Electronic Journal of Statistics
Vol. 5 (2011) 1–17
ISSN: 1935-7524
DOI: 10.1214/10-EJS593

# On the asymptotics of penalized spline smoothing

**Xiao Wang***

*Department of Statistics*
*Purdue University*
*West Lafayette, IN 47909, U.S.A.*
*e-mail:* wangxiao@purdue.edu

**Jinglai Shen†**

*Department of Mathematics and Statistics*
*University of Maryland Baltimore County*
*Baltimore, MD 21250, U.S.A.*
*e-mail:* shenj@umbc.edu

**and**

**David Ruppert***

*School of Operations Research and Information Engineering*
*Cornell University*
*1170 Comstock Hall*
*Ithaca, NY 14853, U.S.A.*
*e-mail:* dr24@cornell.edu

**Abstract:** This paper performs an asymptotic analysis of penalized spline estimators. We compare $P$-splines and splines with a penalty of the type used with smoothing splines. The asymptotic rates of the supremum norm of the difference between these two estimators over compact subsets of the interior and over the entire interval are established. It is shown that a P-spline and a smoothing spline are asymptotically equivalent provided that the number of knots of the P-spline is large enough, and the two estimators have the same equivalent kernels for both interior points and boundary points.

**AMS 2000 subject classifications:** Primary 62G08, 62G20; secondary 62G05.
**Keywords and phrases:** Boundary kernel, difference penalty, equivalent kernel, Green's function, $P$-spline.

Received June 2010.

## 1. Introduction

Consider the problem of estimating the function $f : [0, 1] \to \mathbb{R}$ under a univariate regression model

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

*Research is supported by the NSF grants DMS-1042967 and CMMI-1030246.
†Research is supported by the NSF grants DMS-1042916 and CMMI-1030804.
‡Research is supported by the NSF grant DMS-0805975.

where the $t_i$ are pre-specified design points and the $\epsilon_i$ are iid normal random variables with mean 0 and variance $\sigma^2$. This paper compares Eilers and Marx's [7] P-spline estimator with the corresponding smoothing spline estimator, and establishes the asymptotic rate of the supremum norm of the difference between these two estimators. Our findings show that the P-spline and smoothing spline estimators are asymptotically equivalent, and they have the same equivalent kernels at both interior and boundary points, providing sufficiently large number of knots is taken.

Penalized spline regression estimators, which use fewer knots than that of the classic smoothing spline, have been studied at least as far back as O'Sullivan [19]. One special case is the $P$-spline estimator introduced by Eilers and Marx [7], which uses a difference penalty and a flexible number of knots. Penalized spline smoothing has become popular over the last decade and the use of low rank bases leads to simple and highly efficient computation. (It is worth mentioning that certain splines, such as smoothing splines, also admit efficient numerical methods, e.g., the Kalman filter (Eubank [9]) for computation of the GCV score for selecting the smoothing parameter.) The methodology and applications of penalized splines are discussed extensively in Ruppert, Wand and Carroll [22], but asymptotic properties of the penalized spline estimators have been less explored. A few exceptions include the recent papers such as Hall and Opsomer [11], Li and Ruppert [13], and Claeskens, Krivobokova, and Opsomer [2]. Hall and Opsomer [11] placed knots continuously over a design set and established consistency of the estimator. Li and Ruppert [13] developed an asymptotic theory of penalized splines for piecewise constant and linear B-splines with the first and second order difference penalties. Claeskens, Krivobokova, and Opsomer [2] studied bias, variance, and asymptotic rates of the penalized spline estimator under different choices of the number of knots and penalty parameters. We refer the interested reader to Wahba [25], Eubank [8], Gu [10], and Eggermont and LaRicci [6] for extensive discussions on general spline regression.

The penalized spline model studied here approximates the regression function by $f^{[p]}(x) = \sum_{k=1}^{K_n+p} b_k\ B_k^{[p]}(x)$, where $\left\{B_k^{[p]} : k = 1, \ldots, K_n + p\right\}$ is the $p$th degree B-spline basis with knots $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{K_n} = 1$. The value of $K_n$ will depend upon $n$ as discussed below.

Various types of roughness penalties are in use to prevent overfitting. In Eilers and Marx's P-spline, the spline coefficients $\hat{b} = \{\hat{b}_k, k = 1, \ldots, K_n + p\}$ are subject to the $m$th-order difference penalty, that is, they are chosen to minimize

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

where $\lambda^* > 0$ and $\Delta$ is the backward difference operator, i.e., $\Delta b_k \equiv b_k - b_{k-1}$ and

$$\Delta^m b_k = \Delta\Delta^{m-1} b_k = \cdots = \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} b_{k-m+j}. \tag{3}$$

The P-spline estimator is given by $\hat{f}^{[p]}(x) = \sum_{k=1}^{K_n+p} \hat{b}_k B_k^{[p]}(x)$. On the other hand, Wand and Ormerod [26] studied splines which replace the difference penalty in (2) by a smoothing spline type penalty, so that

$$\sum_{i=1}^{n} \left[ y_i - \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(t_i) \right]^2 + \tilde{\lambda}^* \int_0^1 \left[ \frac{d^m}{dt^m} \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(t) \right]^2 dt \qquad (4)$$

is minimized where $m \le p$ and $\tilde{\lambda}^* > 0$.

We use the name "P-spline" for the minimizer of (2), "smoothing spline" for the minimizer of (4), and "classic smoothing spline" for the minimizer of (4) when there is a knot at each unique design point. The term "penalized spline" will be used for any estimator using a roughness penalty, so that penalized splines includes P-splines and smoothing splines as special cases. It is somewhat non-standard to call the minimizer of (4) without the full set of knots a smoothing spline, but this terminology agrees with that of the `smooth.spline()` function in `R`.

Initially, we assume that both the design points and the knots are equally spaced on the interval $[0, 1]$ and $n/K_n$ is an integer denoted by $M_n$; a more general case will be discussed in Section 4.

It should be noted that other bases are often used for penalized splines; for example, the truncated polynomials are used extensively in Ruppert et al. [22]. As discussed in Section 3.7.1 of Ruppert et al. [22], a penalized spline in one basis will be algebraically identical to a penalized spline in a second basis, if the two bases span the same vector space of functions and if they use identical penalties.

The contributions of the present paper are twofold: (i) The paper provides a rigorous proof that penalized splines and smoothing splines are asymptotically equivalent, and they have the same equivalent kernels at both interior and boundary points. Therefore, both the estimators have the same asymptotic distribution for all $t \in [0, 1]$ under the optimal choices of $K_n$ and $\lambda^*$. The asymptotic distribution of the general penalized spline estimator can be easily obtained by using the existing results on smoothing splines. It is worth mentioning that using equivalent kernels to perform asymptotic analysis of smoothing splines has been studied by Rice and Rosenblatt [20], Silverman [23], Messer [16], Nychka [18], and Abramovich and Grinshtein [1]. (ii) Compared with the results based on matrix techniques, e.g. Li and Ruppert [13], our approach considerably simplifies the development and yields an instrumental alternative to establish the equivalent kernels for general penalized splines. Moreover, our approach also leads to the observation that the convergence rates are independent of the splines' degrees and the number of knots for an arbitrary penalized spline estimator. While this observation was pointed out by Li and Ruppert [13] for piecewise constant and piecewise linear P-splines and was conjectured for general penalized splines, no rigorous justification has been given for general penalized splines; the current paper offers a satisfactory answer to this issue in a general setting and, in particular, provides results for the common choices of quadratic and cubic splines which Li and Ruppert did not analyze.

The paper is organized as follows. In Section 2, we give the characterization of the penalized spline estimator, and state the main result that establishes the asymptotic equivalence between the penalized spline estimator and the smoothing spline estimator. The asymptotic distributions for the cases of $p = m$ and $p \neq m$ are presented in Section 3. Discussions are given in Section 4. The Appendix contains proofs for all technical developments.

## 2. Main results

We first focus on the case when $p = m$. This is the easiest case to analyze, and splines with $p \neq m$ will be studied later by approximating them using splines with $p = m$; see Lemma 3.1. It follows from the following derivative formula for B-spline functions (de Boor [3])

$$\frac{d^l}{dx^l} \sum_{k=1}^{K_n+p} b_k B_k^{[m]}(x) = \sum_{k=l+1}^{K_n+m} K_n^l \Delta^l b_k \ B_{k-l}^{[m-l]}(x), \quad l \leq m, \tag{5}$$

that

$$\Delta^m b_{m+k} = \frac{1}{K_n^m} \frac{d^m}{dx^m} f^{[m]}(x), \qquad x \in (\kappa_{k-1}, \kappa_k], \quad k = 1, \dots, K_n. \tag{6}$$

Therefore, when $p = m$, the problems (2) and (4) are equivalent if we use equally spaced knots. Both optimization problems can be written as

$$\text{minimize} \quad \frac{1}{n} \sum_{i=1}^{n} \Big[ y_i - f(t_i) \Big]^2 + \lambda \int_0^1 \big[ f^{(m)}(t) \big]^2 dt \quad \text{ over all } f \in S_m, \tag{7}$$

where $\lambda = \lambda^*/(nK_n^{m-1})$ and $S_m = \text{span}\{B_k^{[m]} : k = 1, \dots, K_n + m\}$ is the B-spline space of order $m$. Let $W_2^m = \big\{ f : f^{(m-1)}$ absolutely continuous and $f^{(m)} \in L_2[0,1] \big\}$ be the Sobolev space of order $m$. The smoothing spline estimator is the function $\phi \in W_2^m$ that minimizes the functional

$$\frac{1}{n} \sum_{i=1}^{n} [y_i - \phi(t_i)]^2 + \lambda \int_0^1 \big[ \phi^{(m)}(t) \big]^2 dt. \tag{8}$$

Let $\hat{f}^{[m]}$ and $\hat{\phi}$ be the optimal solutions for (7) and (8), respectively. For a function $h : [0,1] \to \mathbb{R}$, define $\|h\| \equiv \sup_{t \in [0,1]} |h(t)|$ and the subsequent norms are defined in the same way.

It is easy to see that the optimal solution $\hat{f}^{[m]}$ exists and is unique for any given data. To characterize $\hat{f}^{[m]}$, we will show that $\hat{f}^{[m]}$ is an approximate solution to a certain differential equation (see Theorem 2.1), and to do this we introduce some variables and functions. Let $\omega_1$ be the uniform distribution on $\{t_1, \dots, t_n\}$ and $\omega_2$ be the uniform distribution on $\{\kappa_1, \dots, \kappa_{K_n}\}$. Let $g$ be a piecewise constant function for which $g(x_k) = y_k$ for $k = 1, \ldots, n$. Define

$$\begin{aligned} G_1(x) &= \int_0^x g(t) d\omega_1(t) = \frac{1}{n} \sum_{i=1}^n y_i I\{t_i \le x\}, \\ \check{F}_1(x) &= \int_0^x \hat{f}^{[m]}(t) d\omega_1(t) = \frac{1}{n} \sum_{i=1}^n \hat{f}^{[m]}(t_i) I\{t_i \le x\}, \end{aligned}$$

where $I$ is the indicator function of a set, and for $k \ge 2$, define

$$\begin{aligned} G_k(x) &= \int_0^x G_{k-1}(t) d\omega_2(t) = \frac{1}{K_n} \sum_{j=1}^{K_n} G_{k-1}(\kappa_j) I\{\kappa_j \le x\}, \\ \check{F}_k(x) &= \int_0^x \check{F}_{k-1}(t) d\omega_2(t) = \frac{1}{K_n} \sum_{j=1}^{K_n} \check{F}_{k-1}(\kappa_j) I\{\kappa_j \le x\}. \end{aligned}$$

We also define

$$\hat{F}_1(x) = \int_0^x \hat{f}^{[m]}(t) dt, \qquad \hat{F}_k(x) = \int_0^x \hat{F}_{k-1}(t) dt, \quad k \ge 2.$$

Let $X = [B_k(x_i)] \in \mathbb{R}^{n \times (K_n+p)}$ be the design matrix, and let $D_m \in \mathbb{R}^{(K_n+p-m)\times(K_n+p)}$ be the $m$th-order difference matrix such that

$$D_m b = [\Delta^m(b_{m+1}), \ldots, \Delta^m(b_{K_n+p})]^T.$$

The minimizer $\hat{b}$ of (2) is given by

$$(X^T X + \lambda^* D_m^T D_m)\hat{b} = X^T y, \tag{9}$$

where $y = (y_1, \ldots, y_n)^T$. Define $C \in \mathbb{R}^{(K_n+m)\times(K_n+m)}$ and $\tilde{C} \in \mathbb{R}^{(K_n+m)\times n}$, respectively, as

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \quad \text{and} \quad \tilde{C} = \begin{bmatrix} \mathbf{1}^T & 0 & 0 & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{1}^T & \mathbf{1}^T & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{1}^T & \mathbf{1}^T & \mathbf{1}^T & \cdots & \mathbf{1}^T & \mathbf{0} \\ \mathbf{1}^T & \mathbf{1}^T & \mathbf{1}^T & \cdots & \mathbf{1}^T & \mathbf{1}^T \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{1}^T & \mathbf{1}^T & \mathbf{1}^T & \cdots & \mathbf{1}^T & \mathbf{1}^T \end{bmatrix},$$

where $\mathbf{0} = [0, 0, \ldots, 0]$, $\mathbf{1} = [1, 1, \ldots, 1]^T \in \mathbb{R}^{M_n \times 1}$, and the last $m$ rows of $\tilde{C}$ are all ones. Left multiplication by $C$ and $\tilde{C}$ are discrete analogs of integration. Since $C$ is invertible, (9) is equivalent to

$$\lambda^* C^m D_m^T D_m \hat{b} + C^m X^T \hat{f} = C^m X^T y, \tag{10}$$

where $\hat{f} = [\hat{f}^{[m]}(x_1), \ldots, \hat{f}^{[m]}(x_n)]^T$ and $C^k$ denotes the product of $k$ copies of $C$. In the following development, the difference equation (10) is replaced by its analogous differential equation, in which the term $C^m D_m^T D_m$ is replaced by the differentiation operator and $C^m X^T$ is replaced by the integration operator. Let

$$R = \frac{1}{nK_n^{m-1}} \Big[ C^m \ X^T - C^{m-1} \ \tilde{C} \Big],$$

and $\check{R}$ be a piecewise constant function such that $\check{R}(\kappa_j)$ is the $j$th row of $R\big(y - \hat{f}^{[m]}\big)$.

The following result states that the optimal solution $\hat{f}^{[m]}$ can be approximated by the solution of an ordinary differential equation (ODE); its proof is given in the Appendix.

**Theorem 2.1.** *The necessary and sufficient conditions for $\hat{f}^{[m]}$ to minimize (7) are*

$$(-1)^m \ \lambda \ \frac{d^m}{dx^m} \hat{f}^{[m]}(x) = G_m(x) - \check{F}_m(x) + \check{R}(x), \quad a.e. \ \ x \in [0,1], \tag{11}$$

*and*

$$\check{F}_k(1) = G_k(1), \qquad k = 1, \ldots, m, \tag{12}$$

*where the asymptotic order of $\|\check{R}\|$ is*

$$\|\check{R}\| = O_p\left(\frac{\lambda^{1/2}}{K_n}\right) + O_p\left(\left(\frac{\log K_n}{nK_n}\right)^{1/2}\right).$$

It is well-known that smoothing splines satisfy the natural boundary conditions that the $m$th derivative of $\hat{\phi}$ is zero between 0 and the first design point and between the last design point and 1. The issue as to whether the penalized splines satisfy natural boundary conditions is very interesting. Since $G_m(x) = \check{F}(x) = 0$ for $x \in [0, t_1)$, and $G_m(x) - \check{F}(x) = 0$ for $x \in (t_{n-1}, 1]$ from (12), we have

$$\frac{d^m}{dx^m} \hat{f}^{[m]}(x) = (-1)^m \lambda^{-1} \check{R}(x), \quad x \in [0, t_1) \cup [t_{n-1}, 1].$$

Therefore, $\hat{f}^{[m]}$ does not satisfy the natural boundary conditions on $[0, t_1)$ and $(t_{n-1}, 1]$ in general, i.e., $(d^m/dx^m) f^{[m]}(x) \neq 0$ for $x \in [0, t_1) \cup (t_{n-1}, 1]$. However, under the optimal choices of $\lambda$ and $K_n$ such that $\lambda$ is of order $n^{-2m/(4m+1)}$ and $K_n \sim n^\gamma$ with $\gamma > (2m-1)/(4m+1)$, we have $\|\check{R}\|/\lambda \to 0$. This shows that $(d^m/dx^m) f^{[m]}(x) \to 0$ for $x \in [0, t_1) \cup (t_{n-1}, 1]$. Therefore, $\hat{f}^{[m]}$ satisfies the natural boundary conditions asymptotically.

The next result establishes the asymptotic equivalence of $\hat{f}^{[m]}$ and $\hat{\phi}$; its proof is in the Appendix.

**Theorem 2.2.** *For any fixed $\varrho > 0$, we have*

$$\sup_{x \in [\varrho, 1-\varrho]} |\hat{f}^{[m]}(x) - \hat{\phi}(x)| = O_p\left(\frac{\lambda^{1/2}}{K_n}\right) + O_p\left(\left(\frac{\log K_n}{n\lambda K_n}\right)^{1/2}\right). \tag{13}$$

*Furthermore,*

$$\|\hat{f}^{[m]} - \hat{\phi}\| = O_p\Big(\frac{1}{K_n}\Big) + O_p\bigg(\Big(\frac{\log K_n}{n\lambda K_n}\Big)^{1/2}\bigg). \tag{14}$$

Theorem 2.2 gives the convergence rates for the difference between $\hat{f}^{[m]}$ and $\hat{\phi}$ over any compact subset of the interior of [0, 1] and over the whole interval. It is observed from Theorem 2.2 that if $\lambda$ is of order $n^{-2m/(4m+1)}$ and $K_n \sim n^{\gamma}$ with $\gamma > (2m-1)/(4m+1)$, then for any $x \in (0,1)$, $\hat{f}^{[m]}(x) - \hat{\phi}(x)$ is of order $n^{-\varsigma}\log n$ with $\varsigma > 2m/(4m+1)$. It is known that the optimal convergence rate of $\hat{\phi}$ at any given inner point is of order $n^{2m/(4m+1)}$ under the optimal choice of $\lambda$ which is of order $n^{-2m/(4m+1)}$ (Eggermont and LaRicca [6]). This shows that $\hat{f}^{[m]}(x)$ and $\hat{\phi}(x)$ have the same asymptotic distribution for all inner points. When $t$ is close to the boundary and $K_n \sim n^{\gamma}$ with $\gamma > 2m/(4m+1)$, we have $\|\hat{f}^{[m]} - \hat{\phi}\| = O_p(n^{-\varsigma})$ with $\varsigma > 2m/(4m+1)$. The convergence rate of $\hat{\phi}$ is slower than $n^{2m/(4m+1)}$ at boundary points. Under this circumstance, $\hat{f}^{[m]}$ and $\hat{\phi}$ are asymptotically equivalent and they have the same asymptotic distributions for any $x \in [0,1]$.

## 3. Applications

It is well-known that the smoothing spline estimator $\hat{\phi}$ is asymptotically equivalent to the kernel smoothing (Silverman [23]). Specifically, Eggermont and LaRiccia [4, 6] have shown that, for any $t \in [0,1]$,

$$\hat{\phi}(t) = \int_0^1 K_\lambda(t,s)f(s)ds + \frac{1}{n}\sum_{i=1}^{n} K_\lambda(t,t_i)\epsilon_i + \text{higher order terms}, \tag{15}$$

where the equivalent kernel $K_\lambda(t,s)$ is the corresponding Green's function for the following ordinary differential equation with boundary conditions and given $v(t)$:

$$\begin{aligned}&(-1)^m\lambda u^{(2m)}(t) + u(t) = v(t), \quad \text{on} \quad [0,1],\\ &\text{subject to} \quad u^{(k)}(0) = u^{(k)}(1) = 0, \text{ for } k = m,\dots,2m-1.\end{aligned}$$

The equivalent kernel $K_\lambda(t,s)$ can be computed explicitly for an equidistant design, see e.g., Messer and Goldstein [17]. The higher order terms in (15) are negligible since they converge to zero at faster rates. Theorem 2.2 indicates that the $P$-spline or splines that minimize (4) are also approximately kernel regression estimators. The equivalent kernels for both interior points and boundary points are the same as the equivalent kernels of smoothing splines.

**Corollary 3.1.** *Let $\lambda$ satisfy $\lambda n^{2m/(4m+1)} \to 0$ and $\lambda^{-(2m-1)/2m}\log K_n/K_n \to 0$. Suppose that the true regression function $f$ is $2m$th order continuously differentiable with bounded $2m$th derivative. Define $\beta = \lambda^{-1/(2m)}$. Then for each fixed $t \in (0,1)$,*

$$\sqrt{\frac{n}{\beta}}\,[\hat{f}^{[m]}(t) - f(t)] \to^d N\big(0, \sigma_K^2(t)\big), \tag{16}$$

*where* $\lambda^{1/2m}\int_0^1 K_\lambda^2(t,s)ds \to \sigma_K^2(t)$ *as* $n\to\infty$. *However, if* $\lambda = c^{2m}n^{-\frac{2m}{4m+1}}$ *for* $c>0$ *and if* $K_n \sim n^\gamma$ *with* $\gamma > (2m-1)/(4m+1)$, *then*

$$n^{2m/(4m+1)}\ \left[\hat{f}^{[m]}(t) - f(t)\right] \to^d N\Big((-1)^{m-1}c^{2m}f^{(2m)}(t),\ c^{-1}\ \sigma_K^2(t)\ \Big). \tag{17}$$

The proof of Corollary 3.1 follows from a direct application of (15) and is thus omitted. The asymptotic results given by Corollary 3.1 provide theoretical justification of the observation that the number of knots is not important, as long as it is above some minimal level (Ruppert [21]). It is easy to find that the mean squared error of the $P$-spline estimator is of order $n^{-4m/4m+1}$, which achieves the optimal rate of convergence given in Stone [24].

In the following, we study the asymptotic properties of $\hat{f}^{[p]}(t) = \sum_{k=1}^{K_n+p}\hat{b}_k$ $B_k^{[p]}(t)$ when $p\neq m$. We first define a piecewise $m$th degree polynomial $\tilde{f}^{[m]}$, where $\hat{f}^{[p]}$ and $\tilde{f}^{[m]}$ share the same set of spline coefficients. In particular, define

$$\tilde{f}^{[m]}(t) = \begin{cases} \sum_{k=1}^{K_n+m}\hat{b}_k B_k^{[m]}(t), & \text{if}\ \ p>m \\ \sum_{k=1}^{K_n+p}\hat{b}_k B_k^{[m]}(t), & \text{if}\ \ p<m \end{cases}$$

Note that, if $p<m$, then $\tilde{f}^{[m]}$ is defined on $[0, 1-\frac{m-p}{K_n}]$. The following lemma, whose proof is given in the Appendix, characterizes the difference between $\hat{f}^{[p]}$ and $\tilde{f}^{[m]}$.

**Lemma 3.1.** *For any fixed* $t\in(0,1)$, *let* $d = \lfloor K_n t\rfloor + 1$. *Let* $\hat{\gamma}(t) = \hat{f}^{[p]}(t) - \tilde{f}^{[m]}(t)$. *Then, if* $p>m$,

$$\hat{\gamma}(t) = \sum_{q=m+1}^{p}\sum_{i=d+1}^{d+q}\left(\frac{K_n}{q}(t-\kappa_{i-q})\right)B_i^{[q-1]}(t)\sum_{l=1}^{p}a_{i+1-d,l}K_n^{-l}\frac{d^l}{dt^l}\hat{f}^{[p]}(t), \tag{18}$$

*and if* $p<m$,

$$\hat{\gamma}(t) = -\sum_{q=p+1}^{m}\sum_{i=d+1}^{d+m}\left(\frac{K_n}{q}(t-\kappa_{i-q})\right)B_i^{[q-1]}(t)\sum_{l=1}^{m}b_{i+1-d,l}K_n^{-l}\frac{d^l}{dt^l}\tilde{f}^{[m]}(t), \tag{19}$$

*where the coefficients* $\{a_{ij}\}$ *and* $\{b_{ij}\}$ *are constants.*

Following the similar discussion as above, we can establish the asymptotic distribution for $\tilde{f}^{[m]}$ as in (16) and (17), respectively, under different admissible ranges of $\lambda$ and $K_n$. Since $\hat{f}^{[p]} = \tilde{f}^{[m]} + \hat{\gamma}(t)$, we have the following asymptotic distribution for $\hat{f}^{[p]}$ for any $p\neq m$ at a fixed interior point.

**Corollary 3.2.** *Suppose that* $f$ *is* $2m$*th order continuously differentiable with bounded* $2m$*th derivative on* $[0,1]$. *Let* $\lambda$ *satisfy* $\lambda n^{2m/(4m+1)} \to 0$ *and* $\lambda^{-(2m-1)/2m}\log K_n/K_n \to 0$. *Then, for each fixed* $t\in(0,1)$ *and with* $\beta = \lambda^{-1/(2m)}$ *as before,*

$$\sqrt{\frac{n}{\beta}}\ [\hat{f}^{[p]}(t) - f(t) - \hat{\gamma}(t)] \to^d N\big(0, \sigma_K^2(t)\big), \tag{20}$$

*where $\gamma(t)$ is given by (18) if $p < m$ or by (19) if $p > m$. However, if $\lambda = c^{2m} n^{-\frac{2m}{4m+1}}$ for $c > 0$, and let $K_n \sim n^{\gamma}$ with $\gamma > (2m-1)/(4m+1)$, then*

$$n^{2m/(4m+1)}\ [\hat{f}^{[p]}(t) - f(t) - \hat{\gamma}(t)] \rightarrow^d N\Big((-1)^{m-1} c^{2m} f^{(2m)}(t),\ c^{-1}\ \sigma_K^2(t)\ \Big). \tag{21}$$

It can be seen from the above corollary that when $p$ is not equal to $m$, the asymptotic bias has an additional term $\hat{\gamma}(t)$, which is of order $O_p(1/K_n)$. When $K_n$ grows sufficiently fast with respect to $n$, this term is asymptotically negligible.

## 4. Discussions

We have so far focused on the equally spaced design case and equally spaced knots. When the design is not equally spaced and we use equidistant knots, under similar arguments in Section 2, problems (7) and (8) are still asymptotically equivalent, and the problem (8) is asymptotically equivalent to

$$\text{minimize} \quad \int_0^1 [\phi(t) - f(t)]^2 \omega(t) dt + \frac{2}{n} \sum_{i=1}^{n} (\phi(t_i) - f(t_i))\epsilon_i + \lambda \int_0^1 \left[\phi^{(m)}(t)\right]^2 dt,$$

where $\omega(t)$ is the asymptotic design density, and the rest is as the same as in Chapter 21 of Eggermont and LaRicca [6].

We have assumed that the random errors $\{\epsilon_i : i = 1, \ldots, n\}$ in the regression model satisfy a normal distribution, and this assumption can be relaxed. A crucial step in the proofs of the asymptotic properties of the estimators is the order of $\max_{i=1,\ldots,n} |\epsilon_i|$. Indeed, when the $\epsilon_i$'s are independent normal random variables, $\max_{i=1,\ldots,n} |\epsilon_i|$ is of order $O_p((2\log n)^{1/2})$. If the $\epsilon_i$'s satisfy other distributions, then the order of $\max_{i=1,\ldots,n} |\epsilon_i|$ can be determined by the tail probability $\Pr(\epsilon_i > x)$. By making use of assumptions of this tail probability, all derivations for asymptotic properties can be obtained in a similar fashion.

One may ask "what is the interpretation of cases $m > p$?" These cases are, of course, impossible for the smoothing spline penalty, since if $m > p$, then the $m$th derivative will not exist at the knots and will be zero elsewhere. For the discrete $P$-spline penalty, the cases $m > p$ are valid and indeed were allowed in Eilers and Marx [7]. To interpret these cases, it is useful to look at the simple case when $p = 0$, i.e. piecewise constant splines, under the assumption of equally spaced knots. In this case, $\Delta b_k$ is the jump of the function at the knot $\kappa_k$. Hence when $m = 1$, any deviations from a constant function are penalized. This effect is similar to what it would be if the first derivative existed and was penalized. Similarly, when $m = 2$, $\Delta^2 b_k$ is the difference between the jumps at two consecutive knots. The functions that are unpenalized are step function approximations to linear functions. This pattern persists for higher values of $m$ and $p$. For example, if $p = 1$, then the functions that are unpenalized are piecewise linear approximations to polynomials of degree $m - 1$, because the coefficients will follow a polynomial trend of the same degree.

The univariate $P$-splines can be naturally extended to multivariate $P$-splines Marx and Eilers [15]. The asymptotic properties can be studied along the same line. Our conjecture is that the multivariate $P$-spline smoothing is asymptotically equivalent to multivariate kernel smoothing and the equivalent kernel is the Green's function corresponding to a related partial differential equation. Further study of this issue is beyond the scope of this paper and shall be reported in a future publication.

## Appendix

### *Proof of Theorem 2.1*

Since $C$ is invertible, for any $k \in \mathbb{N}$, (9) is equivalent to

$$\lambda^* C^k D_m^T D_m \hat{b} + C^k X^T \hat{f} = C^k X^T y, \tag{22}$$

where $\hat{f} = [\hat{f}^{[m]}(x_1), \ldots, \hat{f}^{[m]}(x_n)]^T$ and $C^k$ denotes the product of $k$ copies of $C$.

The matrix $D_m^T D_m$ is a banded symmetric matrix. Except for the first $m$ and last $m$ rows, every row of $D_m^T D_m$ has the form $(0, \ldots, 0, \omega_0^*, \omega_1^*, \ldots, \omega_{2m}^*, 0, \ldots, 0)$, where $\omega_j^* = (-1)^m (-1)^{2m-j} \binom{2m}{j}$, $j = 0, \ldots, 2m$. Moreover, except for the first $m - k$ and last $m$ rows, the $i$th row of $C^k D_m^T D_m$ has the form

$$\Big( \underbrace{0, \ldots, 0,}_{(i-m+k-1)-\text{copies}} \quad \tilde{\omega}_0, \ldots, \tilde{\omega}_{2m-k}, \quad \underbrace{0, \ldots, 0}_{(K_n+p)-(i+m)-\text{copies}} \Big),$$

where

$$\tilde{\omega}_j = (-1)^m (-1)^{2m-k-j} \binom{2m-k}{j}, \; j = 0, \ldots, 2m-k.$$

Further, the elements of the last $k$ rows of $C^k D_m^T D_m$ are all zeros. In particular, when $k = m$,

$$C^m D_m^T D_m \hat{b} = (-1)^m \left[ \Delta^m \hat{b}_{m+1}, \Delta^m \hat{b}_{m+2}, \ldots, \Delta^m \hat{b}_{K_n+p}, 0, \ldots, 0 \right]^T. \tag{23}$$

From (5),

$$\Delta^m \hat{b}_{m+k} = \frac{1}{K_n^m} \frac{d^m}{dx^m} \hat{f}^{[m]}(x), \qquad x \in (\kappa_{k-1}, \kappa_k], \quad k = 1, \ldots, K_n. \tag{24}$$

Since the elements of the last $k$ rows of $C^k D_m^T D_m$ are all zeros for $k = 1, \ldots, m$, we have, from (22),

$$\check{F}_k(1) = G_k(1), \qquad k = 1, \ldots, m. \tag{25}$$

Also note that

$$\frac{1}{n} \tilde{C} \hat{f} = \left[ \check{F}_1(\kappa_1), \check{F}_1(\kappa_2), \ldots, \check{F}_1(1), \ldots, \check{F}_1(1) \right]^T,$$

and from (25),

$$\frac{1}{nK_n^{m-1}} C^{m-1}\tilde{C}\ (\hat{f}-y) = \Big[\check{F}_m(\kappa_1) - G_m(\kappa_1), \check{F}_m(\kappa_2) - G_m(\kappa_2), \ldots, \\ \check{F}_m(1) - G_m(1), 0, \ldots, 0\Big]^T.$$

Let

$$R = \frac{1}{nK_n^{m-1}}\left[C^m\ X^T - C^{m-1}\ \tilde{C}\right],$$

and $\check{R}$ be a piecewise constant function such that $\check{R}(\kappa_j)$ is the jth row of $R\ (y - \hat{f}^{[m]})$. Therefore, the $j$th row of (22), when $k = m$, can be written as

$$(-1)^m \frac{\lambda^*}{nK_n^{m-1}} \Delta^m b_{m+j} + \check{F}_m(\kappa_j) = G_m(\kappa_j) + \check{R}(\kappa_j), \quad j = 1, \ldots, K_n. \tag{26}$$

Combining (24) and (26) gives

$$(-1)^m \lambda \hat{F}_m^{(2m)}(x) + \check{F}_m(x) = G_m(x) + \check{R}(x), \quad x \in [0, 1],$$

where $\lambda = \lambda^*/(nK_n^{2m-1})$. The asymptotic order of $\check{R}$ is given in Lemma A.1. □

**Lemma A.1.** *The following holds:*

$$\|\check{R}\| = O_p\left(\frac{\lambda^{1/2}}{K_n}\right) + O_p\left(\left(\frac{\log K_n}{nK_n}\right)^{1/2}\right).$$

*Proof.* Let $\bar{y} = \frac{K_n}{n} X^T y$ and $\alpha = \lambda^* K_n/n$. Claeskens et al. [2] showed that $\|H^{-1}\|_\infty = O(1)$, where $H = \frac{K_n}{n} X^T X + \alpha D_m^T D_m$. Thus, $\hat{b}$ is stochastically bounded, so is $\hat{f}^{[m]}$. Thus, $\|\hat{F}_m - \check{F}_m\|$ is of order $O_p(1/n)$. Let $\bar{b}$ solve $(X^T X + \lambda^* D_m^T D_m)\bar{b} = X^T f$ and denote $\bar{f}^{[m]}(x) = \sum_{k=1}^{K_n+m} \bar{b}_k B_k^{(m)}(x)$. Note that $\bar{f}^{[m]}$ is the estimator when there is no noise in the regression model (1). We have

$$\|\hat{f}^{[m]} - \bar{f}^{[m]}\| \ \le\ \|\hat{b} - \bar{b}\|_\infty \ \le\ \|H^{-1}\|_\infty \,\|\, \bar{y} - \mathbb{E}[\bar{y}]\,\|_\infty \ =\ O_p\left(\sqrt{\frac{K_n}{n}}\sqrt{2\log K_n}\right). \tag{27}$$

It is shown that $\|\bar{f}^{[m]} - f\| = O(\lambda^{1/2})$ and the development of this rate is similar to Theorem 7 in Eggermont and LaRiccia [4]. Thus,

$$\begin{aligned} \|R\ (y - \hat{f}^{[m]})\|_\infty &\le \|R\ (y - f)\|_\infty + \|R\ (f - \bar{f}^{[m]})\|_\infty + \|R\ (\bar{f} - \hat{f}^{[m]})\|_\infty \\ &= O_p\left(\sqrt{\frac{\log K_n}{nK_n}}\right) + O_p\left(\frac{\lambda^{1/2}}{K_n}\right) + O_p\left(\left(\frac{\log K_n}{nK_n}\right)^{1/2}\right). \end{aligned}$$

Hence, the lemma follows. □

***Proof of Theorem 2.2***

Define

$$\tilde{R}(x) = (-1)^m \alpha \hat{F}_m^{(2m)}(x) + \hat{F}_m(x) - G_m(x).$$

Then, $\tilde{R} = \hat{F}_m - \check{F}_m + \check{R}$. Hence, $\hat{F}_m$ solves the ordinary differential equation

$$(-1)^m \lambda \hat{F}_m^{(2m)}(x) + \hat{F}_m(x) = G_m(x) + \tilde{R}(x), \qquad 0 \le x \le 1, \tag{28}$$

with $2m$ boundary conditions from (25):

$$\hat{F}_m^{(k)}(0) = 0, \quad \hat{F}_m^{(k)}(1) = G_{m-k}(1) + e_{m-k}, \quad k = 0, \dots, m-1, \tag{29}$$

where $e_{m-k} = \hat{F}_{m-k}(1) - \check{F}_{m-k}(1)$. Lemma A.1 indicates that $\hat{f}^{[m]}$ is stochastically bounded. Therefore $e_k$ are small with an order of $O_p(1/n)$. Lemma A.1 also indicates that $\|\tilde{R}\|$ has the same rate as that of $\|\check{R}\|$ since $\|\hat{F}_m - \check{F}_m\|$ is of order $O_p(1/n)$. Hence, $\|\tilde{R}\| = O_p(\lambda^{1/2}/K_n) + O_p\big((\log K_n/nK_n)^{1/2}\big)$.

Next, consider the smoothing spline problem (8).

**Lemma A.2.** *The necessary and sufficient conditions for $\hat{\phi}$ to minimize (8) are*

$$(-1)^m \ \lambda \ \hat{\phi}_m^{(m)}(x) + \check{\Phi}_m(x) = \check{G}_m(x), \qquad a.e. \ \ x \in [0,1]$$

*and*

$$\check{\Phi}_k(1) = \check{G}_k(1), \quad k = 1, \dots, m,$$

*where*

$$\begin{aligned} \check{\Phi}_1(x) &= \int_0^x \hat{\phi}(t) d\omega_1(t), \qquad \check{\Phi}_k(x) = \int_0^x \Phi_{k-1}(t)dt, \quad k \ge 2, \\ \check{G}_1(x) &= G_1(x), \qquad \check{G}_k(x) = \int_0^x \check{G}_{k-1}(t)dt, \quad k \ge 2. \end{aligned}$$

*Proof.* Denote the functional (8) as $H(\phi)$. For any $\phi, \bar{\phi} \in W_2^m$ and $\delta \in \mathbb{R}$,

$$H(\phi+\delta\bar{\phi}) - H(\phi) = 2\delta H_1(\phi, \bar{\phi}) + \delta^2 \Big\{ \int_0^1 \bar{\phi}^2(t) d\omega_1(t) + \lambda \int_0^1 \{\bar{\phi}^{(m)}(t)\}^2 dt \Big\}, \tag{30}$$

where

$$H_1(\phi, \bar{\phi}) = \int_0^1 [\phi(t) - g(t)] \bar{\phi}(t) d\omega_1(t) + \lambda \int_0^1 \phi^{(m)}(t) \bar{\phi}^{(m)}(t) dt. \tag{31}$$

Then, $\phi \in W_2^m$ minimizes $H(\phi)$, if and only if, $H_1(\phi, \bar{\phi}) = 0$ for all $\bar{\phi} \in W_2^m$. The reason is as follows. If $\phi \in W_2^m$ minimizes $H(\phi)$, $H(\phi + \delta\bar{\phi}) - H(\phi) \ge 0$ for all $\bar{\phi} \in W_2^m$ and any $\delta \in \mathbb{R}$. Then $H_1(\phi, \bar{\phi}) = 0$ follows since $\delta$ can be either negative or positive. On the other hand, if $H_1(\phi, \bar{\phi}) = 0$, we have $H(\phi + \delta\bar{\phi}) - H(\phi) \ge 0$ by (30). Thus, $\phi$ minimizes $H(\phi)$.

Letting $v(t) = t^k$, $k = 0, \ldots, m-1$ in (31) shows that if $\phi$ minimizes $H(f)$, then,

$$\int_0^1 [\phi(t) - g(t)]\ t^k d\omega_1(t) = 0, \quad k = 0, 1, \ldots, m-1.$$

We first have

$$\check{\Phi}_1(1) - \check{G}_1(1) = \int_0^1 [f(t) - g(t)] d\omega_1(t) = 0.$$

Further,

$$\begin{aligned} \check{\Phi}_2(1) - \check{G}_2(1) &= \int_0^1 \int_0^s [f(t) - g(t)] d\omega_1(t) ds \\ &= \int_0^1 [f(t) - g(t)]\ t\ d\omega_1(t) = 0. \end{aligned}$$

Similarly, it is shown that $\check{\Phi}_k(1) = \check{G}_k(1)$ for $k = 1, \ldots, m$.

If $\phi \in W_2^m$ satisfies $\check{\Phi}_k(1) = \check{G}_k(1)$, $k = 1, \ldots, m$, we have

$$\begin{aligned} \int_0^1 [\phi(t) - g(t)] \bar{\phi}(t) d\omega_1(t) &= \int_0^1 [\phi(t) - g(t)]\ [\bar{\phi}(t) - \bar{\phi}(1)] d\omega_1(t) \\ &= -\int_0^1 [\phi(t) - g(t)]\ \int_t^1 \bar{\phi}'(s) ds d\omega_1(t) \\ &= -\int_0^1 [\check{F}_1(s) - G_1(s)]\ g'(s) ds \\ &= \cdots\cdots \\ &= (-1)^m \int_0^1 [\check{\Phi}_m(s) - \check{G}_m(s)]\ \bar{\phi}^{(m)}(s) ds. \end{aligned}$$

Hence,

$$H(\phi, \bar{\phi}) = \int_0^1 H_2(\phi)\ \bar{\phi}^{(m)}(t) dt, \tag{32}$$

where

$$H_2(\phi) = \lambda\ \phi^{(m)}(t) + (-1)^m\ [\check{\Phi}_m(t) - \check{G}_m(t)]. \tag{33}$$

If $H_1(\phi, \bar{\phi}) = 0$ for all $\bar{\phi} \in W_2^m$, letting $B = \{t \in [0,1] : H_2(\phi) \neq 0\}$ and $\bar{\phi}^{(m)}(t) = -I_B(t)$ gives

$$H_1(\phi, \bar{\phi}) = \int_B H_2(\phi) dt \neq 0,$$

unless $B$ is of measure zero. This shows $H_2(\phi) = 0$ almost everywhere. This completes the proof of this lemma. □

Define

$$\hat{\Phi}_1(x) = \int_0^x \hat{\phi}(t) dt, \quad \hat{\Phi}_k(x) = \int_0^x \Phi_{k-1}(t) dt, \quad k \geq 2.$$

Let $\mathcal{R} = (-1)^m \lambda \hat{\Phi}_m^{(m)} + \hat{\Phi}_m - \check{G}_m = \hat{\Phi}_m - \check{\Phi}_m$. Hence, $\hat{\Phi}_m$ solves the ordinary differential equation

$$(-1)^m \lambda \hat{\Phi}_m^{(2m)}(x) + \hat{\Phi}_m(x) = \check{G}_m(x) + \mathcal{R}(x), \quad 0 \le x \le 1, \tag{34}$$

with $2m$ boundary conditions:

$$\hat{\Phi}_m^{(k)}(0) = 0, \quad \hat{\Phi}_m^{(k)}(1) = \check{G}_{m-k}(1) + \check{e}_{m-k}, \quad k = 0, \ldots, m-1, \tag{35}$$

where $\check{e}_{m-k} = \hat{\Phi}_k(1) - \check{\Phi}_k(1)$. Since $\hat{\phi}$ is stochastically bounded, it is easy to see that $\|\mathcal{R}\|$ and $|\check{e}_{m-k}|$, $k = 0, \ldots, m-1$ are all of order $O_p(1/n)$.

It is interesting to note that the ordinary differential equations (28) and (34) share many similarities. To obtain the relationship between $\hat{f}^{[m]}$ and $\hat{\phi}$, we further introduce a few variables and functions related to the true regression function $f$. Define

$$\begin{aligned}
\Psi_1(x) &= \int_0^x f(t)dt, \quad \Psi_k(x) = \int_0^x \Psi_{k-1}(t)dt, \quad k \ge 2, \\
\check{\Psi}_1(x) &= \int_0^x f(t)d\omega_1(t), \quad \check{\Psi}_k(x) = \int_0^x \check{\Psi}_{k-1}(t)dt, \quad k \ge 2, \\
\tilde{\Psi}_1(x) &= \check{\Psi}_1(x), \quad \tilde{\Psi}_k(x) = \int_0^x \tilde{\Psi}_{k-1}(t)d\omega_2(t), \quad k \ge 2.
\end{aligned}$$

Let $\delta = \hat{f}^{[m]} - \hat{\phi}$ and $\Delta_m = \hat{F}_m - \hat{\Phi}_m$. It is observed from (28) and (34) that $\Delta_m$ solves the ordinary differential equation

$$(-1)^m \lambda \Delta_m^{(2m)}(x) + \Delta_m(x) = \eta(x), \quad x \in [0,1], \tag{36}$$

with $2m$ boundary conditions:

$$\Delta_m^{(k)}(0) = 0, \quad \Delta_m^{(k)}(1) = \zeta_k, \quad k = 0, \ldots, m-1, \tag{37}$$

where

$$\eta = G_m + \tilde{R} - \check{G}_m,$$

and

$$\zeta_k = G_{m-k}(1) + e_{m-k} - \check{G}_{m-k}(1) - \check{e}_{m-k}.$$

Note that

$$\eta = (G_m - \tilde{\Psi}_m - \check{G}_m + \check{\Psi}_m) + (\tilde{\Psi}_m - \Psi_m) + \tilde{R},$$

in which $\|G_k - \tilde{\Psi}_k - \check{G}_k + \check{\Psi}_k\|$ for $k \ge 2$, which are of order $O_p(\log n/\sqrt{n}K_n)$ by the strong approximation theorem (Komlós et al. [12]), and $\|\tilde{\Psi}_m - \Psi_m\|$ is of order $O(1/n)$. Hence $\|\eta\|$ is of order $O_p(\lambda^{1/2}/K_n) + O_p\big((\log K_n/nK_n)^{1/2}\big)$, and $\|\zeta\|_\infty$ is of order $O_p(1/K_n)$ with $\zeta = (\zeta_0, \ldots, \zeta_{m-1})$.

Let $K_\lambda(t,s)$ be the Green's function corresponding to (36). Then,

$$\Delta_m(x) = \int_0^1 K_\lambda(x,s)\eta(s)ds,$$

and

$$\delta(x) = \int_0^1 \frac{\partial^m}{\partial x^m} K_\lambda(x,s)\eta(s)ds. \tag{38}$$

Eggermont and LaRiccia [4] showed that $\lambda^{\frac{l}{2m}} \frac{\partial^l}{\partial x^l} K_\lambda(x,s)$ is kernel-like for $l = 0, \ldots, m$. In particular,

$$\int_0^1 \left| \lambda^{\frac{1}{2}} \frac{\partial^m}{\partial x^m} K_\lambda(x,s) \right| ds$$

is uniformly bounded for $x \in [0,1]$. Combining this with (38) shows that $\|\delta\|$ has an order of $O_p(1/K_n) + O_p\big((\log K_n/n\lambda K_n)^{1/2}\big)$.

The proof for the case where both $\hat{f}^{[m]}$ and $\hat{\phi}$ are restricted to a compact subinterval $[\varrho, 1-\varrho]$ is similar, where $\varrho \in (0, 1/2)$. The only difference is that the rate for $\check{R}$ in Lemma A.1 becomes $O_p\big(\frac{\lambda}{K_n}\big) + O_p\big(\big(\frac{\log K_n}{nK_n}\big)^{1/2}\big)$. This is because the bias term $\sup_{x\in[\varrho,1-\varrho]} |\bar{f}^{[m]}(x) - f(x)|$ is of order $O(\lambda)$ in the latter case. $\square$

### ***Proof of Lemma 3.1***

The B-spline basis functions satisfy the following recurrence relationship

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

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

$$\begin{aligned} f^{[p]}(t) - f^{[p-1]}(t) &= \sum_{i=d+1}^{d+p} \left[ b_{i+1} \frac{K_n}{p}(t - \kappa_{i-p}) + b_i \left( \frac{K_n}{p}(\kappa_i - t) - 1 \right) \right] B_i^{[p-1]}(t) \\ &= \sum_{i=d+1}^{d+p} (b_{i+1} - b_i) \left( \frac{K_n}{p}(t - \kappa_{i-p}) \right) B_i^{[p-1]}(t). \end{aligned} \tag{39}$$

From (39), if $p > m$,

$$\hat{f}^{[p]}(t) = \tilde{f}^{[m]}(t) + \sum_{q=m+1}^{p} \sum_{i=d+1}^{d+q} \Delta b_{i+1} \left( \frac{K_n}{q}(t - \kappa_{i-q}) \right) B_i^{[q-1]}(t).$$

From (3), we have $\Delta^l b_k = c_l^T (\Delta b_{k-l+1}, \Delta b_{k-l+2}, \ldots, \Delta b_k)$, where

$$c_l = \left[ \; (-1)^{l-1} \binom{l-1}{0}, (-1)^{l-2} \binom{l-1}{1}, \ldots, (-1)^0 \binom{l-1}{l-1} \; \right]^T.$$

Combining this with (5), it is easy to show that there exists $C_d \in \mathbb{R}^{p\times p}$ such that

$$\left[ \; \Delta b_{d+2}, \Delta b_{d+2}, \ldots, \Delta b_{d+p+1} \; \right]^T = C_d \left[ \; K_n^{-1} \frac{d}{dt} f^{[p]}(t), \ldots, K_n^{-p} \frac{d^p}{dt^p} f^{[p]}(t) \; \right]^T.$$

Hence, we can write

$$\Delta b_{d+k} = \sum_{l=1}^{p} a_{kl} K_n^{-l} \frac{d^l}{dx^l} f^{[p]}(t), \ k = 2, \ldots, p+1, \tag{40}$$

which gives (18). (19) can be established similarly. Thus the lemma follows. □

### Acknowledgements

We would like to thank the guest Editor, Professor Matt Wand, and two referees for their constructive comments and suggestions which helped improve our presentation greatly.

### References

[1] Abramovich, F. and Grinshtein, V. (1999). Deriviation of equivalent kernel for general spline smoothing: a systematic approach. *Bernoulli*, **5**, 359–379. MR1681703

[2] Claeskens, G., Krivobokova, T. and Opsomer, J. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, **96**, 529–544. MR2538755

[3] de Boor, C. (2001). *A Practical Guide to Splines*. Springer. MR1900298

[4] Eggermont, P. P. B. and LaRicci, V. N. (2006a). Equivalent kernels for smoothing splines. *Journal of Integral Equations and Applications*, **18**, 197–225. MR2273348

[5] Eggermont, P. P. B. and LaRicci, V. N. (2006b). Uniform error bounds for smoothing splines. *IMS Lecture Notes—Monograph Series High Dimensional Probability* **51**, 220–237. MR2387772

[6] Eggermont, P. P. B. and LaRicci, V. N. (2009). *Maximum Penalized Likelihood estimation*. Volume II: Regression. New York: Springer. MR1837879

[7] Eiler, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science*, **11**, 89–121. MR1435485

[8] Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcek Dekker. MR1680784

[9] Eubank, R. L. (1999). *A Kalman Filter Primer*, CRC Press, Boca Raton. MR2193537

[10] Gu, C. (2002). *Smoothing Spline: ANOVA Models*. New York: Springer. MR1876599

[11] Hall, P. and Opsomer, J.D. (2005). Theory for penalised spline regression. *Biometrika*, **92**, 105–118. MR2158613

[12] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f., *Z. Wahrsch. Verw. Gebiete*, **32**, 111–131. MR0375412

[13] Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika*, **95**, 415–436. MR2521591

[14] Marx, B. and Eilers, P. (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science*, **11**, 89–121. MR1435485

[15] Marx, B. and Eilers, P. (2005). Multidimensional penalized signal regression. *Technometrics*, **47**, 13–22. MR2135789

[16] Messer, K. (1991). A comparison of a spline estimate to its equivelent kernel estimate. *Annals of Statistics*, **19**, 817–829. MR1105846

[17] Messer, K. and Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation. *Annals of Statistics*, **21**, 179–196. MR1212172

[18] Nychka, D. (1995). Splines as local smoothers. *Annals of Statistics*, **23**, 1175–1197. MR1353501

[19] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with Discussion), *Statistical Science*, **1**, 505–527. MR0874480

[20] Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Annals of Statistics*, **11**, 141–156. MR0684872

[21] Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statisitcs*, **11**, 735–757. MR1944261

[22] Ruppert, D., Wand, M.P., and Carroll, R.J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press. MR1998720

[23] Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, **12**, 898–916. MR0751281

[24] Stone, C.J. (1982). Optimal rate of convergence for nonparametric regression. *Annals of Statistics*, **10**, 1040–1053. MR0673642

[25] Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia, PA: SIAM. MR1045442

[26] Wand, M.P. and Ormerod, J.T. (2008) On semiparametric regression with O'Sullivan penalized splines. *Austral. New Zeal. J. Statist.*, **50**, 179–198. MR2431193