

# A two stage $k$ -monotone B-spline regression estimator: Uniform Lipschitz property and optimal convergence rate

Teresa M. Lebair\*

*Department of Applied Mathematics and Statistics  
Johns Hopkins University  
3400 N Charles Street  
Baltimore, MD 21218, U.S.A.  
e-mail: [tlebair1@jhu.edu](mailto:tlebair1@jhu.edu)*

and

Jinglai Shen\*

*Department of Mathematics and Statistics  
University of Maryland, Baltimore County  
1000 Hilltop Circle  
Baltimore, MD 21250, U.S.A.  
e-mail: [shenj@umbc.edu](mailto:shenj@umbc.edu)*

**Abstract:** This paper considers  $k$ -monotone estimation and the related asymptotic performance analysis over a suitable Hölder class for general  $k$ . A novel two stage  $k$ -monotone B-spline estimator is proposed: in the first stage, an unconstrained estimator with optimal asymptotic performance is considered; in the second stage, a  $k$ -monotone B-spline estimator is constructed (roughly) by projecting the unconstrained estimator onto a cone of  $k$ -monotone splines. To study the asymptotic performance of the second stage estimator under the sup-norm and other risks, a critical *uniform Lipschitz property* for the  $k$ -monotone B-spline estimator is established under the  $\ell_\infty$ -norm. This property uniformly bounds the Lipschitz constants associated with the mapping from a (weighted) first stage input vector to the B-spline coefficients of the second stage  $k$ -monotone estimator, independent of the sample size and the number of knots. This result is then exploited to analyze the second stage estimator performance and develop convergence rates under the sup-norm, pointwise, and  $L_p$ -norm (with  $p \in [1, \infty)$ ) risks. By employing recent results in  $k$ -monotone estimation minimax lower bound theory, we show that these convergence rates are optimal.

**MSC 2010 subject classifications:** Primary 62G08; secondary 62G20.

**Keywords and phrases:** Asymptotic analysis, B-splines, convergence rates,  $k$ -monotone estimation, nonparametric regression, shape constraints.

Received May 2017.

---

\*Supported in part by the NSF grants CMMI-1030804 and DMS-1042916.

## 1. Introduction

For fixed  $k \in \mathbb{N}$ , a real-valued univariate function is said to be  $k$ -monotone if its  $(k - 1)$ -th derivative is increasing. Examples of  $k$ -monotone functions include monotone functions (with  $k = 1$ ) and convex functions (with  $k = 2$ ). The study of  $k$ -monotone regression concerns the nonparametric estimation of an underlying  $k$ -monotone function such that a constructed estimator preserves the  $k$ -monotone shape. Belonging to the general framework of shape constrained estimation in nonparametric statistics, this problem has garnered substantial attention in statistics and approximation theory, due to the vast number of applications; see [10, 25, 32, 35, 36] and references therein. The surging interest in shape constrained estimation is driven by the observation that incorporating shape constraints into estimator construction often improves estimator accuracy and efficiency [31].

### 1.1. Literature review and motivation

Current literature on  $k$ -monotone regression focuses mostly on monotone regression (when  $k = 1$ ) [5, 7, 29, 37, 40] and convex regression (when  $k = 2$ ) [6, 11, 17, 18, 24, 38, 41]. For  $k > 2$ , asymptotic results on regression estimation of  $(k - 1)$  times differentiable functions are limited to unconstrained estimators [26, 39] and least-squares estimators [25]. Statistical analysis results on  $k$ -monotone estimation are restricted to [1, 2, 13, 14, 15, 25], in which  $k$ -monotone density estimation (both continuous and discrete), rather than  $k$ -monotone regression is discussed. A continuous  $k$ -monotone density is defined on the positive orthant  $(0, \infty)$  and governed by a related but different notion of  $k$ -monotonicity. It is shown in [2] that a continuous density is  $k$ -monotone in the sense of [1, 2, 13, 14] if and only if its  $(k - 1)$ -th derivative is monotone (either increasing or decreasing, depending on  $k$ ), relating our  $k$ -monotone regression problem to  $k$ -monotone density estimation.

Despite extensive research on monotone and convex regression, no asymptotic performance analysis results are available on  $k$ -monotone regression estimators when  $k > 2$ , especially under the sup-norm risk, even though the (shape preserving) spline approximation of  $k$ -monotone functions for larger  $k$  has been studied in approximation theory [19, 20, 21, 30] and numerous applications are available. Such applications include insurance (with  $k = 3, 4$ ) [12], qualitative simulation (with  $k = 3$ ) [22], and attitude control (with  $k = 3$ ) [32]. Particularly, a prudent utility function  $u(\cdot)$ , in the context of risk management [12], has a nonnegative third derivative (and is thus, 3-monotone), whereas a temperate utility function  $u(\cdot)$  has a nonpositive fourth derivative (and thus,  $-u(\cdot)$  is 4-monotone). Additionally, when solving systems of qualitative differential equations [22], nonnegative third order derivative constraints are used to reduce spurious distinctions between qualitative behaviors. Finally, in the attitude control system treated in [32], the control force  $F$  is a nonlinear function of input voltage  $V$ , and experimental studies show that  $F^{(3)}(V) \geq 0$ .

Motivated by the higher-order  $k$ -monotone constraints arising in applications and the lack of statistical performance analysis, we consider a two stage  $k$ -monotone B-spline regression estimator for arbitrary  $k \in \mathbb{N}$  and study its asymptotic performance under the sup-norm, pointwise, and  $L_p$ -norm (with  $p \in [1, \infty)$ ) risks. B-splines are popular in approximation and estimation theory due to their numerical advantages [8, 10]. One can easily impose  $k$ -monotone constraints on B-spline coefficients, which can be efficiently computed via quadratic programs in polynomial time. Moreover, recent results show that when  $k = 1, 2$ ,  $k$ -monotone B-spline estimators outperform many other constrained estimators [37, 38, 41]. For instance, it is known that the popular monotone and convex least squares estimators lack smoothness and cannot achieve uniform convergence since they are inconsistent on the boundary with non-negligible asymptotic bias [28, 42]; these deficiencies are overcome by the monotone and convex B-spline estimators [37, 41]. In spite of numerical efficiency and performance advantages, the asymptotic analysis of  $k$ -monotone B-spline estimators is non-trivial, particularly when the sup-norm risk is considered.

### 1.2. Challenges in $k$ -monotone estimation

A major challenge in  $k$ -monotone spline estimation arises from the difficulty of accurately approximating certain  $k$ -monotone functions by splines that preserve the  $k$ -monotone shape. It is known that when  $k > 3$ , the accuracy of a  $k$ -monotone spline approximation associated with a given knot sequence is hampered by the  $k$ -monotone constraint [21]. Consequently, the bias of a  $k$ -monotone B-spline estimator associated with a given knot sequence is unsatisfactorily large if the knot sequence is coarse. In particular, knot sequences typically utilized to produce unconstrained B-spline estimators with optimal (asymptotic) performance are too coarse to be directly used in the production of  $k$ -monotone B-spline estimators. Further, the noisy data prevents us from choosing (asymptotically) finer knot sequences to obtain a more accurate  $k$ -monotone B-spline estimator. Hence, a single stage  $k$ -monotone B-spline estimator cannot achieve the optimal asymptotic performance under the sup-norm risk when  $k > 3$ .

Another major challenge in the  $k$ -monotone B-spline estimator asymptotic analysis emerges from the requirement of a deep understanding of the mapping from a (weighted) input vector to the corresponding B-spline coefficient vector associated with the  $k$ -monotone constraint, where the size of the input vector is related to that of the knot sequence. For fixed knot and design point sequences, it is known in optimization theory that this mapping is a Lipschitz piecewise linear function due to the  $k$ -monotone inequality constraint. As the sample and knot sequence sizes increase and tend to infinity, an infinite family of size-varying piecewise linear functions arises. A fundamental question in the asymptotic analysis asks how the Lipschitz constants of these piecewise linear functions behave as the knot sequence size tends to infinity. A critical *uniform Lipschitz property* has been established for monotone P-splines (corresponding to  $k = 1$ ) [37] and convex B-splines (corresponding to  $k = 2$ ) [41]. This property

states that the size-varying piecewise linear mappings from the (weighted) input vector to the constrained B-spline coefficient vector attain a uniform Lipschitz constant under the  $\ell_\infty$ -norm, independent of the number of linear pieces, design points, and knots. It leads to many important results in the asymptotic performance analysis, e.g., uniform convergence, optimal convergence rates, and the sup-norm and pointwise mean squared risks [41]. It has been conjectured that this property can be extended to  $k > 2$  [41]. However, this extension encounters several difficulties: the proof of this property for the monotone and convex cases ( $k = 1$  or  $2$ ), relies on the diagonal dominance of certain matrices that no longer holds for larger  $k$ . In addition, the results in [37, 41] are based on a restrictive assumption of evenly spaced design points and knots, and the extension to the unevenly spaced case is nontrivial.

### 1.3. Contributions

This paper addresses the aforementioned challenges and presents new developments in  $k$ -monotone estimation. We summarize these results and contributions as follows.

(1) As one of the key contributions of this paper, we establish the uniform Lipschitz property for general  $k \in \mathbb{N}$  via a new technique (cf. Theorem 3.1). This technique relies on a deep result in B-spline theory (dubbed de Boor’s conjecture) first proved by A. Shardin [34]; see [16] for a recent, simpler proof. Informally speaking, this result says that the  $\ell_\infty$ -norm of the inverse of the Gramian formed by the normalized B-splines of order  $k$  is uniformly bounded, independent of the knot sequence and the number of B-splines (cf. Theorem 7.1). Inspired by this result, we construct (nontrivial) coefficient matrices for the piecewise linear functions corresponding to the  $k$ -monotone constraint and relate these matrices to suitable B-splines. This yields the uniform bounds on the Lipschitz constants in the  $\ell_\infty$ -norm for arbitrary  $k$  and possibly unevenly spaced design points and knots; see Section 7.1 for an overview of the proof.

(2) We propose a novel two stage  $k$ -monotone B-spline estimator. In the first stage, an unconstrained estimator with optimal performance is used; in the second stage, a  $k$ -monotone B-spline estimator is constructed by projecting the first stage estimator onto the cone of  $k$ -monotone B-splines with a suitable knot sequence. The second stage estimator can be quickly produced by solving a strictly convex quadratic program with a small number of variables; see Section 3.2. The two stage estimator has several notable advantages: it is simple, flexible, and can be efficiently computed. More importantly, the optimal first stage estimator allows the second stage constrained B-spline estimator to obtain optimal bias and stochastic error, a key motivation for the two stage estimator.

(3) By leveraging the uniform Lipschitz property,  $k$ -monotone approximation theory, and statistical techniques, we derive (asymptotically) optimal uniform bounds on the estimator risk with respect to the sup-norm, pointwise, and  $L_p$ -norm (with  $p \in [1, \infty)$ ) errors over a Hölder class. These bounds lead to the uniform convergence and convergence rates under the sup-norm, pointwise, and

$L_p$ -norm (with  $p \in [1, \infty)$ ) risks, including boundary consistency, noting that many constrained estimators, such as the monotone least-squares estimators, fail to achieve the boundary consistency [28, 42].

(4) Finally, we show that the proposed  $k$ -monotone estimator obtains an optimal convergence rate, and thus the optimal minimax risk, over a suitable Hölder class in the sup-norm, by exploiting the recent development in minimax lower bound theory for  $k$ -monotone estimators [23, Chapter V]. This result is the first of its kind to the best of our knowledge.

The present paper not only recovers the monotone and convex estimation results and leads to new asymptotic analysis results for arbitrary  $k \in \mathbb{N}$ , but also treats  $k$ -monotone estimation in a unified framework and paves the way for the study of constrained estimation under a broader class of shape constraints.

#### 1.4. Organization and notation

The paper is organized as follows. In Section 2, we introduce the  $k$ -monotone estimation problem and  $k$ -monotone splines. In Section 3, we present the two stage  $k$ -monotone B-spline estimator, characterize the second stage estimator, and state the uniform Lipschitz property. With the help of the uniform Lipschitz property, we establish optimal convergence rates for the two stage estimator in Section 4. Two numerical examples illustrate the performance of the second stage estimator in Section 5. Concluding remarks and the proof of the uniform Lipschitz property are given in Sections 6 and 7 respectively.

*Notation.* We introduce some notation used in this paper. Define the function  $\delta_{ij}$  on  $\mathbb{N} \times \mathbb{N}$  so that  $\delta_{ij} = 1$  if  $i = j$ , and  $\delta_{ij} = 0$  otherwise. Let  $\mathbf{I}_S$  be the indicator function for a set  $S$  and let  $\mathbf{1}_n \in \mathbb{R}^n$  and  $\mathbf{1}_{n_1 \times n_2} \in \mathbb{R}^{n_1 \times n_2}$  be the column vector and matrix of all ones, respectively. For an index set  $\alpha$ , let  $\bar{\alpha}$  be its complement and  $|\alpha|$  be its cardinality. For  $j \in \mathbb{N}$ , let the set  $\alpha + j := \{i + j : i \in \alpha\}$ . For a column vector  $v \in \mathbb{R}^n$ , let  $v_i$  be its  $i$ th component. For a matrix  $A$ , let  $(A)_{ij}$  or  $(A)_{i,j}$  be its  $(i, j)$ -entry, let  $(A)_{i\bullet}$  be its  $i$ th row, and  $(A)_{\bullet j}$  be its  $j$ th column. Given an index set  $\alpha$ , let  $v_\alpha \in \mathbb{R}^{|\alpha|}$  be the vector formed by the components of  $v$  indexed by elements of  $\alpha$ , and  $(A)_{\alpha\bullet}$  be the matrix formed by the rows of  $A$  indexed by elements of  $\alpha$ . Let  $\binom{n}{p}$  denote the binomial coefficient indexed by  $n, p \in \mathbb{N}$  with  $n \geq p$ . For sequences of positive numbers  $(a_n)$  and  $(b_n)$  we write  $a_n \asymp b_n$  if there exist positive  $c, C \in \mathbb{R}^n$  such that  $c \leq \liminf_{n \rightarrow \infty} a_n/b_n \leq \limsup_{n \rightarrow \infty} a_n/b_n \leq C$ . For  $p \in [1, \infty)$ , let  $\|\cdot\|_{L_p}$  denote the  $L_p$ -norm on  $[0, 1]$  (or on some other interval  $I \subseteq \mathbb{R}$  when the context is clear), so that  $\|f\|_{L_p} = \left[ \int_0^1 |f(x)|^p dx \right]^{\frac{1}{p}}$  for any piecewise continuous function  $f : [0, 1] \rightarrow \mathbb{R}$ . Additionally, let  $\langle \cdot, \cdot \rangle_{L_2}$  denote the  $L_2$ -inner product on  $[0, 1]$  (or on some other interval  $I \subseteq \mathbb{R}$  when the context is clear), i.e.,  $\langle f, g \rangle_{L_2} := \int_0^1 f(x)g(x) dx$  for any piecewise continuous functions  $f : [0, 1] \rightarrow \mathbb{R}$  and  $g : [0, 1] \rightarrow \mathbb{R}$ . Lastly, let  $\mathbb{E}(\cdot)$  denote the expectation operator.

**2. *k*-monotone estimation and *k*-monotone splines**

In this section, we introduce the *k*-monotone estimation problem, and collect some key properties of *k*-monotone splines to be used in the sequel.

**2.1. *k*-monotone estimation**

Fix  $k \in \mathbb{N}$ . The class of *k*-monotone univariate functions on the interval  $[0, 1]$  is

$$\mathcal{S}_k := \left\{ f : [0, 1] \rightarrow \mathbb{R} \mid \begin{array}{l} \text{the } (k - 1)\text{th derivative } f^{(k-1)} \text{ exists almost} \\ \text{everywhere on } [0, 1], \text{ and } \left( f^{(k-1)}(x_1) - f^{(k-1)}(x_2) \right) \cdot (x_1 - x_2) \geq 0 \quad (1) \\ \text{when } f^{(k-1)}(x_1), f^{(k-1)}(x_2) \text{ exist} \end{array} \right\}.$$

When  $k = 1$ ,  $\mathcal{S}_k$  represents the set of increasing functions on  $[0, 1]$ . Similarly, when  $k = 2$ ,  $\mathcal{S}_k$  denotes the set of continuous convex functions on  $[0, 1]$ . Let  $r \in (k - 1, k]$  and  $L > 0$ . Set  $\gamma := r - (k - 1)$ , and define the Hölder class  $H_L^r$  with Hölder exponent  $r$  and Hölder constant  $L$ :

$$H_L^r := \left\{ f : [0, 1] \rightarrow \mathbb{R} \mid |f^{(k-1)}(x_1) - f^{(k-1)}(x_2)| \leq L|x_1 - x_2|^\gamma, \forall x_1, x_2 \in [0, 1] \right\}. \quad (2)$$

Additionally, given  $n \in \mathbb{N}$ , consider the family of design point sequences for fixed constants  $c_{\omega,1}$  and  $c_{\omega,2}$  such that  $0 < c_{\omega,1} \leq 1 \leq c_{\omega,2}$ , where  $c_{\omega,1}$  is used for the minimax lower bound analysis:

$$\mathcal{P}_n := \left\{ (x_i)_{i=0}^n \mid \begin{array}{l} 0 = x_0 < x_1 < \dots < x_n = 1, \text{ and} \\ \frac{c_{\omega,1}}{n} \leq x_i - x_{i-1} \leq \frac{c_{\omega,2}}{n}, \forall i = 1, \dots, n \end{array} \right\}. \quad (3)$$

This paper focuses on the estimation of functions in  $\mathcal{S}_{k,H}(r, L) := \mathcal{S}_k \cap H_L^r$ . Specifically, given a sequence of design points  $(x_i)_{i=0}^n \in \mathcal{P}_n$  on the interval  $[0, 1]$ , consider the regression problem

$$y_i = f(x_i) + \sigma \varepsilon_i, \quad i = 0, 1, \dots, n, \quad (4)$$

where  $f$  is an unknown true function in  $\mathcal{S}_{k,H}(r, L)$ , the  $\varepsilon_i$ 's are independent standard normal errors,  $\sigma$  is a positive constant, and the  $y_i$ 's are samples. The goal of *k*-monotone estimation is to construct an estimator  $\hat{f}$  that preserves the *k*-monotonicity of the true function characterized by  $\mathcal{S}_k$ . With regards to the estimator asymptotic analysis, we are particularly interested in the estimator uniform convergence on the entire interval  $[0, 1]$ , as well as the convergence rates of (i)  $\sup_{f \in \mathcal{S}_{k,H}(r,L)} \mathbb{E}(\|\hat{f} - f\|_\infty)$ , (ii)  $\sup_{f \in \mathcal{S}_{k,H}(r,L)} [\mathbb{E}(|\hat{f}(x) - f(x)|^p)]^{\frac{1}{p}}$  for fixed  $x \in [0, 1]$ ,  $p \in [1, \infty)$ , and (iii)  $\sup_{f \in \mathcal{S}_{k,H}(r,L)} \mathbb{E}(\|\hat{f} - f\|_{L_p})$  for  $p \in [1, \infty)$ , when the sample size  $n$  is sufficiently large. Note that it is possible to relax the

assumption of the normality of the errors  $\varepsilon_i$ 's in the model (4). However, we assume normal errors to simplify technical developments in this paper.

A two stage  $k$ -monotone B-spline regression estimator will be developed for  $k$ -monotone estimation problems (cf. Section 3). Since this estimator is given by a  $k$ -monotone spline, we give a detailed discussion on such splines in the following subsections.

**2.2. B-spline properties**

We first provide a brief review of B-splines as follows; see the monograph [8] for more details. For a given  $K \in \mathbb{N}$ , let  $\mathbf{t} := \{0 = t_0 < t_1 < \dots < t_K = 1\}$  be a sequence of  $(K + 1)$  knots in  $\mathbb{R}$ . Given  $p \in \mathbb{N}$ , let  $\{B_{p,j}^{\mathbf{t}}\}_{j=1}^{K+p-1}$  denote the  $(K + p - 1)$  B-splines of order  $p$  (or equivalently degree  $(p - 1)$ ) with knots at  $t_0, t_1, \dots, t_K$ , and the usual extension  $t_{1-p} = \dots = t_{-1} = 0$  on the left and  $t_{K+1} = \dots = t_{K+p-1} = 1$  on the right, scaled such that  $\sum_{j=1}^{K+p-1} B_{p,j}^{\mathbf{t}}(x) = 1$  for any  $x \in [0, 1]$ . The support of  $B_{p,j}^{\mathbf{t}}$  is given by (i)  $[t_{j-p}, t_j]$  when  $p = 1$  and  $1 \leq j \leq K - 1$ ; (ii)  $[t_{j-p}, t_j]$  when  $p = 1$  and  $j = K$  or for each  $j = 1, \dots, K + p - 1$  when  $p \geq 2$ . We summarize some properties of B-splines to be used in the subsequent development below:

- (i) Nonnegativity, upper bound, and partition of unity: for each  $p, j$ , and  $\mathbf{t}$ ,  $0 \leq B_{p,j}^{\mathbf{t}}(x) \leq 1$  and  $\sum_{j=1}^{K+p-1} B_{p,j}^{\mathbf{t}}(x) = 1$  for any  $x \in [0, 1]$ .
- (ii) Continuity and differentiability: when  $p = 1$ , each  $B_{p,j}^{\mathbf{t}}(x)$  is a (discontinuous) piecewise constant function given by  $\mathbf{I}_{[t_{j-1}, t_j]}(x)$  for  $1 \leq j \leq K - 1$  or  $\mathbf{I}_{[t_{K-1}, t_K]}(x)$  for  $j = K$ ; also,  $B_{p,i}^{\mathbf{t}}(x) \cdot B_{p,j}^{\mathbf{t}}(x) = 0, \forall x$  if  $i \neq j$ . When  $p = 2$ , the  $B_{p,j}^{\mathbf{t}}$ 's are continuous piecewise linear functions, and there are at most three points in  $\mathbb{R}$  where each  $B_{p,j}^{\mathbf{t}}$  is not differentiable. When  $p > 2$ , each  $B_{p,j}^{\mathbf{t}}$  is differentiable on  $\mathbb{R}$ . For  $p \geq 2$ , the derivative of  $B_{p,j}^{\mathbf{t}}$  (when it exists) is

$$\left(B_{p,j}^{\mathbf{t}}(x)\right)' = \frac{p-1}{t_{j-1} - t_{j-p}} B_{p-1,j-1}^{\mathbf{t}}(x) - \frac{p-1}{t_j - t_{j-p+1}} B_{p-1,j}^{\mathbf{t}}(x), \quad (5)$$

where we define  $\frac{p-1}{t_j - t_{j-p+1}} B_{p-1,j}^{\mathbf{t}}(x) := 0, \forall x \in [0, 1]$  for  $j = 0$  and  $j = K + p - 1$ .

- (iii)  $L_1$ -norm: for each  $j$ , the  $L_1$ -norm of  $B_{p,j}^{\mathbf{t}}$  is known to be [8, Chapter IX, eqns.(5) and (7)]

$$\|B_{p,j}^{\mathbf{t}}\|_{L_1} := \int_{\mathbb{R}} |B_{p,j}^{\mathbf{t}}(x)| dx = \frac{t_j - t_{j-p}}{p}. \quad (6)$$

**2.3.  $k$ -monotone B-splines**

In this subsection, we characterize the class of  $k$ -monotone splines via B-spline coefficients. Let

$$\mathbf{t} := \{0 = t_0 < t_1 < \dots < t_{K_n} = 1\}$$

be a given sequence of  $(K_n + 1)$  knots in  $[0, 1]$ , and let  $g_{b,\mathbf{t}} : [0, 1] \rightarrow \mathbb{R}$  be such that  $g_{b,\mathbf{t}}(x) = \sum_{j=1}^{K_n+k-1} b_j B_{k,j}^{\mathbf{t}}(x)$ , where the  $b_j$ 's are real coefficients of B-splines,  $b := (b_1, \dots, b_{K_n+k-1})^T$  is the spline coefficient vector, and the subscript  $n$  in  $K_n$  corresponds to the number of design points to be used in the subsequent sections.

To derive a necessary and sufficient condition for  $g_{b,\mathbf{t}} \in \mathcal{S}_k$ , we introduce some matrices. Let  $\tilde{D}^{(j)} \in \mathbb{R}^{j \times (j+1)}$  denote the first order difference matrix, i.e.,

$$\tilde{D}^{(j)} := \begin{bmatrix} -1 & 1 & & & & & \\ & -1 & 1 & & & & \\ & & & \ddots & \ddots & & \\ & & & & \ddots & \ddots & \\ & & & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{j \times (j+1)}. \tag{7}$$

If  $k = 1$ , let  $D_{k,\mathbf{t}} := \tilde{D}^{(K_n-1)}$ . In what follows, consider  $k > 1$ . For the given knot sequence  $\mathbf{t}$  with the usual extension  $t_j = 0$  for any  $j < 0$  and  $t_j = 1$  for any  $j > K_n$ , define the following diagonal matrices:  $\tilde{\Delta}_{0,\mathbf{t}} := I_{K_n-1}$ , and for each  $p = 1, \dots, k$ , define each  $\tilde{\Delta}_{p,\mathbf{t}} \in \mathbb{R}^{(K_n+p-1) \times (K_n+p-1)}$  as follows:

$$\tilde{\Delta}_{p,\mathbf{t}} := \frac{1}{p} \text{diag}(t_1 - t_{1-p}, t_2 - t_{2-p}, \dots, t_{K_n+p-1} - t_{K_n-1}). \tag{8}$$

Furthermore, define the matrices  $D_{p,\mathbf{t}} \in \mathbb{R}^{(K_n+k-1-p) \times (K_n+k-1)}$  inductively:

$$D_{0,\mathbf{t}} := I_{K_n+k-1}, \quad \text{and} \quad D_{p,\mathbf{t}} := \tilde{\Delta}_{k-p,\mathbf{t}}^{-1} \cdot \tilde{D}^{(K_n+k-1-p)} \cdot D_{p-1,\mathbf{t}}, \tag{9}$$

for each  $p = 1, \dots, k$ . To simplify notation, we often drop the subscript  $\mathbf{t}$  and write  $D_{p,\mathbf{t}}$  as  $D_p$  when the context is clear. Roughly speaking,  $D_{p,\mathbf{t}}$  denotes the  $p$ th order difference matrix weighted by the knots of  $\mathbf{t}$ . When the knots are evenly spaced, it is almost identical to a standard difference matrix (except on the boundary). Since  $\tilde{\Delta}_{k-p,\mathbf{t}}^{-1}$  is invertible and  $\tilde{D}^{(K_n+k-1-p)}$  has full row rank, it can be shown via induction that  $D_{p,\mathbf{t}}$  is of full row rank for any  $p$  and  $\mathbf{t}$ .

In what follows, define  $N := K_n + k - 1$  for a fixed spline order  $k \in \mathbb{N}$ . Note that  $N$  depends on  $K_n$  and thus on  $n$ .

**Lemma 2.1.** *Fix  $k \in \mathbb{N}$ . Let  $\mathbf{t} = \{0 = t_0 < t_1 < \dots < t_{K_n} = 1\}$  be a sequence of  $(K_n + 1)$  knots, and for each  $p = 1, \dots, k$ , let  $\{B_{p,j}^{\mathbf{t}}\}_{j=1}^{K_n+p-1}$  denote the B-splines of order  $p$  defined by  $\mathbf{t}$ , with the usual extension. Then the following hold:*

- (1) *For any given  $b \in \mathbb{R}^N$  and  $p = 0, 1, \dots, k - 1$ , the  $p$ th derivative of  $g_{b,\mathbf{t}} := \sum_{j=1}^N b_j B_{k,j}^{\mathbf{t}}$  is  $\sum_{j=1}^{N-p} (D_p b)_j B_{k-p,j}^{\mathbf{t}}$ , except at (at most) finitely many points on  $[0, 1]$ ;*
- (2)  *$g_{b,\mathbf{t}} \in \mathcal{S}_k$  if and only if  $D_k b \geq 0$ .*

*Proof.* We write  $g_{b,\mathbf{t}}$  as  $g$  in the following proof for notational simplicity.

(1) We prove statement (1) by induction on  $p = 0, 1, \dots, k - 1$ . Clearly, the statement holds for  $p = 0$ . Consider  $p$  with  $1 \leq p \leq k - 1$ , and assume the



statement holds for  $(p - 1)$ . It follows from (5), the induction hypothesis, and the definitions of  $\tilde{\Delta}_{k-p,t}$  and  $D_p$  that

$$\begin{aligned} g^{(p)} &= \left( g^{(p-1)} \right)' = \left( \sum_{j=1}^{N-p+1} (D_{p-1}b)_j B_{k-p+1,j}^t \right)' \\ &= (k-p) \sum_{j=1}^{N-p} \frac{(D_{p-1}b)_{j+1} - (D_{p-1}b)_j}{t_j - t_{j-k+p}} B_{k-p,j}^t \\ &= \sum_{j=1}^{N-p} \left( \tilde{\Delta}_{k-p,t}^{-1} \tilde{D}^{(N-p)} D_{p-1}b \right)_j B_{k-p,j}^t = \sum_{j=1}^{N-p} (D_p b)_j B_{k-p,j}^t, \end{aligned}$$

whenever  $g^{(p)}$  and  $g^{(p-1)}$  exist. Hence, statement (1) holds for  $p$ .

(2) It is easy to see that  $g^{(k-1)}$  exists on  $[0, 1]$  except at (at most) finitely many points in  $[0, 1]$ . It thus follows from statement (1) that  $g^{(k-1)}$  is a piecewise constant function on  $[0, 1]$ . Therefore,  $g \in \mathcal{S}_k$  if and only if the spline coefficients of  $g^{(k-1)}$  are increasing, i.e.,  $(D_{k-1}b)_j \leq (D_{k-1}b)_{j+1}$  for each  $j = 1, \dots, K_n - 1$ . This is equivalent to  $\tilde{D}^{(K_n-1)} D_{k-1}b \geq 0$ , which is further equivalent to  $D_k b \geq 0$ , in view of  $\tilde{\Delta}_{0,t} = I$  and  $D_k = \tilde{D}^{(K_n-1)} D_{k-1}$ . This gives rise to statement (2).  $\square$

### 3. Two stage $k$ -monotone B-spline estimator and the uniform Lipschitz property

In this section, we present the two stage  $k$ -monotone B-spline estimator, study the second stage  $k$ -monotone B-spline estimator, and state the uniform Lipschitz property. This property is critical to the asymptotic performance analysis.

#### 3.1. Two stage $k$ -monotone B-spline estimator

The proposed  $k$ -monotone B-spline estimator consists of two stages: in the first stage, we choose any unconstrained estimator that converges to the true function at the optimal rate over the Hölder class  $H_L^r$  in the minimax sense; the second stage B-spline estimator is constructed roughly by projecting the first stage estimator onto the cone of order  $k$  splines associated with  $\mathcal{S}_k$  and a suitable knot sequence. In what follows, we describe these two stages in detail.

##### 3.1.1. First stage unconstrained estimator

Given a design point sequence  $P = (x_i)_{i=0}^n \in \mathcal{P}_n$ , where  $c_{\omega,1}, c_{\omega,2}$  are fixed (cf. (3)), we choose the first stage estimator  $\hat{f}_P^{[1]}$  to be any unconstrained estimator such that it converges to the true function at the optimal asymptotic

rate uniformly over  $H_L^r$  for a given performance metric, independent of  $P \in \mathcal{P}_n$ . Specifically, when the sup-norm risk is considered,  $\widehat{f}_P^{[1]}$  achieves

$$\sup_{f \in H_L^r, P \in \mathcal{P}_n} \mathbb{E} \left( \|\widehat{f}_P^{[1]} - f\|_\infty \right) \asymp \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}}. \quad (10)$$

Equivalently, there exists a positive constant  $c_*$  such that  $\widehat{f}_P^{[1]}$  satisfies

$$\sup_{f \in H_L^r, P \in \mathcal{P}_n} \mathbb{E} \left( \|\widehat{f}_P^{[1]} - f\|_\infty \right) \leq c_* \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}}, \quad (11)$$

for all large  $n$ , where  $c_*$  depends only on  $k, r, L, \sigma, c_{\omega,1}$ , and  $c_{\omega,2}$ .

Alternatively, when the pointwise  $p$ -risk or  $L_p$ -norm risk with  $p \in [1, \infty)$  are considered, the first stage estimator achieves

$$\sup_{f \in H_L^r, P \in \mathcal{P}_n} \left[ \mathbb{E} \left( \left| \widehat{f}_P^{[1]}(x) - f(x) \right|^p \right) \right]^{\frac{1}{p}} \asymp n^{-\frac{r}{2r+1}} \text{ for any fixed } x \in [0, 1], \quad (12)$$

or equivalently, there exists a positive constant  $c'_*$  such that for any fixed  $x \in [0, 1]$ ,

$$\sup_{f \in H_L^r, P \in \mathcal{P}_n} \left[ \mathbb{E} \left( \left| \widehat{f}_P^{[1]}(x) - f(x) \right|^p \right) \right]^{\frac{1}{p}} \leq c'_* n^{-\frac{r}{2r+1}} \quad (13)$$

for all large  $n$ , where  $c'_*$  depends only on  $p, k, r, L, \sigma, c_{\omega,1}$ , and  $c_{\omega,2}$ . The asymptotic error bound on the second stage estimator  $L_p$ -norm risk, established in Corollary 4.1, does not follow from an asymptotic error bound on the first stage estimator  $L_p$ -norm risk. This is a result of the fact that the second stage estimator depends on pointwise evaluations of the first stage estimator at each of the design points (cf. (15) and (16)). Hence, we require (12) and (13) to hold when the  $L_p$ -norm risk is the performance metric under consideration, rather than analogous bounds on the first stage estimator  $L_p$ -norm risk.

It is known from nonparametric statistical theory [27, 39] that many (unconstrained) estimators meet the above conditions. One example is the local polynomial estimator of order  $(k - 1)$ , with bandwidth  $h_n \asymp \left( \frac{\log n}{n} \right)^{\frac{1}{2r+1}}$  for the sup-norm risk (resp.  $h_n \asymp n^{-\frac{1}{2r+1}}$  for the pointwise and  $L_p$ -norm (with  $p \in [1, \infty)$ ) risks); see [39, Section 1.6]. Another example is the (unconstrained) B-spline estimator. This gives rise to great flexibility in the construction of the two stage  $k$ -monotone estimator.

### 3.1.2. Second stage $k$ -monotone B-spline estimator

In this subsection, we describe how the first stage unconstrained estimator  $\widehat{f}_P^{[1]}$  is utilized to produce the second stage  $k$ -monotone B-spline estimator.

Given a sample data vector  $y = (y_0, y_1, \dots, y_n)^T \in \mathbb{R}^{(n+1)}$  corresponding to the to the design point sequence  $P = (x_i)_{i=0}^n \in \mathcal{P}_n$ , compute  $\widehat{f}_P^{[1]}$ . Let  $(K_n)$  be a sequence of integers depending on the sample size  $n$  to be specified later (cf. Theorems 4.1, 4.2, and Corollary 4.1). Additionally, fix constants  $c_{t,1}$  and  $c_{t,2}$  with  $0 < c_{t,1} \leq 1 \leq c_{t,2}$ . For each  $K_n \in \mathbb{N}$ , define the set of sequences of  $(K_n + 1)$  knots on  $[0, 1]$  with the usual extension on the left and right boundary:

$$\mathcal{T}_{K_n} := \left\{ (t_j)_{j=0}^{K_n} \mid 0 = t_0 < t_1 < \dots < t_{K_n} = 1, \text{ and} \right. \quad (14)$$

$$\left. \frac{c_{t,1}}{K_n} \leq t_j - t_{j-1} \leq \frac{c_{t,2}}{K_n}, \forall j = 1, \dots, K_n \right\}.$$

Recall that  $N = K_n + k - 1$ . For a given knot sequence  $\mathbf{t} := (t_j)_{j=0}^{K_n} \in \mathcal{T}_{K_n}$ , let  $\{B_{k,j}^{\mathbf{t}}\}_{j=1}^N$  denote the set of order  $k$  B-splines associated with  $\mathbf{t}$  and the usual extension, scaled such that  $\sum_{j=1}^N B_{k,j}^{\mathbf{t}}(x) = 1$  for all  $x \in [0, 1]$  (see Section 2.2 for a relevant discussion of the properties of B-splines). The second stage constrained B-spline estimator characterized by  $\mathcal{S}_k$  is given by

$$\widehat{f}_{P,\mathbf{t}}(x) := \sum_{j=1}^N \widehat{b}_j B_{k,j}^{\mathbf{t}}(x), \quad (15)$$

where the coefficient vector  $\widehat{b}_{P,\mathbf{t}} := (\widehat{b}_1, \widehat{b}_2, \dots, \widehat{b}_N)^T \in \mathbb{R}^N$  is determined by the constrained least squares regression problem

$$\widehat{b}_{P,\mathbf{t}} := \arg \min_{D_{k,\mathbf{t}} b \geq 0} \sum_{i=0}^n (x_{i+1} - x_i) \left( \widehat{f}_P^{[1]}(x_i) - \sum_{j=1}^N b_j B_{k,j}^{\mathbf{t}}(x_i) \right)^2, \quad (16)$$

where in (16)  $x_{n+1} := 1$ , and the matrix  $D_{k,\mathbf{t}}$  is defined in (9). It follows from Lemma 2.1 that the constraint  $D_{k,\mathbf{t}} \widehat{b}_{P,\mathbf{t}} \geq 0$  is equivalent to the  $k$ -monotonicity of the spline  $\widehat{f}_{P,\mathbf{t}} \in \mathcal{S}_k$ . Hence, the second stage B-spline estimator  $\widehat{f}_{P,\mathbf{t}}$  preserves the  $k$ -monotone constraint characterized by  $\mathcal{S}_k$ .

Algorithm 1 below summarizes the construction of the two stage  $k$ -monotone B-spline estimator. Note that the second stage estimator  $\widehat{f}_{P,\mathbf{t}}$  can be computed in polynomial time as a function of  $N \ll n$  by solving a strictly convex quadratic program; see Section 3.2 for more details.

### 3.2. Characterization of the second stage $k$ -monotone B-spline estimator

In this subsection, we further characterize the second stage constrained B-spline estimator  $\widehat{f}_{P,\mathbf{t}}$  introduced in (15). Toward this end, we summarize all of the assumptions that will be used throughout the rest of the paper for a fixed  $k \in \mathbb{N}$  as follows:

- (A.1) The design point sequence  $P = (x_i)_{i=0}^n$  lies in  $\mathcal{P}_n$ , where  $\mathcal{P}_n$  is defined by the fixed constants  $c_{\omega,1}, c_{\omega,2}$  satisfying  $0 < c_{\omega,1} \leq 1 \leq c_{\omega,2}$  (cf. (3)).

---

**Algorithm 1** Two Stage *k*-monotone B-spline Estimator Construction

---

**Stage 1:**

Fix constants  $c_{\omega,1}, c_{\omega,2}$  with  $0 < c_{\omega,1} \leq 1 \leq c_{\omega,2}$ .

Obtain sample data  $y = (y_0, y_1, \dots, y_n)^T$  associated with design point sequence  $P = (x_i)_{i=0}^n \in \mathcal{P}_n$ .

Compute the first stage unconstrained estimator  $\widehat{f}_P^{[1]}$  to satisfy (10) or (12).

**Stage 2:**

Fix constants  $c_{t,1}, c_{t,2}$  with  $0 < c_{t,1} \leq 1 \leq c_{t,2}$ .

Choose a suitable sequence of integers  $(K_n)$  (cf. Theorems 4.1, 4.2, and Corollary 4.1).

Pick a knot sequence  $\mathbf{t} \in \mathcal{T}_{K_n}$ .

Compute the second stage *k*-monotone B-spline estimator  $\widehat{f}_{P,\mathbf{t}}$  via (15) and (16).

---

(A.2) The sample  $y = (y_i)_{i=0}^n$  is produced from the regression model (4), i.e.,  $y_i = f(x_i) + \sigma \varepsilon_i, \forall i$ , where the true function  $f \in \mathcal{S}_{k,H}(r, L)$  with constants  $L > 0, r \in (k-1, k]$  and  $\sigma > 0$ ; the  $\varepsilon_i$ 's are independent standard normal errors.

(A.3) The knot sequence  $\mathbf{t} = (t_j)_{j=0}^{K_n}$  for the second stage *k*-monotone B-spline estimator lies in  $\mathcal{T}_{K_n}$ , where  $\mathcal{T}_{K_n}$  is defined by the fixed constants  $c_{t,1}, c_{t,2}$  satisfying  $0 < c_{t,1} \leq 1 \leq c_{t,2}$  (cf. (14)).

Consider the second stage estimator  $\widehat{f}_{P,\mathbf{t}}$  given by (15) and (16). Let  $\widehat{f}$  and  $\widehat{b}$  denote  $\widehat{f}_{P,\mathbf{t}}$  and  $\widehat{b}_{P,\mathbf{t}}$  respectively, noting that  $\widehat{f}$  and  $\widehat{b}$  depend on  $P$  and  $\mathbf{t}$ . It follows from Lemma 2.1 that  $\widehat{f} \in \mathcal{S}_k$ .

Define the diagonal matrix

$$\Theta := \text{diag}(x_1 - x_0, x_2 - x_1, \dots, x_{n+1} - x_n) \in \mathbb{R}^{(n+1) \times (n+1)},$$

and the design matrix  $\widehat{X} \in \mathbb{R}^{(n+1) \times N}$  with  $(\widehat{X})_{(i+1),j} := B_{k,j}^{\mathbf{t}}(x_i) \forall i, j$ . Further, let  $\Lambda_{K_n, P, \mathbf{t}} := K_n \cdot \widehat{X}^T \Theta \widehat{X} \in \mathbb{R}^{N \times N}$ , where we write  $\Lambda$  for  $\Lambda_{K_n, P, \mathbf{t}}$  for notational simplicity. Let

$$\vec{f}^{[1]} := \left( \widehat{f}_P^{[1]}(x_0), \widehat{f}_P^{[1]}(x_1), \dots, \widehat{f}_P^{[1]}(x_n) \right)^T \in \mathbb{R}^{(n+1)}, \tag{17}$$

where  $\widehat{f}_P^{[1]}$  is the first stage estimator (cf. Section 3.1.1); define the weighted input vector  $\vec{y} := K_n \cdot \widehat{X}^T \Theta \vec{f}^{[1]} \in \mathbb{R}^N$ . The quadratic optimization problem in (16) for  $\widehat{b}_{P,\mathbf{t}}$  or simply  $\widehat{b}$  can be written as

$$\widehat{b} := \arg \min_{D_k b \geq 0} \frac{1}{2} b^T \Lambda b - b^T \vec{y}. \tag{18}$$

Note that  $\Lambda$  is positive definite for any given  $K_n, P$ , and  $\mathbf{t}$ . Hence, problem (18) (or equivalently (16)) is a strictly convex quadratic program, which can be solved in polynomial time by many effective numerical solvers.

Define  $\bar{\Lambda} \in \mathbb{R}^{N \times N}$  and  $\bar{y} \in \mathbb{R}^N$  such that

$$(\bar{\Lambda})_{ij} = K_n \langle B_{k,i}^{\mathbf{t}}, B_{k,j}^{\mathbf{t}} \rangle_{L_2} \quad \text{and} \quad (\bar{y})_i = K_n \langle B_{k,i}^{\mathbf{t}}, \widehat{f}_P^{[1]} \rangle_{L_2}$$

for  $i, j = 1, \dots, N$ . We note that for  $n$  sufficiently large,  $\bar{\Lambda}$  approximates  $\Lambda$  and  $\bar{\mathbf{y}}$  approximates  $\mathbf{y}$ . Hence, (18) is closely related to

$$\bar{\mathbf{b}} := \arg \min_{D_k \mathbf{b} \geq 0} \frac{1}{2} \mathbf{b}^T \bar{\Lambda} \mathbf{b} - \mathbf{b}^T \bar{\mathbf{y}} = \arg \min_{D_k \mathbf{b} \geq 0} \left\| \hat{f}_P^{[1]} - \sum_{j=1}^N b_j B_{k,j}^{\mathbf{t}} \right\|_{L_2}^2. \quad (19)$$

Because a spline  $g := \sum_{j=1}^N b_j B_{k,j}^{\mathbf{t}}$  is  $k$ -monotone if and only if  $D_{k,\mathbf{t}} \mathbf{b} \geq 0$  (cf. Lemma 2.1), it follows that the estimator  $\bar{f}_{P,\mathbf{t}}(x) := \sum_{j=1}^N \bar{b}_j B_{k,j}^{\mathbf{t}}(x)$ , with  $\bar{b}_j$  given by (19), is the  $L_2$ -projection of the first stage estimator onto the cone of  $k$ -monotone splines with knot sequence  $\mathbf{t}$ . Hence, the second stage estimator given by (15) and (16) roughly corresponds to the  $L_2$ -projection of the first stage estimator onto the cone of  $k$ -monotone splines with knot sequence  $\mathbf{t}$ .

We call a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  *piecewise linear* if there are finitely many linear functions  $\{h_s\}_{s=1}^r$  on  $\mathbb{R}^n$  such that for any  $x \in \mathbb{R}^n$ ,  $f(x) = h_s(x)$  for some  $h_s$ , where each linear function  $h_s$  is called a *linear piece* of  $f$ . It is well known that a (continuous) piecewise linear function is globally Lipschitz continuous [33]. Note that  $\hat{\mathbf{b}}$  in (18) can be treated as a function of the weighted input vector  $\bar{\mathbf{y}}$ . In fact, since the matrix  $\Lambda$  is positive definite for any given  $K_n$ ,  $P$ , and  $\mathbf{t}$ , it follows from the standard theory of quadratic programming that the function  $\hat{\mathbf{b}} : \mathbb{R}^N \rightarrow \mathbb{R}^N$  from  $\bar{\mathbf{y}}$  to  $\hat{\mathbf{b}}(\bar{\mathbf{y}})$  is piecewise linear and globally Lipschitz continuous [33]. To establish the piecewise linear formulation of  $\hat{\mathbf{b}}$  or equivalently the linear pieces of  $\hat{\mathbf{b}}$ , we consider the KKT conditions for (18) given by:

$$\Lambda \hat{\mathbf{b}}(\bar{\mathbf{y}}) - \bar{\mathbf{y}} - D_k^T \chi = 0, \quad 0 \leq \chi \perp D_k \hat{\mathbf{b}}(\bar{\mathbf{y}}) \geq 0, \quad (20)$$

where  $\chi \in \mathbb{R}^{K_n-1}$  is the Lagrange multiplier, and  $u \perp v$  means that the vectors  $u$  and  $v$  are orthogonal. Similar to [37, 38, 41], define the index set for any given  $\bar{\mathbf{y}} \in \mathbb{R}^N$ ,

$$\alpha := \left\{ i : (D_k \hat{\mathbf{b}}(\bar{\mathbf{y}}))_i = 0 \right\} \subseteq \{1, \dots, K_n - 1\}, \quad (21)$$

where  $\alpha$  may be the empty set. For each  $\alpha$  associated with  $\bar{\mathbf{y}}$ , the KKT conditions in (20) become

$$(D_k)_{\alpha \bullet} \hat{\mathbf{b}}(\bar{\mathbf{y}}) = 0, \quad \chi_{\bar{\alpha}} = 0, \quad \Lambda \hat{\mathbf{b}}(\bar{\mathbf{y}}) - \bar{\mathbf{y}} - ((D_k)_{\alpha \bullet})^T \chi_{\alpha} = 0. \quad (22)$$

For a given index set  $\alpha$ , let  $F_{\alpha}^T \in \mathbb{R}^{N \times (|\bar{\alpha}|+k)}$  be a matrix whose columns form a basis for the null space of  $(D_{k,\mathbf{t}})_{\alpha \bullet}$  or simply  $(D_k)_{\alpha \bullet}$ . This implies that  $F_{\alpha}$  depends on both  $\alpha$  and  $\mathbf{t}$ . In particular, if  $\alpha$  is the empty set, then  $N = K_n + k - 1 = |\bar{\alpha}| + k$ , and  $F_{\alpha}^T$  is the identity matrix of order  $N$ . To express  $\hat{\mathbf{b}}(\bar{\mathbf{y}})$  in terms of  $F_{\alpha}$  for an arbitrary  $\alpha$ , we deduce from  $(D_k)_{\alpha \bullet} \hat{\mathbf{b}}(\bar{\mathbf{y}}) = 0$  in (22) that  $\hat{\mathbf{b}}(\bar{\mathbf{y}}) = F_{\alpha}^T \tilde{\mathbf{b}}^{\alpha}$ , where  $\tilde{\mathbf{b}}^{\alpha}$  is the vector of the free variables in the linear equation  $(D_k)_{\alpha \bullet} \hat{\mathbf{b}}(\bar{\mathbf{y}}) = 0$ . To determine  $\tilde{\mathbf{b}}^{\alpha}$ , we substitute  $\hat{\mathbf{b}}(\bar{\mathbf{y}}) = F_{\alpha}^T \tilde{\mathbf{b}}^{\alpha}$  into the

rightmost equation of (22) and obtain that

$$\begin{aligned} \Lambda F_\alpha^\top \tilde{b}^\alpha - \bar{y} - ((D_k)_{\alpha\bullet})^\top \chi_\alpha = 0 &\implies F_\alpha \Lambda F_\alpha^\top \tilde{b}^\alpha - F_\alpha \bar{y} = 0 \\ &\implies \tilde{b}^\alpha = (F_\alpha \Lambda F_\alpha^\top)^{-1} F_\alpha \bar{y} \\ &\implies \hat{b}(\bar{y}) = F_\alpha^\top \tilde{b}^\alpha = F_\alpha^\top (F_\alpha \Lambda F_\alpha^\top)^{-1} F_\alpha \bar{y}, \end{aligned}$$

where we use the fact that  $F_\alpha \Lambda F_\alpha^\top$  is invertible, since  $F_\alpha$  has full row rank and  $\Lambda$  is positive definite.

For any index set  $\alpha$  given in (21), define the linear function  $\hat{b}_{P,\mathbf{t}}^\alpha : \mathbb{R}^N \rightarrow \mathbb{R}^N$  as

$$\hat{b}_{P,\mathbf{t}}^\alpha(\bar{y}) := F_\alpha^\top (F_\alpha \Lambda F_\alpha^\top)^{-1} F_\alpha \bar{y}, \quad (23)$$

where  $F_\alpha$  depends on  $\mathbf{t}, \alpha$ , and  $\Lambda$  depends on  $K_n, P, \mathbf{t}$ . Clearly,  $\hat{b}_{P,\mathbf{t}}^\alpha$  is a linear piece of the piecewise linear function  $\hat{b} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ , and  $\hat{b}$  has  $2^{K_n-1}$  linear pieces for a given  $K_n$ . Finally, observe that for any invertible matrix  $R \in \mathbb{R}^{(|\bar{\alpha}|+k) \times (|\bar{\alpha}|+k)}$ ,

$$(RF_\alpha)^\top ((RF_\alpha) \Lambda (RF_\alpha)^\top)^{-1} (RF_\alpha) = F_\alpha^\top (F_\alpha \Lambda F_\alpha^\top)^{-1} F_\alpha.$$

Thus any choice of  $F_\alpha$  leads to the same  $\hat{b}_{P,\mathbf{t}}^\alpha$ , provided that the columns of  $F_\alpha^\top$  form a basis of the null space of  $(D_k)_{\alpha\bullet}$ .

### 3.3. *k-monotone B-spline estimator: Uniform Lipschitz property*

As indicated in the previous subsection, the piecewise linear function  $\hat{b}_{P,\mathbf{t}}(\cdot)$  (or simply  $\hat{b}(\cdot)$ ) is Lipschitz continuous for fixed  $K_n, P, \mathbf{t}$ . An important question asks whether the Lipschitz constants of size-varying  $\hat{b}$  with respect to the  $\ell_\infty$ -norm are uniformly bounded, independent of  $K_n, P$ , and  $\mathbf{t}$ , as long as the numbers of design points and knots are sufficiently large. If so, we say that  $\hat{b}$  satisfies the *uniform Lipschitz property*. Originally introduced and studied in [37, 38, 40, 41] for monotone P-splines ( $k = 1$ ) and convex B-splines ( $k = 2$ ) with equally spaced design points and knots, this property is shown to play a crucial role in the uniform convergence and asymptotic analysis of  $k$ -monotone B-spline estimators. In this paper, we establish this property for  $k$ -monotone B-spline estimators with general  $k$ , under relaxed conditions on the design points and knots.

For any  $p, K_n \in \mathbb{N}$ , and  $\mathbf{t} \in \mathcal{T}_{K_n}$ , it is noted that for any  $t_j \in \mathbf{t}$ , we have,

$$\frac{t_j - t_{j-p}}{p} = \frac{1}{p} \sum_{s=j}^{j-p+1} (t_s - t_{s-1}) \leq \frac{1}{p} \cdot p \cdot \frac{c_{t,2}}{K_n} \leq \frac{c_{t,2}}{K_n}.$$

Moreover, in view of  $t_{j-p} = 0$  for any  $j \leq p$  and  $t_j = 1$  for any  $j \geq K_n$ , it can be shown that for each  $1 \leq j \leq K_n + p - 1$ ,  $\frac{t_j - t_{j-p}}{p} \geq c_{t,1}/(p \cdot K_n)$  such that  $\frac{p}{t_j - t_{j-p}} \leq p \cdot K_n / c_{t,1}$ . In summary, we have, for each  $j = 1, \dots, K_n + p - 1$ ,

$$\frac{c_{t,1}}{p \cdot K_n} \leq \frac{t_j - t_{j-p}}{p} \leq \frac{c_{t,2}}{K_n}, \quad \text{and} \quad \frac{K_n}{c_{t,2}} \leq \frac{p}{t_j - t_{j-p}} \leq \frac{p \cdot K_n}{c_{t,1}}. \quad (24)$$

Using the above notation, the following theorem states one of the the main results of the paper, i.e., the uniform Lipschitz property of  $\widehat{b}_{P,\mathbf{t}}(\cdot)$ . Since the proof of this property is technical, we postpone it to Section 7 to maintain a smooth paper flow.

**Theorem 3.1.** *Let  $k \in \mathbb{N}$  and constants  $c_{\omega,1}, c_{\omega,2}, c_{t,1}, c_{t,2}$  be fixed, where  $0 < c_{\omega,1} \leq 1 \leq c_{\omega,2}$  and  $0 < c_{t,1} \leq 1 \leq c_{t,2}$ . For any  $n, K_n \in \mathbb{N}$ , let  $\widehat{b}_{P,\mathbf{t}} : \mathbb{R}^{K_n+k-1} \rightarrow \mathbb{R}^{K_n+k-1}$  be the piecewise linear function in (18) corresponding to the  $k$ th order B-splines defined by the design point sequence  $P \in \mathcal{P}_n$  and the knot sequence  $\mathbf{t} \in \mathcal{T}_{K_n}$ . Then there exists a positive constant  $c_\infty$ , depending on  $k, c_{t,1}$  only, such that for any increasing sequence  $(K_n)$  with  $K_n \rightarrow \infty$  and  $K_n/n \rightarrow 0$  as  $n \rightarrow \infty$ , there exists  $n_* \in \mathbb{N}$ , depending on  $(K_n)$  (and the fixed constants  $k, c_{\omega,2}, c_{t,1}, c_{t,2}$ ) only, such that for any  $P \in \mathcal{P}_n$  and  $\mathbf{t} \in \mathcal{T}_{K_n}$  with  $n \geq n_*$ ,*

$$\|\widehat{b}_{P,\mathbf{t}}(u) - \widehat{b}_{P,\mathbf{t}}(v)\|_\infty \leq c_\infty \|u - v\|_\infty, \quad \forall u, v \in \mathbb{R}^{K_n+k-1}.$$

The above result can be refined for a particular sequence of  $P$  and  $\mathbf{t}$ .

**Corollary 3.1.** *Let  $(K_n)$  be an increasing sequence with  $K_n \rightarrow \infty$  and  $K_n/n \rightarrow 0$  as  $n \rightarrow \infty$ , and  $(P_n, \mathbf{t}_n)$  be a sequence in  $\mathcal{P}_n \times \mathcal{T}_{K_n}$ . Then there exists a positive constant  $c'_\infty$ , independent of  $n$ , such that for each  $n$ ,*

$$\|\widehat{b}_{P_n, \mathbf{t}_n}(u) - \widehat{b}_{P_n, \mathbf{t}_n}(v)\|_\infty \leq c'_\infty \|u - v\|_\infty, \quad \forall u, v \in \mathbb{R}^{K_n+k-1}.$$

This corollary recovers the past results on the uniform Lipschitz property for  $k = 1, 2$  (e.g., [37, 41]) for equally spaced design points and knots.

#### 4. Two stage $k$ -monotone estimator risk and optimal convergence rate

In this section, we exploit the uniform Lipschitz property to establish the sup-norm, pointwise, and  $L_p$ -norm (with  $p \in [1, \infty)$ ) risks of the second stage  $k$ -monotone B-spline estimator. In particular, we show that the second stage  $k$ -monotone B-spline estimator achieves the same asymptotic convergence rate as the first stage (unconstrained) estimator over  $\mathcal{S}_{k,H}(r, L)$  under the sup-norm risk, pointwise risk, and  $L_p$ -norm risk. This result leads to the optimal convergence rate of the two stage  $k$ -monotone B-spline estimator  $\widehat{f}_{P,\mathbf{t}}$  over  $\mathcal{S}_{k,H}(r, L)$  in the sup-norm (cf. Theorem 4.4).

##### 4.1. Second stage $k$ -monotone B-spline estimator risks

To develop the risks of the second stage  $k$ -monotone B-spline estimator  $\widehat{f}_{P,\mathbf{t}}$ , we first establish the following lemma, which concerns the approximation of  $k$ -monotone functions  $f \in \mathcal{S}_{k,H}(r, L)$  by  $k$ -monotone splines associated with a

given knot sequence. It will be used to bound the second stage estimator risks in the subsequent development. For a given knot sequence  $\mathbf{t} \in \mathcal{T}_{K_n}$ , let

$$\mathbb{S}_{+,k}^{\mathbf{t}} := \left\{ f_B : [0, 1] \rightarrow \mathbb{R} \mid f_B \text{ is an order } k \text{ B-spline with knots in } \mathbf{t} \right. \\ \left. \text{and } f_B \in \mathcal{S}_k \right\}.$$

Recall that  $r \in (k - 1, k]$  for a fixed  $k \in \mathbb{N}$ , and define  $q := \min\{r - k + 3, r\}$ .

**Lemma 4.1.** *Let the positive constant  $c_B := \frac{3}{2}Lc_{t,2}^3$ . Then for each  $K_n \in \mathbb{N}$ ,  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and  $f \in \mathcal{S}_{k,H}(r, L)$ , there exists an order  $k$  spline  $f_B \in \mathbb{S}_{+,k}^{\mathbf{t}}$  satisfying  $\|f_B - f\|_\infty \leq c_B K_n^{-q}$ .*

*Proof.* Consider the following two cases:

**Case 1:**  $q = r$ . In this case, we have that  $r \leq r - k + 3$  so that  $k \leq 3$ , i.e.,  $k = 1, 2$ , or  $3$ . It is known in [4, 9] that the lemma holds for  $k = 1$ , by taking  $f_B$  to be the piecewise constant least squares approximation or interpolant of  $f$ . Similarly, when  $k = 2$ , the lemma holds, by taking  $f_B$  to be the piecewise linear interpolant of  $f$  at the knots in  $\mathbf{t}$  [3, 4, 41]. Finally, when  $k = 3$ , it is shown in [20] that the lemma holds for evenly spaced knots and in [30] that a similar result holds (where  $c_B$  is taken to be a slightly larger constant) for unevenly spaced knots. An alternative construction of the  $k$ -monotone constrained  $f_B$  is given in [23, Appendix A], demonstrating that when  $k = 3$ , the lemma holds for unevenly spaced knots with  $c_B = \frac{3}{2}Lc_{t,2}^3$ . Hence, the lemma holds in Case 1.

**Case 2:**  $q < r$ . In this case, we have that  $r > r - k + 3 = q$  so that  $k > 3$ . Define  $g := f^{(k-3)} \in \mathcal{S}_{3,H}(r + 3 - k, L)$ . By the results of Case 1, there exists a spline  $g_B \in \mathbb{S}_{+,k}^{\mathbf{t}}$  such that  $\|g_B - g\|_\infty \leq c_B K_n^{-q}$ . Define the function  $f_B : [0, 1] \rightarrow \mathbb{R}$  as

$$f_B(x) := \sum_{p=0}^{k-4} \frac{f^{(p)}(0)}{p!} x^p + \int_0^x \int_0^{t_1} \cdots \int_0^{t_{k-4}} g_B(t_{k-3}) dt_{k-3} \cdots dt_2 dt_1,$$

Then  $f_B \in \mathcal{S}_{k,H}(r, L)$  and

$$\begin{aligned} |f_B(x) - f(x)| &\leq \int_0^x \int_0^{t_1} \cdots \int_0^{t_{k-4}} |g_B(t_{k-3}) - g(t_{k-3})| dt_{k-3} \cdots dt_2 dt_1 \\ &\leq \int_0^x \int_0^{t_1} \cdots \int_0^{t_{k-4}} c_B K_n^{-q} dt_{k-3} \cdots dt_2 dt_1 \\ &\leq \frac{c_B}{(k-3)!} K_n^{-q} \leq c_B K_n^{-q}, \end{aligned}$$

for all  $x \in [0, 1]$ , where, if  $k > 4$ , we define, for any integrable function  $h$ ,

$$\int_0^x \int_0^{t_1} \cdots \int_0^{t_{k-4}} h(t_{k-3}) dt_{k-3} \cdots dt_2 dt_1 := \int_0^x h(t_1) dt_1.$$

Consequently, we have  $\|f_B - f\|_\infty \leq c_B K_n^{-q}$  in either case. □



The above lemma establishes a uniform bound on the error that arises from approximating a  $k$ -monotone function  $f \in \mathcal{S}_{k,H}(r, L)$  by a  $k$ -monotone spline for a fixed knot sequence. With the help of this result, we obtain the following theorem, which bounds the risk and establishes a convergence rate for the second stage  $k$ -monotone B-spline estimator under the sup-norm over this Hölder class.

**Theorem 4.1.** *Let  $K_n := \left\lceil \left(\frac{n}{\log n}\right)^{\frac{r}{q(2r+1)}} \right\rceil$ . Let  $\widehat{f}_P^{[1]}$  denote any suitable first stage estimator as described in Section 3.1.1 (cf. (11)). Then there exists a constant  $C_k > 0$ , depending only on  $k, r, L, c_{t,1}, c_{t,2}, c_{\omega,1}, c_{\omega,2}$ , and  $\sigma$  such that for all  $n$  sufficiently large,*

$$\sup_{f \in \mathcal{S}_{k,H}(r,L), P \in \mathcal{P}_n, \mathbf{t} \in \mathcal{T}_{K_n}} \mathbb{E} \left( \|\widehat{f}_{P,\mathbf{t}} - f\|_\infty \right) \leq C_k \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}}.$$

*Proof.* Fix  $f \in \mathcal{S}_{k,H}(r, L)$ . For notational simplicity, let  $\widehat{f}$  and  $\widehat{f}^{[1]}$  denote  $\widehat{f}_{P,\mathbf{t}}$  and  $\widehat{f}_P^{[1]}$  respectively for any given  $P \in \mathcal{P}_n$  and  $\mathbf{t} \in \mathcal{T}_{K_n}$ . By Lemma 4.1, for each  $K_n \in \mathbb{N}$  and  $\mathbf{t} \in \mathcal{T}_{K_n}$ , there exists  $f_B \in \mathbb{S}_{+,k}^{\mathbf{t}}$  such that  $\|f - f_B\|_\infty \leq c_B K_n^{-q} \leq c_B \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}$ . Hence,

$$\|\widehat{f} - f\|_\infty \leq \|\widehat{f} - f_B\|_\infty + \|f_B - f\|_\infty \leq \|\widehat{f} - f_B\|_\infty + c_B \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}, \quad (25)$$

and we bound  $\|\widehat{f} - f_B\|_\infty$  as follows. In view of  $f_B \in \mathbb{S}_{+,k}^{\mathbf{t}}$  and Lemma 2.1, there exists  $\tilde{b} := (\tilde{b}_1, \dots, \tilde{b}_N)^T \in \mathbb{R}^N$  such that

$$f_B(x) = \sum_{j=1}^N \tilde{b}_j B_{k,j}^{\mathbf{t}}(x) \quad \text{and} \quad D_{k,\mathbf{t}} \tilde{b} \geq 0.$$

Define  $\vec{f}_B := (f_B(x_0), f_B(x_1), \dots, f_B(x_n))^T \in \mathbb{R}^{(n+1)}$ . In light of the results in Section 3, we have

$$\frac{1}{2} \|K_n^{1/2} \Theta^{1/2} (\widehat{X} \tilde{b} - \vec{f}_B)\|_2^2 = 0,$$

and therefore,

$$\tilde{b} = \arg \min_{D_{k,\mathbf{t}} b \geq 0} \frac{1}{2} b^T \Lambda b - b^T (K_n \widehat{X}^T \Theta \vec{f}_B).$$

By the uniform Lipschitz property stated in Theorem 3.1, there exist a uniform positive constant  $c_\infty$  and a large  $n_* \in \mathbb{N}$  such that for all  $n \geq n_*$ ,

$$\begin{aligned} \|\widehat{f} - f_B\|_\infty &\leq \|\widehat{b}_{P,\mathbf{t}} - \tilde{b}\|_\infty \leq c_\infty \|K_n \widehat{X}^T \Theta (f^{[1]} - \vec{f}_B)\|_\infty \\ &\leq c_\infty \|K_n \widehat{X}^T \Theta\|_\infty \cdot \|f^{[1]} - f_B\|_\infty \\ &\leq c_\infty \|K_n \widehat{X}^T \Theta\|_\infty \cdot (\|\widehat{f}^{[1]} - f\|_\infty + \|f - f_B\|_\infty) \\ &\leq c_\infty \|K_n \widehat{X}^T \Theta\|_\infty \cdot (\|\widehat{f}^{[1]} - f\|_\infty + c_B \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}), \end{aligned} \quad (26)$$

where  $\widehat{f}^{[1]}$  is defined in (17). Additionally, it follows from the definition of  $\widehat{X}^T$  and the non-negativity, upper bound, and support of the  $B_{k,j}^t$ 's shown in Section 2.2 that for each  $j = 1, \dots, N$ ,

$$\begin{aligned} \left\| (K_n \widehat{X}^T \Theta)_{j\bullet} \right\|_\infty &= K_n \sum_{i=0}^n B_{k,j}^t(x_i) \cdot (x_{i+1} - x_i) \\ &\leq K_n \sum_{i=0}^n \mathbf{I}_{[t_{j-k}, t_j]}(x_i) \cdot (x_{i+1} - x_i) \\ &\leq K_n \left( t_j - t_{j-k} + \frac{2c_{\omega,2}}{n} \right) \leq k \cdot c_{t,2} + \frac{2c_{\omega,2}K_n}{n} \\ &\leq c_{t,2}(k+1), \end{aligned} \tag{27}$$

whenever  $n$  is sufficiently large such that  $(2c_{\omega,2}K_n/n) \leq c_{t,2}$ . Combining (25)–(27), we obtain

$$\|\widehat{f} - f\|_\infty \leq c_\infty c_{t,2}(k+1) \cdot \|f - \widehat{f}^{[1]}\|_\infty + c_B \left( c_\infty c_{t,2}(k+1) + 1 \right) \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}}.$$

In view of (11) and taking the expectation of both sides of the above inequality, we have

$$\mathbb{E} \left( \|\widehat{f} - f\|_\infty \right) \leq \left[ (c_B + c_*) c_\infty c_{t,2}(k+1) + c_B \right] \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}}$$

for all  $n$  sufficiently large. This yields the desired result with  $C_k := (c_B + c_*) c_\infty c_{t,2}(k+1) + c_B$ .  $\square$

The above theorem establishes the uniform convergence of the estimator  $\widehat{f}_{P,t}$  on the entire interval  $[0, 1]$ , as well as this estimator's consistency, including the consistency at the two boundary points, over a Hölder class. Note that the monotone and convex least-squares estimators are inconsistent at the boundary points due to a non-negligible asymptotic bias [17, 28, 29, 42], which is known as the spiking problem. Furthermore, the asymptotic convergence rate established in Theorem 4.1 will be shown to be optimal in Section 4.2.

An analogous result can be established under the pointwise risk by using the uniform Lipschitz property and piecewise linear function theory [33].

**Theorem 4.2.** *Let  $K_n := \left\lceil n^{\frac{r}{q(2r+1)}} \right\rceil$  and  $p \in [1, \infty)$ . For any  $x_* \in [0, 1]$ , let  $\widehat{f}_P^{[1]}$  denote any suitable first stage estimator as described in Section 3.1.1 (cf. (13)). Then there exists a positive constant  $\overline{C}_{p,k}$ , depending only on  $p, k, r, L, c_{t,1}, c_{t,2}, c_{\omega,1}, c_{\omega,2}$ , and  $\sigma$ , such that for all  $n$  sufficiently large,*

$$\sup_{f \in S_{k,H}(r,L), P \in \mathcal{P}_n, t \in \mathcal{T}_{K_n}} \left[ \mathbb{E} \left( \left| \widehat{f}_{P,t}(x_*) - f(x_*) \right|^p \right) \right]^{\frac{1}{p}} \leq \overline{C}_{p,k} n^{-\frac{r}{2r+1}}.$$

*Proof.* Fix arbitrary  $f \in \mathcal{S}_{k,H}(r, L)$  and  $x_* \in [0, 1]$ . For any  $P \in \mathcal{P}_n$  and  $\mathbf{t} \in \mathcal{T}_{K_n}$ , we again let  $\widehat{f}$  and  $\widehat{f}^{[1]}$  denote  $\widehat{f}_{P,\mathbf{t}}$  and  $\widehat{f}_P^{[1]}$ , respectively. By an argument similar to that in the proof of Theorem 4.1, there exists  $f_B \in \mathbb{S}_{+,k}^{\mathbf{t}}$  such that  $\|f - f_B\|_\infty \leq c_B K_n^{-q}$ , where  $f_B(x) = \sum_{j=1}^N \widetilde{b}_j B_{k,j}^{\mathbf{t}}(x)$  for all  $x \in [0, 1]$  for some  $\widetilde{\mathbf{b}} := (\widetilde{b}_1, \dots, \widetilde{b}_N)^T \in \mathbb{R}^N$  satisfying  $D_{k,\mathbf{t}} \widetilde{\mathbf{b}} \geq 0$ . Note that for  $p \in [1, \infty)$ ,

$$\begin{aligned} |\widehat{f}(x_*) - f(x_*)|^p &\leq \left( |\widehat{f}(x_*) - f_B(x_*)| + |f_B(x_*) - f(x_*)| \right)^p \\ &\leq \left[ \max \left( 2|\widehat{f}(x_*) - f_B(x_*)|, 2|f_B(x_*) - f(x_*)| \right) \right]^p \\ &\leq 2^p |\widehat{f}(x_*) - f_B(x_*)|^p + 2^p |f_B(x_*) - f(x_*)|^p \quad (28) \\ &\leq 2^p |\widehat{f}(x_*) - f_B(x_*)|^p + 2^p c_B^p n^{-\frac{pr}{2r+1}}. \end{aligned}$$

In what follows, we establish a uniform bound on  $|\widehat{f}(x_*) - f_B(x_*)|^p$ .

Define  $\vec{f}_B := (f_B(x_0), f_B(x_1), \dots, f_B(x_n))^T \in \mathbb{R}^{(n+1)}$ . As in the proof of Theorem 4.1, we have

$$\widetilde{\mathbf{b}} = \arg \min_{D_{k,\mathbf{t}} \mathbf{b} \geq 0} \frac{1}{2} \mathbf{b}^T \Lambda \mathbf{b} - \mathbf{b}^T (K_n \widehat{X}^T \Theta \vec{f}_B).$$

Now, define  $h := \left( B_{k,1}^{\mathbf{t}}(x_*), B_{k,2}^{\mathbf{t}}(x_*), \dots, B_{k,N}^{\mathbf{t}}(x_*) \right)^T \in \mathbb{R}^N$ , and for any given  $P$  and  $\mathbf{t}$ , let  $W_\alpha := F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha$  for any index set  $\alpha$  defined in (21). Since  $\widehat{b}_{P,\mathbf{t}}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^N$  is a Lipschitz piecewise linear function on  $\mathbb{R}^N$ , it follows from piecewise linear function theory [33] that  $\widehat{b}_{P,\mathbf{t}}(\cdot)$  admits a conic subdivision of  $\mathbb{R}^N$ , which consists of  $q_N$  polyhedral cones that partition  $\mathbb{R}^N$ . On the  $s$ th polyhedral cone where  $s = 1, \dots, q_N$ , the function  $\widehat{b}_{P,\mathbf{t}}(\cdot)$  coincides with a linear function whose coefficient matrix is given by  $W_{\alpha^s}$  for some index set  $\alpha^s$  defined in (21). By an argument similar to that in [41, Proposition 4.3], we see that for any  $z, z' \in \mathbb{R}^N$ , there exist real numbers  $\mu_1, \dots, \mu_{q_N} \in [0, 1]$  (depending on  $z$  and  $z'$ ) whose sum is one such that

$$\widehat{b}_{P,\mathbf{t}}(z') - \widehat{b}_{P,\mathbf{t}}(z) = \left( \sum_{s=1}^{q_N} \mu_s W_{\alpha^s} \right) (z' - z).$$

Note that  $\widehat{\mathbf{b}} = \widehat{b}_{P,\mathbf{t}}(K_n \widehat{X}^T \Theta \vec{f}^{[1]})$  and  $\widetilde{\mathbf{b}} = \widehat{b}_{P,\mathbf{t}}(K_n \widehat{X}^T \Theta \vec{f}_B)$ . Thus for fixed  $\vec{f}^{[1]}$  and  $\vec{f}_B$ , there exist  $\mu_1, \dots, \mu_{q_N} \in [0, 1]$ , whose sum is one, such that

$$\begin{aligned} |\widehat{f}(x_*) - f_B(x_*)|^p &= |h^T (\widehat{\mathbf{b}} - \widetilde{\mathbf{b}})|^p \\ &= \left| h^T \left( \widehat{b}_{P,\mathbf{t}}(K_n \widehat{X}^T \Theta \vec{f}^{[1]}) - \widehat{b}_{P,\mathbf{t}}(K_n \widehat{X}^T \Theta \vec{f}_B) \right) \right|^p \\ &= \left| h^T \left( \sum_{s=1}^{q_N} \mu_s W_{\alpha^s} \right) (K_n \widehat{X}^T \Theta) (\vec{f}^{[1]} - \vec{f}_B) \right|^p \quad (29) \\ &\leq 2^p \left| h^T \left( \sum_{s=1}^{q_N} \mu_s W_{\alpha^s} \right) (K_n \widehat{X}^T \Theta) (\vec{f}^{[1]} - \vec{f}) \right|^p \end{aligned}$$

$$+ 2^p \left| h^T \left( \sum_{s=1}^{q_N} \mu_s W_{\alpha^s} \right) (K_n \widehat{X}^T \Theta) (\vec{f} - \vec{f}_B) \right|^p,$$

where the inequality follows from an argument similar to that in (28), and  $\vec{f} := (f(x_0), f(x_1), \dots, f(x_n))^T \in \mathbb{R}^{n+1}$  denotes the noiseless data vector. Furthermore, by (27), we have, for all  $n$  sufficiently large,

$$\|K_n \Theta \widehat{X}\|_1 = \max_{j=1, \dots, N} \|(K_n \Theta \widehat{X})_{\bullet j}\|_1 = \max_{j=1, \dots, N} \|(K_n \widehat{X}^T \Theta)_{j \bullet}\|_\infty \leq c_{t,2}(k+1).$$

Now, the uniform Lipschitz property (cf. Theorem 3.1) implies  $\|W_{\alpha^s}\|_\infty \leq c_\infty$  for any  $\alpha^s$ . Therefore, in view of  $\|h\|_1 = 1$  and the symmetry of  $W_{\alpha^s}$ , we have, for all  $n$  sufficiently large and any fixed  $q_N$ -tuple  $(\mu_1, \dots, \mu_{q_N})$  satisfying  $\mu_s \in [0, 1], \forall s = 1, \dots, q_N$  and  $\sum_{s=1}^{q_N} \mu_s = 1$ ,

$$\begin{aligned} \left\| K_n \Theta \widehat{X} \left( \sum_{s=1}^{q_N} \mu_s W_{\alpha^s} \right) h \right\|_1 &\leq \|K_n \Theta \widehat{X}\|_1 \cdot \left( \sum_{s=1}^{q_N} \mu_s \|W_{\alpha^s}\|_\infty \right) \cdot \|h\|_1 \\ &\leq c_{t,2}(k+1)c_\infty. \end{aligned} \tag{30}$$

To simplify notation, define the random vectors  $v := K_n \Theta \widehat{X} \left( \sum_{s=1}^{q_N} \mu_s W_{\alpha^s} \right) h \in \mathbb{R}^{n+1}$  and  $u := \vec{f}^{[1]} - \vec{f} \in \mathbb{R}^{(n+1)}$ . By (30) and the bound on  $\|f - f_B\|_\infty$ ,

$$\begin{aligned} \mathbb{E} \left[ \left| v^T (\vec{f} - \vec{f}_B) \right|^p \right] &\leq \mathbb{E} \left[ \left| \|v\|_1 \cdot \|\vec{f} - \vec{f}_B\|_\infty \right|^p \right] \leq \mathbb{E} \left[ \|v\|_1^p \right] \cdot \|f - f_B\|_\infty^p \\ &\leq [c_{t,2}(k+1)c_\infty]^p c_B^p n^{-\frac{pr}{2r+1}}. \end{aligned}$$

By (13), (30), Jensen's inequality, the law of total expectation, we obtain

$$\begin{aligned} \mathbb{E} \left[ |v^T u|^p \right] &\leq \mathbb{E} \left[ \left| \sum_{i=1}^{n+1} |v_i| \cdot |u_i| \right|^p \right] \leq \mathbb{E} \left[ \|v\|_1^p \cdot \sum_{i=1}^{n+1} \frac{|v_i|}{\|v\|_1} |u_i|^p \right] \\ &\leq [c_{t,2}(k+1)c_\infty]^p \cdot \mathbb{E} \left[ \mathbb{E} \left[ \sum_{i=1}^{n+1} \frac{|v_i|}{\|v\|_1} |u_i|^p \mid v \right] \right] \\ &\leq [c_{t,2}(k+1)c_\infty]^p (c'_*)^p n^{-\frac{pr}{2r+1}}, \end{aligned} \tag{31}$$

noting that the positive constant  $c'_*$  given in (13) is independent of  $x \in [0, 1]$ . Combining the above results and (29), we obtain

$$\mathbb{E} \left( \left| \widehat{f}(x_*) - f_B(x_*) \right|^p \right) \leq 2^p [c_{t,2}(k+1)c_\infty]^p ((c'_*)^p + c_B^p) n^{-\frac{pr}{2r+1}}.$$

It then follows from the above display and (28) that for all  $n$  sufficiently large,

$$\left[ \mathbb{E}(|\widehat{f}(x_*) - f(x_*)|^p) \right]^{\frac{1}{p}} \leq \overline{C}_{p,k} n^{-\frac{r}{2r+1}},$$

where  $\overline{C}_{p,k} := 2 \left[ [c_{t,2}(k+1)c_\infty]^p ((c'_*)^p + c_B^p) + c_B^p \right]^{\frac{1}{p}}$ . □

We finally bound the  $L_p$ -norm risk of the second stage  $k$ -monotone B-spline estimator, over a Hölder class, in the following corollary.

**Corollary 4.1.** *Let  $K_n := \left\lceil n^{\frac{r}{2r+1}} \right\rceil$  and  $p \in [1, \infty)$ . Let  $\widehat{f}_P^{[1]}$  denote any suitable first stage estimator as described in Section 3.1.1 (cf. (13)). Then there exists a positive constant  $\overline{C}_{p,k}$ , depending only on  $k, r, L, c_{t,1}, c_{t,2}, c_{\omega,1}, c_{\omega,2}, \sigma$ , and  $p$ , such that for all  $n$  sufficiently large,*

$$\sup_{f \in \mathcal{S}_{k,H}(r,L), P \in \mathcal{P}_n, \mathbf{t} \in \mathcal{T}_{K_n}} \mathbb{E} \left( \|\widehat{f}_{P,\mathbf{t}} - f\|_{L_p} \right) \leq \overline{C}_{p,k} n^{-\frac{r}{2r+1}}.$$

*Proof.* By the concavity of the function  $\sqrt[p]{\cdot}$  and the Fubini-Tonelli Theorem,

$$\begin{aligned} \mathbb{E} \left( \|\widehat{f}_{P,\mathbf{t}} - f\|_{L_p} \right) &= \mathbb{E} \left[ \left( \int_0^1 |\widehat{f}_{P,\mathbf{t}}(x) - f(x)|^p dx \right)^{\frac{1}{p}} \right] \\ &\leq \left[ \mathbb{E} \left( \int_0^1 |\widehat{f}_{P,\mathbf{t}}(x) - f(x)|^p dx \right) \right]^{\frac{1}{p}} \\ &= \left[ \int_0^1 \mathbb{E} (|\widehat{f}_{P,\mathbf{t}}(x) - f(x)|^p) dx \right]^{\frac{1}{p}} \leq \overline{C}_{p,k} n^{-\frac{r}{2r+1}}. \end{aligned}$$

The last inequality follows from Theorem 4.2, as  $\overline{C}_{p,k}$  in Theorem 4.2 does not depend on  $x_* \in [0, 1]$ .  $\square$

#### 4.2. Optimal convergence rate in the sup-norm

In this subsection, we show that the attained convergence rate indicates the optimal asymptotic performance in the sup-norm. Consider the  $k$ -monotone regression problem given by (4). For simplicity, we assume that for each  $n \in \mathbb{N}$ , the design points  $(x_i)_{i=0}^n$  are evenly spaced, i.e.,  $x_i = i/n$ , although this assumption is not crucial and can be relaxed. In what follows, we denote an estimator as  $\widehat{f}_n$  to emphasize the dependence on the sample size  $n$ .

Theorem 4.3 below establishes a lower bound under the sup-norm on the minimax risk associated with  $k$ -monotone estimators over  $\mathcal{S}_{k,H}(r, L)$ . Its proof relies on minimax lower bound theory from nonparametric estimation [39, Section 2] and a careful construction of a suitable collection of functions (called hypotheses) that satisfy the  $k$ -monotone constraint. In particular, such a collection of hypothesis functions from  $\mathcal{S}_{k,H}(r, L)$  satisfies a certain sup-norm separation order and a small total  $L_2$ -distance order, based upon information theoretical results on distances between probability measures in minimax lower bound theory [27, 39]. See [23, Chapter V] for the construction and complete proof; a simpler proof using a similar idea for convex estimation is given in [24].

**Theorem 4.3.** *Fix  $k \in \mathbb{N}$  and  $r \in (k-1, k]$ . Consider the constrained regression problem (4). There exists a positive constant  $c_k$  such that*

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

where  $\inf_{\hat{f}_n}$  denotes the infimum over all *k*-monotone estimators  $\hat{f}_n \in \mathcal{S}_k$  on the interval  $[0, 1]$ .

It follows from Theorems 4.1 and 4.3 that the two stage *k*-monotone B-spline estimator  $\hat{f}_{P,t}$  satisfies: for all *n* sufficiently large,

$$\begin{aligned} c_k \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}} &\leq \inf_{\hat{f}_n} \sup_{f \in \mathcal{S}_{k,H}(r,L)} \mathbb{E}(\|\hat{f}_n - f\|_\infty) \\ &\leq \sup_{f \in \mathcal{S}_{k,H}(r,L)} \mathbb{E}(\|\hat{f}_{P,t} - f\|_\infty) \leq C_k \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}}. \end{aligned}$$

In light of the above result, we obtain the following theorem demonstrating that  $\hat{f}_{P,t}$  is an (asymptotically) optimally performing estimator over the constrained function class  $\mathcal{S}_{k,H}(r, L)$ , and the optimal rate of convergence is given by  $\left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}$ . This result not only unifies the optimal minimax risk theory for monotone and convex estimation in the sup-norm, but also extends such developments to the general *k*-monotone case with unevenly spaced design points and/or knots, which is the first result of its kind to the best of our knowledge.

**Theorem 4.4.** *Fix  $k \in \mathbb{N}$ ,  $r \in (k - 1, k]$ , and  $L > 0$ . Consider the regression problem given by (4). Then*

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{S}_{k,H}(r,L)} \mathbb{E}(\|\hat{f}_n - f\|_\infty) \asymp \left( \frac{\log n}{n} \right)^{\frac{r}{2r+1}},$$

where  $\inf_{\hat{f}_n}$  denotes the infimum over all *k*-monotone estimators  $\hat{f}_n \in \mathcal{S}_k$  on the interval  $[0, 1]$ .

### 5. Numerical examples

In this section, two numerical examples are presented, in order to illustrate the performance of the second stage *k*-monotone B-spline estimator under the sup-norm risk, the pointwise root mean squared error, and the  $L_2$ -norm risk. In both examples, the first stage unconstrained estimator is given by a minimax optimal unconstrained B-spline estimator, with evenly spaced knots; the second stage *k*-monotone/constrained B-spline estimator utilizes evenly spaced knots as well. For performance comparison, the *k*-monotone least squares estimator is also computed. The design points are taken to be evenly spaced in both examples. The sample data vector *y* is generated via the model in (4). In Example 1, we consider the 3-monotone function  $f \in \mathcal{S}_{3,H}(3, 54)$  given by  $f(t) := 9 \min((t - 1/3), 0)^3$  with  $k = r = 3$  and  $L = 54$  and take  $\sigma = 0.1$ . In Example 2, we consider the 4-monotone function  $f(t) := e^t + 16(\max(0, t - 1/2))^4 \in \mathcal{S}_{4,H}(4, 384 + e)$  with  $k = r = 4$  and  $L = 384 + e$ , and take  $\sigma = 1$ . The number of subintervals for the second stage constrained B-spline estimator is given by  $K_n := \lceil 2 \left(\frac{n}{\log n}\right)^{\frac{1}{2q+1}} \rceil$ , when analyzing the sup-norm risk, and  $K_n := \lceil 2n^{\frac{1}{2q+1}} \rceil$ , when analyzing the

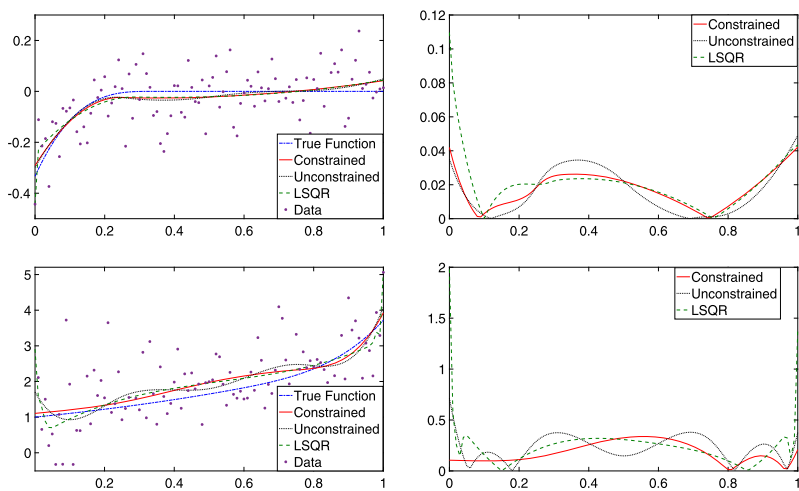


FIG 1. When  $n = 100$ , left column: estimator performance for 3-monotone (top) and 4-monotone (bottom) functions; right column: the corresponding absolute error.

pointwise root mean squared error and the  $L_2$ -norm risk, with  $q = \min(r - k + 3, r) = 3$  (cf. Section 4.1) in both cases.

Plots of the second stage  $k$ -monotone/constrained B-spline estimator, the unconstrained first stage estimator, and the  $k$ -monotone least squares estimator are given on the left in Figure 1 for Example 1 and Example 2, with  $n = 100$ . The corresponding absolute errors for these three estimators are then plotted on the right in the same figure. To further compare the performance of the three estimators, simulations were run 200 times, and the average performance of each estimator over these simulations was recorded in each case. Three performance metrics were considered, i.e., the sup-norm risk, the pointwise root mean squared error (at  $x = 0$ ), and the  $L_2$ -norm risk. Table 1 summarizes the performance of the estimators for different sample sizes, where  $\hat{f}$  denotes the computed estimator. It is seen in these examples, the second stage constrained estimator either performs approximately the same as or noticeably outperforms the other two estimators. In particular, the second stage estimator outperforms the other two estimators with respect to the sup-norm risk (in all cases). It is also observed that the  $k$ -monotone least squares estimator performs poorly with respect to both the sup-norm risk and the root mean squared error at  $x = 0$ . This is expected, as least squares constrained estimators are inconsistent on the boundary [17, 42].

## 6. Conclusions

This paper establishes a critical uniform Lipschitz property for a two stage  $k$ -monotone B-spline estimator with possibly unevenly spaced design points

TABLE 1  
 Performance of the 2nd Stage Estimator (con.), the 1st Stage Estimator (unc.), and the *k*-Monotone Least Squares Estimator (LSQR)

n	$\mathbb{E} [\ f - \hat{f}\ _\infty]$			$\sqrt{\mathbb{E} [(f(0) - \hat{f}(0))^2]}$			$\mathbb{E} [\ f - \hat{f}\ _{L_2}]$		
	con.	unc.	LSQR	con.	unc.	LSQR	con.	unc.	LSQR
	Example 1			Example 1			Example 1		
10 <sup>2</sup>	0.0548	0.0607	0.0915	0.0488	0.0481	0.0757	0.0100	0.0101	0.0101
10 <sup>3</sup>	0.0192	0.0223	0.0838	0.0185	0.0188	0.0779	0.0033	0.0033	0.0033
10 <sup>4</sup>	0.0076	0.0083	0.0856	0.0071	0.0072	0.0737	0.0010	0.0010	0.0010
	Example 2			Example 2			Example 2		
10 <sup>2</sup>	0.5217	0.6325	1.0524	0.3805	0.5525	0.8695	0.1006	0.1009	0.1002
10 <sup>3</sup>	0.1858	0.2380	1.0207	0.1192	0.2155	0.8710	0.0327	0.0327	0.0327
10 <sup>4</sup>	0.0736	0.0849	1.0307	0.0485	0.0759	0.8206	0.0100	0.0100	0.0100

and/or knots. This property is exploited, in combination with statistical techniques and B-spline approximation theory, to study the asymptotic performance of the two stage *k*-monotone B-spline estimator under the sup-norm, pointwise, and  $L_p$ -norm (with  $p \in [1, \infty)$ ) risks. The optimal convergence rate under the sup-norm risk is established. Future research topics include estimation under more general constraints, e.g., multiple derivative constraints or constraints on a linear combination of derivatives of different orders. Additional extensions to Sobolev classes and multivariate regression will also be considered.

**7. Appendix: Proof of the uniform Lipschitz property**

In this section, we prove the uniform Lipschitz property stated in Theorem 3.1.

**7.1. Overview of the proof**

The proof of Theorem 3.1 is somewhat technical. To facilitate the reading, we outline its key ideas and provide a road map of the proof as follows. In view of the piecewise linear formulation of  $\hat{b}_{P,t}(\cdot)$  in (23), we see that for any given  $K_n$ ,  $P$ , and  $\mathbf{t}$ , the Lipschitz constant of  $\hat{b}_{P,t}(\cdot)$  with respect to the  $\ell_\infty$ -norm is

$$\max_\alpha \|F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha\|_\infty,$$

where  $\max_\alpha$  denotes the maximum over all the index sets  $\alpha$  defined in (21). Therefore it suffices to establish a uniform bound on  $\|F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha\|_\infty$  for all large  $n$ , independent of  $K_n$ ,  $\alpha$ ,  $P \in \mathcal{P}_n$ , and  $\mathbf{t} \in \mathcal{T}_{K_n}$ . Motivated by [41], we observe that for any diagonal matrix  $\Xi'_\alpha$  (possibly depending on  $\alpha$ ) with positive diagonal entries,

$$\|F_\alpha^T (F_\alpha \Lambda F_\alpha^T)^{-1} F_\alpha\|_\infty \leq \|F_\alpha^T\|_\infty \cdot \|(\Xi'_\alpha F_\alpha \Lambda F_\alpha^T)^{-1}\|_\infty \cdot \|\Xi'_\alpha F_\alpha\|_\infty. \tag{32}$$

Hence, a natural idea is to choose suitable matrices  $F_\alpha$  and  $\Xi'_\alpha$  so that the quantities  $\|F_\alpha^T\|_\infty$ ,  $\|(\Xi'_\alpha F_\alpha \Lambda F_\alpha^T)^{-1}\|_\infty$ , and  $\|\Xi'_\alpha F_\alpha\|_\infty$  are all uniformly bounded.



A critical technique for establishing a uniform bound on  $\|(\Xi'_\alpha F_\alpha \Lambda F_\alpha^\top)^{-1}\|_\infty$  relies on a deep result in B-spline theory (dubbed de Boor's conjecture) proved by Shardin [34]. Roughly speaking, this result says that the  $\ell_\infty$ -norm of the inverse of the Gramian of the normalized B-splines of order  $k$  is uniformly bounded, independent of the knot sequence and the number of B-splines.

**Theorem 7.1.** [34, Theorem I] *Fix a spline order  $k \in \mathbb{N}$ . Let  $a, b \in \mathbb{R}$  with  $a < b$ , and  $\{B_{k,j}^{\mathbf{t}'}\}_{j=1}^{K+k-1}$  be the  $k$ th order B-splines on  $[a, b]$  defined by a knot sequence  $\mathbf{t}' := \{a = t'_0 < t'_1 < \dots < t'_K = b\}$  for some  $K \in \mathbb{N}$ . Let  $G \in \mathbb{R}^{(K+k-1) \times (K+k-1)}$  be the Gramian matrix given by*

$$(G)_{ij} := \frac{\langle B_{k,i}^{\mathbf{t}'}, B_{k,j}^{\mathbf{t}'} \rangle}{\|B_{k,i}^{\mathbf{t}'}\|_{L_1}}, \quad \forall i, j = 1, \dots, K+k-1.$$

Then there exists a positive constant  $\rho_k$ , independent of  $a, b, \mathbf{t}'$ , and  $K$ , such that  $\|G^{-1}\|_\infty \leq \rho_k$ .

Inspired by this theorem, we intend to approximate  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^\top$  by an appropriate B-spline Gramian matrix with a uniform approximation error bound. To achieve this goal, we construct suitable matrices  $F_\alpha$  and  $\Xi'_\alpha$  in Section 7.2; see  $F_\alpha := F_{\alpha,k,\mathbf{t}}$  defined in (39) and  $\Xi'_\alpha := \frac{k}{K_n} \cdot \Xi_{k,V_{\alpha,\mathbf{t}}}$  defined in (37). In view of  $F_\alpha \Lambda F_\alpha^\top = K_n \cdot (F_\alpha \widehat{X}^\top) \cdot \Theta \cdot (F_\alpha \widehat{X}^\top)^\top$ , we compute the entries of  $F_\alpha \widehat{X}^\top$  in Section 7.3 and show that they are related to suitable B-splines (cf. Proposition 7.2). It is shown in Section 7.4 that  $F_\alpha^\top$  and  $\Xi'_\alpha F_\alpha$  attain uniform bounds (cf. Proposition 7.4) and that  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^\top$  attains a uniform approximation error bound (cf. Proposition 7.5), all with respect to the  $\ell_\infty$ -norm. We then apply Theorem 7.1 to establish a uniform bound on  $(\Xi'_\alpha F_\alpha \Lambda F_\alpha^\top)^{-1}$  (cf. Corollary 7.1). Using these bounds, the uniform Lipschitz property is proven in Section 7.5.

## 7.2. Construction of the matrix $F_\alpha$

In this subsection, we construct a suitable matrix  $F_\alpha$  used for the piecewise linear formulation of  $\widehat{b}_{P,\mathbf{t}}$  in (23). For  $K_n \in \mathbb{N}$ , let  $\mathbf{t} \in \mathcal{T}_{K_n}$  be a knot sequence of  $(K_n + 1)$  knots, and  $\alpha \subseteq \{1, \dots, K_n - 1\}$  be an index set defined in (21). The complement of  $\alpha$  is  $\bar{\alpha} = \{i_1, \dots, i_{|\bar{\alpha}|}\}$  with  $1 \leq i_1 < \dots < i_{|\bar{\alpha}|} \leq K_n - 1$ . For notational simplicity, define  $q_\alpha := |\bar{\alpha}| + k$ .

We introduce the following matrix which is of full row rank:

$$F_{\alpha,0,\mathbf{t}} := (I_N)_{(\bar{\alpha} \cup \{K_n, K_n+1, \dots, N\}) \bullet} \in \mathbb{R}^{q_\alpha \times N}, \quad (33)$$

where we recall  $N = K_n + k - 1$ . For the given index set  $\alpha$ , define the knots

$$\tau_{\alpha,\mathbf{t},j} := \begin{cases} 0, & \text{for } j = 1 - k, \dots, 0 \\ t_{i_j}, & \text{for } j = 1, \dots, |\bar{\alpha}| \\ 1, & \text{for } j = |\bar{\alpha}| + 1, \dots, q_\alpha, \end{cases} \quad (34)$$

and the associated knot sequence

$$V_{\alpha, \mathbf{t}} := \left\{ \tau_{\alpha, \mathbf{t}, 0}, \tau_{\alpha, \mathbf{t}, 1}, \dots, \tau_{\alpha, \mathbf{t}, |\bar{\alpha}|+1} \right\}, \quad (35)$$

with the usual extension. To simplify notation, we write  $\tau_{\alpha, \mathbf{t}, j}$  and  $V_{\alpha, \mathbf{t}}$  as  $\tau_j$  and  $V$  respectively if the context is clear. Note that if  $\alpha$  is empty, then  $\bar{\alpha} = \{1, \dots, K_n - 1\}$ ,  $F_{\alpha, 0, \mathbf{t}} = I_N$ , and  $V = \mathbf{t}$ .

For any fixed  $K \in \mathbb{N}$  and any knot sequence  $\mathbf{t}' := (t'_j)_{j=0}^K$  on  $[0, 1]$  with  $(K + 1)$  distinct knots and the usual extension, consider the diagonal matrices of order  $(K + k - 1)$ :  $\Delta_{0, \mathbf{t}'} := I_{K+k-1}$  and for each  $p = 1, \dots, k$ ,

$$\begin{aligned} \Delta_{p, \mathbf{t}'} &:= \begin{bmatrix} p \cdot \tilde{\Delta}_{p, \mathbf{t}'} & 0 \\ 0 & I_{k-p} \end{bmatrix} \\ &= \text{diag} \left( (t'_1 - t'_{1-p}), (t'_2 - t'_{2-p}), \dots, (t'_{K+p-1} - t'_{K-1}), \underbrace{1, \dots, 1}_{(k-p)\text{-copies}} \right), \end{aligned} \quad (36)$$

where  $\tilde{\Delta}_{p, \mathbf{t}'}$  is defined in (8). Related to these, we further introduce the following matrices of order  $(K + k - 1)$ : for  $p = 0$ , let  $\Xi_{0, \mathbf{t}'} := (\Delta_{0, \mathbf{t}'})^{-1} = I_{K+k-1}$ , and for each  $p = 1, \dots, k$ ,

$$\begin{aligned} \Xi_{p, \mathbf{t}'} &:= (\Delta_{p, \mathbf{t}'})^{-1} \\ &= \text{diag} \left( (t'_1 - t'_{1-p})^{-1}, (t'_2 - t'_{2-p})^{-1}, \dots, (t'_{K+p-1} - t'_{K-1})^{-1}, \underbrace{1, \dots, 1}_{(k-p)\text{-copies}} \right). \end{aligned} \quad (37)$$

In addition, define the following square matrices of order  $r \in \mathbb{N}$ :

$$\widehat{S}^{(r)} := \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \dots & 1 \end{bmatrix}, \quad \text{and} \quad \widehat{D}^{(r)} := (\widehat{S}^{(r)})^{-1} = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}. \quad (38)$$

Note that  $\widehat{S}^{(r)}$  acts as a summation operator, while  $\widehat{D}^{(r)}$  is similar to a difference matrix.

With the above notation, we now define  $F_{\alpha, p, \mathbf{t}}$  inductively:  $F_{\alpha, 0, \mathbf{t}}$  is as defined in (33), and

$$F_{\alpha, p, \mathbf{t}} := \widehat{D}^{(q_\alpha)} \cdot \Xi_{p-1, V_{\alpha, \mathbf{t}}} \cdot F_{\alpha, p-1, \mathbf{t}} \cdot \Delta_{p-1, \mathbf{t}} \cdot \widehat{S}^{(N)} \in \mathbb{R}^{q_\alpha \times N}, \quad p = 1, \dots, k. \quad (39)$$

For notational simplicity, we also write  $F_{\alpha, p, \mathbf{t}}$  as  $F_{\alpha, p}$  for each  $p = 0, 1, \dots, k$ . Note that  $F_{\alpha, p}$  depends on both  $\alpha$  and  $\mathbf{t}$  for  $p \geq 1$ , since  $\Xi_{p-1, V} := \Xi_{p-1, V_{\alpha, \mathbf{t}}}$  depends on both  $\alpha$  and  $\mathbf{t}$ , and  $\Delta_{p-1, \mathbf{t}}$  depends on  $\mathbf{t}$ . Additionally, since  $\widehat{D}^{(q_\alpha)}$ ,  $\Xi_{p-1, V}$ ,  $\Delta_{p-1, \mathbf{t}}$ , and  $\widehat{S}^{(N)}$  are all invertible, each  $F_{\alpha, p}$  has full row rank by induction. Furthermore, it is easy to see that if  $\alpha$  is the empty set, then, in view of  $F_{\alpha, 0} = I_N$  and using induction,  $F_{\alpha, p} = I_N$  for each  $p = 1, \dots, k$ , regardless of the choice of  $\mathbf{t}$ .

For any  $K \in \mathbb{N}$  and any knot sequence  $\mathbf{t}' := (t'_j)_{j=0}^K$  with  $(K + 1)$  knots, we also define the following matrices inductively:

$$S_{0,\mathbf{t}'} := I_{K+k-1}, \text{ and } S_{p,\mathbf{t}'} := \Delta_{k-p,\mathbf{t}'} \cdot \widehat{S}^{(K+k-1)} \cdot S_{p-1,\mathbf{t}'}, \quad p = 1, \dots, k. \tag{40}$$

A useful observation is that  $F_{\alpha,k,\mathbf{t}}$  can be expressed in terms of  $S_{k,\mathbf{t}}$ . In fact, it is easy to verify that  $F_{\alpha,k,\mathbf{t}} = Q \cdot F_{\alpha,0,\mathbf{t}} \cdot S_{k,\mathbf{t}}$  for a suitable matrix  $Q$ . It is shown below that  $F_{\alpha,k,\mathbf{t}}$  (or simply  $F_{\alpha,k}$ ) constructed above is a suitable choice for  $F_\alpha$  in the piecewise linear formulation of the function  $\widehat{b}_{P,\mathbf{t}}$  in (23).

**Proposition 7.1.** Fix  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and let  $\alpha$  be any index set defined in (21). If  $\alpha$  is nonempty, then the columns of  $F_{\alpha,k,\mathbf{t}}^\mathbb{T}$  form a basis of the null space of  $(D_{k,\mathbf{t}})_{\alpha\bullet}$ ; otherwise,  $F_{\alpha,k,\mathbf{t}}^\mathbb{T} = I_N$ .

*Proof.* When  $\alpha$  is the empty set, it has been shown that  $F_{\alpha,k,\mathbf{t}}^\mathbb{T} = I_N$ ; see the discussion below (39). Therefore, in what follows, consider the case when  $\alpha$  is nonempty. To simplify notation, we drop the subscript  $\mathbf{t}$  in  $\widetilde{\Delta}_{p,\mathbf{t}}$ ,  $D_{p,\mathbf{t}}$ ,  $\Delta_{p,\mathbf{t}}$ , and  $S_{p,\mathbf{t}}$  for  $p = 0, 1, \dots, k$ ; see (8), (9), (36), and (40) for the definitions of  $\widetilde{\Delta}_{p,\mathbf{t}}$ ,  $D_{p,\mathbf{t}}$ ,  $\Delta_{p,\mathbf{t}}$ , and  $S_{p,\mathbf{t}}$  respectively.

We first show via induction on  $p$  that there exists a constant  $c_{p,k} \in \mathbb{R}$ , depending only on  $p$  and  $k$ , such that

$$D_p \cdot S_p^\mathbb{T} = c_{p,k} [I_{N-p} \quad 0_{(N-p) \times p}], \quad \forall p = 0, 1, \dots, k. \tag{41}$$

Clearly, this result holds for  $p = 0$  with  $c_{0,k} = 1$ . Given  $p \geq 1$ , and assuming that (41) holds for  $(p - 1)$ , it follows from (7)-(9), (36), (38), (40), and the induction hypothesis that

$$\begin{aligned} D_p \cdot S_p^\mathbb{T} &= \left( \widetilde{\Delta}_{k-p}^{-1} \widetilde{D}^{(N-p)} D_{p-1} \right) \cdot \left( S_{p-1}^\mathbb{T} (\widehat{S}^{(N)})^\mathbb{T} \Delta_{k-p} \right) \\ &= c_{p-1,k} \widetilde{\Delta}_{k-p}^{-1} \widetilde{D}^{(N-p)} [I_{N-(p-1)} \quad 0_{(N-(p-1)) \times (p-1)}] (\widehat{S}^{(N)})^\mathbb{T} \Delta_{k-p} \\ &= c_{p-1,k} \widetilde{\Delta}_{k-p}^{-1} \widetilde{D}^{(N-p)} \left[ (\widehat{S}^{(N-(p-1))})^\mathbb{T} \quad \mathbf{1}_{(N-(p-1)) \times (p-1)} \right] \Delta_{k-p} \\ &= -c_{p-1,k} \widetilde{\Delta}_{k-p}^{-1} [I_{N-p} \quad 0_{(N-p) \times p}] \Delta_{k-p} \\ &= c_{p,k} \widetilde{\Delta}_{k-p}^{-1} [\widetilde{\Delta}_{k-p} \quad 0_{(N-p) \times p}] \\ &= c_{p,k} [I_{N-p} \quad 0_{(N-p) \times p}], \end{aligned}$$

where  $c_{p,k} := -(k - p)c_{p-1,k}$  for  $p = 1, \dots, k - 1$ , and  $c_{k,k} := -c_{k-1,k}$ . This gives us (41).

Recalling that  $N = K_n + k - 1$ , it is easy to see via (9) and (33), that

$$(D_0)_{\alpha\bullet} \cdot F_{\alpha,0}^\mathbb{T} = (I_N)_{\alpha\bullet} (I_N)_{\bullet(\bar{\alpha} \cup \{K_n, K_n+1, \dots, N\})} = 0 \in \mathbb{R}^{|\alpha| \times q_\alpha}.$$

Moreover, it follows from the discussions below (40) that  $F_{\alpha,k} = Q \cdot F_{\alpha,0} \cdot S_k$  for a suitable matrix  $Q$ . Combining the above results, we have

$$\begin{aligned}
 (D_k)_{\alpha\bullet} \cdot F_{\alpha,k}^T &= \left( (I_{K_n-1})_{\alpha\bullet} D_k \right) \cdot \left( Q F_{\alpha,0} S_k \right)^T \\
 &= (I_{K_n-1})_{\alpha\bullet} \cdot D_k \cdot S_k^T \cdot F_{\alpha,0}^T \cdot Q^T \\
 &= c_{k,k} (I_{K_n-1})_{\alpha\bullet} [I_{K_n-1} \quad 0_{(K_n-1) \times k}] F_{\alpha,0}^T \cdot Q^T \\
 &= c_{k,k} (I_N)_{\alpha\bullet} \cdot F_{\alpha,0}^T \cdot Q^T = c_{k,k} (D_0)_{\alpha\bullet} \cdot F_{\alpha,0}^T \cdot Q^T \\
 &= 0.
 \end{aligned}$$

Since  $F_{\alpha,k}$  has full row rank, the  $q_\alpha$  columns of  $F_{\alpha,k}^T$  are linearly independent. Additionally, since  $D_k$  is of full row rank, as indicated after (9), so is  $(D_k)_{\alpha\bullet}$ . Therefore,  $\text{rank}[(D_k)_{\alpha\bullet}] = |\alpha|$  and the null space of  $(D_k)_{\alpha\bullet}$  has dimension  $(K_n + k - 1 - |\alpha|)$ , which is equal to  $q_\alpha$  in light of  $|\alpha| + |\bar{\alpha}| = K_n - 1$ . Therefore the columns of  $F_{\alpha,k}^T$  form a basis for the null space of  $(D_k)_{\alpha\bullet}$ .  $\square$

The above proposition shows that for any knot sequent  $\mathbf{t} \in \mathcal{T}_{K_n}$  and any index set  $\alpha$ , the matrix  $F_{\alpha,k,\mathbf{t}}$  defined in (39) is a suitable choice for  $F_\alpha$  that determines the linear piece  $\hat{b}_{P,\mathbf{t}}^\alpha$  associated with  $\alpha$  shown in (23). Hence we conclude this subsection by setting

$$F_\alpha := F_{\alpha,k} = F_{\alpha,k,\mathbf{t}}. \tag{42}$$

### 7.3. Properties of the matrix $F_\alpha \hat{X}^T$

As discussed in Section 7.1, the proof of the uniform Lipschitz property boils down to establishing certain uniform bounds in the  $\ell_\infty$ -norm, including bounds for  $\|(\Xi'_\alpha F_\alpha \Lambda F_\alpha^T)^{-1}\|_\infty$ , where  $\Xi'_\alpha$  is a diagonal matrix with positive diagonal entries to be specified later. Recall that  $\Lambda := \Lambda_{K_n,P,\mathbf{t}} = K_n \cdot \hat{X}^T \Theta \hat{X} \in \mathbb{R}^{N \times N}$  for a design point sequence  $P$  and a knot sequence  $\mathbf{t} \in \mathcal{T}_{K_n}$ ; see the definition above (18). Note that the design matrix  $\hat{X}$  depends on  $P$  and  $\mathbf{t}$ , and

$$F_\alpha \Lambda F_\alpha^T = K_n \cdot (F_\alpha \hat{X}^T) \cdot \Theta \cdot (F_\alpha \hat{X}^T)^T.$$

Therefore in order to study the matrix  $F_\alpha \Lambda F_\alpha^T$ , it is essential to know about  $F_\alpha \hat{X}^T$ . In Proposition 7.2, we determine the values of most of the entries of  $F_\alpha \hat{X}^T$  in terms of certain B-splines. For this purpose, we first establish a technical lemma, which relates the B-spline design matrix  $\hat{X}$  to certain truncated power functions.

Let  $P = (x_i)_{i=0}^n$  be a sequence of design points on the interval  $[0, 1]$ . For each  $i = 0, 1, \dots, n$ , consider the truncated power function  $\varphi_i : [0, 1] \rightarrow \mathbb{R}$  given by

$$\varphi_i(x) := [(x - x_i)_+]^{k-1}. \tag{43}$$

In what follows,  $\varphi_i[s_0, s_1, \dots, s_p]$  denotes the  $p$ th order Newton divided difference of  $\varphi_i$  [8, page 3] at the points  $s_0, s_1, \dots, s_p \in [0, 1]$ . The following lemma relates  $\varphi_i$  and  $\hat{X}$ .

**Lemma 7.1.** *Let  $\mathbf{t}' := (t'_j)_{j=0}^K$  be a sequence of  $(K + 1)$  distinct knots on  $[0, 1]$  with  $t'_0 = 0, t'_K = 1$ , and the usual extension. Let  $P := (x_i)_{i=0}^n$  be a sequence*

of distinct design points with  $x_0 = 0$  and  $x_n = 1$ , so that the design matrix  $\widehat{X}$  is given by  $(\widehat{X}^T)_{j(i+1)} = B_{j,k}^{\mathbf{t}'}(x_i)$  for each  $i$  and  $j$ . For any  $i = 1, \dots, n - 1$ ,  $j = 1, \dots, N := K + k - 1$ , and  $p = 0, 1, \dots, k$ ,

$$\left( S_{p,\mathbf{t}'} \widehat{X}^T \right)_{j(i+1)} = \begin{cases} (\Delta_{k-p,\mathbf{t}'} )_{jj} \cdot \varphi_i[t'_{j-k+p}, \dots, t'_j] & j = 1, \dots, N - p, \\ \sum_{\ell=0}^{j-N+p-1} \frac{\varphi_i^{(k-p+\ell)}(1)}{(k-p+\ell)!} \binom{j-N+p-1}{\ell} & \text{otherwise,} \end{cases} \tag{44}$$

where the matrix  $S_{p,\mathbf{t}'}$  is defined in (40),

*Proof.* In what follows, we drop the subscript  $\mathbf{t}'$  in  $S_{p,\mathbf{t}'}$  and  $\Delta_{p,\mathbf{t}'}$  to simplify notation. Fix  $i = 1, \dots, n - 1$  and  $j = 1, \dots, N$ . We proceed by induction on  $p$ .

Consider  $p = 0$  first. By the definition of  $\widehat{X}$  and [8, page 87], we have

$$\begin{aligned} (S_0 \widehat{X}^T)_{j(i+1)} &= (\widehat{X}^T)_{j(i+1)} = B_{k,j}^{\mathbf{t}'}(x_i) = (t'_j - t'_{j-k}) \cdot \varphi_i[t'_{j-k}, \dots, t'_j] \\ &= (\Delta_k)_{jj} \cdot \varphi_i[t'_{j-k}, \dots, t'_j]. \end{aligned}$$

Hence, (44) holds for  $p = 0$ .

Fix  $p = 1, \dots, k$ , and assume that the result holds for  $(p - 1)$ . First, note that since  $x_i > 0$  for all  $i = 1, \dots, n$ , we have, in view of (43), that for each  $q \in \mathbb{Z}_+$ , the  $q$ -th derivative  $\varphi_i^{(q)}(0) = 0$ . By [8, (viii) on page 6], we have, for all  $p = 0, 1, \dots, k$ ,

$$\varphi_i[t'_{p-k}, \dots, t'_0] = \varphi_i[\underbrace{0, \dots, 0}_{(k-p+1) \text{ copies}}] = \frac{\varphi_i^{(k-p)}(0)}{(k-p)!} = 0. \tag{45}$$

We consider the following two cases.

**Case 1:**  $j = 1, \dots, N - p$ . In this case, by the definition of  $\Delta_{k-p+1}$  in (36), the structure of  $\widehat{S}^{(N)}$  in (38) and  $S_{p-1}$  in (40), the induction hypothesis, and (45), we have

$$\begin{aligned} \left( S_p \widehat{X}^T \right)_{j(i+1)} &= \left( \Delta_{k-p} \widehat{S}^{(N)} S_{p-1} \widehat{X}^T \right)_{j(i+1)} \\ &= (\Delta_{k-p})_{jj} (\widehat{S}^{(N)})_{j\bullet} (S_{p-1} \widehat{X}^T)_{\bullet(i+1)} \\ &= (\Delta_{k-p})_{jj} \sum_{\ell=1}^j (S_{p-1} \widehat{X}^T)_{\ell(i+1)} \\ &= (\Delta_{k-p})_{jj} \sum_{\ell=1}^j (\Delta_{k-p+1})_{\ell\ell} \cdot \varphi_i[t'_{\ell-k+p-1}, \dots, t'_\ell] \\ &= (\Delta_{k-p})_{jj} \sum_{\ell=1}^j (t'_\ell - t'_{\ell-k+p-1}) \cdot \varphi_i[t'_{\ell-k+p-1}, \dots, t'_\ell] \tag{46} \\ &= (\Delta_{k-p})_{jj} \sum_{\ell=1}^j \left( \varphi_i[t'_{\ell-k+p}, \dots, t'_\ell] - \varphi_i[t'_{\ell-k+p-1}, \dots, t'_{\ell-1}] \right) \end{aligned}$$

$$\begin{aligned}
 &= (\Delta_{k-p})_{jj} \cdot \left( \varphi_i[t'_{j-k+p}, \dots, t'_j] - \varphi_i[t'_{p-k}, \dots, t'_0] \right) \\
 &= (\Delta_{k-p})_{jj} \cdot \varphi_i[t'_{j-k+p}, \dots, t'_j],
 \end{aligned}$$

where the derivation of (46) follows from [8, eqn.(20) on page 9]. Thus (44) holds in Case 1.

**Case 2:**  $j = N - p + 1, \dots, N$ . We show that (44) holds for  $p$  in this case by induction on  $j$ .

Consider  $j = N - p + 1$  first. Similar to the argument for Case 1, we use  $(\Delta_{k-p})_{jj} = 1$  for  $j = N - p + 1, \dots, N$  and [8, eqn.(20) on page 9], to obtain

$$\begin{aligned}
 &\left( S_p \widehat{X}^T \right)_{(N-p+1)(i+1)} \\
 &= (\Delta_{k-p} \widehat{S}^{(N)} S_{p-1} \widehat{X}^T)_{(N-p+1)(i+1)} \\
 &= (\Delta_{k-p})_{(N-p+1)(N-p+1)} \cdot (\widehat{S}^{(N)})_{(N-p+1)\bullet} \cdot (S_{p-1} \widehat{X}^T)_{\bullet(i+1)} \\
 &= \sum_{\ell=1}^{N-p+1} (S_{p-1} \widehat{X}^T)_{\ell(i+1)} \\
 &= \sum_{\ell=1}^{N-p+1} (\Delta_{k-p+1})_{\ell\ell} \cdot \varphi_i[t'_{\ell-k+p-1}, \dots, t'_\ell] \\
 &= \sum_{\ell=1}^{N-p+1} \left( \varphi_i[t'_{\ell-k+p}, \dots, t'_\ell] - \varphi_i[t'_{\ell-k+p-1}, \dots, t'_{\ell-1}] \right) \\
 &= \varphi_i[t'_{N-k+1}, \dots, t'_{N-p+1}] - \varphi_i[t'_{p-k}, \dots, t'_0] \\
 &= \varphi_i[\underbrace{1, \dots, 1}_{(k-p+1) \text{ copies}}] = \frac{\varphi_i^{(k-p)}(1)}{(k-p)!},
 \end{aligned}$$

where the last equality follows from [8, (viii) on page 6] and the fact that for any  $i = 1, \dots, n - 1$ ,  $x_i < 1$  such that  $\varphi_i(\cdot)$  is smooth at  $x = 1$ . Hence (44) holds for  $p$  in Case 2, when  $j = N - p + 1$ .

Now, fix  $j = N - p + 2, \dots, N$ , and assume that (44) holds for  $(j - 1)$ . Since  $(\Delta_{k-p})_{jj} = 1$  for  $j = N - p + 1, \dots, N$ , we obtain via (40) that

$$\begin{aligned}
 &\left( S_p \widehat{X}^T \right)_{j(i+1)} \\
 &= (\Delta_{k-p} \widehat{S}^{(N)} S_{p-1} \widehat{X}^T)_{j(i+1)} \\
 &= (\Delta_{k-p})_{jj} (\widehat{S}^{(N)})_{j\bullet} (S_{p-1} \widehat{X}^T)_{\bullet(i+1)} \\
 &= \sum_{\ell=1}^j (S_{p-1} \widehat{X}^T)_{\ell(i+1)} = \sum_{\ell=1}^{j-1} (S_{p-1} \widehat{X}^T)_{\ell(i+1)} + (S_{p-1} \widehat{X}^T)_{j(i+1)} \\
 &= (\Delta_{k-p})_{(j-1)(j-1)} (\widehat{S}^{(N)})_{(j-1)\bullet} (S_{p-1} \widehat{X}^T)_{\bullet(i+1)} + (S_{p-1} \widehat{X}^T)_{j(i+1)} \\
 &= (\Delta_{k-p} \cdot \widehat{S}^{(N)} \cdot S_{p-1} \widehat{X}^T)_{(j-1)(i+1)} + (S_{p-1} \widehat{X}^T)_{j(i+1)}
 \end{aligned}$$

$$= (S_p \widehat{X}^T)_{(j-1)(i+1)} + (S_{p-1} \widehat{X}^T)_{j(i+1)}.$$

By the respective induction hypotheses on  $j$  and  $p$ , we further have

$$\begin{aligned} (S_p \widehat{X}^T)_{j(i+1)} &= \sum_{\ell=0}^{j-N+p-2} \frac{\varphi_i^{(k-p+\ell)}(1)}{(k-p+\ell)!} \binom{j-N+p-2}{\ell} \\ &\quad + \sum_{\ell=0}^{j-N+p-2} \frac{\varphi_i^{(k-p+1+\ell)}(1)}{(k-p+1+\ell)!} \binom{j-N+p-2}{\ell} \\ &= \sum_{\ell=0}^{j-N+p-2} \frac{\varphi_i^{(k-p+\ell)}(1)}{(k-p+\ell)!} \binom{j-N+p-2}{\ell} \\ &\quad + \sum_{\ell=1}^{j-N+p-1} \frac{\varphi_i^{(k-p+\ell)}(1)}{(k-p+\ell)!} \binom{j-N+p-2}{\ell-1} \\ &= \frac{\varphi_i^{(k-p)}(1)}{(k-p)!} + \sum_{\ell=1}^{j-N+p-2} \frac{\varphi_i^{(k-p+\ell)}(1)}{(k-p+\ell)!} \left[ \binom{j-N+p-2}{\ell} + \binom{j-N+p-2}{\ell-1} \right] \\ &\quad + \frac{\varphi_i^{(k+j-N-1)}(1)}{(k+j-N-1)!} = \sum_{\ell=0}^{j-N+p-1} \frac{\varphi_i^{(k-p+\ell)}(1)}{(k-p+\ell)!} \binom{j-N+p-1}{\ell}. \end{aligned}$$

Hence, (44) holds for  $p$  in Case 2 by induction. This completes the proof.  $\square$

Based on Lemma 7.1, we show next that most of the entries of  $F_\alpha \widehat{X}^T$  are characterized by certain B-splines defined by the knot sequence  $V_{\alpha, \mathbf{t}}$  (or simply  $V$ ).

**Proposition 7.2.** Fix a knot sequence  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and an index set  $\alpha$  defined in (21). Let  $P := (x_i)_{i=0}^n$  be a sequence of distinct design points with  $x_0 = 0$  and  $x_n = 1$ , so that the design matrix  $\widehat{X}$  is given by  $(\widehat{X}^T)_{j(i+1)} = B_{k,j}^{\mathbf{t}}(x_i)$  for each  $i$  and  $j$ . Let  $V := V_{\alpha, \mathbf{t}}$  be the knot sequence given by (34) and (35), and  $\{B_{k,j}^V\}_{j=1}^{q_\alpha}$  be the B-splines of order  $k$  with knot sequence  $V$  and the usual extension, where  $q_\alpha = |\overline{\alpha}| + k$ . For each  $i = 1, \dots, n-1$ ,

$$(F_\alpha \widehat{X}^T)_{\bullet(i+1)} = (F_{\alpha, k, \mathbf{t}} \widehat{X}^T)_{\bullet(i+1)} = \left( B_{k,1}^V(x_i), B_{k,2}^V(x_i), \dots, B_{k,q_\alpha}^V(x_i) \right)^T. \tag{47}$$

*Proof.* Recall that  $N = K_n + k - 1$ . Let  $\mathbf{t} = (t_j)_{j=0}^{K_n}$ . Setting  $p = k$ , we have via Lemma 7.1 that for each  $i = 1, \dots, n-1$ ,

$$(S_{k, \mathbf{t}} \widehat{X}^T)_{j(i+1)} = \begin{cases} \varphi_i(t_j) & \text{if } j = 1, \dots, K_n - 1, \\ \sum_{\ell=0}^{j-K_n} \frac{\varphi_i^{(\ell)}(1)}{\ell!} \binom{j-K_n}{\ell} & \text{if } j = K_n, \dots, N. \end{cases}$$

Therefore, in view of the definition of  $F_{\alpha, 0, \mathbf{t}}$  (33) and that of  $\tau_j = \tau_{\alpha, \mathbf{t}, j}$  (34),

$$\left(F_{\alpha,0,t} S_{k,t} \widehat{X}^T\right)_{j(i+1)} = \begin{cases} \varphi_i(\tau_j) & j = 1, \dots, |\bar{\alpha}|, \\ \sum_{\ell=0}^{j-|\bar{\alpha}|-1} \frac{\varphi_i^{(\ell)}(1)}{\ell!} \binom{j-|\bar{\alpha}|-1}{\ell} & j = |\bar{\alpha}| + 1, \dots, q_\alpha. \end{cases} \quad (48)$$

For the knot sequence  $V := V_{\alpha,t} = \{\tau_0, \tau_1, \dots, \tau_{|\bar{\alpha}|+1}\}$  of  $(|\bar{\alpha}| + 2)$  knots with  $\tau_0 = 0$  and  $\tau_{|\bar{\alpha}|+1} = 1$  given in (35) and the design point sequence  $P = (x_i)_{i=0}^n$ , let  $\widehat{X}_V$  be the design matrix given by  $((\widehat{X}_V)^T)_{j(i+1)} = B_{k,j}^V(x_i)$  for each  $i = 0, 1, \dots, n$  and  $j = 1, \dots, q_\alpha$ . Applying Lemma 7.1 with  $\mathbf{t}' = V$  and  $p = k$ , we see via (48) that for each  $i = 1, \dots, n - 1$ ,

$$S_{k,V} \cdot ((\widehat{X}_V)^T)_{\bullet(i+1)} = (S_{k,V} (\widehat{X}_V)^T)_{\bullet(i+1)} = \left(F_{\alpha,0,t} S_{k,t} \widehat{X}^T\right)_{\bullet(i+1)}. \quad (49)$$

Moreover, it follows from (37)-(40) that

$$\begin{aligned} F_{\alpha,k,t} \widehat{X}^T &= \left(\widehat{D}^{(q_\alpha)} \cdot \Xi_{k-1,V} \cdots \widehat{D}^{(q_\alpha)} \cdot \Xi_{1,V} \cdot \widehat{D}^{(q_\alpha)} \Xi_{0,V}\right) \cdot \left(F_{\alpha,0,t} S_{k,t} \widehat{X}^T\right) \\ &= \left(\Delta_{0,V} \cdot \widehat{S}^{(q_\alpha)} \cdot \Delta_{1,V} \cdot \widehat{S}^{(q_\alpha)} \cdots \Delta_{k-1,V} \cdot \widehat{S}^{(q_\alpha)} \cdot S_{0,V}\right)^{-1} \cdot \left(F_{\alpha,0,t} S_{k,t} \widehat{X}^T\right) \\ &= (S_{k,V})^{-1} \cdot \left(F_{\alpha,0,t} S_{k,t} \widehat{X}^T\right). \end{aligned}$$

Hence  $S_{k,V} F_{\alpha,k,t} \widehat{X}^T = F_{\alpha,0,t} S_{k,t} \widehat{X}^T$  and for each  $i = 1, \dots, n - 1$ ,  $S_{k,V} \cdot (F_{\alpha,k,t} \widehat{X}^T)_{\bullet(i+1)} = (F_{\alpha,0,t} S_{k,t} \widehat{X}^T)_{\bullet(i+1)}$ . Since the matrix  $S_{k,V}$  is invertible, we have, for each  $i = 1, \dots, n - 1$ ,  $(F_{\alpha,k,t} \widehat{X}^T)_{\bullet(i+1)} = ((\widehat{X}_V)^T)_{\bullet(i+1)}$ , via (49).

Therefore by the definition of  $\widehat{X}_V$  and (42), we obtain

$$(F_{\alpha} \widehat{X}^T)_{\bullet(i+1)} = (F_{\alpha,k,t} \widehat{X}^T)_{\bullet(i+1)} = \left(B_{k,1}^V(x_i), B_{k,2}^V(x_i), \dots, B_{k,q_\alpha}^V(x_i)\right)^T$$

for each  $i = 1, \dots, n - 1$ . □

The following proposition reveals further properties of  $F_{\alpha,k,t}$  (or simply  $F_\alpha$ ) based on Proposition 7.2. In particular, it interprets the result in (47) in terms of B-spline functions and shows that (47) holds for  $i = 0, n$  using the continuity of B-splines.

**Proposition 7.3.** *For any given  $k, K_n \in \mathbb{N}$ , a knot sequence  $\mathbf{t} \in \mathcal{T}_{K_n}$ , an index set  $\alpha$  defined in (21), and the knot sequence  $V := V_{\alpha,t}$  given by (35), the B-splines  $\{B_{k,\ell}^V\}_{\ell=1}^{q_\alpha}$  and  $\{B_{k,j}^t\}_{j=1}^N$  satisfy*

$$\sum_{j=1}^N (F_\alpha)_{\ell j} B_{k,j}^t(x) = B_{k,\ell}^V(x), \quad \forall x \in [0, 1], \quad (50)$$

for each  $\ell = 1, \dots, q_\alpha$ , where  $F_\alpha := F_{\alpha,k,t}$  is given by (39).

*Proof.* Let  $x_*$  be an arbitrary but fixed real number in the open interval  $(0, 1)$ . Consider a sequence of design points  $P = (x_i)_{i=0}^n$  with  $x_0 = 0$  and  $x_n = 1$



such that  $x_{i_*} = x_*$  for some  $i_* \in \{1, \dots, n-1\}$ . As before, let  $\widehat{X}$  be the design matrix defined by the B-splines  $B_{k,j}^{\mathbf{t}}$  and  $P$ . By Proposition 7.2, we have  $(F_\alpha \widehat{X}^T)_{\ell(i_*+1)} = B_{k,\ell}^V(x_*)$  for any  $\ell = 1, \dots, q_\alpha$ . In light of  $(F_\alpha \widehat{X}^T)_{\ell(i_*+1)} = \sum_{j=1}^N (F_\alpha)_{\ell j} B_{k,j}^{\mathbf{t}}(x_*)$ , we deduce that for any  $\ell = 1, \dots, q_\alpha$ ,

$$\sum_{j=1}^N (F_\alpha)_{\ell j} B_{k,j}^{\mathbf{t}}(x_*) = (F_\alpha \widehat{X}^T)_{\ell(i_*+1)} = B_{k,\ell}^V(x_*).$$

This shows that (50) holds for all  $x \in (0, 1)$ . Since  $B_{k,j}^{\mathbf{t}}(\cdot)$  and  $B_{k,\ell}^V(\cdot)$  are right continuous on  $[0, 1)$  for any  $j, \ell, \mathbf{t}$ , and  $\alpha$ , we conclude that (50) holds for  $x = 0$ . Finally, the continuity of  $B_{k,j}^{\mathbf{t}}(\cdot)$  and  $B_{k,\ell}^V(\cdot)$  on  $[t_{K_n-1}, t_{K_n}]$  also implies that (50) holds at  $x = 1$ . Setting  $x = x_0, x_n$  demonstrates that (47) holds for  $i = 0, n$ , via the definition of  $\widehat{X}$ .  $\square$

#### 7.4. Preliminary uniform bounds

This subsection establishes uniform bounds and uniform approximation error bounds of several of the constructed matrices.

We first derive uniform bounds for  $\|F_\alpha^T\|_\infty$  and  $\|\frac{k}{K_n} \Xi_{k,V} F_\alpha\|_\infty$  using Proposition 7.3. We introduce more notation. Let  $\mathbf{e}_\ell$  be the  $\ell$ th standard basis vector in the Euclidean space, i.e.,  $[\mathbf{e}_\ell]_i = \delta_{\ell i}$ . Moreover, for a given vector  $v = (v_1, \dots, v_j) \in \mathbb{R}^j$ , the number of sign changes of  $v$  is defined as the largest integer  $r_v \in \mathbb{Z}_+$  such that for some  $1 \leq j_1 < \dots < j_{r_v} \leq j$ ,  $v_{j_i} \cdot v_{j_{i+1}} < 0$  for each  $i = 1, \dots, r_v$  [8, page 138]. Clearly,  $\mathbf{e}_\ell$  has zero sign changes for each  $\ell$ .

**Proposition 7.4.** Fix  $k, K_n \in \mathbb{N}$ ,  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and an index set  $\alpha$  defined in (21). Let  $F_\alpha := F_{\alpha,k,\mathbf{t}}$  be given by (42) and  $V := V_{\alpha,\mathbf{t}}$  be given by (35). Then the following hold:

- (1)  $F_\alpha$  is a nonnegative matrix,  $\|F_\alpha^T\|_\infty = 1$ , and
- (2)  $\|\frac{k}{K_n} \Xi_{k,V} F_\alpha\|_\infty \leq \frac{k}{c_{\mathbf{t},1}}$ .

*Proof.* (1) Observe that the knot sequence  $\mathbf{t}$  either can be formed by inserting additional knots into the knot sequence  $V$  (if  $\alpha \neq \emptyset$ ), or is the same as  $V$  (if  $\alpha = \emptyset$ ). Fix  $\ell = 1, \dots, q_\alpha$ . By Proposition 7.3, we see that

$$\sum_{j=1}^N (F_\alpha)_{\ell j} B_{k,j}^{\mathbf{t}}(x) = B_{k,\ell}^V(x) = \sum_{i=1}^{q_\alpha} [\mathbf{e}_\ell]_i B_{k,i}^V(x), \quad \forall x \in [0, 1].$$

Since  $\mathbf{e}_\ell$  has zero sign changes, we deduce via [8, Lemma 27, Chapter XI] that  $(F_\alpha)_{\ell \bullet}$  has zero sign changes. This shows that either  $(F_\alpha)_{\ell \bullet} \geq 0$  or  $(F_\alpha)_{\ell \bullet} \leq 0$ . In view of the nonnegativity of  $B_{k,j}^{\mathbf{t}}$  and (50), the latter implies that  $B_{k,\ell}^V(x) \leq 0$  for all  $x \in [0, 1]$ . But this contradicts the fact that  $B_{k,\ell}^V(x) > 0$  when  $x$  is in the interior of the support of  $B_{k,\ell}^V$ . Therefore,  $F_\alpha$  is a nonnegative matrix.

Next, define  $b \in \mathbb{R}^N$  such that  $b := F_\alpha^T \cdot \mathbf{1}_{q_\alpha}$ , i.e.,  $b_j = \sum_{\ell=1}^{q_\alpha} (F_\alpha)_{\ell,j}$  for each  $j = 1, \dots, N$ . By the nonnegativity of  $F_\alpha$ , we have  $\|F_\alpha^T\|_\infty = \|F_\alpha^T \cdot \mathbf{1}_{q_\alpha}\|_\infty = \|b\|_\infty$ . Further, in view of item (i) in Section 3 and Proposition 7.3,

$$\begin{aligned} \sum_{j=1}^N b_j B_{k,j}^{\mathbf{t}}(x) &= \sum_{j=1}^N \sum_{\ell=1}^{q_\alpha} (F_\alpha)_{\ell,j} B_{k,j}^{\mathbf{t}}(x) = \sum_{\ell=1}^{q_\alpha} \sum_{j=1}^N (F_\alpha)_{\ell,j} B_{k,j}^{\mathbf{t}}(x) \\ &= \sum_{\ell=1}^{q_\alpha} B_{k,\ell}^V(x) = 1, \quad \forall x \in [0, 1]. \end{aligned}$$

Since the  $B_{k,j}^{\mathbf{t}}$ 's form a partition of unity and a basis of the space of order  $k$  splines with the knot sequence  $\mathbf{t}$ , we must have  $b = \mathbf{1}_N$  by the above display. Thus  $\|F_\alpha^T\|_\infty = \|b\|_\infty = 1$ . This completes the proof of statement (1).

(2) It follows from (50) that for each  $\ell = 1, \dots, q_\alpha$ ,  $\sum_{j=1}^N (F_\alpha)_{\ell,j} B_{k,j}^{\mathbf{t}}(x) = 0$  except on the support of  $B_{k,\ell}^V$  given by  $[\tau_{\ell-k}, \tau_\ell] := [t_{i_{\ell-k}}, t_{i_\ell}]$  for some  $i_\ell$  and  $i_{\ell-k}$ . Note that  $i_\ell := \ell$  if  $\ell < 0$ , and  $i_{|\bar{\alpha}|+p} := K_n + p - 1$  for any  $p = 1, \dots, k$ ; furthermore,  $t_{i_\ell} = 0$  if  $\ell \leq 0$  and  $t_{i_\ell} = 1$  for  $\ell \geq |\bar{\alpha}| + 1$ . Additionally, for any  $r = 1, \dots, K_n$ , the B-splines  $B_{k,j}^{\mathbf{t}}$  that are not identically zero on  $[t_{r-1}, t_r]$  are linearly independent when restricted to  $[t_{r-1}, t_r]$ . Hence, if  $\sum_{j=1}^N (F_\alpha)_{\ell,j} B_{k,j}^{\mathbf{t}}(x) = B_{k,\ell}^V(x) = 0, \forall x \in [t_{r-1}, t_r]$  for some  $r$ , then  $(F_\alpha)_{\ell,j} = 0$  for each  $j = r, r + 1, \dots, r + k - 1$ . This, along with the fact that  $B_{k,\ell}^V(x) = 0$  except on  $[t_{i_{\ell-k}}, t_{i_\ell}]$ , shows that  $(F_\alpha)_{\ell,j} = 0$  for all  $j = 1, 2, \dots, i_{\ell-k} + k - 1$  and  $j = i_\ell + 1, i_\ell + 2, \dots, N$ . Moreover, by the nonnegativity of  $F_\alpha$  and  $\|F_\alpha^T\|_\infty = 1$ , we see that all nonzero entries of  $F_\alpha$  are less than or equal to one. Therefore, by the definition of  $\Xi_{k,V}$  (cf. (37)), we have, for each  $\ell = 1, \dots, q_\alpha$ ,

$$\begin{aligned} \left\| \left( \frac{k}{K_n} \Xi_{k,V} F_\alpha \right)_{\ell \bullet} \right\|_\infty &= \frac{1}{K_n} \cdot \frac{k}{t_{i_\ell} - t_{i_{\ell-k}}} \sum_{j=i_{\ell-k}+k}^{i_\ell} (F_\alpha)_{\ell,j} \\ &\leq \frac{1}{K_n} \cdot \frac{k}{t_{i_\ell} - t_{i_{\ell-k}}} \cdot (i_\ell - i_{\ell-k} - k + 1). \end{aligned}$$

By virtue of the discussions before (24), we also have

$$t_{i_\ell} - t_{i_{\ell-k}} \geq \frac{c_{t,1}}{K_n} (i_\ell - i_{\ell-k} - k + 1) > 0.$$

Consequently, we obtain, for each  $\ell = 1, \dots, q_\alpha$ ,  $\left\| \left( \frac{k}{K_n} \Xi_{k,V} F_\alpha \right)_{\ell \bullet} \right\|_\infty \leq k/c_{t,1}$ . This completes the proof of statement (2).  $\square$

Motivated by Proposition 7.4, we define the following diagonal matrix with positive diagonal entries for the given  $\alpha, \mathbf{t}$ , and  $V := V_{\alpha,\mathbf{t}}$ :

$$\Xi'_\alpha := \frac{k}{K_n} \cdot \Xi_{k,V}, \tag{51}$$

where the diagonal matrix  $\Xi_{k,V}$  is given in (37). By Proposition 7.4, both  $\|F_\alpha^T\|_\infty$  and  $\|\Xi'_\alpha F_\alpha\|_\infty$  are uniformly bounded. As discussed in Section 7.1, in order to establish the uniform Lipschitz property, it remains to establish a uniform bound

on  $\|(\Xi'_\alpha F_\alpha \Lambda F_\alpha^T)^{-1}\|_\infty$ . We will demonstrate that  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^T$  approximates a certain B-spline Gramian matrix to obtain such a bound.

Consider the  $k$ th order B-splines  $\{B_{k,j}^V\}_{j=1}^{|\bar{\alpha}|+k}$  corresponding to the knot sequence  $V := V_{\alpha,\mathbf{t}}$  defined in (35), associated with  $\mathbf{t} \in \mathcal{T}_{K_n}$  and any index set  $\alpha$  defined in (21). Recall that  $q_\alpha = |\bar{\alpha}| + k$ . Define the Gramian matrix  $G_{\alpha,\mathbf{t}} \in \mathbb{R}^{q_\alpha \times q_\alpha}$  as

$$(G_{\alpha,\mathbf{t}})_{i,j} := \frac{\langle B_{k,i}^V, B_{k,j}^V \rangle}{\|B_{k,i}^V\|_{L_1}}, \quad \forall i, j = 1, \dots, q_\alpha. \tag{52}$$

Proposition 7.5 given below shows that  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^T$  approximates  $G_{\alpha,\mathbf{t}}$  with a uniform approximation error bound. Its proof relies on the following technical lemma, which characterizes the difference between an integral of a continuous function and its discrete approximation.

**Lemma 7.2.** *Let  $\tilde{n} \in \mathbb{N}$ ,  $[a, b] \subset \mathbb{R}$  with  $a < b$ , and the sequence of points  $(s_i)_{i=0}^{\tilde{n}}$  with  $a = s_0 < s_1 < \dots < s_{\tilde{n}} = b$  be such that  $\max_{i=1, \dots, \tilde{n}} |s_i - s_{i-1}| \leq \varrho$  for some  $\varrho > 0$ . Let a continuous function  $f : [a, b] \rightarrow \mathbb{R}$  be differentiable on  $[a, b]$  except at finitely many points in  $[a, b]$ . Let the positive constant  $\mu$  be such that  $|f'(x)| \leq \mu$  for any  $x \in [a, b]$  where  $f'(x)$  exists. Then*

$$\left| \sum_{i=1}^{\tilde{n}} f(s_{i-1}) (s_i - s_{i-1}) - \int_a^b f(x) dx \right| \leq \frac{3}{2} \mu \varrho (b - a).$$

*Proof.* Fix an arbitrary  $i \in \{1, \dots, \tilde{n}\}$ . Suppose that  $\tilde{s}_1, \dots, \tilde{s}_{\ell-1} \in (s_{i-1}, s_i)$  with  $s_{i-1} := \tilde{s}_0 < \tilde{s}_1 < \tilde{s}_2 < \dots < \tilde{s}_{\ell-1} < \tilde{s}_\ell := s_i$  are the only points where  $f$  is not differentiable on the interval  $(s_{i-1}, s_i)$ . It follows from the continuity of  $f$  and the Mean-value Theorem that for each  $j = 1, \dots, \ell$ , there exists  $\xi_j \in (\tilde{s}_{j-1}, \tilde{s}_j)$  such that

$$f(\tilde{s}_j) = f(\tilde{s}_{j-1}) + f'(\xi_j)(\tilde{s}_j - \tilde{s}_{j-1}) = f(s_{i-1}) + \sum_{r=1}^j f'(\xi_r)(\tilde{s}_r - \tilde{s}_{r-1}).$$

Since  $f$  is continuous and piecewise differentiable, we have

$$\begin{aligned} & \left| \int_{s_{i-1}}^{s_i} f(x) dx - (s_i - s_{i-1}) f(s_{i-1}) \right| \\ &= \left| \sum_{j=1}^{\ell} \int_{\tilde{s}_{j-1}}^{\tilde{s}_j} \left[ f(\tilde{s}_{j-1}) + f'(\xi_x)(x - \tilde{s}_{j-1}) \right] dx - (s_i - s_{i-1}) f(s_{i-1}) \right| \\ &\leq \left| \sum_{j=1}^{\ell} \left[ f(s_{i-1}) + \sum_{r=1}^{j-1} f'(\xi_r)(\tilde{s}_r - \tilde{s}_{r-1}) \right] (\tilde{s}_j - \tilde{s}_{j-1}) - (s_i - s_{i-1}) f(s_{i-1}) \right| \\ &\quad + \left| \sum_{j=1}^{\ell} \int_{\tilde{s}_{j-1}}^{\tilde{s}_j} f'(\xi_x)(x - \tilde{s}_{j-1}) dx \right| \end{aligned}$$

$$\begin{aligned} &\leq \left| \sum_{j=1}^{\ell} f'(\xi_j)(\tilde{s}_j - \tilde{s}_{j-1})(s_i - \tilde{s}_j) \right| + \frac{\mu}{2} \sum_{j=1}^{\ell} (\tilde{s}_j - \tilde{s}_{j-1})^2 \\ &\leq \frac{3\mu}{2}(s_i - s_{i-1})^2 \leq \frac{3\mu\varrho}{2}(s_i - s_{i-1}). \end{aligned}$$

Consequently,

$$\begin{aligned} \left| \sum_{i=1}^{\tilde{n}} f(s_{i-1})(s_i - s_{i-1}) - \int_a^b f(x) dx \right| &\leq \sum_{i=1}^{\tilde{n}} \left| f(s_{i-1})(s_i - s_{i-1}) - \int_{s_{i-1}}^{s_i} f(x) dx \right| \\ &\leq \sum_{i=1}^{\tilde{n}} \frac{3\mu\varrho}{2}(s_i - s_{i-1}) = \frac{3\mu\varrho}{2}(b - a). \end{aligned}$$

This completes the proof. □

**Proposition 7.5.** *Let  $(K_n)$  be an increasing sequence with  $K_n \rightarrow \infty$  and  $\frac{K_n}{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $F_\alpha := F_{\alpha,k,t}$ ,  $V := V_{\alpha,t}$ ,  $\Xi'_\alpha := (k/K_n) \cdot \Xi_{k,V}$ , and  $\Lambda := \Lambda_{K_n,P,t}$  be defined for  $K_n$ ,  $P \in \mathcal{P}_n$ ,  $t \in \mathcal{T}_{K_n}$  and the index set  $\alpha$ . Then there exists  $n_1 \in \mathbb{N}$ , independent of  $P$ ,  $t$ , and  $\alpha$ , such that for any  $P \in \mathcal{P}_n$  and  $t \in \mathcal{T}_{K_n}$  with  $n \geq n_1$ , and any index set  $\alpha$  defined in (21),*

$$\left\| G_{\alpha,t} - \Xi'_\alpha F_\alpha \Lambda F_\alpha^T \right\|_\infty \leq \frac{16k^3 c_{\omega,2} K_n}{c_{t,1} n}, \quad \forall n \geq n_1.$$

*Proof.* Let  $P = (x_i)_{i=0}^n$ . In light of (6), (37), the definitions of  $\widehat{X}$ ,  $\Theta$ , and  $\Lambda$ , (47), and (51), we have, for any  $j, r = 1, \dots, q_\alpha$ ,

$$\begin{aligned} \left( \Xi'_\alpha F_\alpha \Lambda F_\alpha^T \right)_{j,r} &= \left( \frac{k}{K_n} \Xi_{k,V} F_\alpha \Lambda F_\alpha^T \right)_{j,r} = \frac{k}{t_{i_j} - t_{i_{j-k}}} (F_\alpha \widehat{X}^T)_{j,\bullet} \Theta (F_\alpha \widehat{X}^T)_{r,\bullet}^T \\ &= \|B_{k,j}^V\|_{L_1}^{-1} \sum_{\ell=1}^n B_{k,j}^V(x_{\ell-1}) \cdot B_{k,r}^V(x_{\ell-1}) \cdot (x_\ell - x_{\ell-1}). \end{aligned} \tag{53}$$

Consider  $k = 1$  first. In this case,  $G_{\alpha,t}$  is diagonal (see the summary of B-spline properties in Section 2.2 for the reason), and

$$(G_{\alpha,t})_{j,j} = \|\mathbf{I}_{[t_{i_{j-1}}, t_{i_j}]}\|_{L_1}^{-1} \cdot \left\langle \mathbf{I}_{[t_{i_{j-1}}, t_{i_j}]}, \mathbf{I}_{[t_{i_{j-1}}, t_{i_j}]} \right\rangle = 1.$$

Additionally,  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^T$  is also diagonal in view of the definitions of  $\Lambda$  and  $\widehat{X}$ , (47), and (51). Since  $P \in \mathcal{P}_n$ ,  $V \subseteq t \in \mathcal{T}_{K_n}$ , and  $(\|B_{1,j}^V\|_{L_1})^{-1} = \frac{1}{t_{i_j} - t_{i_{j-1}}} = (\Xi_{1,V})_{j,j}$ , we have, by (53), that for each  $j$ ,

$$\begin{aligned} \left| \left( \Xi'_\alpha F_\alpha \Lambda F_\alpha^T \right)_{j,j} - 1 \right| &= \left| \left( K_n^{-1} \Xi_{1,V} F_\alpha \Lambda F_\alpha^T \right)_{j,j} - 1 \right| \\ &= \left| \sum_{\ell=1}^n \frac{\mathbf{I}_{[t_{i_{j-1}}, t_{i_j}]}(x_{\ell-1})}{t_{i_j} - t_{i_{j-1}}} (x_\ell - x_{\ell-1}) - 1 \right| \leq \frac{2c_{\omega,2} K_n}{c_{t,1} n}. \end{aligned}$$

Thus the proposition holds for  $k = 1$ .

We consider  $k \geq 2$  next. Note that  $B_{k,j}^V(\cdot) \times B_{k,r}^V(\cdot)$  is continuous and differentiable on  $[0, 1]$  except at (at most) finitely many points in  $[0, 1]$ . In light of

items (i) and (ii) in Section 3, as well as the discussion following (14), we have, for any  $x \in [0, 1]$  where the derivative exists,

$$\left| \left( B_{k,j}^V(x) \right)' \right| \leq 2 \max_{\ell} \left| \frac{k-1}{t_{i_\ell} - t_{i_{\ell-k+1}}} B_{k-1,\ell}^V(x) \right| \leq \frac{2(k-1)}{c_{t,1}} K_n.$$

Hence, we also have that

$$\left| \left( B_{k,j}^V(x) B_{k,r}^V(x) \right)' \right| = \left| \left( B_{k,j}^V \right)'(x) B_{k,r}^V(x) + B_{k,j}^V(x) \left( B_{k,r}^V \right)'(x) \right| \leq \frac{4(k-1)}{c_{t,1}} K_n. \quad (54)$$

For a fixed  $j$ , let  $x_{a_j-1}$  and  $x_{b_j-1}$  be the first and last points in  $P$  such that  $x_{a_j-1}, x_{b_j-1} \in [t_{i_{j-k}}, t_{i_j}]$ . Since  $P \in \mathcal{P}_n$ , we have  $|x_{a_j-1} - t_{i_{j-k}}| \leq \frac{c_{\omega,2}}{n}$  and  $|x_{b_j} - t_{i_j}| \leq \frac{c_{\omega,2}}{n}$ . Therefore, by (6), we obtain

$$|x_{b_j} - x_{a_j-1}| \leq |x_{b_j} - t_{i_j}| + |t_{i_j} - t_{i_{j-k}}| + |x_{a_j-1} - t_{i_{j-k}}| \leq k \|B_{k,j}^V\|_{L_1} + \frac{2c_{\omega,2}}{n}. \quad (55)$$

Since  $\frac{K_n}{n} \rightarrow 0$  as  $n \rightarrow \infty$ , we deduce via (6) and the discussion following (14) that there exists  $n_1$  sufficiently large (depending on  $(K_n)$  only) such that

$$\frac{2c_{\omega,2}}{n} \leq \frac{c_{t,1}}{kK_n} \leq \|B_{k,j}^V\|_{L_1}, \quad \forall n \geq n_1. \quad (56)$$

Hence, combining (55) and (56), we have  $|x_{b_j} - x_{a_j-1}| \leq (k+1) \|B_{k,j}^V\|_{L_1}$ ,  $\forall n \geq n_1$ . By this result, (53)-(54), (56) and Lemma 7.2, with  $a := x_{a_j-1}$ ,  $b := x_{b_j}$ ,  $s_i := x_{a_j-1+i}$ ,  $\varrho := c_{\omega,2}/n$ ,  $f(x) := B_{k,j}^V(x) \cdot B_{k,r}^V(x)$  satisfying  $\|f\|_\infty \leq 1$ , and  $\mu := \frac{4(k-1)}{c_{t,1}} K_n$ , we obtain that for any  $j, r = 1, \dots, q_\alpha$ ,

$$\begin{aligned} & \left| \left( \Xi'_\alpha F_\alpha \Lambda F_\alpha^T \right)_{j,r} - (G_{\alpha,t})_{j,r} \right| \\ & \leq \|B_{k,j}^V\|_{L_1}^{-1} \cdot \left| \sum_{\ell=a_j}^{b_j} f(x_{\ell-1})(x_\ell - x_{\ell-1}) - \int_{x_{a_j-1}}^{x_{b_j}} f(t) dt \right| \\ & \quad + \|B_{k,j}^V\|_{L_1}^{-1} \cdot \left( \left| \int_{t_{i_{j-k}}}^{x_{a_j-1}} f(x) dx \right| + \left| \int_{x_{b_j}}^{t_{i_j}} f(x) dx \right| \right) \\ & \leq \|B_{k,j}^V\|_{L_1}^{-1} \cdot \frac{3}{2} \frac{4(k-1)K_n}{c_{t,1}} \frac{c_{\omega,2}}{n} \cdot |x_{b_j} - x_{a_j-1}| \\ & \quad + \|B_{k,j}^V\|_{L_1}^{-1} \cdot (|x_{b_j} - t_{i_j}| + |x_{a_j-1} - t_{i_{j-k}}|) \\ & \leq \|B_{k,j}^V\|_{L_1}^{-1} \cdot \frac{3}{2} \frac{4(k-1)K_n}{c_{t,1}} \frac{c_{\omega,2}}{n} \cdot (k+1) \|B_{k,j}^V\|_{L_1} + \|B_{k,j}^V\|_{L_1}^{-1} \cdot \frac{2c_{\omega,2}}{n} \\ & \leq \frac{6k^2 c_{\omega,2} K_n}{c_{t,1} n} + \frac{2c_{\omega,2} k K_n}{c_{t,1} n} \leq \frac{8k^2 c_{\omega,2} K_n}{c_{t,1} n}, \quad \forall n \geq n_1. \end{aligned}$$

Based on the support of each  $B_{k,j}^V$ , each row of  $G_{\alpha,t}$  and  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^T$  has at most  $(2k-1)$  nonzero entries. Thus, we deduce that for any  $\alpha, \mathbf{t} \in \mathcal{T}_{K_n}$ , and  $P \in \mathcal{P}_n$ ,

$$\left\| G_{\alpha, \mathbf{t}} - \Xi'_\alpha F_\alpha \Lambda F_\alpha^T \right\|_\infty \leq (2k - 1) \cdot \frac{8k^2 c_{\omega, 2} K_n}{c_{t, 1} n} \leq \frac{16k^3 c_{\omega, 2} K_n}{c_{t, 1} n}, \quad \forall n \geq n_1.$$

This yields the uniform approximation error bound. □

The following technical lemma is instrumental to attaining a uniform bound on the inverses of size-varying matrices in the next corollary.

**Lemma 7.3.** *Let  $\{A_i \in \mathbb{R}^{n_i \times n_i} : i \in \mathcal{I}\}$  and  $\{B_i \in \mathbb{R}^{n_i \times n_i} : i \in \mathcal{I}\}$  be two families of invertible matrices for a (possibly infinite) index set  $\mathcal{I}$ , where  $n_i \in \mathbb{N}$  need not be the same for different  $i \in \mathcal{I}$ . Suppose that  $\mu := \sup_{i \in \mathcal{I}} \|A_i^{-1}\|_\infty < \infty$  and  $\sup_{i \in \mathcal{I}} \|A_i - B_i\|_\infty \leq \frac{1}{2\mu}$ . Then for all  $i \in \mathcal{I}$ ,  $\|B_i^{-1}\|_\infty \leq 2\mu$ .*

*Proof.* Define  $C_i := B_i - A_i$  so that  $\|C_i\|_\infty < \frac{1}{2\mu}$  for each  $i \in \mathcal{I}$ . Since  $B_i = A_i + C_i$  and  $A_i$  is invertible, we have  $A_i^{-1}B_i = I + A_i^{-1}C_i$ . Hence, we obtain

$$\|A_i^{-1}C_i\|_\infty \leq \|A_i^{-1}\|_\infty \cdot \|C_i\|_\infty \leq \mu \cdot \frac{1}{2\mu} = \frac{1}{2}, \forall i \in \mathcal{I}.$$

Furthermore,  $\|(I + A_i^{-1}C_i)^{-1}\|_\infty \leq \frac{1}{1 - \|A_i^{-1}C_i\|_\infty} \leq 2, \forall i \in \mathcal{I}$ . Using  $A_i^{-1}B_i = I + A_i^{-1}C_i$  again, we further have that

$$\|B_i^{-1}\|_\infty = \|(I + A_i^{-1}C_i)^{-1} \cdot A_i^{-1}\|_\infty \leq \|(I + A_i^{-1}C_i)^{-1}\|_\infty \cdot \|A_i^{-1}\|_\infty \leq 2\mu,$$

for all  $i \in \mathcal{I}$ . This yields the desired bound for  $B_i^{-1}$ . □

An immediate consequence of Proposition 7.5 and Lemma 7.3 is a uniform bound on the inverse of  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^T$  in the  $\ell_\infty$ -norm. Note that each  $\Xi'_\alpha F_\alpha \Lambda F_\alpha^T$  is invertible, since  $\Lambda$  is positive definite (see Section 3.2),  $F_\alpha$  has full row rank (see Section 7.2), and  $\Xi'_\alpha$  is invertible.

**Corollary 7.1.** *Let  $(K_n)$  be an increasing sequence with  $K_n \rightarrow \infty$  and  $\frac{K_n}{n} \rightarrow 0$  as  $n \rightarrow \infty$ . For any  $P \in \mathcal{P}_n$ ,  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and index set  $\alpha$  defined in (21), let  $F_\alpha := F_{\alpha, k, \mathbf{t}}$ ,  $\Xi'_\alpha := (k/K_n) \Xi_{k, V}$  for the knot sequence  $V := V_{\alpha, \mathbf{t}}$ , and  $\Lambda := \Lambda_{K_n, P, \mathbf{t}}$ . There exists  $n_* \in \mathbb{N}$ , which depends on  $(K_n)$  only, such that for any  $P \in \mathcal{P}_n$  and  $\mathbf{t} \in \mathcal{T}_{K_n}$  with  $n \geq n_*$ , and any index set  $\alpha$ ,*

$$\left\| (\Xi'_\alpha F_\alpha \Lambda F_\alpha^T)^{-1} \right\|_\infty \leq 2\rho_k,$$

where  $\rho_k$  is a positive constant depending on  $k$  only.

*Proof.* Choose arbitrary  $\alpha$ ,  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and  $P \in \mathcal{P}_n$ . By [34, Theorem I] (cf. Theorem 7.1), the Gramian matrix  $G_{\alpha, \mathbf{t}}$  is invertible and there exists a positive constant  $\rho_k$ , independent of  $\alpha$ ,  $\mathbf{t}$ ,  $P$ , such that  $\|(G_{\alpha, \mathbf{t}})^{-1}\|_\infty \leq \rho_k$ . Also, it follows from Proposition 7.5 that  $\|G_{\alpha, \mathbf{t}} - \Xi'_\alpha F_\alpha \Lambda F_\alpha^T\|_\infty \leq \frac{16k^3 c_{\omega, 2} K_n}{c_{t, 1} n}$ , provided that  $n \geq n_1$ . Since  $\frac{K_n}{n} \rightarrow 0$  as  $n \rightarrow \infty$ , we deduce from Lemma 7.3 that there exists  $n_* \in \mathbb{N}$  with  $n_* \geq n_1$  such that for any  $P \in \mathcal{P}_n$  and  $\mathbf{t} \in \mathcal{T}_{K_n}$  with  $n \geq n_*$ , and any index set  $\alpha$ ,  $\left\| (\Xi'_\alpha F_\alpha \Lambda F_\alpha^T)^{-1} \right\|_\infty \leq 2\rho_k$ . □

### 7.5. Proof of Theorem 3.1

We apply the results established in the previous subsections to show the uniform Lipschitz property of  $\widehat{b}_{P,\mathbf{t}}(\cdot)$  stated in Theorem 3.1. Fix the B-spline order  $k \in \mathbb{N}$ . Let the strictly increasing sequence  $(K_n)$  be such that  $K_n \rightarrow \infty$  and  $K_n/n \rightarrow 0$  as  $n \rightarrow \infty$ .

Let  $n \geq n_*$  (cf. Corollary 7.1). Fix  $P \in \mathcal{P}_n$ ,  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and an index set  $\alpha$  defined in (21). Recall that  $q_\alpha = |\overline{\alpha}| + k$ ,  $N = K_n + k - 1$ ,  $\Lambda := \Lambda_{K_n, P, \mathbf{t}} \in \mathbb{R}^{N \times N}$ , and  $V := V_{\alpha, \mathbf{t}}$ . We then construct the following matrices based on the previous developments:  $F_\alpha := F_{\alpha, k, \mathbf{t}} \in \mathbb{R}^{q_\alpha \times N}$  (cf. (39)) and  $\Xi'_\alpha := (k/K_n) \Xi_{k, V} \in \mathbb{R}^{q_\alpha \times q_\alpha}$  (cf. (37)). In light of Proposition 7.1 and (23), the linear piece of  $\widehat{b}_{P,\mathbf{t}}(\cdot)$  corresponding to  $\alpha$  is  $\widehat{b}_{P,\mathbf{t}}^\alpha(\bar{y}) = F_\alpha^T \cdot (F_\alpha \Lambda F_\alpha^T)^{-1} \cdot F_\alpha \bar{y}$ . By combining Proposition 7.4 and Corollary 7.1, we have that

$$\begin{aligned} \left\| F_\alpha^T \cdot (F_\alpha \Lambda F_\alpha^T)^{-1} \cdot F_\alpha \right\|_\infty &\leq \|F_\alpha^T\|_\infty \cdot \|(\Xi'_\alpha F_\alpha \Lambda F_\alpha^T)^{-1}\|_\infty \cdot \|\Xi'_\alpha F_\alpha\|_\infty \\ &\leq 1 \cdot 2\rho_k \cdot \frac{k}{c_{t,1}} =: c_\infty, \end{aligned}$$

Since  $n \geq n_*$ ,  $P \in \mathcal{P}_n$ ,  $\mathbf{t} \in \mathcal{T}_{K_n}$ , and  $\alpha$  are arbitrary, the uniform Lipschitz property of  $\widehat{b}_{P,\mathbf{t}}(\cdot)$  holds.

### References

- [1] Balabdaoui, F. and Wellner, J.A. (2010) Estimation of a  $k$ -monotone density: characterizations, consistency and minimax lower bounds. *Stat. Neerl.*, **64**, 45–70. [MR2830965](#)
- [2] Balabdaoui, F. and Wellner, J.A. (2007) Estimation of a  $k$ -monotone density: limit distribution theory and spline connection. *Ann. Statist.*, **35**, 2536–2564. [MR2382657](#)
- [3] Beatson, R.K. (1981) Convex approximation by splines. *SIAM J. Math. Anal.*, **12**, 549–559. [MR0617714](#)
- [4] Beatson, R.K. (1982) Monotone and convex approximation by splines: error estimates and a curve fitting algorithm. *SIAM J. Numer. Anal.*, **19**, 1278–1285. [MR0679666](#)
- [5] Bellec, P. and Tsybakov, A. (2015) Sharp oracle bounds for monotone and convex regression through aggregation. *J. Mach. Learn. Res.*, **16**, 1879–1892. [MR3417801](#)
- [6] Birke, M. and Dette, J. (2007) Estimating a convex function in nonparametric regression. *Scand. J. Stat.*, **34**, 384–404. [MR2346646](#)
- [7] Cator, E. (2011) Adaptivity and optimality of the monotone least squares estimator. *Bernoulli*, **17**, 714–735. [MR2787612](#)
- [8] De Boor, C. (2001) *A Practical Guide to Splines*. Springer. [MR1900298](#)
- [9] DeVore, C. (1977) Monotone approximation by splines. *SIAM J. Math. Anal.*, **8** 891–905. [MR0510725](#)
- [10] DeVore, R.A. and Lorentz, G.G. (1993) *Constructive Approximation*. Springer. [MR1261635](#)

- [11] Dümbgen, L., Freitag, S., Jongbloed, G. (2004) Consistency of concave regression with an application to current-status data. *Math. Methods Statist.*, **13**, 69–81. [MR2078313](#)
- [12] Eeckhoudt, L. and Schlesinger, H. (2013) Higher-order risk attitudes. In *Handbook of Insurance* 41–57. Springer.
- [13] Gao, F.C. (2008) Entropy estimate for  $k$ -monotone functions via small ball probability of integrated Brownian motions. *Electron. Commun. Probab.*, **13**, 121–130. [MR2386068](#)
- [14] Gao, F.C. and Wellner, J.A. (2009) On the rate of convergence of the maximum likelihood estimator of a  $k$ -monotone density. *Sci. China Ser. A*, **52**, 1–14. [MR2520591](#)
- [15] Giguelay, J. (2017) Estimation of a discrete probability under constraint of  $k$ -monotonicity. *Electron. J. Stat.*, **11**, 1–49. [MR3592697](#)
- [16] Golitschek, M.V. (2014) On the  $L_\infty$ -norm of the orthogonal projector onto splines: A short proof of A. Shardin’s theorem. *J. Approx. Theory*, **181**, 30–42. [MR3182739](#)
- [17] Groeneboom, P., Jongbloed, F., and Wellner, J.A. (2001) Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.*, **29**, 1653–1698. [MR1891742](#)
- [18] Guntuboyina, A. and Sen, B. (2015) Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields*, **63**, 379–411. [MR3405621](#)
- [19] Hu, Y.K. (1991) Convexity preserving approximation by free knot splines. *SIAM J. Math. Anal.*, **22** 1183–1191. [MR1112074](#)
- [20] Konovalov, V.N. and Leviatan, D. (2001) Estimates on the approximation of 3-monotone functions by 3-monotone quadratic splines. *East J. Approx.*, **7**, 333–349. [MR1861396](#)
- [21] Konovalov, V.N. and Leviatan, D. (2003) Shape preserving widths of Sobolev-type classes of  $s$ -monotone functions on a finite interval. *Israel J. Math.*, **133**, 239–268. [MR1968430](#)
- [22] Kuipers, B.J., Chiu, C., Dalle Molle, D.T., and Throop, D.R. (1991). Higher-order derivative constraints in qualitative simulation. *Artificial Intelligence*, **51**, 343–379. [MR1144605](#)
- [23] Lebair, T.M. (2016) *Constrained Estimation and Approximation using Control, Optimization, and Spline Theory*. Ph.D. Dissertation, University of Maryland, Baltimore County. [http://pages.jh.edu/~tlebair1/Lebair\\_thesis.pdf](http://pages.jh.edu/~tlebair1/Lebair_thesis.pdf) [MR3579487](#)
- [24] Lebair, T.M., Shen, J., and Wang, X. (2016) Minimax lower bound and optimal estimation of convex functions in the sup-norm. *IEEE Trans. Automat. Control*, Vol. 62(7), pp. 3482–3487, 2017. [MR3669469](#)
- [25] Mammen, E. (1991) Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.*, **19**, 741–759. [MR1105842](#)
- [26] Mammen, E. and van de Geer, S. (1997) Locally adaptive regression splines. *Ann. Statist.*, **25**, 387–413. [MR1429931](#)



- [27] Nemirovski, A. (2000) Topics in Non-parametric Statistics. In *Ecole d'Été de Probabilités de Saint-Flour*, Berlin, Germany: Springer, **1738**, Lecture Notes in Mathematics. [MR1775640](#)
- [28] Pal, J.K. (2008) Spiking problem in monotone regression: Penalized residual sum of squares. *Statist. Probab. Lett.*, **78**, 1548–1556. [MR2453808](#)
- [29] Pal, J.K. and Woodroffe, M. (2007) Large sample properties of shape restricted regression estimators with smoothness adjustments. *Statist. Sinica*, **17**, 1601–1616. [MR2398605](#)
- [30] Prymak, A.V. (2002) Three-convex approximation by quadratic splines with arbitrary fixed knots. *East J. Approx.*, **8**, 185–196. [MR1983850](#)
- [31] Robertson, T., Wright, F.T., and Dykstra, R.L. (1988) *Order Restricted Statistical Inference*. John Wiley & Sons. [MR0961262](#)
- [32] Sanyal, A.K., Chellappa, M., Ahmed, J., Shen, J., and Bernstein, D.S. (2003) Globally convergent adaptive tracking of spacecraft angular velocity with inertia identification and adaptive linearization. *Proc. of the 42nd IEEE Conf. on Dec. and Contr.*, 2704–2709, Hawaii, 2003.
- [33] Scholtes, S. (2012) *Introduction to piecewise differentiable equations*. Springer. [MR2953259](#)
- [34] Shadrin, A.Y. (2001) The  $L_\infty$ -norm of the  $L_2$ -spline projector is bounded independently of the knot sequence: A proof of de Boor's conjecture. *Acta Math.*, **87**, 59–137. [MR1864631](#)
- [35] Shen, J. and Lehair, T.M. (2015) Shape restricted smoothing splines via constrained optimal control and nonsmooth Newton's methods. *Automatica*, **53**, 216–224. [MR3318591](#)
- [36] Shen, J. and Wang, X. (2010) Estimation of shape constrained functions in dynamical systems and its application to genetic networks. *Proc. of Am. Control Conf.*, 5948–5953, Baltimore.
- [37] Shen, J. and Wang, X. (2011) Estimation of monotone functions via  $P$ -splines: A constrained dynamical optimization approach. *SIAM J. Control Optim.*, **49**, 646–671. [MR2784704](#)
- [38] Shen, J. and Wang, X. (2012) Convex regression via penalized splines: A complementarity approach. *Proc. of Am. Control Conf.*, 332–337, Montreal, Canada.
- [39] Tsybakov, A.B. (2010) *Introduction to Nonparametric Estimation*. Springer. [MR2013911](#)
- [40] Wang, X. and Shen, J. (2010) A class of grouped Brunk estimators and penalized spline estimators for monotone regression. *Biometrika*, **97**, 585–601. [MR2672485](#)
- [41] Wang, X. and Shen, J. (2013) Uniform convergence and rate adaptive estimation of convex functions via constrained optimization. *SIAM J. Control Optim.*, **51**, 2753–2787. [MR3072224](#)
- [42] Woodroffe, M. and Sun, J. (1993) A penalized maximum likelihood estimate of  $f(0_+)$  when  $f$  is nonincreasing. *Statist. Sinica*, **3**, 501–515. [MR1243398](#)