + h_1(1+2\eta) - \eta\, p_{K_n - h_1} \ \ge\ \theta - \eta p_{K_n} \ >\ 0,$$

where the second inequality follows from (18). Hence, $c^\alpha_{11} \le (\Lambda^{-1})_{11}$. Similarly $c^\alpha_{\ell\ell} \le (\Lambda^{-1})_{11}$. Now consider $i \in \{2, \cdots, \ell-1\}$. It is seen from (20) that we have two sequences $\{w_j\}_{j=1}^{i-1}$ defined by $h_j \in \mathbb{Z}_+$ and $\{q_j\}_{j=1}^{\ell-i}$ defined by $h_{\ell-j+1} \in \mathbb{Z}_+$. Let $s_1 \equiv \sum_{j=1}^{i-1}(1 + h_j)$ and $s_2 \equiv \sum_{j=1}^{\ell-i}(1 + h_{\ell-j+1})$. Hence, $s_1 + s_2 = K_n - h_i$. By (b) of Lemma 3.3 we have $w_{i-1} \ge p_{s_1}$ and $q_{\ell-i} \ge p_{s_2}$. As a result,

$$\begin{aligned}
d_{ii}(h_i) - \eta\big(w_{i-1} + q_{\ell-i}\big) \ &\ge\ 1 + h_i(1+2\eta) - \eta\left(p_{s_1} + p_{(K_n - h_i) - s_1}\right) \\
&\ge\ 1 + h_i(1+2\eta) - \eta\left(p_{K_n - h_i} - \frac{1-\theta}{-\eta}\right) \quad \text{(via (b) of Lemma 3.2)} \\
&\ge\ \theta + h_i(1+2\eta) - \eta\, p_{K_n - h_i} \\
&\ge\ \theta - \eta\, p_{K_n} \qquad \text{(via (18))}
\end{aligned}$$

This yields $c^\alpha_{ii} \le (\Lambda^{-1})_{11}$ for all $i \in \{2, \cdots, \ell-1\}$. □

Recall that for each piecewise linear function $\bar{b}_i(z)$ with $i \in \{1, \cdots, K_n + p\}$, its selection function (corresponding to α) is given by $\bar{b}^\alpha_i(z) = \sum_{j=1}^{K_n+p} a^\alpha_{ij} z$. According to Proposition 3.1, we conclude that each a^α_{ij} is positive for any i, j and α. The following proposition establishes an upper bound of $\sum_{j=1}^{K_n} |a^\alpha_{ij}|$.

Proposition 3.2. *Let $p = 0$ and $\lambda > 0$. For each index set $\alpha \subseteq \{1, \cdots, K_n - 1\}$ with $K_n \geq 2$, $\sum_{j=1}^{K_n} |a_{ij}^\alpha| = \sum_{j=1}^{K_n} a_{ij}^\alpha = 1 + 2\lambda$.*

Proof. Let $z^* \equiv (1 + 2\lambda)^{-1} \mathbf{1}$, where $\mathbf{1}$ denotes the vector of ones. By making use of the matrix $\widetilde{\Lambda}^\alpha$ defined in Lemma 3.1 and the formula of d_{ii} in (13), it can be verified via straightforward computation that $\widetilde{b}^\alpha(z^*) = \mathbf{1}$, and thus $\overline{b}^\alpha(z^*) = \mathbf{1}$, for each $\alpha \subseteq \{1, \cdots, K_n - 1\}$. The proposition thus follows in view of $\overline{b}_i^\alpha(z^*) = \sum_{j=1}^{K_n} a_{ij}^\alpha / (1 + 2\lambda)$ for each $i = 1, \cdots, K_n$. □

In what follows, we focus on $p = 1$ and establish a uniform bound for the sums of the coefficients of each selection function $\overline{b}_i^\alpha(z)$, regardless of i, K_n, λ, and α. Note that for each i, the sum of the positive coefficients of $\overline{b}_i^\alpha$ is equal to $\overline{b}_i^\alpha(\mathbf{1})$ and further equals to the ith element of $\big(\widetilde{\Lambda}^\alpha(h^\alpha)\big)^{-1}\big(\mathbf{1} + h^\alpha\big)$, where $h^\alpha = (h_1^\alpha, \cdots, h_\ell^\alpha)^T \in \mathbb{R}_+^\ell$. We show below that each element of $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{1} + h\big)$ attains its maximum at $h = 0$. This implies that the sum of the positive coefficients of $\overline{b}_i^\alpha$ is bounded above by $\|\big(\widetilde{\Lambda}^\alpha(0)\big)^{-1}\|_\infty$.

Proposition 3.3. *Let $p = 1$, and let n/K_n and $\lambda > 0$ be sufficiently large with $K_n \geq 2$. For a given index set α and each $i \in \{1, \cdots, \ell\}$,*

$$\Big[\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{1} + h\big)\Big]_i < \Big[\big(\widetilde{\Lambda}^\alpha(0)\big)^{-1}\mathbf{1}\Big]_i, \qquad \forall \;\; 0 \neq h \in \mathbb{R}_+^\ell.$$

Proof. The case of $\ell = 1$ is easy to verify and thus omitted. We assume that $\ell \geq 2$ in the sequel. For the given α and i, define the real-valued function $g(h) \equiv \mathbf{e}_i^T \big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{1} + h\big)$. Clearly $g(h)$ is continuously differentiable on an open covering of $\mathbb{R}_+^\ell$. For each j,

$$\begin{aligned}
\frac{\partial g(h)}{\partial h_j} &= \mathbf{e}_i^T \big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_j + \mathbf{e}_i^T\Big[-\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\frac{\partial \widetilde{\Lambda}^\alpha(h)}{\partial h_j}\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\Big]\big(\mathbf{1} + h\big) \\
&= \Big(\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_i\Big)_j - (1 + 2\eta)\Big(\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_i\Big)_j \Big(\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{1} + h\big)\Big)_j \\
&= \Big(\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_i\Big)_j \Big[1 - (1 + 2\eta)\Big(\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{1} + h\big)\Big)_j\Big]
\end{aligned}$$

We deduce from (a) of Proposition 3.1 that $\Big(\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\mathbf{e}_i\Big)_j > 0$ for all $h \in \mathbb{R}_+^\ell$. Noticing $\Lambda^\alpha(h)\mathbf{1} = \big(\theta + \eta + h_1(1 + 2\eta),\, (1 + h_2)(1 + 2\eta), \cdots, (1 + h_{\ell-1})(1 + 2\eta),\, \theta + \eta + h_\ell(1 + 2\eta)\big)^T$, we further have

$$\begin{aligned}
\mathbf{1} - (1 + 2\eta)\Big(\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{1} + h\big)\Big) &= \big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\Big[\widetilde{\Lambda}^\alpha(h)\mathbf{1} - (1 + 2\eta)(\mathbf{1} + h)\Big] \\
&= (\theta - \eta - 1)\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{e}_1 + \mathbf{e}_\ell\big) < 0, \qquad (21)
\end{aligned}$$

where we use the facts that $\theta - \eta - 1 < 0$ for all large $n/K_n, \lambda$ and $\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{e}_1 + \mathbf{e}_\ell\big) > 0$ for all $h \geq 0$. As a result, the gradient $\nabla g(h) = \big(\frac{\partial g(h)}{\partial h_1}, \cdots, \frac{\partial g(h)}{\partial h_\ell}\big)^T < 0$ for all $h \in \mathbb{R}_+^\ell$. For a nonzero h^* in the convex cone $\mathbb{R}_+^\ell$, let $(0, h^*) \equiv \{\mu h^* : \mu \in (0, 1)\}$ be the open line segment joining 0 and h^*. It follows from the Mean Value Theorem that there exists $z \in (0, h^*)$ such that $g(h^*) = g(0) + \langle \nabla g(z), h^* \rangle$. Consequently, $g(h^*) < g(0)$. □

In view of the above proposition, it suffices to show that each element of $\big(\widetilde{\Lambda}^\alpha(0)\big)^{-1}\mathbf{1}$ is uniformly bounded, regardless of K_n, λ and α. This is proven in the following proposition, where $\widetilde{\Lambda}^\alpha(0)$ is denoted by Λ_0 for notational convenience.

Proposition 3.4. *Let $p = 1$, and n/K_n and $\lambda > 0$ be sufficiently large. Then for any $\ell \in \mathbb{N}$, the following holds for each $\Lambda_0 \in \mathbb{R}^{\ell \times \ell}$:*

$$0 < \mathbf{e}_i^T (\Lambda_0)^{-1} \mathbf{1} \le \frac{1+2\lambda}{\widetilde{\theta}_n + \widetilde{\eta}_n}, \qquad \forall\, i = 1, \cdots, \ell$$

Proof. When $\ell = 1$, we have $\Lambda_0 \ge 2(\theta + \eta) > 0$ (cf. (13)). Hence the proposition follows. Consider $\ell \ge 2$. Let $v_i \equiv \mathbf{e}_i^T (\Lambda_0)^{-1} \mathbf{1}$ for each i. Let $\rho \equiv (1 - \sqrt{1 - 4\eta^2})/(-2\eta)$ be the root of the equation $\eta x^2 + x + \eta = 0$ with $0 < \rho < 1$. It is known from [7] that

$$v_1 = v_\ell = \frac{\left(\sum_{j=0}^{\ell-1} \rho^j\right) \left[(\theta + \eta\rho) - \rho^{\ell-2}(\eta + \theta\rho)\right]}{(\theta + \eta\rho)^2 - \rho^{2(\ell-2)}(\eta + \theta\rho)^2} = \frac{\sum_{j=0}^{\ell-1} \rho^j}{(\theta + \eta\rho) + \rho^{\ell-2}(\eta + \theta\rho)}.$$

Moreover, it can be shown by direct calculation that for n/K_n sufficiently large, $\theta + \eta\rho > 0$ and $\eta + \theta\rho < 0$. Hence, $(\theta + \eta\rho) + \rho^{\ell-2}(\eta + \theta\rho) \ge (\theta + \eta\rho) + (\eta + \theta\rho) = (1 + \rho)(\theta + \eta) > 0$ for all $\ell \ge 2$. This shows $v_1 = v_\ell > 0$. Furthermore, notice

$$\left(\sum_{j=0}^{\ell-1} \rho^j\right)(\theta + \eta) - \left[(\theta + \eta\rho) + \rho^{\ell-2}(\eta + \theta\rho)\right]$$
$$= \frac{1 - \rho^{\ell-2}}{1 - \rho}(\eta + \theta\rho) - (1 - \rho^{\ell-2})\eta\rho = \frac{1 - \rho^{\ell-2}}{1 - \rho}\left[\eta + \theta\rho - \eta\rho + \eta\rho^2\right] = \frac{1 - \rho^{\ell-2}}{1 - \rho}\Big(\theta - \eta - 1\Big)\rho,$$

where $\eta\rho^2 + \rho + \eta = 0$ is used. Since $\theta - \eta - 1 < 0$ for n/K_n sufficiently large, we have $v_1 = v_\ell < 1/(\theta + \eta) = (1 + 2\lambda)/(\widetilde{\theta}_n + \widetilde{\eta}_n)$.

To obtain the desired result for other i's, we see from (21) that

$$(1 + 2\eta)\, v_i = 1 + (1 + \eta - \theta)\Big[\Big((\Lambda_0)^{-1}\mathbf{e}_1\Big)_i + \Big((\Lambda_0)^{-1}\mathbf{e}_\ell\Big)_i\Big]$$

It follows from Proposition 3.1 that $\left[(\Lambda_0)^{-1}\mathbf{e}_1\right]_i > 0$ and $\left[(\Lambda_0)^{-1}\mathbf{e}_\ell\right]_i > 0$. Hence $v_i > 0$. Furthermore, $\left[(\Lambda_0)^{-1}\mathbf{e}_\ell\right]_i = \left[(\Lambda_0)^{-1}\mathbf{e}_1\right]_{\ell-i}$. Therefore, $\left[(\Lambda_0)^{-1}\mathbf{e}_1\right]_i + \left[(\Lambda_0)^{-1}\mathbf{e}_\ell\right]_i \le \mathbf{e}_1^T (\Lambda_0)^{-1}\mathbf{1} = v_1 \le (\theta + \eta)^{-1}$. Consequently,

$$v_i \le \frac{1}{1 + 2\eta}\left[1 + (\eta + 1 - \theta)\frac{1}{\theta + \eta}\right] \le \frac{1 + 2\lambda}{1 + 2\widetilde{\eta}_n}\left[1 + \frac{\widetilde{\eta}_n + 1 - \widetilde{\theta}_n}{\widetilde{\theta}_n + \widetilde{\eta}_n}\right] \le \frac{1 + 2\lambda}{\widetilde{\theta}_n + \widetilde{\eta}_n}$$

□

Combining Propositions 3.3 and 3.4 and recalling $\widetilde{\theta}_n \to \widetilde{\theta}_*$ and $\widetilde{\eta}_n \to \widetilde{\eta}_*$ as $n/K_n \to \infty$, where $\widetilde{\theta}_* > 0$ and $\widetilde{\eta}_* > 0$, we have

Proposition 3.5. *Let $p = 1$. For all sufficiently large n/K_n and $\lambda > 0$, and for each index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$, the coefficients a_{ij}^α of each selection function $\overline{b}_i^\alpha(z)$ satisfy*

$$\sum_{j=1}^{K_n+p} \left|a_{ij}^\alpha\right| = \sum_{j=1}^{K_n+p} a_{ij}^\alpha \le \frac{2(1 + 2\lambda)}{\widetilde{\theta}_* + \widetilde{\eta}_*}, \qquad \forall\, i.$$

Remark 3.1. We point out two observations to be used in the following subsection.

(1) In view of Propositions 3.3 and 3.4, we conclude that for all sufficiently large n/K_n and $\lambda > 0$, $\|\big(\widetilde{\Lambda}^\alpha(h)\big)^{-1}\big(\mathbf{1} + h\big)\|_\infty \le 2(1 + 2\lambda)(\widetilde{\theta}_* + \widetilde{\eta}_*)^{-1}$ for all α and $h \ge 0$;

(2) All the results in this subsection remain true if $\widetilde{\eta}_*$ is replaced by an arbitrary positive number (with $\widetilde{\theta}_* = 1/2$). This observation is instrumental to the case $p \ge 2$ as shown in Proposition 3.6.

3.2 The Case of $p \geq 2$

In this case, $X^T X$ is a $(2p+1)$-diagonal matrix of order $(K_n + p)$, i.e.,

$$X^T X =$$

$$\begin{bmatrix}
\alpha_{1n} & \nu_{(2,1)n} & \cdots & \nu_{(p,1)n} & \gamma_{1n} & 0 & \cdots & 0 & \vdots & 0 & \cdots & 0 \\
\nu_{(2,1)n} & \alpha_{2n} & & \vdots & \vdots & \gamma_{1n} & \cdots & 0 & \vdots & 0 & \cdots & 0 \\
\vdots & & \ddots & \nu_{(p,p-1)n} & \vdots & & \ddots & & 0 & 0 & \cdots & 0 \\
\nu_{(p,1)n} & \cdots & \nu_{(p,p-1)n} & \alpha_{p\,n} & \gamma_{p\,n} & \cdots & \cdots & \gamma_{1n} & 0 & \cdots & \cdots & 0 \\
\gamma_{1n} & \cdots & \cdots & \gamma_{p\,n} & \beta_n & \gamma_{pn} & \cdots & \cdots & \gamma_{1n} & 0 & \cdots & 0 \\
 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & 0 & \cdots & 0 \\
 & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & 0 & \cdots & 0 \\
0 & \cdots & 0 & \gamma_{1n} & \cdots & \cdots & \gamma_{pn} & \beta_n & \gamma_{pn} & \cdots & \cdots & \gamma_{1n} \\
0 & \cdots & 0 & 0 & \gamma_{1n} & \ddots & \ddots & \gamma_{pn} & \alpha_{p\,n} & \nu_{(p,p-1)n} & \cdots & \nu_{(p,1)n} \\
0 & \cdots & 0 & 0 & 0 & \ddots & \ddots & \vdots & \nu_{(p,p-1)n} & \ddots & & \vdots \\
0 & \cdots & 0 & 0 & 0 & 0 & \gamma_{1n} & \vdots & \vdots & & \alpha_{2n} & \nu_{(2,1)n} \\
0 & \cdots & 0 & 0 & 0 & 0 & 0 & \gamma_{1n} & \nu_{(p,1)n} & \cdots & \nu_{(2,1)n} & \alpha_{1n}
\end{bmatrix}$$

Define

$$\lambda = \frac{\lambda^*}{\beta_n}, \qquad \Gamma_n = \frac{1}{\beta_n} X^T X, \qquad \text{and} \qquad \bar{y} = \frac{1}{\beta_n} X^T y$$

and

$$\widetilde{\theta}_{k\,n} \equiv \frac{\alpha_{kn}}{\beta_n}, \qquad \widetilde{\tau}_{(j,\,k)n} \equiv \frac{\nu_{(j,\,k)n}}{\beta_n}, \qquad \widetilde{\eta}_{kn} \equiv \frac{\gamma_{kn}}{\beta_n},$$

which satisfy $0 < \widetilde{\theta}_{1n} < \cdots < \widetilde{\theta}_{pn} < 1$, $0 < \widetilde{\eta}_{1n} < \cdots < \widetilde{\eta}_{pn} < 1$ with $\sum_{j=1}^{p} \widetilde{\eta}_{jn}$ dependent on p only, and $0 < \widetilde{\tau}_{(i,j)n} < \widetilde{\tau}_{(i-1,j)n} < 1$ and $0 < \widetilde{\tau}_{(i,j)n} < \widetilde{\tau}_{(i,j+1)n} < 1$. Moreover, define Λ (whose subscript n is dropped as before) as

$$\Lambda = (1+2\lambda)^{-1}(\Gamma_n + \lambda D^T D) =$$

$$\begin{bmatrix}
\theta_{1n} & \tau_{(2,1)n} & \cdots & \tau_{(p,1)n} & \eta_{1n} & 0 & \cdots & 0 & \vdots & 0 & \cdots & 0 \\
\tau_{(2,1)n} & \theta_{2n} & & \vdots & \vdots & \eta_{1n} & \cdots & 0 & \vdots & 0 & \cdots & 0 \\
\vdots & & \ddots & \tau_{(p,p-1)n} & \vdots & & \ddots & & 0 & 0 & \cdots & 0 \\
\tau_{(p,1)n} & \cdots & \tau_{(p,p-1)n} & \theta_{p\,n} & \eta_{p\,n} & \cdots & \cdots & \eta_{1n} & 0 & \cdots & \cdots & 0 \\
\eta_{1n} & \cdots & \cdots & \eta_{p\,n} & 1 & \eta_{pn} & \cdots & \cdots & \eta_{1n} & 0 & \cdots & 0 \\
 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & 0 & \cdots & 0 \\
 & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & 0 & \cdots & 0 \\
0 & \cdots & 0 & \eta_{1n} & \cdots & \cdots & \eta_{pn} & 1 & \eta_{pn} & \cdots & \cdots & \eta_{1n} \\
0 & \cdots & 0 & 0 & \eta_{1n} & \ddots & \ddots & \eta_{pn} & \theta_{p\,n} & \tau_{(p,p-1)n} & \cdots & \tau_{(p,1)n} \\
0 & \cdots & 0 & 0 & 0 & \ddots & \ddots & \vdots & \tau_{(p,p-1)n} & \ddots & & \vdots \\
0 & \cdots & 0 & 0 & 0 & 0 & \eta_{1n} & \vdots & \vdots & & \theta_{2n} & \tau_{(2,1)n} \\
0 & \cdots & 0 & 0 & 0 & 0 & 0 & \eta_{1n} & \tau_{(p,1)n} & \cdots & \tau_{(2,1)n} & \theta_{1n}
\end{bmatrix},$$

where

$$\theta_{k\,n} \equiv \frac{\widetilde{\theta}_{kn} + 2\lambda}{1+2\lambda}, \qquad \tau_{(j,\,k)n} \equiv \begin{cases} \dfrac{\widetilde{\tau}_{(j,\,k)n} - \lambda}{1+2\lambda}, & \text{if } k = j-1 \\ \dfrac{\widetilde{\tau}_{(j,\,k)n}}{1+2\lambda}, & \text{otherwise} \end{cases}, \qquad \eta_{kn} \equiv \begin{cases} \dfrac{\widetilde{\eta}_{kn} - \lambda}{1+2\lambda}, & \text{if } k = p \\ \dfrac{\widetilde{\eta}_{kn}}{1+2\lambda}, & \text{otherwise} \end{cases}$$

Following the similar discussions near Lemma 3.1, we obtain the matrix $\widetilde{\Lambda}^\alpha$ from Λ pertaining to each index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$. Specifically, let $\{\beta_1^\alpha, \cdots, \beta_\ell^\alpha\}$ be the partition of $\{1, \cdots, K_n + p\}$ corresponding to α. Then the elements of $\widetilde{\Lambda}^\alpha$ are given by $(\widetilde{\Lambda}^\alpha)_{ij} = \sum_{s \in \beta_i^\alpha, t \in \beta_j^\alpha} \Lambda_{st}$. For a given $p \geq 2$, choose $\widehat{\theta}_* = 1/2$ and let $\widehat{\eta}_* \equiv \sum_{j=1}^p \widetilde{\eta}_{j\,n} > 0$ (dependent on p only). Define $\theta_* \equiv (\widehat{\theta}_* + \lambda)/(1+2\lambda)$ and $\eta_* \equiv (\widehat{\eta}_* - \lambda)/(1+2\lambda)$. Hence $\eta_* = \sum_{j=1}^p \eta_{jn}$. Moreover, for each given K_n, define a tridiagonal matrix Λ_* of order $(K_n + p)$ with the similar structure as that of Λ defined in (12) but with θ and η replaced by θ_* and η_* respectively. Thus for each index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$, we obtain the corresponding matrix $\widetilde{\Lambda}_*^\alpha$ as in the prior subsection. Let $\Delta\widetilde{\Lambda}^\alpha \equiv \widetilde{\Lambda}^\alpha - \widetilde{\Lambda}_*^\alpha$. The lemma below shows that each element of $\Delta\widetilde{\Lambda}^\alpha$ is of order λ^{-1}.

Lemma 3.4. *For any given $p \geq 2$, there exists a positive constant χ_p, dependent on p only, such that for each given K_n and index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$, any row and column of $\Delta\widetilde{\Lambda}^\alpha$ has at most $(2p+1)$ nonzero elements, each of which satisfies $\left|(\Delta\widetilde{\Lambda}^\alpha)_{ij}\right| \leq \chi_p\,(1+2\lambda)^{-1}$.*

Proof. If $\Delta\widetilde{\Lambda}^\alpha$ is of order no less than $(2p+1)$, then it is a $(2p+1)$-diagonal matrix. Hence a row or column of $\Delta\widetilde{\Lambda}^\alpha$ has at most $(2p+1)$ nonzero elements. To establish the desired upper bound for nonzero elements, we consider two cases: (1) $i \neq j$; and (2) $i = j$. For case (1), recall that $\beta_i^\alpha \cap \beta_j^\alpha = \emptyset$. Hence, $\Lambda_{\beta_i^\alpha\,\beta_j^\alpha}$ is a sub-block of Λ above or below the diagonal of Λ and $\Lambda_{\beta_i^\alpha\,\beta_j^\alpha}$ contains at most $p(p+1)/2$ nonzero elements of Λ. Moreover, for $s \in \beta_i^\alpha$, $t \in \beta_j^\alpha$, we have: (i) if $|s-t| \geq 2$, then $|\Lambda_{st} - (\Lambda_*)_{st}| = |\Lambda_{st}| \leq (1+2\lambda)^{-1}$; and (ii) if $|s-t| = 1$, then $|\Lambda_{st} - (\Lambda_*)_{st}| \leq \widehat{\eta}_*(1+2\lambda)^{-1}$, where $\widehat{\eta}_* > 0$ depends on p only. Consequently, $|(\Delta\widetilde{\Lambda}^\alpha)_{ij}| \leq p(p+1)\max(1, \widehat{\eta}_*)/[2(1+2\lambda)]$ once $i \neq j$. In what follows, we consider case (2) where $i = j$, i.e., $\Lambda_{\beta_i^\alpha\,\beta_j^\alpha}$ is a principal submatrix of Λ. Recall $h_i = |\beta_i^\alpha| - 1$. Letting $d \equiv \min(p, h_i)$, it is noticed that $(\widetilde{\Lambda}^\alpha)_{ii} = D_0 + 2\sum_{k=1}^d D_k$, where D_0 is the sum of the diagonal entries of $\Lambda_{\beta_i^\alpha\,\beta_j^\alpha}$, and D_k is the sum of the (one-sided) kth off-diagonal entries of $\Lambda_{\beta_i^\alpha\,\beta_j^\alpha}$. Since at most $2p$ diagonal entries are different from 1 and each difference is bounded by $(1+2\lambda)^{-1}$, $D_0 = (1+h_i) + e_0$ with $|e_0| \leq 2p(1+2\lambda)^{-1}$. Similarly, at most $(p-1)$ 1st off-diagonal entries are different from $\eta_{p\,n}$ and each difference is bounded by $(1+2\lambda)^{-1}$. Thus $D_1 = h_i\,\eta_{p\,n} + e_1$ with $|e_1| \leq (p-1)(1+2\lambda)^{-1}$. In general, we have $D_k = (h_i + 1 - k)\,\eta_{(p+1-k)\,n} + e_k$ with $|e_k| \leq (p-k)(1+2\lambda)^{-1}$ for each $1 \leq k \leq d$. Consequently, by observing $d \leq p$, $(\widetilde{\Lambda}^\alpha)_{ii} = (1+h_i) + \sum_{k=1}^d 2(h_i + 1 - k)\,\eta_{(p+1-k)\,n} + \widetilde{e}$, where $|\widetilde{e}| \leq \zeta_p(1+2\lambda)^{-1}$ for some constant $\zeta_p > 0$, dependent on p only. In light of (13), we have $(\widetilde{\Lambda}_*^\alpha)_{ii} = (1+h_i) + 2h_i\,\eta_* + e'$, where $|e'| \leq 2(1+2\lambda)^{-1}$. Using $\eta_* = \sum_{k=1}^p \eta_{k\,n}$ and $|\eta_{(p+1-k)\,n}| \leq (1+2\lambda)^{-1}$ for all $k \geq 2$, we obtain $|(\Delta\widetilde{\Lambda}^\alpha)_{ii}| \leq \sum_{k=1}^d 2(k-1)\,|\eta_{(p+1-k)\,n}| + (\zeta_p + 2)(1+2\lambda)^{-1} \leq (p^2 + \zeta_p + 2)(1+2\lambda)^{-1}$. Hence $\chi_p \equiv \max\big(p(p+1)\max(1, \widehat{\eta}_*)/2,\ p^2 + \zeta_p + 2\big)$ is the desired upper bound, dependent on p only. □

We introduce more notation for the subsequent development. For a given matrix $A = [a_{ij}]$, let $|A| \equiv [|a_{ij}|]$ denote the matrix formed by the absolute values of the elements of A. It is easy to verify that for matrices A and B, $\||A|\|_\infty = \|A\|_\infty$ and $[|AB|]_{ij} \leq [|A| \cdot |B|]_{ij}, \forall\, i, j$.

Proposition 3.6. *Let $p \geq 2$, $\varrho \in (0,1)$, and $\gamma \in (\varrho, 1)$. Suppose that $K_n \sim n^\gamma$ and $\lambda \sim n^{2(\gamma-\varrho)}$. Then there exists $\kappa_p > 0$, dependent on p only, such that for all n sufficiently large and for each K_n and any index set $\alpha \subseteq \{1, \cdots, K_n + p - 1\}$, the coefficients a_{ij}^α of each selection function $\overline{b}_i^\alpha(z)$ satisfy $\sum_{j=1}^{K_n+p} |a_{ij}^\alpha| \leq \kappa_p(1+2\lambda)$.*

Proof. It is easy to verify that the given orders of K_n and $\lambda(n)$ satisfy the conditions in (b) of Proposition 3.1. Hence, for the θ_* and η_* corresponding to the given p, as long as n is sufficiently large, each element of $(\widetilde{\Lambda}^\alpha_*)^{-1}$ is positive and is not greater than $(\Lambda_*^{-1})_{11}$ for any K_n and index set α. Letting $\rho_* \in (0,1)$ be the solution of the equation $\eta_* x^2 + x + \eta_* = 0$, it is shown via a similar argument as in [7] that under the given order conditions, $\rho_*^{K_n} \sim \exp(-c\, n^{\varrho})$ with $c > 0$ such that $(\Lambda_*^{-1})_{11} \to (\theta_* + \eta_* \rho_*)^{-1} \sim \sqrt{\lambda(n)}$ as $n \to \infty$. Therefore, for all large n and any index set α, each element of $(\widetilde{\Lambda}^\alpha_*)^{-1}$ is of order no greater than $\sqrt{\lambda(n)}$. Since Lemma 3.4 shows that each column of $\Delta\widetilde{\Lambda}^\alpha$ has at most $(2p+1)$ nonzero elements of order λ^{-1}, we have, for all n sufficiently large, $\|(\widetilde{\Lambda}^\alpha_*)^{-1}(\Delta\widetilde{\Lambda}^\alpha)\|_\infty \sim \lambda^{-1/2}(n)$ for each α, where $\|\cdot\|_\infty$ denotes the induced infinity matrix norm. Hence $(\widetilde{\Lambda}^\alpha)^{-1} = (\widetilde{\Lambda}^\alpha_* + \Delta\widetilde{\Lambda}^\alpha)^{-1} = \big(\sum_{i=0}^\infty \big[-(\widetilde{\Lambda}^\alpha_*)^{-1}(\Delta\widetilde{\Lambda}^\alpha)\big]^i\big)(\widetilde{\Lambda}^\alpha_*)^{-1}$. Let $h^\alpha \in \mathbb{R}^\ell_+$ be defined similarly as in the last subsection, i.e., $h^\alpha_i \equiv |\beta^\alpha_i| - 1, \forall\, i = 1, \cdots, \ell$. For the selection function $\overline{b}^\alpha_i(z) \equiv \sum_{j=1}^{K_n+p} a^\alpha_{ij} z_j$, it follows from the similar argument before Proposition 3.3 that $\sum_{j=1}^{K_n+p} |a^\alpha_{ij}| \le \||(\widetilde{\Lambda}^\alpha)^{-1}|(\mathbf{1}+h^\alpha)\|_\infty$. Since $|(\widetilde{\Lambda}^\alpha)^{-1}|_{ij} \le \big[\big(\sum_{i=0}^\infty |(\widetilde{\Lambda}^\alpha_*)^{-1}(\Delta\widetilde{\Lambda}^\alpha)|^i\big)|(\widetilde{\Lambda}^\alpha_*)^{-1}|\big]_{ij}$ for all i, j and $\mathbf{1} + h^\alpha > 0$, we deduce that

$$\begin{aligned}\Big\||(\widetilde{\Lambda}^\alpha)^{-1}|(\mathbf{1}+h^\alpha)\Big\|_\infty &\le \Big\|\Big(\sum_{i=0}^\infty |(\widetilde{\Lambda}^\alpha_*)^{-1}(\Delta\widetilde{\Lambda}^\alpha)|^i\Big)|(\widetilde{\Lambda}^\alpha_*)^{-1}|(\mathbf{1}+h^\alpha)\Big\|_\infty \\ &\le \Big(\sum_{i=0}^\infty \|(\widetilde{\Lambda}^\alpha_*)^{-1}(\Delta\widetilde{\Lambda}^\alpha)\|^i_\infty\Big)\Big\|(\widetilde{\Lambda}^\alpha_*)^{-1}(\mathbf{1}+h^\alpha)\Big\|_\infty,\end{aligned}$$

where the last inequality is due to the fact that each element of $(\widetilde{\Lambda}^\alpha_*)^{-1}$ is positive. Therefore, we have $\||(\widetilde{\Lambda}^\alpha)^{-1}|(\mathbf{1}+h^\alpha)\|_\infty \le 2\,\|(\widetilde{\Lambda}^\alpha_*)^{-1}(\mathbf{1}+h^\alpha)\|_\infty$ for all K_n, λ and all α as long as n is sufficiently large. Since it is observed from Remark 3.1 that $\|(\widetilde{\Lambda}^\alpha_*)^{-1}(\mathbf{1}+h^\alpha)\|_\infty \le 2(1+2\lambda)(\widehat{\theta}_* + \widehat{\eta}_*)^{-1}$ for all K_n, λ and α (for any large n), where $\widehat{\theta}_* + \widehat{\eta}_*$ depends on p only, the proposition follows. □

3.3 Proof of Theorem 3.1

Since $z = \bar{y}/(1+2\lambda)$, each function $\hat{b}^\alpha_i(\bar{y}) = \sum_{j=1}^{K_n+p} \bar{a}^\alpha_{ij}\, \bar{y}_j$, where $\bar{a}^\alpha_{ij} = a^\alpha_{ij}/(1+2\lambda) > 0, \forall\, i, j = 1, \cdots, K_n + p$ for each index set α. By virtue of Propositions 3.2, 3.5 and 3.6, we have, for any $p \in \mathbb{Z}_+$, under the specified conditions in each proposition, the mapping $\bar{y} \mapsto \hat{b}_i$ is a continuous piecewise linear function whose each selection function $\hat{b}^\alpha_i : \mathbb{R}^{K_n+p} \to \mathbb{R}$ satisfies $\big|\hat{b}^\alpha_i(\bar{y})\big| \le \big(\sum_{j=1}^{K_n+p} |\bar{a}^\alpha_{ij}|\big)\max(|\bar{y}_i|) = \kappa_p \|\,\bar{y}\|_\infty, \forall\, y$, namely, each $\hat{b}^\alpha_i$ has the Lipschitz constant κ_p, regardless of K_n, λ, α and i. Hence, for a given p and a fixed K_n, $\hat{b}_i$ admits a conic subdivision of $\mathbb{R}^{K_n+p}$ [4, 21, 23], i.e., $\mathbb{R}^{K_n+p}$ is partitioned into finitely many polyhedral cones and $\hat{b}_i$ coincides with one of its selection functions on each cone. For arbitrary $u, v \in \mathbb{R}^{K_n+p}$, the line segment joining u and v is partitioned by the conic subdivision into finitely many sub-segments, on each of which $\hat{b}_i$ has the same Lipschitz constant κ_p. It thus follows from the similar proof of [4, Proposition 4.2.2] that $\big|\hat{b}_i(u) - \hat{b}_i(v)\big| \le \kappa_p \|u - v\|_\infty, \forall\, u, v \in \mathbb{R}^{K_n+p}$. □

3.4 Implications of Uniform Lipschitz Property

The uniform Lipschitz property of $\hat{b}$ leads to two important consequences that pave the way to asymptotic analysis: stochastic boundedness and consistency at the boundary. Specifically, let $\check{b} \equiv \hat{b}(\mathbb{E}(\bar{y}_n))$, where $\mathbb{E}(\cdot)$ denotes the expectation operator, and define $\bar{f}^{[p]}(x) \equiv \sum_{k=1}^{K_n+p} \check{b}_k B^{[p]}_k(x)$. According to Theorem 3.1, we have

$$\sup_{x\in[0,1]} |\hat{f}^{[p]}(x) - \bar{f}^{[p]}(x)| \le \|\hat{b}(\bar{y}) - \hat{b}(\mathbb{E}(\bar{y}))\|_\infty \le \kappa_p \|\bar{y} - \mathbb{E}(\bar{y})\|_\infty = O_p\big(\sqrt{n^{-1}K_n \log K_n}\,\big), \quad (22)$$

where "$a = O_p(b)$" means that a/b is bounded in probability. Let $\mu = \lambda^*/(nK_n)$. It shall be shown that, under mild conditions on f,

$$\sup_{x\in[0,1]} |\bar{f}^{[p]}(x) - f(x)| = \begin{cases} O(\mu), & \text{if } p = 1 \\ O(\mu) + O(K_n^{-1}), & \text{if } p \neq 1 \end{cases} \tag{23}$$

The development of (23) is a special case of Theorem 4.1. Combining (22) and (23), we have

$$\sup_{x\in[0,1]} |\hat{f}^{[p]}(x) - f(x)| = \begin{cases} O_p\big(\sqrt{n^{-1}K_n \log K_n}\,\big) + O(\mu), & \text{if } p = 1 \\ O_p\big(\sqrt{n^{-1}K_n \log K_n}\,\big) + O(\mu) + O(K_n^{-1}), & \text{if } p \neq 1 \end{cases} \tag{24}$$

Brunk's estimator (1) is inconsistent at boundary points [31], which is known as the *spiking problem*. In contrast, (24) shows that $\hat{f}^{[p]}$ is stochastically uniformly bounded and $\hat{f}^{[p]}(0)$ is consistent if $n^{-1}K_n \log K_n \to 0$ and $\mu \to 0$ as $K_n \to \infty$ and $n \to \infty$. This result is critical to estimation of error terms in asymptotic analysis in Section 4; see the proof of Lemma 4.1, for example.

4 Asymptotic Properties of Monotone P-Spline Estimator

4.1 Linear Splines: $p = 1$

We first concentrate on $\hat{f}^{[p]}$ with $p = 1$. The closed form representation of $\hat{f}^{[1]}$ is unavailable from (8), which makes it difficult to study its properties. In this subsection, we replace the difference equation (8) by its analogous differential equation to establish its asymptotic distributions.

Let ω be the uniform distribution on $x_1, \ldots, x_n$, and let g be the piecewise constant function for which $g(x_i) = y_i$ for $i = 1, \ldots, n$. Define $\hat{F}(x) \equiv \int_0^x \hat{f}^{[1]}(y)dy$ and

$$G(x) \equiv \int_0^x g(y)d\omega(y) = \frac{1}{n}\sum_{i=1}^n y_i \; I\{x \in \mathbb{R} \,|\, x_i \leq x\},$$

where I denotes the indicator function of a set. Denote $\tilde{G}$ the greatest convex minorant of the cumulative sum diagram G, i.e., $\tilde{G}$ is the supremum of all convex functions lying below G (see [3, p.11] for a similar discussion on a convex hull of a given set of points). Hence, $\tilde{G}$ is a convex and piecewise linear function. It is shown in [15] that G and $\tilde{G}$ are close when the derivative of the true regression function f is bounded away from zero, and $\|G - \tilde{G}\| = O_p((n^{-1}\log n)^{2/3})$, where $\|f\| \equiv \sup_{[0,1]} |f(x)|$. The subsequent norms are defined in the same way. For any $x \in (0, 1)$, let $d = \lfloor K_n x \rfloor$. It is clear that $\hat{F}''(x) = K_n(\hat{b}_{d+2} - \hat{b}_{d+1})$. Let

$$R_1(x) = \big[\hat{F}(x) - G(x)\big] - \Big[\frac{1}{n}\sum_{k=1}^{d+1}\sum_{i=1}^{n} B_k^{[1]}(x_i)\hat{f}^{[1]}(x_i) - \frac{1}{n}\sum_{k=1}^{d+1}\sum_{i=1}^{n} B_k^{[1]}(x_i)y_i\Big]. \tag{25}$$

Recall that $\mu = \lambda^*/(nK_n)$. Thus, the optimality condition (8) becomes the following ODE with a constrained right-hand side and thus a complementarity system [22, 23]: $\mu\hat{F}'' = \big[\hat{F} - G - R_1\big]_+$. Define $R_2 = (\hat{F} - \tilde{G}) - \mu\hat{F}''$. Then, $\hat{F}$ solves the differential equation

$$\mu\hat{F}''(t) = \hat{F}(t) - \tilde{G}(t) - R_2(t), \quad t \in [0, 1], \tag{26}$$

with two boundary conditions $\hat{F}(0) = 0$ and $\hat{F}(1) = \tilde{G}(1) + e_1$ by (9), where $e_1 = \hat{F}(1) - \check{F}(1)$ is of order $O_p(n^{-1})$ since $\hat{f}^{[1]}$ is bounded with probability one according to (22) and (23). The following lemma shows that $\|R_2\|$ is small and of order $O_p((n^{-1}\log n)^{2/3}) + O_p((n^{-1}K_n^{-1}\log K_n)^{1/2})$.

Lemma 4.1. *Assume that either (i) μ and K_n satisfy $\mu n^{2/3} \to \infty$, $\mu n^{2/5} \to 0$, and $\mu^{-1/2}\log K_n/K_n \to 0$; or (ii) $\mu = c^2 n^{-2/5}$ and $K_n \sim n^\gamma$ with $\gamma > 1/5$. Then*

$$\|R_2\| \quad = \quad O_p\Big((\frac{\log n}{n})^{2/3}\Big) + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big).$$

Proof. Let $\tilde d = \lfloor nx \rfloor$. Define $\check F(x) = \int_0^x \hat f^{[1]}(y)d\omega(y)$. Note that

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{d+1}\sum_{i=1}^{n} B_k^{[1]}(x_i)\hat f^{[1]}(x_i) &= \frac{1}{n}\sum_{i=1}^{dM_n}\hat f^{[1]}(x_i) + \frac{1}{n}\sum_{i=dM_n+1}^{(d+1)M_n} B_{d+1}^{[1]}(x_i)\hat f^{[1]}(x_i)\\ &= \check F(x) - \frac{1}{n}\sum_{i=dM_n+1}^{\tilde d}\hat f^{[1]}(x_i) + \frac{1}{n}\sum_{i=dM_n+1}^{(d+1)M_n} B_{d+1}^{[1]}(x_i)\hat f^{[1]}(x_i).\end{aligned}$$

Similarly, we have

$$\frac{1}{n}\sum_{k=1}^{d+1}\sum_{i=1}^{n} B_k^{[1]}(x_i)y_i = G(x) - \frac{1}{n}\sum_{i=dM_n+1}^{\tilde d} y_i + \frac{1}{n}\sum_{i=dM_n+1}^{(d+1)M_n} B_{d+1}^{[1]}(x_i)y_i.$$

Hence, it follows from (25) that

$$\begin{aligned}R_1(x) &= \hat F(x) - \check F(x) + \frac{1}{n}\sum_{i=dM_n+1}^{\tilde d}(\hat f^{[1]}(x_i) - y_i) - \frac{1}{n}\sum_{i=dM_n+1}^{(d+1)M_n} B_{d+1}^{[1]}(x_i)(\hat f^{[1]}(x_i) - y_i)\\ &= \hat F(x) - \check F(x) + W_1(x) + W_2(x),\end{aligned}$$

where

$$\begin{aligned}W_1(x) &= \frac{1}{n}\sum_{i=dM_n+1}^{\tilde d}(\hat f^{[1]}(x_i) - f(x_i)) - \frac{1}{n}\sum_{i=dM_n+1}^{(d+1)M_n} B_{d+1}^{[1]}(x_i)(\hat f^{[1]}(x_i) - f(x_i)),\\ W_2(x) &= \frac{1}{n}\sum_{i=dM_n+1}^{\tilde d}(f(x_i) - y_i) - \frac{1}{n}\sum_{i=dM_n+1}^{(d+1)M_n} B_{d+1}^{[1]}(x_i)(f(x_i) - y_i).\end{aligned}$$

It is clear that $\|\hat F - \check F\| = O_p(n^{-1})$ [16]. From (24), we have $\|W_1\| \le 2n^{-1}M_n\|\hat f^{[1]} - f\| = O_p(\sqrt{\log K_n/(nK_n)})$. Note that $W_2(x)$ is a normal random variable with mean zero and variance of order $O((nK_n)^{-1})$, and hence $\|W_2\|$ is of order $O_p(\sqrt{\log K_n/(nK_n)})$. Therefore, $\|R_1\|$ is of order $O_p(\sqrt{\log K_n/(nK_n)}) + O_p(n^{-1})$. Since $R_2 = \mu\hat F'' - (\hat F - \tilde G)$, we have

$$\begin{aligned}\|R_2\| &= \|(\hat F - G - R_1)_+ - (\hat F - \tilde G)\| \le \|(\hat F - G)_+ - (\hat F - \tilde G)_+\| + \|(\hat F - \tilde G)_-\| + \|R_1\|\\ &\le \|G - \tilde G\| + \|(\check F - \tilde G)_-\| + \|\hat F - \check F\| + \|R_1\|\\ &= O_p\Big((\frac{\log n}{n})^{2/3}\Big) + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big) + O_p\Big(\frac{1}{n}\Big),\end{aligned}$$

where $\|(\check F - \tilde G)_-\| = O_p(n^{-1})$ is given by [16, Lemma 2]. Hence, the lemma follows. $\square$

Denote by $\xi = \mu^{-1/2}$. The solution to (26) can be expressed explicitly by the corresponding Green's function [16]: $\chi_\mu(t,s) = 2^{-1}\,\xi\,\exp(-\xi|t-s|)$, $0 \le t \le 1$. Using this, we have

$$\hat F(x) = \int_0^1 \chi_\mu(x,s)\tilde G(s)ds + \int_0^1 \chi_\mu(x,s)R_2(s)ds + c_0(\xi)e^{-\xi x} + c_1(\xi)e^{-\xi(1-x)}, \qquad (27)$$

where both c_0 and c_1 can be obtained from the boundary conditions and it can be shown that $|c_0(\xi)| + |c_1(\xi)| \le 6\|\tilde G + R_2\| + 4\|\hat F\|$, for $\xi \ge 1$.

Theorem 4.1. *Assume that the true regression function f is twice continuously differentiable. Then, the function $\hat{f}$ is given by*

$$\begin{aligned}\hat{f}^{[1]}(x) &= f(x) + \mu f''(x) + o(\mu) + \frac{\sigma\xi}{2n}\sum_{i=1}^{n} e^{-\xi|x-x_i|}z_i + O_p\Big((\frac{n}{\log n})^{-2/3}\Big)\xi \\ &\quad + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big)\xi + e^{-\xi x(1-x)}O_p(\xi) \qquad (28)\end{aligned}$$

uniformly in λ and $x \in (0,1)$. Moreover, if f is three times continuously differentiable, then

$$\begin{aligned}\frac{d}{dx}\hat{f}^{[1]}(x) &= f'(x) + \mu f'''(x) + o(\mu) + O_p\Big(\frac{1}{\sqrt{n\xi}}\Big) + O_p\Big((\frac{n}{\log n})^{-2/3}\Big)\xi^2 \\ &\quad + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big)\xi^2 + e^{-\xi x(1-x)}O_p(\xi^2) \qquad (29)\end{aligned}$$

uniformly in λ and $x \in (0,1)$.

Proof. Let $F(x) = \int_0^x f(y)dy$ and $\acute{F}(x) = \int_0^x f(y)d\omega(y)$. Obviously, $\|\acute{F} - F\| = O(n^{-1})$. Differentiating pointwise of equation (27), we have

$$\begin{aligned}\hat{f}(x) &= \int_0^1 \frac{\partial}{\partial x}\chi_\mu(x,s)\tilde{G}(s)ds + \int_0^1 \frac{\partial}{\partial x}\chi_\mu(x,s)R_2(s)ds \;-\; \xi e^{-\xi x}c_0(\xi) + \xi e^{-\xi(1-x)}c_1(\xi) \\ &= \int_0^1 \frac{d}{ds}\chi_\mu(x,s)F(s)ds + V_0(x) + V_1(x) + V_2(x), \qquad (30)\end{aligned}$$

where

$$\begin{aligned}V_0(x) &= \int_0^1 \frac{\partial}{\partial x}\chi_\mu(x,s)[G(s) - \acute{F}(s)]ds = -\int_0^1 \frac{\partial}{\partial s}\chi_\mu(x,s)[G(s) - \acute{F}(s)]ds \\ &= -\chi_\mu(x,1)[G(1) - \tilde{F}(1)] + \int_0^1 \chi_\mu(x,s)[dG(s) - d\acute{F}(s)] \\ &= -\frac{\xi}{2n}e^{-\xi(1-x)}\sum_{i=1}^{n}\sigma z_i + \frac{\xi}{2n}\sum_{i=1}^{n} e^{-\xi|x-x_i|}\sigma z_i = O_p\Big(\frac{\xi e^{-\xi(1-x)}}{\sqrt{n}}\Big) + \frac{\xi}{2n}\sum_{i=1}^{n} e^{-\xi|x-x_i|}\sigma z_i, \\ V_1(x) &= \int_0^1 \frac{\partial}{\partial x}\chi_\mu(x,s)\big[\tilde{G}(s) + R_2(s) - F(s) - G(s) + \acute{F}(s)\big]ds, \\ V_2(x) &= -\xi e^{-\xi x}c_0(\xi) + \xi e^{-\xi(1-x)}c_1(\xi) - \frac{1}{2}\xi e^{-\xi(1-x)}F(1) = e^{-\xi x(1-x)}O_p(\xi).\end{aligned}$$

However, since $\|\acute{F} - F\| = O(n^{-1})$,

$$|V_1(x)| < \frac{1}{2}\Big\|\tilde{G} - G + R_2 - F + \acute{F}\Big\|\int_0^1 \xi^2 e^{-\xi|x-s|}ds \;=\; O_p\Big((\frac{n}{\log n})^{-2/3}\Big)\xi + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big)\xi.$$

Furthermore,

$$\begin{aligned}\int_0^1 \frac{\partial}{\partial x}\chi_\mu(x,s)F(s)ds &= -\int_0^1 \frac{\partial}{\partial s}\chi_\mu(x,s)F(s)ds \;=\; -\chi_\mu(x,1)F(1) + \int_0^1 \chi_\mu(x,s)f(s)ds \\ &= -\frac{1}{2}F(1)\xi e^{\xi(x-1)} + \int_0^1 \chi_\mu(x,s)f(s)ds.\end{aligned}$$

Note that $f_0(x) = \int_0^1 \chi_\mu(x,s)f(s)ds$ satisfies the equation $\mu f_0'' = f_0 - f$. By equation (6.4) in [14, Theorem 2.2], we obtain $\int_0^1 \chi_\mu(x,s)f(s)ds = f(x) + \mu f''(x) + o(\mu)$. This gives rise to the representation of $\hat{f}$ in (28). By differentiating (30) again, (29) can be established similarly. □

Theorem 4.1 indicates that the monotone P-spline estimator is approximately a kernel regression estimator. The equivalent kernel is the double-exponential or Laplace kernel and the equivalent bandwidth is of order μ. The asymptotic mean has the bias $\mu f''(x) + o(\mu)$, which is negligible if μ is reasonably small. On the other hand, μ can not be arbitrarily small as that will inflate the random component. The admissible range for μ given in Theorem 4.2 is a compromise between these two.

Theorem 4.2. *Suppose that f is twice continuously differentiable with bounded second order derivative on* $[0, 1]$.

(a) *If μ and K_n satisfy $\mu n^{2/3} \to \infty$, $\mu n^{2/5} \to 0$, and $\mu^{-1/2} \log K_n / K_n \to 0$, then*

$$\sqrt{n\mu^{\frac{1}{2}}}\ (\hat{f}^{[1]}(x) - f(x)) \longrightarrow N\Big(0, \frac{\sigma^2}{4}\Big) \tag{31}$$

in distribution as $n \to \infty$.

(b) *If $\mu = c^2 n^{-2/5}$ and $K_n \sim n^\gamma$ with $\gamma > 1/5$, then*

$$n^{\frac{2}{5}}(\hat{f}^{[1]}(x) - f(x)) \longrightarrow N\Big(c^2 f''(x), \frac{\sigma^2}{4c}\Big) \tag{32}$$

in distribution as $n \to \infty$.

Proof. Let $\Pi_\mu(x) = \sigma\xi/(2n) \sum_{i=1}^n e^{-\xi|x-x_i|} z_i$. For any fixed x, the Lindeberg-Levy Central Limit Theorem gives

$$\sqrt{n\mu^{1/2}}\ \Pi_\mu(t) \longrightarrow N\Big(0, \frac{\sigma^2}{4}\Big),$$

in distribution. When μ and K_n satisfy $\mu n^{2/3} \to \infty$, $\mu n^{2/5} \to 0$, and $\mu^{-1/2} \log K_n / K_n \to 0$, it is easy to see that the remainder terms in (28) are $o_p(1)$. If $\mu = c^2 n^{-2/5}$, we have $\sqrt{n\mu^{1/2}}\mu f''(x) = c^2 f''(x)$. Hence the theorem follows. □

When $\mu \sim n^{-2/3}$, this yields the slowest rate of convergence ($\sim n^{1/3}$) in the limit, which is the same as that of Brunk's estimator in (1). The asymptotic results in Theorem 4.2 provide theoretical justification of the observation that the number of knots is not important, as long as it is above some minimal level [18]. A comparison to [7, Theorem 4] shows that both the unconstrained P-spline estimator and the monotone P-spline estimator share the same asymptotic distribution given in (32). It is also interesting to notice that the monotone linear P-spline estimator and the monotone linear smoothing spline estimator in [16] are asymptotically equivalent. However, many challenges emerge for both algorithms and asymptotic analysis when we shift from linear monotone smoothing splines to higher-degree counterparts. On the other hand, it is relatively easier to obtain monotone P-spline estimators of other degrees. We discuss these estimators in Subsection 4.2.

4.2 Splines of Other Degrees: $p \neq 1$

In this subsection, we study the asymptotic properties of $\hat{f}^{[p]}(x) = \sum_{k=1}^{K_n+p} \hat{b}_k B_k^{[p]}(x)$ when $p \neq 1$. We first define a piecewise linear function $\tilde{f}^{[p]}$, where $\hat{f}^{[p]}$ and $\tilde{f}^{[p]}$ share the same set of spline coefficients. In particular, define $\tilde{f}^{[0]}(x) = \sum_{k=1}^{K_n} \hat{b}_k B_k^{[1]}(x)$, and $\tilde{f}^{[p]}(x) = \sum_{k=1}^{K_n+1} \hat{b}_k^{[p]} B_k^{[1]}(x)$ if

$p \geq 2$. Note that $\tilde{f}^{[0]}$ is defined on $[0, 1-1/K_n]$. Denote $\tilde{F}(x) = \int_0^x \tilde{f}^{[p]}(y)dy$. For any $x \in (0,1)$ and $d = \lfloor K_n x \rfloor$, let

$$R_3(x) = \big[\tilde{F}(x) - G(x)\big] - \frac{1}{n}\Big[\sum_{k=1}^{d+1}\sum_{i=1}^{n} B_k^{[p]}(x_i)\tilde{f}(x_i) - \sum_{k=1}^{d+1}\sum_{i=1}^{n} B_k^{[p]}(x_i)y_i\Big]. \tag{33}$$

Thus, the optimality condition (8) becomes $\mu\tilde{F}''(x) = [\tilde{F} - G - R_3]_+$. Define $R_4 = (\tilde{F} - \tilde{G}) - \mu\tilde{F}''$. Then, $\tilde{F}$ solves the differential equation $\mu\tilde{F}''(t) = \tilde{F}(t) - \tilde{G}(t) - R_4(t),\ t \in [0,1]$, with two boundary conditions $\tilde{F}(0) = 0$ and $\tilde{F}(1) = \tilde{G}(1) + e_2$, where $e_2 = O_p(n^{-1})$. Following the same discussion as in Subsection 4.1, we can establish the asymptotic distribution for $\tilde{f}^{[p]}$ as in (31) and (32), respectively, under different admissible ranges of μ and K_n.

Lemma 4.2. *For any $x \in (0,1)$, let $d = \lfloor K_n x \rfloor$. Then,*

$$\hat{f}^{[0]}(x) = \tilde{f}^{[0]}(x) + \frac{1}{2}\frac{d\tilde{f}^{[0]}(x)}{dx}\Big[(\kappa_{d+1} - x)^2 - (x - \kappa_d)^2 - \frac{1}{K_n}\Big], \tag{34}$$

and for $p \geq 2$,

$$\hat{f}^{[p]}(x) = \tilde{f}^{[p]}(x) + \sum_{q=2}^{p}\sum_{i=d+2}^{d+q+1}\frac{1}{q}\frac{d\tilde{f}^{[p]}(x + \frac{i-d}{K_n})}{dx}(x - \kappa_{i-q})B_i^{[q-1]}(x). \tag{35}$$

Proof. Direct algebra yields

$$\begin{aligned}\hat{f}^{[0]}(x) &= K_n(\hat{F}(\kappa_{d+1}) - \hat{F}(\kappa_d)) \;=\; K_n(\tilde{F}(\kappa_d + 1) - \tilde{F}(\kappa_d)) - \frac{1}{2}(\hat{b}_{d+2} - \hat{b}_{d+1}) \\ &= \tilde{f}^{[0]}(x) + \frac{1}{2}\frac{d\tilde{f}^{[0]}(x)}{dx}\Big[(\kappa_{d+1} - x)^2 - (x - \kappa_d)^2 - \frac{1}{K_n}\Big].\end{aligned}$$

The B-spline basis functions have the recurrence relationship such that

$$B_j^{[p]}(x) = \frac{K_n}{p}\Big(x - \kappa_{j-p-1}\Big)B_{j-1}^{[p-1]}(x) + \frac{K_n}{p}\Big(\kappa_j - x\Big)B_j^{[p-1]}(x).$$

Let $f^{[p-1]}(x) = \sum_{k=1}^{K_n+p-1} b_k B_k^{[p-1]}(x)$ with the same first $(K_n + p - 1)$ coefficients of $\hat{f}^{[p]}$. For $x \in (\kappa_d, \kappa_{d+1})$, the difference between $f^{[p]}(x)$ and $f^{[p-1]}(x)$ is given by

$$\begin{aligned}f^{[p]}(x) - f^{[p-1]}(x) &= \sum_{i=d+2}^{d+p+1}\Big[b_{i+1}\frac{K_n}{p}(x - \kappa_{i-p}) + b_i\big(\frac{K_n}{p}(\kappa_i - x) - 1\big)\Big]B_i^{[p-1]}(x) \\ &= \sum_{i=d+2}^{d+p+1}(b_{i+1} - b_i)\Big(\frac{K_n}{p}(x - \kappa_{i-p})\Big)B_i^{[p-1]}(x).\end{aligned} \tag{36}$$

Therefore,

$$\begin{aligned}\hat{f}^{[p]}(x) &= \tilde{f}^{[p]}(x) + \sum_{q=2}^{p}\sum_{i=d+2}^{d+q+1}(b_{i+1} - b_i)\Big(\frac{K_n}{q}(x - \kappa_{i-q})\Big)B_i^{[q-1]}(x) \\ &= \tilde{f}^{[p]}(x) + \sum_{q=2}^{p}\sum_{i=d+2}^{d+q+1}\frac{d\tilde{f}^{[p]}(x + \frac{i-d}{K_n})}{dx}\Big(\frac{1}{q}(x - \kappa_{i-q})\Big)B_i^{[q-1]}(x).\end{aligned}$$

Hence, the lemma follows. □

Theorem 4.3. *Suppose that f is three times continuously differentiable with bounded third order derivative on $[0,1]$. Let $p \neq 1$.*

(a) *If μ and K_n satisfy $\mu n^{2/3} \to \infty$, $\mu n^{2/5} \to 0$, and $\mu^{-1/2} \log K_n / K_n \to 0$, then*

$$\sqrt{n\mu^{\frac{1}{2}}}\Big(\hat{f}^{[p]}(x) - f(x) - r_n^{[p]}(x)\Big) \longrightarrow N\Big(0, \frac{\sigma^2}{4}\Big) \tag{37}$$

in distribution as $n \to \infty$, where

$$r_n^{[p]}(x) = \begin{cases} -\frac{1}{2K_n} f'(x) & \text{if } p = 0 \\ f'(x) \sum_{q=2}^{p} \sum_{i=d+2}^{d+q+1} \frac{1}{q}(x - \kappa_{i-q}) B_i^{[q-1]}(x) & \text{if } p \geq 2. \end{cases}$$

(b) *If $\mu = c^2 n^{-2/5}$ and $K_n \sim n^{\gamma}$ with $\gamma > 1/5$, then*

$$n^{\frac{2}{5}}\Big(\hat{f}^{[p]}(x) - f(x) - r_n^{[p]}(x)\Big) \longrightarrow N\Big(c^2 f''(x), \frac{\sigma^2}{4c}\Big) \tag{38}$$

in distribution as $n \to \infty$.

Proof. We may go through the same proof as of Theorem 4.1 to establish the asymptotic distribution of $\tilde{f}^{[p]}(x)$. For $x \in (0,1)$, equation (34) shows that the difference between $\hat{f}^{[0]}(x)$ and a piecewise linear function $\tilde{f}^{[0]}(x)$ is

$$\frac{1}{2}\frac{d\tilde{f}^{[0]}(x)}{dx}\Big[\big(\kappa_{d+1} - x\big)^2 - \big(x - \kappa_d\big)^2 - \frac{1}{K_n}\Big]. \tag{39}$$

By (29), we have

$$\begin{aligned} \frac{d}{dx}\tilde{f}^{[0]}(x) &= f'(x) + \mu f'''(x) + o(\mu) + O_p\Big(\frac{1}{\sqrt{n\xi}}\Big) + O_p\Big((\frac{n}{\log n})^{-2/3}\Big)\xi^2 \\ &\quad + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big)\xi^2 + e^{-\xi x(1-x)} O_p(\xi^2). \end{aligned}$$

Hence, (39) is equal to $-f'(x)/2K_n + o_p(1/\sqrt{n\mu^{1/2}}\,)$. Combining the results in Theorem 4.2, part (a) follows. By (29), we also have

$$\begin{aligned} \frac{d}{dx}\tilde{f}^{[p]}(x + \frac{d}{K_n}) &= f'(x) + O\Big(\frac{1}{K_n}\Big) + \mu f'''(x + \frac{d}{K_n}) + o(\mu) + O_p\Big(\frac{1}{\sqrt{n\xi}}\Big) \\ &\quad + O_p\Big((\frac{n}{\log n})^{-2/3}\Big)\xi^2 + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big)\xi^2 + e^{-\xi x(1-x)} O_p(\xi^2). \end{aligned}$$

For any $x \in (0,1)$, the difference between $\hat{f}^{[p]}(x)$ and a piecewise linear function $\tilde{f}^{[p]}(x)$ is given by

$$\begin{aligned} &\sum_{q=2}^{p} \sum_{i=d+2}^{d+q+1} \frac{1}{q} \frac{d\tilde{f}^{[p]}(x + \frac{i-d}{K_n})}{dx}(x - \kappa_{i-q}) B_i^{[q-1]}(x) \\ &= \sum_{q=2}^{p} \sum_{i=d+2}^{d+q+1} \frac{1}{q} f'(x)(x - \kappa_{i-q}) B_i^{[q-1]}(x) + O_p\Big(\frac{1}{K_n^2}\Big) + O_p\Big(\frac{\mu}{K_n}\Big) + O_p\Big(\frac{1}{K_n\sqrt{n\xi}}\Big). \end{aligned}$$

Thus (b) follows easily. □

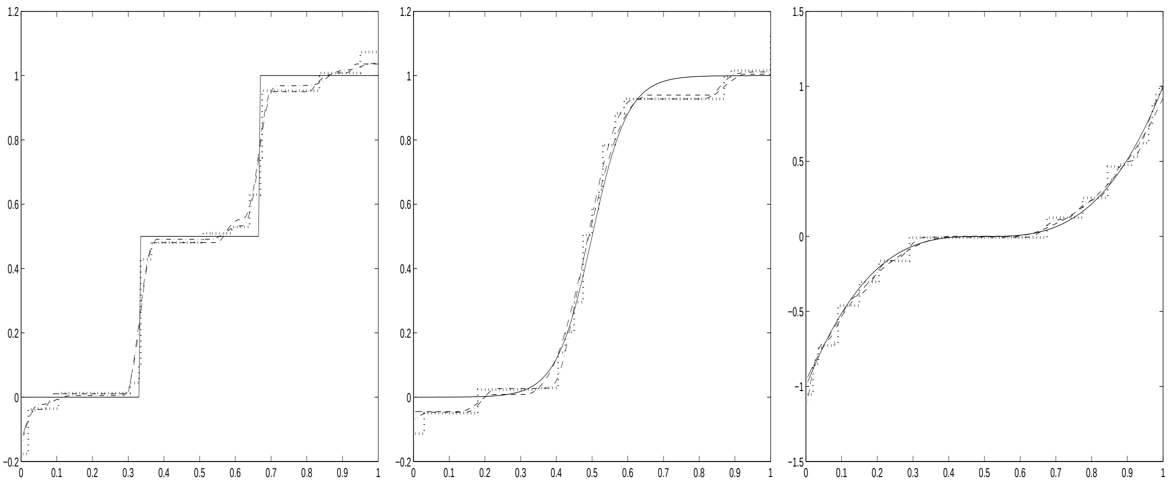

Figure 1: Monotone penalized spline estimator with $p = 2$ (dashes), Brunk's estimate (dots), quadratic regression spline estimator (dot-dash), and the true regression curve (solid).

An interesting observation is that the convergence rate of $\hat{f}^{[p]}$ does not depend on the spline degree p. Comparing Theorem 4.2 with Theorem 4.3, it is clear that the asymptotic distributions are the same except that the asymptotic bias term of $\hat{f}^{[p]}$ has a higher order when $p \neq 1$, where both $r_n^{[0]}$ and $r_n^{[p]}$ defined in Theorem 4.3 are of order $O(K_n^{-1})$. Further, the modeling bias due to approximating f by a spline is asymptotically negligible if $K_n \sim n^\gamma$ with $\gamma > 2/5$. While the similar observation is attained in [7] for the *unconstrained* P-splines, it is the first time that this is rigorously established for the monotone P-spline estimator.

5 Examples and Discussions

5.1 Simulation Examples

In this subsection, simulation results are presented to compare the performance of the following three estimators: the Brunk's estimator (BK), the monotone quadratic regression spline estimator (QUAD) developed by [12], and the proposed monotone penalized spline estimator (PM) with $p = 2$. We choose the number of knots for both quadratic regression splines and monotone penalized splines as $K_n = 20, 60, 100$, respectively. The x_i's are defined in the interval $[0, 1]$ with sample size $n = 200$. The noise distribution is normal with standard deviation 0.2. Figure 1 shows the true functions of the three examples together with the estimates obtained from the estimators. The performance criterion is the mean squared error $n^{-1} \sum_{i=1}^n \{\hat{f}(x_i) - f(x_i)\}^2$, where f and $\hat{f}$ represent the true function and estimating function, respectively. The average value of this criterion over 1000 simulations is computed and summarized in Table 1. In addition, we also compare the performance of the estimators at the boundary point $x = 0$. The average value of the difference between $\hat{f}(0)$ and $f(0)$ over 1000 simulations is summarized in Table 1 as well. The asymptotically optimal penalty parameter μ can be found by minimizing the asymptotic integrated mean square error, which is given by

$$\mu^2 \int_0^1 f''(x)^2 dx + \frac{\sigma^2}{4c} n^{-4/5}.$$

Table 1: Average mean square errors from three estimation methods

		BK	QUAD			PM		
			$K_n = 20$	$K_n = 60$	$K_n = 100$	$K_n = 20$	$K_n = 60$	$K_n = 100$
	Step	0.0028	0.0038	0.0031	0.0030	0.0043	0.0042	0.0042
MSE	Logistic	0.0028	0.0017	0.0023	0.0025	0.0024	0.0024	0.0024
	Cubic	0.0036	0.0023	0.0030	0.0032	0.0023	0.0023	0.0023
	Step	0.0258	0.0074	0.0207	0.0235	0.0086	0.0092	0.0097
MSE	Logistic	0.0268	0.0073	0.0198	0.0240	0.0058	0.0058	0.0057
at 0_+	Cubic	0.0255	0.0143	0.0229	0.0249	0.0092	0.0093	0.0098

Therefore, the asymptotic optimal μ is

$$\tilde{\mu} = \left[16n\sigma^{-2}\int_0^1 f''(x)^2 dx\right]^{-2/5}.$$

Meyer and Woodroofe [13] gave a consistent estimate of σ^2. We also use the kernel estimator of f to obtain an estimate of the second derivative of f in practice. In the following examples, the penalty parameter is chosen as 0.04.

Example 1. In this example, consider an increasing step function

$$f(x) = \begin{cases} 0, & 0 \le x < \frac{1}{3}, \\ \frac{1}{2}, & \frac{1}{3} \le x < \frac{2}{3}, \\ 1, & \frac{2}{3} \le x \le 1. \end{cases}$$

Example 2. This example focuses on the logistic function $f(x) = [1 + \exp(-20x + 10)]^{-1}$.

Example 3. The third example involves the cubic function $f(x) = 10^{-3}(20x - 10)^3$.

Since the true regression function in Example 1 is a step function, the Brunk's estimator outperforms the other two smooth estimators. In this case, the quadratic regression spline estimator shows a slightly better performance than the monotone penalized spline estimator. When the true regression function is smooth and monotone in Examples 2 and 3, the penalized monotone estimator and the quadratic regression spline estimator demonstrate better behaviors than the Brunk's estimator. It is shown that both the quadratic regression splines and the monotone penalized splines are robust to the number of knots. However, they behave quite differently at the boundary. As the number of knot increases, the boundary behavior of the quadratic spline estimator tends to that of the Brunk's estimator. In contrast, the monotone penalized spline estimator demonstrates consistent estimation at the boundary. This agrees with the asymptotic analysis performed before. Finally, to be fair to unpenalized splines, it should be pointed out that unpenalized splines do not use many knots in practice. On the other hand, more knots are expected in penalized splines since the penalty parameter reduces the effective degrees of freedom.

5.2 Discussions

We have so far focused on the equally spaced design points and knots. When the design is not equally spaced, one can use the ideas of [7, 26]. In specific, assume that x_i's are in (a, b). Find a smoothing monotone function H such that $H(x_i) = i/n$ from (a,b) to $(0,1)$. We use the P-spline smoothing to fit $(i/n, y_i)$, and thus the regression function is given by $f \circ H^{-1}$. We place knots at sample quantiles so that there are equal numbers of data points between consecutive knots. Further study of this issue is beyond the scope of this paper and will be reported in the future.

Our methods can be applied to an estimator defined in (3) with a higher-order difference penalty. It is conjectured that this will improve the convergence rate. Nevertheless its development becomes much more complicated and we intend to address it in the future. We have worked on the B-spline bases in this paper while [19] used truncated polynomials as basis functions for unconstrained estimation. As pointed out in [20], these two bases are algebraically identical in the unconstrained setting. For example, the penalty term in the latter case is $\lambda^* \sum_{k=1}^{K_n+p} a_k^2$, where a_k's are the coefficients.

6 Conclusions

This paper develops an asymptotic theory of monotone P-spline estimators with arbitrary spline degrees and the first-order difference penalty from a constrained dynamic optimization perspective. The presence of the monotone constraint complicates asymptotic analysis of the estimator. For example, the optimality conditions of spline coefficients are given by a size-dependent complementarity problem and are approximated by a dynamical complementarity system. Various tools from constrained optimization, ODE and statistical theory are exploited to establish consistency, asymptotic normality, and convergence rates of the estimator. These techniques can be extended to handle additional constraints. Hence, the results developed in this paper open a door to more complex nonparametric estimation problems subject to both dynamical and shape constraints.

Acknowledgments. The authors are grateful to the associate editor and the referees for their constructive comments that improve presentation and quality of the paper. The work of the first author is partially supported by the NSF grants ECCS-0900960, CMMI-1030804 and DMS-1042916; the research of the second author is supported by the NSF grants DMS-1042967 and CMMI-1030246.

References

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

[2] G. Claeskens, T. Krivobokova, and J. Opsomer. Asymptotic properties of penalized spline estimators. *Biometrika*, Vol. 96, pp. 529–544, 2009.

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

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

[5] P. Hall and J.D. Opsomer. Theory for penalised spline regression. *Biometrika*, Vol. 92, pp. 105–118, 2005.

[6] A.L. Juloski, S. Weiland, and W.P.M.H. Heemels. A Bayesian approach to identification of hybrid systems. *IEEE Trans. on Auto. Control*, Vol. 50(10), pp. 1520–1533, 2005.

[7] Y. Li and D. Ruppert. On the asymptotics of penalized splines. *Biometrika*, Vol. 95, pp. 415–436, 2008.

[8] E. Mammen. Estimating a smoothing regression function. *Annals of Statistics*, Vol. 19, pp. 724–740, 1991.

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

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

[11] K. Messer. A comparison of a spline estimate to its equivelent kernel estimate. *Annals of Statistics*, Vol. 19, pp. 817–829, 1991.

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

[13] M. Meyer and M. Woodroofe. On degrees of freedom in shape restricted regression. *Annals of Statistics*, Vol. 28, pp. 1083–104, 2000.

[14] D. Nychka. Splines as local smoothers. *Annals of Statistics*, Vol. 23, pp. 1175-1197, 1995.

[15] J. Pal and M. Woodroofe. On the distance between cumulative sum diagram and its greatest convex minorant for unequally spaced design points. *Scandinavian Journal of Statistics*, Vol. 33, pp. 279–291, 2006.

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

[17] T. Robertson, F.T. Wright, and R.L. Dykstra. *Order Restricted Statistical Inference*. John Wiley & Sons Ltd., 1988.

[18] D. Ruppert. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statisitcs*, Vol. 11, pp. 735–757, 2002.

[19] D. Ruppert and R. Carroll. Spatially-adaptive penalities for spline fitting. *Australian & New Zealand Journal of Statistics*, Vol. 42, pp. 205–224, 2000.

[20] D. Ruppert, M.P. Wand, and R.J. Carroll. *Semiparametric Regression*. Cambridge: Cambridge University Press, 2003.

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

[22] J. Shen and J.S. Pang. Linear complementarity systems: Zeno states. *SIAM Journal on Control and Optimization*, Vol.44, pp. 1040–1066, 2005.

[23] J. Shen and J.S. Pang. Linear complementarity systems with singleton properties: non-Zenoness. *Proceedings of 2007 American Control Conference*, pp. 2769–2774, New York, 2007.

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

[25] B.W. Silverman. Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, Vol. 12, pp. 898–916, 1984.

[26] W. Stute. Asymptotic normality of nearest neighbor regression function estimates. *Annals of Statistics*, Vo. 12, pp. 917–926, 1984.

[27] S. Sun, M.B. Egerstedt, and C.F. Martin. Control theoretic smoothing splines. *IEEE Trans. on Automatic Control*, Vol. 45(12), pp. 2271–2279, 2000.

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

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

[30] X. Wang, M. Woodroofe, M. Walker, M. Mateo, and E. Olszewski. Estimating dark matter distributions. *The Astrophysical Journal*, Vol. 626, pp. 145–158, 2005.

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

[32] I. Wright and E. Wegman. Isotonic, convex and related splines. *Annals of Statistics*, Vol. 8, pp. 1023–1035, 1980.