Convex Regression via Penalized Splines: A Complementarity Approach

Jinglai Shen and Xiao Wang

***Abstract*— Estimation of a convex function is a critical shape restricted nonparametric inference problem with a wide range of applications in many important fields. In this paper, penalized splines (or simply P-splines) are exploited for convex estimation. The paper is devoted to developing an asymptotic theory of a class of P-spline convex estimators using complementarity techniques and asymptotic statistics. In particular, due to the convex constraints, the optimality conditions of P-splines are characterized by nonsmooth complementarity conditions. A critical uniform Lipschitz property is established for optimal spline coefficients via complementarity techniques. This property yields boundary consistency and uniform stochastic boundedness. Using this property, the P-spline estimator is approximated by a two-step estimator based on the corresponding least squares estimator, and its asymptotic behaviors are obtained using asymptotic statistic techniques.**

I. Introduction

Nonparametric estimation of shape restricted functions receives increasing attention in statistics [3], [8], [10], [13], [14], [16], driven by numerous applications in science and engineering. Examples include reliability engineering, biomedical research, finance, and astronomy. A challenge in shape restricted estimation is that an estimator is subject to *inequality* constraints, e.g., monotone and convex constraints. These constraints lead to nonsmooth optimality conditions that complicate performance analysis of estimators.

The polynomial spline models have been extensively studied in approximation theory and statistics, thanks to their computational advantages. The non-penalized polynomial splines are used to develop the shape restricted least squares estimators [5] for monotone and convex functions. However, the least squares estimators suffer several deficiencies. For example, since the least squares estimators are necessarily piecewise constant (resp. linear) functions for the monotone (resp. convex) constraint, they lack of smoothness. Further, the least squares estimators have unsatisfactory performance: they are inconsistent at boundary and have a non-negligible asymptotic bias with low convergence rates and non-normal asymptotic distributions.

In this paper, we consider the penalized polynomial splines (or P-splines for short) for convex estimation and analyze their asymptotic performance, i.e. the estimation performance as the sample size is sufficiently large. The penalty on the difference of splines improves estimation performance, e.g., smoothness and boundary consistency. However, due to the convex constraints and size dependent difference penalty, performance analysis of the P-spline convex estimators is more complicated. In particular, the optimality conditions of the P-spline estimators give rise to a family of size-varying, penalty parameter dependent complementarity conditions. The closed form solution of these complementarity conditions do not exist. To deal with these difficulties, we establish a critical uniform Lipschitz property [11], [12] of the optimal spline coefficients and use this property to approximate the estimator by by a two-step estimator based on the corresponding least squares estimator [5]. By exploiting asymptotic statistic tools, we further approximate this dynamical complementarity system and develop asymptotic behaviors of the P-spline estimators.

The paper is organized as follows. In Section II, we formulate the convex regression problem and derive optimality conditions for the P-spline estimator. Section III establishes a uniform Lipschitz property for a class of P-splines with the first order difference penalty. Asymptotic analysis is performed in Section IV with an example shown in Section V and the conclusion drawn in Section VI.

II. Problem Formulation and Optimality Conditions

Consider the problem of estimating a convex function $f : [0,1] \to \mathbb{R}$ from a univariate regression model $y_i = f(t_i) + \epsilon_i$, $i = 1, \dots, n$, where the pre-specified design points are $t_i = i/n$, $i = 1, \dots, n$, and the ϵ_i are independent random variables with mean zero and variance σ^2. Our goal is to estimate the function f which is assumed to be convex.

We propose a class of convex penalized spline estimators based on binned data and investigate their asymptotic properties. In particular, let $\left\{ B_k^{[p]} : k = 1, \dots, K_n + p \right\}$ be the pth degree B-spline basis with knots $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{K_n} = 1$. For simplicity, we consider equally spaced knots, namely, $\kappa_1 = 1/K_n, \kappa_2 = 2/K_n, \dots, \kappa_{K_n} = 1$. The value of K_n will depend upon n as discussed below. Assume that n/K_n is an integer denoted by M_n. Let $\bar{y}_k$ be the average of all y_i such that $\kappa_{k-1} < t_i \le \kappa_k$, i.e.,

$$\bar{y}_k = \frac{\sum_{i=1}^n y_i \, \mathcal{I}(\kappa_{k-1} < t_i \le \kappa_k)}{\sum_{i=1}^n \mathcal{I}(\kappa_{k-1} < t_i \le \kappa_k)} = \frac{\sum_{i=(\kappa_k-1)M_n+1}^{\kappa_k M_n} y_i}{M_n},$$

where $k = 1, \dots, K_n$, and $\mathcal{I}$ is the indicator function. Denote $\bar{y} = (\bar{y}_1, \dots, \bar{y}_{K_n})^T$. Let the polyhedral cone be

$$\Omega = \{ b \in \mathbb{R}^{K_n} : b_k - 2b_{k+1} + b_{k+2} \ge 0, \ k = 1, \dots, K_n - 2 \}.$$

We consider the following constrained optimization problem:

Jinglai Shen is with the Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA. Email: shenj@umbc.edu.

Xiao Wang is with the Department of Statistics, Purdue University, West Lafayette, IN 47906, USA. Email: wangxiao@purdue.edu.

for $m \in \mathbb{N}$,

$$\hat{b}^{[m]} \equiv \hat{b}^{[m]}(\bar{y}) = \arg\min_{b\in\Omega} \sum_{k=1}^{K_n} (\bar{y}_k - b_k)^2 + \lambda^* \sum_{k=m+1}^{K_n} \big(\Delta^m b_k\big)^2, \tag{1}$$

where $\lambda^* > 0$ and Δ is the backward difference operator, i.e., $\Delta(b_k) := b_k - b_{k-1}$ and $\Delta^m(b_k) = \Delta(\Delta^{m-1}(b_k))$. Define the following convex spline estimator: for $p \geq 1$,

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

where $\hat{b}_{K_n+d}^{[m]} = 2\hat{b}_{K_n+d-1}^{[m]} - \hat{b}_{K_n+d-2}^{[m]}$, $d = 1, \ldots, p$. When the knots are equally spaced, it is easy to verify that if the B-spline coefficient vector $\hat{b}^{[m]}$ is in Ω, then $\hat{f}_p^{[m]}$ is convex.

Let

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

and let $D_m \in \mathbb{R}^{(K_n - m)\times K_n}$ be the mth-order difference matrix such that $D_m b = [\Delta^m(b_{m+1}), \cdots, \Delta^m(b_{K_n})]^T$. Formulating (1) via matrix notation, we obtain the following equivalent constrained quadratic program

$$\hat{b}^{[m]} = \arg\min_{b\in\Omega} \frac{1}{2}\, b^T (I_{K_n} + \lambda D_m^T D_m)\, b - b^T \bar{y}, \tag{2}$$

where $\lambda = \lambda^*/M_n = \lambda^* \cdot K_n/n > 0$ and $I_{K_n} \in \mathbb{R}^{K_n \times K_n}$ is the identity matrix.

We first give the characterization of optimality conditions for $\hat{b}^{[m]}$. The conditions are represented by complementarity conditions, which plays a crucial role in addressing analytic and statistical properties of the estimator. We provide a short introduction of the complementarity condition. Two vectors $u = (u_1, \cdots, u_d)^T$ and $v = (v_1, \cdots, v_d)^T$ in $\mathbb{R}^d$ are said to satisfy the *complementarity condition* [1], [4] if $u_i \geq 0$, $v_i \geq 0$, and $u_i\, v_i = 0$ for all $i = 1, \cdots, d$. This condition can be put in a more compact vector form: $0 \leq u \perp v \geq 0$, where $u \perp v$ means that the two vectors are orthogonal, i.e., $u^T v = 0$. We introduce more notation as follows. Let $\mathbf{1}$ denote the vector of ones, i.e., $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^\ell$. We define the sum operator for $\mathbf{1}$, namely, ${}^0\Delta(\mathbf{1}) = \mathbf{1}$, and ${}^{-m}\Delta(\mathbf{1}) = (1, \sum_{i=1}^2 ({}^{1-m}\Delta(\mathbf{1}))_i, \ldots, \sum_{i=1}^\ell ({}^{1-m}\Delta(\mathbf{1}))_i)^T$ for $m \in \mathbb{N}$. In particular, ${}^{-1}\Delta(\mathbf{1}) = (1, 2, \ldots, \ell)^T$.

Theorem 2.1: The necessary and sufficient conditions for $\hat{b}^{[m]} \in \Omega$ to minimize (2) are

$$0 \leq D_2 \hat{b}^{[m]} \perp C_{\gamma\bullet}\, C \left[(I_{K_n} + \lambda D_m^T D_m)\hat{b}^{[m]} - \bar{y} \right] \geq 0, \tag{3}$$

and

$$\begin{aligned} C_{K_n\bullet} \left[(I_{K_n} + \lambda D_m^T D_m)\hat{b}^{[m]} - \bar{y} \right] &= 0, \\ C_{K_n\bullet}\, C \left[(I_{K_n} + \lambda D_m^T D_m)\hat{b}^{[m]} - \bar{y} \right] &= 0, \end{aligned} \tag{4}$$

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

Proof: For notational simplicity, we drop the subscript $[m]$ in $\hat{b}^{[m]}$ as follows. Write the optimization problem (2) as $\min_{b\in\Omega} g(b)$, where the objective function $g(b) := \frac{1}{2} b^T \big(I_{K_n} + \lambda D_m^T D_m\big) b - b^T \bar{y}$. It is clear that g is coercive on $\mathbb{R}^{K_n}$ and strictly convex on the closed convex set Ω. This ensures the existence and uniqueness of an optimal solution. Furthermore, since Ω is a convex polyhedral cone, it is finitely generated by $\{v^1, -v^1, v^2, -v^2, v^3, v^4, \ldots, v^{K_n}\}$. Here, for each $k = 3, \ldots, K_n$, letting $v_j^k = ({}^{-1}\Delta(\mathbf{1}))_{j-k+1}$ with $j = k, \ldots, K_n$,

$$\begin{aligned} v^k &= \Big(\underbrace{0, \ldots, 0,}_{(k-1)-\text{copies}} \; v_k^k, \ldots, v_{K_n}^k \Big)^T \\ &= \Big(\underbrace{0, \ldots, 0,}_{(k-1)-\text{copies}} \; 1, 2, \ldots, K_n - k + 1 \Big)^T, \end{aligned}$$

and for $k = 1, 2$,

$$\begin{aligned} v^1 &= \Big(1,\, 0,\, -1,\, -2, \ldots, -(K_n - 2) \Big)^T, \\ v^2 &= \Big(0,\, 1,\, 2,\, 3, \ldots, K_n - 1 \Big)^T. \end{aligned} \tag{5}$$

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

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

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

$$0 \leq D_2 \hat{b} \perp \widetilde{C} \nabla g(\hat{b}) \geq 0, \quad \langle v^k, \nabla g(\hat{b}) \rangle = 0, \; \forall\, k = 1, 2, \tag{6}$$

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

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

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

$$\begin{aligned} \widetilde{C} &= \begin{bmatrix} v^3 & \cdots & v^{K_n} \end{bmatrix}^T \\ &= \begin{bmatrix} 0 & 0 & 1 & 2 & \cdots & (K_n-4) & (K_n-3) & (K_n-2) \\ 0 & 0 & 0 & 1 & \cdots & \cdots & (K_n-4) & (K_n-3) \\ & \cdots & & & \cdots & & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

It can be shown via the definitions of v^1 and v^2 in (5) that the second optimality condition in (6) can be equivalently written as

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

where $\nabla g(b) = (\Gamma_n + \lambda D_m^T D_m)b - \bar{y}$. This gives rise to the two boundary conditions. Moreover, noting that for any k, the definitions of v^1 and v^2 in (5) yield

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

we obtain the equivalent condition for the first optimality condition in (6):

$$0 \le D_2\,\hat{b} \perp \left(C^2\right)_{\gamma\bullet}\nabla g(\hat{b}) \ge 0, \tag{7}$$

where $\gamma = \{1,\dots,K_n-2\}$. Finally, in view of $(C^2)_{\gamma\bullet} = C_{\gamma\bullet}C$, the proof is complete. ■

III. Uniform Lipschitz Property of $\hat{b}$

In this section, we characterize a critical property of the optimal solution $b^{[m]}$ with $m = 1$. For notational convenience, we drop the superscript in $\hat{b}^{[1]}$ through this section. We firstly establish a piecewise linear formulation of $\hat{b}$. Let $\Lambda := (I_{K_n} + \lambda D_1^T D_1)/(1+2\lambda)$ and $z := \bar{y}/(1+2\lambda)$. In particular, Λ is the following tri-diagonal matrix

$$\begin{bmatrix} \theta & \eta & 0 & 0 & \cdots & 0 \\ \eta & 1 & \eta & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \eta & 1 & \eta & 0 \\ & & & \eta & 1 & \eta \\ 0 & 0 & \cdots & 0 & \eta & \theta \end{bmatrix}, \tag{8}$$

where $\theta := (1+\lambda)/(1+2\lambda)$, and $\eta := -\lambda/(1+2\lambda)$ with $\lambda > 0$. With this notation, the optimality conditions become the following mixed complementarity conditions

$$\begin{aligned} 0 \le D_2\,\overline{b} \perp C_{\gamma\bullet}C(\Lambda\overline{b} - z) \ge 0, \quad &\text{and} \\ C_{K_n\bullet}[\Lambda\overline{b} - z] = C_{K_n\bullet}C[\Lambda\overline{b} - z] = 0. & \end{aligned} \tag{9}$$

It follows from complementarity theory that the optimal solution $\overline{b}$, and thus $\hat{b}$, is a piecewise linear function of z determined by an index set $\alpha = \{\, i \,|\, (D_2\overline{b})_i = 0\} \subseteq \{1,\dots,K_n-2\}$ (α may be empty). Specifically, for given $\overline{b}$ and α, we define a vector $\widetilde{b}^{\alpha}$ and an associated family of index sets $\{\beta_i^{\alpha}\}$ in the following steps:

(1) let $\ell_1 := \min_{3\le i\le K_n}\{i : \Delta^2(\overline{b}_i) = 0\}$, and $\overline{\ell}_1 := \max_{\ell_1\le k\le K_n}\{k : \Delta^2(\overline{b}_i) = 0,\ \forall i = \ell_1,\dots,k\}$. Then inductively define, for $j \ge 1$,

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

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

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

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

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

(4) in this step, we regroup the index sets $\widehat{\beta}_{\ell_j}^{\alpha}$ in a way that preserves desired structural properties to be used in the subsequent development. Define $p_0 := 0$ and $p_1 := \max\big(1, \quad \max\{k \ge 1 : \widetilde{\beta}_j^{\alpha} \cap \widetilde{\beta}_{j+1}^{\alpha} \neq \emptyset,\ \forall j = 1,\dots,k-1\}\big)$, and $\beta_1^{\alpha} := \cup_{j=1}^{p_1}\widetilde{\beta}_j^{\alpha}$, the companion index set $\vartheta_1 := \{\min(\widetilde{\beta}_j^{\alpha}), \forall j = 1,\dots,p_1\} \cup \{\max(\widetilde{\beta}_{p_1}^{\alpha})\}$. Recursively, define, for each $s \ge 1$, $p_{s+1} := \max\big(p_s+1, \quad \max\{k \ge p_s+1 : \widetilde{\beta}_j^{\alpha}\cap\widetilde{\beta}_{j+1}^{\alpha} \neq \emptyset,\ \forall j = p_s+1,\dots,k-1\}\big)$, and $\beta_{s+1}^{\alpha} := \cup_{j=p_s+1}^{p_{s+1}}\widetilde{\beta}_j^{\alpha}$, the companion index set $\vartheta_{s+1} := \{\min(\widetilde{\beta}_j^{\alpha}), \forall j = p_s+1,\dots,p_{s+1}\} \cup \{\max(\widetilde{\beta}_{p_{s+1}}^{\alpha})\}$. Without loss of generality, we assume that the index elements of each ϑ_s are in the strictly increasing order. Hence, any two consecutive index sets in ϑ_s correspond to ℓ_j and $\overline{\ell}_j$ defined in Step (1) with $\ell_{j+1} = \overline{\ell}_j$.

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

It is clear from the above construction that $\{\beta_i^{\alpha}\}$ forms a finite and disjoint partition of $\{1,\dots,K_n\}$, namely, $\bigcup_{i=1}^{L}\beta_i^{\alpha} = \{1,\dots,K_n\}$ and $\beta_j^{\alpha}\cap\beta_k^{\alpha} = \emptyset$ whenever $j \neq k$. Algebraically, the vector $\widetilde{\beta}^{\alpha}$ corresponds to the free variables of a linear equation subject to the constraints defined by α. Moreover, it can be shown that $\widetilde{b}^{\alpha}$, and thus $\overline{b}^{\alpha}(z)$ which denotes $\overline{b}(z)$ corresponding to the index set α, is a linear function of z (cf. Lemma 3.1). Hence, for any $z \in \mathbb{R}^{K_n}$, $\overline{b}(z) \in \{\overline{b}^{\alpha}(z)\}_{\alpha}$, where $\overline{b}^{\alpha}(z)$ is a selection function of $\overline{b}(z)$. Therefore, the solution mapping $z \mapsto \overline{b}$ is a (continuous) piecewise linear function with $2^{(K_n-2)}$ selection functions. The same holds true for the mapping $\bar{y} \mapsto \hat{b}$. In what follows, we characterize each linear selection function of $\widetilde{b}^{\alpha}$ or equivalently $\overline{b}^{\alpha}$.

Lemma 3.1: For each index set $\alpha \subseteq \{1,\dots,K_n-2\}$, let $\ell := K_n - |\alpha|$. Then $\widetilde{b}^{\alpha}$ is the (unique) solution of the linear equation $\widetilde{\Lambda}^{\alpha}\widetilde{b}^{\alpha} = \widetilde{z}^{\alpha}$, where the $\ell\times\ell$ tri-diagonal matrix $\widetilde{\Lambda}^{\alpha}$ and the ℓ-vector $\widetilde{z}^{\alpha}$ are given by

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

and $\widetilde{z}^\alpha = \begin{bmatrix} F_{\alpha,1} z_{\beta_1^\alpha} \\ \vdots \\ F_{\alpha,L} z_{\beta_L^\alpha} \end{bmatrix}^T$, where $F_{\alpha,i}$, d_{ii} and $\widetilde{\eta}_i$ are given in the following proof. Moreover, $\widetilde{\Lambda}^\alpha$ is invertible.

Proof: We introduce some notation first. Let $m_i^\alpha := |\beta_i^\alpha|$ and $h_i^\alpha := m_i^\alpha - 1$, where $i = 1, \ldots, L$. Note that if $m_i^\alpha > 1$, then $m_i^\alpha \geq 3$ such that $h_i^\alpha \geq 2$ and $|\vartheta_i| \geq 2$. It follows from the definition of β_i^α that $\overline{b}^\alpha = (F_\alpha)^T \widetilde{b}^\alpha$, where the matrix

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

and each matrix block corresponding to β_k^α is given as follows: if $m_k^\alpha = 1$, then $F_{\alpha,k} = 1$; otherwise, assuming that the index elements in ϑ_k are in the strictly increasing order without loss of generality, and letting $h_{k,j}^\alpha := \vartheta_k(j+1) - \vartheta_k(j) \geq 2$ for each $j = 1, \ldots, |\vartheta_k| - 1$, we have $F_{\alpha,k} \in \mathbb{R}^{|\vartheta_k| \times m_k^\alpha}$ given in (16) at the end of the paper, where $w_k := |\vartheta_k| - 1$. Here $F_{\alpha,k}$ is determined from β_k^α constructed in Steps (1)-(5).

For notational simplicity, let $v := \Lambda \overline{b} - z$. In view of the complementarity condition in (9), we have $(D_2 \overline{b})^T C_{\gamma\bullet} C v = 0$. Since $\overline{b} = (F_\alpha)^T \widetilde{b}^\alpha$, $(\widetilde{b}^\alpha)^T F_\alpha (D_2^T C_{\gamma\bullet} C v) = 0$. Moreover, it can be verified that

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

where

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

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

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

Letting $\widetilde{\Lambda}^\alpha = F_\alpha \, \Lambda (F_\alpha)^T$ and $\widetilde{z}^\alpha = F_\alpha \, z$, we have a linear equation for $\widetilde{b}^\alpha$. Since F_α is of full row rank and Λ is positive definite, $\widetilde{\Lambda}^\alpha$ is positive definite and hence is invertible.

In what follows, we determine the entries of $\widetilde{\Lambda}^\alpha$. Fix $k \in \{1, \ldots, L\}$. If $m_k^\alpha = 1$, then $F_{\alpha,k} \Lambda_{\beta_k^\alpha \beta_k^\alpha} F_{\alpha,k}^T$ is a real number that appears on the diagonal of $\widetilde{\Lambda}^\alpha$. Denoting this number by d_{ss} (i.e., $d_{ss} = \widetilde{\Lambda}^\alpha_{ss}$), we have

$$d_{ss} = F_{\alpha,k} \Lambda_{\beta_k^\alpha \beta_k^\alpha} F_{\alpha,k}^T = \begin{cases} \theta, & \text{if } k \in \{1, L\} \\ 1, & \text{otherwise} \end{cases}$$

and $\widetilde{\Lambda}^\alpha_{s(s+1)} = \widetilde{\Lambda}^\alpha_{(s+1)s} = \eta$, $\widetilde{\Lambda}^\alpha_{sj} = 0$ for all $j \leq s-2$ and $j \geq s+2$. If $m_k^\alpha > 1$, then $F_{\alpha,k} \Lambda_{\beta_k^\alpha \beta_k^\alpha} F_{\alpha,k}^T$ is a symmetric, positive definite matrix of order $|\vartheta_k|$ that forms a diagonal block of $\widetilde{\Lambda}^\alpha$. Making use of the structure of $F_{\alpha,k}$ given in (16) and somewhat lengthy computation, we obtain the following results in two separate cases (recalling $w_k := |\vartheta_k| - 1$).

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

$$d_{11} = \theta + \eta - \frac{\eta}{h_{1,1}^\alpha} + (1+2\eta) \frac{(h_{1,1}^\alpha - 1)(2h_{1,1}^\alpha - 1)}{6h_{1,1}^\alpha}, \tag{10}$$

$$\widetilde{\eta}_s = \widetilde{\Lambda}^\alpha_{s(s+1)} = \widetilde{\Lambda}^\alpha_{(s+1)s} = \frac{\eta}{h_{1,s}^\alpha} + (1+2\eta) \frac{(h_{1,s}^\alpha)^2 - 1}{6h_{1,s}^\alpha}, \quad \forall\ s = 1, \ldots, w_1,$$

$$d_{ss} = (1+2\eta) \Big[\frac{2(h_{1,s-1}^\alpha)^2 + 1}{6h_{1,s-1}^\alpha} + \frac{2(h_{1,s}^\alpha)^2 + 1}{6h_{1,s}^\alpha} \Big] - \Big(\frac{1}{h_{1,s-1}^\alpha} + \frac{1}{h_{1,s}^\alpha} \Big) \eta, \qquad \forall\ s = 2, \ldots, w_1,$$

$$d_{(w_1+1)(w_1+1)} = (1+2\eta) \frac{(h_{1,w_1}^\alpha + 1)(2h_{1,w_1}^\alpha + 1)}{6h_{1,w_1}^\alpha} - \Big(1 + \frac{1}{h_{1,w_1}^\alpha}\Big)\eta. \tag{11}$$

Besides, $\widetilde{\Lambda}^\alpha_{(w_1+1)(w_1+2)} = \widetilde{\Lambda}^\alpha_{(w_1+2)(w_1+1)} = \eta$ and for each $s = 1, \ldots, w_t$, $\widetilde{\Lambda}^\alpha_{sj} = 0, \forall j \geq s+2$ and $j \leq s-2$. For $k = L$, the similar results can be established by using the symmetry of the rows of $F_{\alpha,L}$.

(2) $k \in \{2, \ldots, L-1\}$. In this case, suppose that the $(1,1)$-element of $F_{\alpha,k} \Lambda_{\beta_k^\alpha \beta_k^\alpha} F_{\alpha,k}^T$ is a diagonal entry of $\widetilde{\Lambda}^\alpha$ denoted by d_{tt}. Then we have

$$d_{tt} = 1 + \eta - \frac{\eta}{h_{k,1}^\alpha} + (1+2\eta) \frac{(h_{k,1}^\alpha - 1)(2h_{k,1}^\alpha - 1)}{6h_{k,1}^\alpha}, \tag{12}$$

$$\widetilde{\eta}_{t+s} = \widetilde{\Lambda}^\alpha_{(t+s)(t+s+1)} = \frac{\eta}{h_{k,s}^\alpha} + (1+2\eta) \frac{(h_{k,s}^\alpha)^2 - 1}{6h_{k,s}^\alpha}, \quad \forall\ s = 1, \ldots, w_k,$$

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

$$d_{(t+w_k)(t+w_k)} = (1+2\eta) \frac{(h_{k,w_k}^\alpha + 1)(2h_{k,w_k}^\alpha + 1)}{6h_{k,w_k}^\alpha} - \Big(1 + \frac{1}{h_{k,w_k}^\alpha}\Big)\eta. \tag{13}$$

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

Due to the identity $\widetilde{\Lambda}^\alpha_{t(t-1)} = \eta$ and the symmetry of $\widetilde{\Lambda}^\alpha$, we further conclude that if a diagonal entry $d_{tt} = \widetilde{\Lambda}^\alpha_{tt}$ with $t \geq 2$ corresponds to a scalar $F_{\alpha,k} \Lambda_{\beta_k^\alpha \beta_k^\alpha} F_{\alpha,k}^T$ (i.e., $m_k^\alpha = 1$), then $\widetilde{\Lambda}^\alpha_{(t-1)t} = \eta$. (Recall that $\widetilde{\Lambda}^\alpha_{t(t+1)} = \eta$ has been obtained before.) Similarly, if d_{tt} is the first diagonal entry of a matrix $F_{\alpha,k} \Lambda_{\beta_k^\alpha \beta_k^\alpha} F_{\alpha,k}^T$, then $\widetilde{\Lambda}^\alpha_{(t-1)t} = \eta$. ∎

In the following, we prove the uniform Lipschitz property of the optimal solution $\hat{b}^{[1]}$. This property implies that $\hat{b}^{[1]}$

is Lipschitz in $\bar{y}$ (in the sense of ℓ_∞-norm) with the same Lipschitz constant, regardless of K_n, λ. It plays a critical role in proving boundary consistency and asymptotic analysis.

Theorem 3.1: Let $m = 1$. For any K_n and any $\lambda > 0$, $\|\hat{b}^{[1]}(\bar{y}^1) - \hat{b}^{[1]}(\bar{y}^2)\|_\infty \le 3\|\bar{y}^1 - \bar{y}^2\|_\infty$ for all $\bar{y}^1, \bar{y}^2 \in \mathbb{R}^{K_n}$.

Proof: Recall that for a given index set α, $\overline{b}^\alpha(z) = F_\alpha^T \widetilde{b}^\alpha(z) = F_\alpha^T(\widetilde{\Lambda}^\alpha)^{-1} F_\alpha z$, where $z = \bar{y}/(1+2\lambda)$. We shall show that $\|F_\alpha^T(\widetilde{\Lambda}^\alpha)^{-1}F_\alpha\|_\infty$ is uniformly bounded, regardless of α, λ and K_n. We break the proof into the following steps.

(1) We first show that for any α, the matrix $\widetilde{\Lambda}^\alpha$ is strictly diagonally dominant and obtain bounds characterizing such dominance. Given $\widetilde{\Lambda}^\alpha \in \mathbb{R}^{\ell\times\ell}$, define $\xi_1 := d_{11} - |\widetilde{\eta}_1|$, $\xi_i := d_{ii} - |\widetilde{\eta}_{i-1}| - |\widetilde{\eta}_i|$ with $i \in \{2,\dots,\ell-1\}$, and $\xi_\ell := d_\ell - |\widetilde{\eta}_{\ell-1}|$. In light of the structure of $\widetilde{\Lambda}^\alpha$ shown in the proof of Lemma 3.1, we obtain, for each $k \in \{1,\dots,L\}$,

(1.1) if $m_k^\alpha = 1$, then (i) the corresponding $\xi_i = \theta - |\eta| = 1/(1+2\lambda)$ if $k \in \{1, L\}$; and (ii) otherwise, the corresponding $\xi_i = 1 - 2|\eta| = 1/(1+2\lambda)$.

(1.2) if $m_k^\alpha > 1$ with $k = 1$, then (i) the corresponding $\xi_i = d_{11} - |\eta| \ge \left(\frac{1}{2} + \frac{h_{1,1}^\alpha}{6}\right)/(1+2\lambda)$; (ii) for $s = 2,\dots,w_1$, the corresponding $\xi_i = d_{ss} - |\widetilde{\Lambda}^\alpha_{s(s-1)}| - |\widetilde{\Lambda}^\alpha_{s(s+1)}| \ge \left(h_{1,s-1}^\alpha + h_{1,s}^\alpha\right)/[6(1+2\lambda)]$; and (iii) the corresponding $\xi_i = d_{(w_1+1)(w_1+1)} - |\widetilde{\Lambda}^\alpha_{(w_1+1)w_1}| - |\widetilde{\Lambda}^\alpha_{(w_1+1)(w_1+2)}| \ge \left(\frac{1}{2} + \frac{h_{1,w_1}^\alpha}{6}\right)/(1+2\lambda)$. The similar results can be obtained for $m_k^\alpha > 1$ with $k = L$ using symmetry.

(1.3) if $m_k^\alpha > 1$ with $k \in \{2,\dots,L-1\}$, then (i) the corresponding $\xi_i = d_{tt} - |\widetilde{\Lambda}^\alpha_{t(t-1)}| - |\widetilde{\Lambda}^\alpha_{t(t+1)}| \ge \left(\frac{1}{2} + \frac{h_{k,1}^\alpha}{6}\right)/(1+2\lambda)$; (ii) for $s = 1,\dots,w_k-1$, the corresponding $\xi_i = d_{(t+s)(t+s)} - |\widetilde{\Lambda}^\alpha_{(t+s)(t+s-1)}| - |\widetilde{\Lambda}^\alpha_{(t+s)(t+s+1)}| \ge \left(h_{k,s}^\alpha + h_{k,s+1}^\alpha\right)/[6(1+2\lambda)]$; and (iii) the corresponding $\xi_i = d_{(t+w_k)(t+w_k)} - |\widetilde{\Lambda}^\alpha_{(t+w_k)(t+w_k-1)}| - |\widetilde{\Lambda}^\alpha_{(t+w_k)(t+w_k+1)}| \ge \left(\frac{1}{2} + \frac{h_{k,w_k}^\alpha}{6}\right)/(1+2\lambda)$.

Consequently, $\xi_i > 0$ for all ξ_i such that $\widetilde{\Lambda}^\alpha$ is strictly diagonally dominant.

(2) For a given $\widetilde{\Lambda}^\alpha$ and ξ_i's obtained in the last step, define the diagonal matrix $\Xi = \mathrm{diag}(\xi_1^{-1},\dots,\xi_\ell^{-1}) \in \mathbb{R}^{\ell\times\ell}$. Clearly, Ξ is invertible. We thus have

$$\begin{aligned}\|F_\alpha^T(\widetilde{\Lambda}^\alpha)^{-1}F_\alpha\|_\infty &= \|F_\alpha^T \cdot (\Xi\widetilde{\Lambda}^\alpha)^{-1} \cdot (\Xi F_\alpha)\|_\infty \\ &\le \|F_\alpha^T\|_\infty \cdot \|(\Xi\widetilde{\Lambda}^\alpha)^{-1}\|_\infty \cdot \|\Xi F_\alpha\|_\infty,\end{aligned}$$

where it is easy to verify $\|F_\alpha^T\|_\infty = 1$. Noting that $G := \Xi\widetilde{\Lambda}^\alpha$ is strictly diagonally dominant with $G_{ii} - \sum_{j=1, j\ne i}^{\ell} |G_{ij}| = 1$ for each i, it follows from the Ahlberg-Nilson-Varah bound [15] that $\|(\Xi\widetilde{\Lambda}^\alpha)^{-1}\|_\infty = \|G^{-1}\|_\infty \le 1$. Furthermore, we have:

(2.1) if $m_k^\alpha = 1$, then the absolute sum of the entries in the corresponding row in ΞF_α is given by $1/\xi_i \le (1+2\lambda)$.

(2.2) if $m_k^\alpha > 1$ with $k = 1$, then (i) the absolute sum of the entries in the row in ΞF_α corresponding to d_{11} is given by $\frac{1+h_{1,1}^\alpha}{2\xi_i} \le \frac{1+h_{1,1}^\alpha}{2\left(\frac{1}{2}+\frac{h_{1,1}^\alpha}{6}\right)/(1+2\lambda)} \le 3(1+2\lambda)$; (ii) for $s = 2,\dots,w_1$, the absolute sum of the entries in the row in ΞF_α corresponding to d_{ss} is given by $\frac{h_{1,s-1}^\alpha + h_{1,s}^\alpha}{2\xi_i} \le \frac{h_{1,s-1}^\alpha + h_{1,s}^\alpha}{2\left(h_{1,s-1}^\alpha + h_{1,s}^\alpha\right)/[6(1+2\lambda)]} \le 3(1+2\lambda)$; and (iii) the absolute sum of the entries in the row in ΞF_α corresponding to $d_{(w_1+1)(w_1+1)}$ is given by $\frac{1+h_{1,w_1}^\alpha}{2\xi_i} \le \frac{1+h_{1,w_k}^\alpha}{2\left(\frac{1}{2}+\frac{h_{1,w_k}^\alpha}{6}\right)/(1+2\lambda)} \le 3(1+2\lambda)$. The same results can be obtained for $m_k^\alpha > 1$ with $k = L$.

(2.3) if $m_k^\alpha > 1$ with $k \in \{2,\dots,L-1\}$, then (i) the absolute sum of the entries in the row in ΞF_α corresponding to d_{tt} is given by $\frac{1+h_{k,1}^\alpha}{2\xi_i} \le \frac{1+h_{k,1}^\alpha}{2\left(\frac{1}{2}+\frac{h_{k,1}^\alpha}{6}\right)/(1+2\lambda)} \le 3(1+2\lambda)$; (ii) for $s = 1,\dots,w_k-1$, the absolute sum of the entries in the row in ΞF_α corresponding to $d_{(t+s)(t+s)}$ is given by $\frac{h_{k,s}^\alpha + h_{k,s+1}^\alpha}{2\xi_i} \le 3(1+2\lambda)$; and (iii) the absolute sum of the entries in the row in ΞF_α corresponding to $d_{(t+w_k)(t+w_k)}$ is given by $\frac{1+h_{k,w_k}^\alpha}{2\xi_i} \le 3(1+2\lambda)$.

In view of the above results, we deduce that $\|\Xi F_\alpha\|_\infty \le 3(1+2\lambda)$, which in turn implies that $\|F_\alpha^T(\widetilde{\Lambda}^\alpha)^{-1}F_\alpha\|_\infty \le 3(1+2\lambda)$, regardless of α, λ, and K_n. Since $z = \bar{y}/(1+2\lambda)$, we have $\|\hat{b}(\bar{y})\|_\infty \le 3\|\bar{y}\|_\infty$ for any $\bar{y} \in \mathbb{R}^{K_n}$. The uniform Lipschitz property thus follows from the piecewise linear property of $\hat{b}$. ∎

A. Consequences of the Uniform Lipschitz Property

The uniform Lipschitz property yields several crucial implications of the P-spline estimator and lays a rigorous foundation for asymptotic analysis. We show stochastic uniform boundedness and boundary consistency here, following the similar line in [11]. Let $\mathbb{E}$ denote the expectation operator and $\check{b}^{[1]} := \hat{b}^{[1]}(\mathbb{E}(\bar{y}))$. Define the companion estimator $\bar{f}_p^{[1]}(t) = \sum_{k=1}^{K_n+p} \check{b}^{[1]} B_k^{[p]}(t)$. Let the norm $\|g\| := \sup_{t\in[0,1]} |g(t)|$ for a function $g \in C([0,1])$. Consequently,

$$\begin{aligned}\|\hat{f}_p^{[1]} - f\| &\le \|\hat{f}_p^{[1]} - \bar{f}_p^{[1]}\| + \|\bar{f}_p^{[1]} - f\| \\ &\le 3\|\bar{y} - \mathbb{E}(\bar{y})\|_\infty + O(\alpha) + O(K_n^{-1}) \\ &\le O_p(\sqrt{n^{-1}K_n \log K_n}) + O(\alpha) + O(K_n^{-1}),\end{aligned}$$

where $\alpha := \lambda^*/(nK_n)$. Hence, under suitable order conditions on n and K_n, we obtain stochastic uniform boundedness and boundary consistency in particular.

IV. Asymptotic Analysis of $\hat{f}^{[1]}$

In this section, we study the asymptotic distribution of $\hat{f}^{[m]}$ with $m = 1$. We first define the *invelope function* of an integrated Brownian motion. The invelope function, denoted by H, is studied in depth in [5] and its definition is as follows. Let $X(t) = W(t) + 4t^3$, where W is a standard two-sided Brownian motion starting from 0, and let $Y = \int_0^t X(s)ds$. The invelope function H satisfies the following conditions: (i) the function H is everywhere above the function Y; (ii) the function H has a convex second derivative, and with probability 1, H is three times differentiable at $t = 0$; (iii) the function H satisfies $\int[H(t) - Y(t)]dH^{(3)}(t) = 0$.

Theorem 4.1: For any fixed $t \in [\delta, 1-\delta]$ with $0 < \delta < 1/2$, assume that f'' is continuous in a neighborhood of t

and $f''(t) > 0$. If $n^{2/5}/K_n \to 0$, then

$$n^{2/5}\{\hat{f}^{[1]}(t) - f(t)\} \longrightarrow \frac{k_2(t)}{k_1(t)} \int_{-\infty}^{\infty} e^{-k_2(t)|u|} H(u) du,$$

in distribution, where $k_1(t) = 24^{-3/5}\sigma^{-8/5} f''(t)^{3/5}$, $k_2(t) = 24^{2/5}\sigma^{2/5} f''(t)^{-2/5}$, and H is the invelope function of the integrated Brownian motion.

Proof: Recall that the optimality conditions (3) and (4) for $m = 1$ are, respectively,

$$0 \le D_2 \hat{b}^{[1]} \perp C_\alpha C\big[(I + \lambda^* D_1^T D_1)\hat{b}^{[1]} - \bar{y}\big] \ge 0, \quad (14)$$

and

$$\sum_{k=1}^{K_n} \hat{b}_k^{[1]} = \sum_{k=1}^{K_n} \bar{y}_k, \quad \sum_{k=1}^{K_n} (K_n - k + 1)\hat{b}_k^{[1]} = \sum_{k=1}^{K_n} (K_n - k + 1)\bar{y}_k. \quad (15)$$

For notational simplicity, we drop the subscript $[1]$ in $\hat{b}^{[1]}$. Consider a two-step estimator. At the first step, we use the least squares estimator by finding $\vec{b}$ to minimize $\sum_{k=1}^{K_n}(\bar{y}_k - \vec{b}_k)^2$, subject to the constraint $D_2\vec{b} \ge 0$. At the second step, we find an unconstrained penalized spline estimator. Let $\tilde{b}$ solve $(I + \lambda^* D_1^T D_1)\tilde{b} = \vec{b}$. For any $t_0 \in [\delta, 1-\delta]$ where $0 < \delta < 1/2$, let $\ell_n = \lfloor K_n t_0 \rfloor$. In the following, we show that $\tilde{b}$ satisfies (15) and $\tilde{b}_{\ell_n}$ satisfies (14) asymptotically.

First note that the optimality conditions for $\vec{b}$ are $0 \le D_2\vec{b} \perp C_\alpha C\big[\vec{b} - \bar{y}\big] \ge 0$, and $\sum_{k=1}^{K_n} \vec{b}_k = \sum_{k=1}^{K_n} \bar{y}_k$ and $\sum_{k=1}^{K_n}(K_n - k + 1)\vec{b}_k = \sum_{k=1}^{K_n}(K_n - k + 1)\bar{y}_k$. It is easy to see that $\sum_{k=1}^{K_n} \vec{b}_k = \sum_{k=1}^{K_n} \hat{b}_k$ and $\sum_{k=1}^{K_n}(K_n - k + 1)\vec{b}_k = \sum_{k=1}^{K_n}(K_n - k + 1)\hat{b}_k$. Hence, $\tilde{b}$ satisfies the condition (15).

Let $\vec{f}$ be a piecewise linear function such that $\vec{f}(\kappa_k) = \vec{b}_k$ and $\tilde{f}$ be another piecewise linear function such that $\tilde{f}(\kappa_k) = \tilde{b}_k$. As shown in [6] and [17], the penalized spline estimator is asymptotically equivalent to the kernel estimator. More specifically, when K_n is of order n^γ with $\gamma > 2/5$ and λ is of order $n^{-2/5}$, for any $t \in [\delta, 1-\delta]$,

$$\begin{aligned}\tilde{f}(t) &= \int_0^1 K(t,s)\vec{f}(s)ds + \int_0^1 K(t,s)\vec{R}(s)ds \\ &\quad + e^{-\beta t(1-t)} O_p(\beta^m),\end{aligned}$$

where $K(t,s)$ is the equivalent kernel when $m = 1$ such that $K(t,s) = \frac{\beta}{2}e^{-\beta|t-s|}$, $0 \le t, s \le 1$, $\beta = \lambda^{-1/2}$ which is of order $n^{1/5}$, and the remainder R satisfies

$$\sup_{s\in[0,1]} |\vec{R}(s)| = O_p\Big(\sqrt{\frac{\log K_n}{n\lambda K_n}}\Big).$$

In particular,

$$\int_0^1 K(t,s)\vec{f}(s)ds = \int_{-\infty}^{\infty} \frac{1}{2}e^{-|x|}\vec{f}\Big(t - \frac{\beta}{u}\Big)du + O_p(e^{-\beta t(1-t)}).$$

Therefore, for any $t \in [\delta, 1-\delta]$, $\tilde{f}''(t) \ge 0$ with probability tending to one. So $\Delta^2\tilde{b}_{\ell_n} \ge 0$ with probability tending to one. Further, $\Delta^2\tilde{b}_{\ell_n} \to 0$ with probability tending to one if and only if $\Delta^2\vec{b}_{\ell_n} = 0$. Therefore, $\tilde{b}_{\ell_n}$ satisfies the optimal condition (14) asymptotically.

In the following, we study the asymptotic distribution of $\tilde{f}(t)$ for a ***fixed*** $t \in [\delta, 1-\delta]$. The asymptotic property of $\vec{f}$ can be studied along the same line as in [5]. When $n^{2/5}/K_n \to \infty$, $\vec{f}$ and the least squares estimator in [5] are asymptotically equivalent. In particular, let ω be the uniform distribution on $\{t_1, \dots, t_n\}$, and g be a piecewise constant function such that $g(t_i) = y_i$. Define

$$\begin{aligned}S_n(t) &= \int_0^t g(s)d\omega(s), \quad R_n(t) = \int_0^t \vec{f}(s)d\omega(s), \\ \tilde{R}_n(t) &= \int_0^t \vec{f}(s)ds, \quad Y_n(t) = \int_0^t S_n(s)ds, \\ H_n(t) &= \int_0^t R_n(s)ds, \quad \tilde{H}_n(t) = \int_0^t \tilde{R}_n(s)ds.\end{aligned}$$

Moreover, define their "local counterparts" at the fixed t:

$$\begin{aligned}&Y_n^{\rm loc}(u) = n^{4/5}\int_t^{t+n^{-1/5}u} \big\{S_n(s) - S_n(t) \\ &- \int_t^s (f(t) + (v-t)f'(t))d\omega(v)\big\}ds, \\ &H_n^{\rm loc}(u) = n^{4/5}\int_t^{t+n^{-1/5}u} \{R_n(s) - R_n(t) \\ &- \int_t^s (f(t) + (v-t)f'(t))d\omega(v)\}ds + A_n u + B_n, \\ &\tilde{H}_n^{\rm loc}(u) = n^{4/5}\int_t^{t+n^{-1/5}u} \{\tilde{R}_n(s) - \tilde{R}_n(t) \\ &- \int_t^s (f(t) + (v-t)f'(t))dv\}ds + A_n u + B_n,\end{aligned}$$

where $A_n = n^{3/5}\{R_n(t) - S_n(t)\}$ and $B_n = n^{4/5}\{H_n(t) - Y_n(t)\}$ which are of order $O_p(1)$ following [7, Lemma 8]. It is shown that

$$Y_n^{\rm loc}(u) \longrightarrow \sigma\int_0^u W(s)ds + \frac{1}{24}f''(t)u^4$$

in distribution uniformly on the compact set $|u| \le c$. Letting $k_1 \equiv k_1(t) = 24^{-3/5}\sigma^{-8/5}f''(t)^{3/5}$ and $k_2 \equiv k_2(t) = 24^{2/5}\sigma^{2/5}f''(t)^{-2/5}$, then

$$k_1 Y_n^{\rm loc}(k_2 u) \longrightarrow Y(u) \equiv \int_0^u W(s)ds + u^4$$

in distribution. Choosing $\beta = n^{-1/5}$, observe that

$$\begin{aligned}&\int_0^1 K(t,s)\vec{f}(s)ds = \int_{-\beta t}^{\beta(1-t)} \frac{1}{2}e^{-|u|}\vec{f}(t + n^{-1/5}u)du \\ &= n^{-2/5}\int_{-\beta t}^{\beta(1-t)} \frac{1}{2}e^{-|u|}(\tilde{H}_n^{\rm loc})''(u)du + f(t) \\ &\qquad + e^{-\beta t(1-t)}O_p(\beta^2) \\ &= n^{-2/5}\int_{-\infty}^{\infty} \frac{1}{2}e^{-|u|}\tilde{H}_n^{\rm loc}(u)du + f(t) + e^{-\beta t(1-t)}O_p(\beta^2).\end{aligned}$$

Since

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-|u|}\tilde{H}_n^{\rm loc}(u)du &= \frac{k_2}{k_1}\int_{-\infty}^{\infty} e^{-k_2|u|}k_1\tilde{H}_n^{\rm loc}(k_2 u)du \\ &\longrightarrow \frac{k_2}{k_1}\int_{-\infty}^{\infty} e^{-k_2|u|}H(u)du,\end{aligned}$$

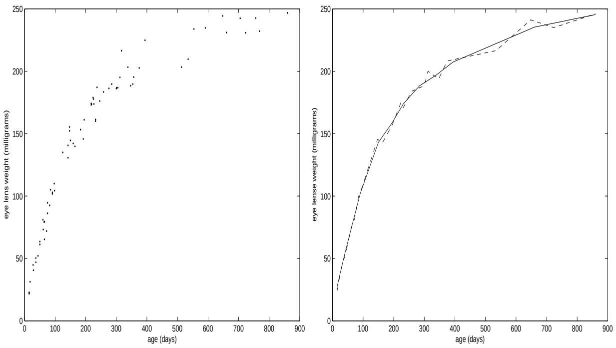

Fig. 1. Left: Rabbit data scatter plot; Right: Unconstrained penalized spline estimator (dashed line) and concave penalized spline estimator (solid line).

in distribution, we have

$$n^{2/5}\left(\int_0^1 K(t,s)\vec{f}(s)ds - f(t)\right) \longrightarrow \frac{k_2(t)}{k_1(t)}\int_{-\infty}^{\infty} e^{-k_2(t)|u|}H(u)du$$

in distribution. ■

V. Application: Rabbit Data

We illustrate two estimators by using the rabbit data from [2]. These data are available at `http://www.statsci.org/data/oz/rabbit.html`. The scatter-plot of the data is illustrated in the upper left of Figure 1. Here, the x-axis is the age measured in days and the y-axis is the eye lens weight for rabbits in Australia. The sample size is $n = 71$. [9] has used a parametric model to fit the data by assume the true regression function having the form $f(x) = ae^{-b/(x+c)}$, which is obvious a concave function. In the right panel of Figure 1, we fit two different estimators: the unconstrained penalized spline estimator and the proposed concave penalized spline estimator. We use piecewise linear function with the first order difference to fit the data. The number of knots we choose is $K_n = 20$. The unconstrained penalized spline estimator in the dashed line has the wiggle behavior and seems not a reasonable fit. Instead, the concave penalized spline estimator in the solid line gives a more reasonable concave fit.

VI. Conclusion

In this paper, we have studied the asymptotic properties of the convex spline estimator with the first order difference penalty. In particular, we have established a critical uniform Lipschitz property for the optimal spline coefficients via complementarity techniques. The pointwise asymptotic distribution of the estimator is also established.

An extension currently under investigation is to peform the asymptotic analysis for a general m. The main difficulty is to establish a similar uniform Lipschitz property for the optimal spline coefficients. Since the design matrix and difference matrix are more complicated, this becomes highly nontrivial and shall be reported in the future.

References

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

[2] K. L. Dudzinski and R. Mykytowycz. The eye lens as an indicator of age in the wild rabbit in Australia, *CSIRO Wildl. Res.*, **6**, pp. 156-159, 1961.

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

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

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

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

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

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

[9] D.A. Ratkowsky. *Nonlinear Regression Modeling.* Marcel Dekker Inc., New York, 1983.

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

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

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

[13] J. Shen and X. Wang. A constrained optimal control approach to smoothing splines. To appear in *the 50th IEEE Conf. Decision and Control*, Orlando, FL, 2011.

[14] C. Tantiyaswasdikul and M. Woodroofe. Isotonic smoothing splines under sequential designs. *Journal of Statistical Planning and Inference*, Vol. 38, pp. 75-88, 1994.

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

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

[17] X. Wang, J. Shen, and D. Ruppert On the asymptotics of penalized spline smoothing, *Electronical Journal of Statistics*, **5**, pp. 1–17, 2011.

$$
F_{\alpha,k} = \begin{bmatrix}
1 & \frac{h^{\alpha}_{k,1}-1}{h^{\alpha}_{k,1}} & \frac{h^{\alpha}_{k,1}-2}{h^{\alpha}_{k,1}} & \cdots & \frac{1}{h^{\alpha}_{k,1}} & 0 & & \cdots & & & & & & & \\
0 & \frac{1}{h^{\alpha}_{k,1}} & \frac{2}{h^{\alpha}_{k,1}} & \cdots & \frac{h^{\alpha}_{k,1}-1}{h^{\alpha}_{k,1}} & 1 & \frac{h^{\alpha}_{k,2}-1}{h^{\alpha}_{k,2}} & \cdots & \frac{1}{h^{\alpha}_{k,2}} & 0 & 0 & \cdots & & & \\
0 & & \cdots & & & 0 & \frac{1}{h^{\alpha}_{k,2}} & \cdots & \frac{h^{\alpha}_{k,2}-1}{h^{\alpha}_{k,2}} & 1 & \frac{h^{\alpha}_{k,3}-1}{h^{\alpha}_{k,3}} & \cdots & & & \\
0 & & \cdots & & & 0 & & \cdots & & 0 & \frac{1}{h^{\alpha}_{k,3}} & \cdots & & & \\
 & & & \cdots & & & & & & \cdots & & & \cdots & & \\
 & & & \cdots & & & & & & \cdots & & & \cdots & & \\
 & & & & \cdots & & & & & \cdots & & & \cdots & \frac{1}{h^{\alpha}_{k,w_k}} & 0 \\
 & & & & \cdots & & & & & \cdots & & & \cdots & \frac{h^{\alpha}_{k,w_k}-1}{h^{\alpha}_{k,w_k}} & 1
\end{bmatrix}
\tag{16}
$$

where $w_k := |\vartheta_k| - 1$.