Mem. Gra. Sci. Eng. Shimane Univ.
Series B: Mathematics
47 (2014), pp. 63–71

ASYMPTOTICS FOR PENALIZED SPLINES IN ADDITIVE MODELS

TAKUMA YOSHIDA

Communicated by Kanta Naito

(Received: December 19, 2013)

Abstract. This paper is based on the author's thesis, "Asymptotic theory of penalized spline regression". The focus of the present paper is on the penalized spline estimators obtained by the backfitting algorithm in additive models. The convergence of the algorithm as well as the uniqueness of its solution are shown. Asymptotic equivalence between the penalized spline estimators by the backfitting algorithm and the convenient estimators proposed by Marx and Eilers [9] is addressed. Asymptotic normality of the estimators is also developed.

1. Introduction

The additive model is a typical regression model with multidimensional covariates and is usually expressed as

$$y_i = f_1(x_{i1}) + \cdots + f_D(x_{iD}) + \varepsilon_i,$$

for given data $\{(y_i, x_{i1}, \cdots, x_{iD}) : i = 1, \cdots, n\}$, where each $f_d (d = 1, \cdots, D)$ is a univariate function with a certain degree of smoothness.

The additive model has become a popular smoothing technique and its fundamental properties have been summarized in literature such as Buja et al. [2] and Hastie and Tibshirani [7]. Buja et al. [2] proposed the so-called backfitting algorithm, which is efficient for nonparametric estimation of $f_d (d = 1, \cdots, D)$. The backfitting algorithm is a repetition update algorithm and its convergence and the uniqueness of its solution are not always assured. Buja et al. [2] showed the sufficient condition for convergence of the backfitting algorithm and the uniqueness of its solution for some smoothing methods.

In this paper, we discuss the asymptotic properties of the penalized spline estimator for the additive model. Penalized spline estimators have been discussed in O'Sullivan [10], Eilers and Marx [5], Marx and Eilers [9], Aerts et al. [1] and

2000 *Mathematics Subject Classification.* 62G08, 62G20, 65D07.

Key words and phrases. Additive model, Asymptotic normality, Backfitting algorithm, Penalized spline.

Ruppert et al. [11]. Despite its richness of application, the asymptotics for spline smoothing seem to have not yet been sufficiently developed.

For the univariate model ($D = 1$), Hall and Opsomer [6] gave the mean squared error and proved consistency of the penalized spline estimator. The asymptotic bias and variance of the penalized spline estimator were obtained in Claeskens et al. [3]. Kauermann et al. [8] worked with the generalized linear model. Wang et al. [14] showed that the penalized spline estimator is asymptotically equivalent to a Nadaraya-Watson estimator. Thus, it seems that developments of the asymptotic theories of the penalized splines are relatively recent events and we note that those works are mainly regarding the univariate model.

The penalized spline estimators for the additive models are obtained by using the penalized least squares method. However there is one problem in that the loss function L is not strictly convex. Therefore it is difficult to find the minimizer of L since its Hessian is not invertible. Though the backfitting algorithm yields a solution that makes the gradient of L equal to zero, there is no guarantee that the solution obtained by this method minimizes L. It has been known that the backfitting algorithm does converge, but the uniqueness of the obtained solution cannot be proved in general.

On the other hand, Marx and Eilers [9] proposed a new loss function to avoid this singularity problem. They proposed to use L_γ which consists of L plus an additional small ridge penalty. This L_γ is strictly convex, hence the global minimum of L_γ is equivalent to a unique local minimum which can be easily obtained. Of course there is a gap between the minimizer of L_γ and that of L. In this paper, the estimator obtained by minimizing L_γ is called the ridge corrected penalized spline estimator (RCPS) and the penalized spline estimator obtained by the backfitting algorithm is denoted as the backfitting penalized spline estimator (BPS). Because we are interested in the estimator obtained by using L, we mainly focus on BPS.

The aim of this paper is to derive the asymptotic distribution of the BPS in the general D-variate additive model. First we show the asymptotic distribution of the RCPS. Next, it is shown that the difference of the RCPS and the BPS asymptotically vanishes. As a result, it demonstrates that the asymptotic normality of the BPS and that the solution of the backfitting algorithm is asymptotically unique. As will be seen in the subsequent section, although the closed form of the BPS can not be written, its asymptotic properties can be shown in each iteration of the algorithm. The properties of the band matrices play an important role as a mathematical tool in asymptotic considerations.

This paper is organized as follows. In Section 2, our model settings and estimating equation in the penalized least squares method are discussed, and the RCPS and the BPS are constructed. Section 3 provides the asymptotic bias and variance of the RCPS, and then its asymptotic normality is developed. Furthermore, we show that the BPS is asymptotically equivalent to the RCPS and the solution of the backfitting algorithm is asymptotically unique. In Section 4, we give some comments. Proofs of all mathematical results are omitted.

2. Model setting and penalized spline estimator

2.1. **Additive spline model.** Consider a D-variate additive regression model

$$y_i = f_1(x_{i1}) + \cdots + f_D(x_{iD}) + \varepsilon_i \tag{1}$$

for the data $\{(y_i, \boldsymbol{x}_i) : i = 1, \cdots, n\}$, where $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{iD})$ is the D-variate explanatory variables, $f_j(\cdot)$ is an unknown regression function and ε_i's are independent random errors with $E[\varepsilon_i|\boldsymbol{X}_i = \boldsymbol{x}_i] = 0$ and $V[\varepsilon_i|\boldsymbol{X}_i = \boldsymbol{x}_i] = \sigma^2(\boldsymbol{x}_i) < \infty$. We assume $E[f_j(X_j)] = 0 (j = 1, \cdots, D)$ to ensure identifiability of f_j. Let $q_j(x_j)$ be the density of X_j and $q(\boldsymbol{x}) = q(x_1, \cdots, x_D)$ be the joint density of $\boldsymbol{X} = (X_1, \cdots, X_D)$. We assume without loss of generality that $\boldsymbol{x}_i \in (0,1)^D$ for all $i \in \{1, \cdots, n\}$, where $(0,1)^D$ is the D-variate unit cube.

Now we consider the B-spline model

$$s_j(x_j) = \sum_{k=-p+1}^{K_n} B_k^{[p]}(x_j) b_{j,k}$$

as an approximation to $f_j(x_j)$ at any $x_j \in (0,1)$ for $j = 1, \cdots, D$. Here, $B_k^{[p]}(x) (k = -p+1, \cdots, K_n)$ are the pth degree B-spline basis functions defined recursively as

$$\begin{aligned} B_k^{[0]}(x) &= \begin{cases} 1, & \kappa_{k-1} < x \le \kappa_k, \\ 0, & \text{otherwise}, \end{cases} \\ B_k^{[p]}(x) &= \frac{x - \kappa_{k-1}}{\kappa_{k+p-1} - \kappa_{k-1}} B_k^{[p-1]}(x) + \frac{\kappa_{k+p} - x}{\kappa_{k+p} - \kappa_k} B_{k+1}^{[p-1]}(x), \end{aligned}$$

where $\kappa_k = k/K_n (k = -p+1, \cdots, K_n + p)$ are knots and $b_{j,k} (j = 1, \cdots, D, k = -p+1, \cdots, K_n)$ are unknown parameters. We denote $B_k^{[p]}(x)$ as $B_k(x)$ in what follows since only the pth degree is treated. The details and many properties of the B-spline function are clarified in de Boor [4]. We aim to obtain an estimator of f_j via the B-spline additive regression model

$$y_i = s_1(x_{i1}) + \cdots + s_D(x_{iD}) + \varepsilon_i \tag{2}$$

instead of model (1). Model (2) can be expressed as

$$\boldsymbol{y} = Z_1 \boldsymbol{b}_1 + \cdots + Z_D \boldsymbol{b}_D + \boldsymbol{\varepsilon} = Z\boldsymbol{b} + \boldsymbol{\varepsilon}$$

by using the notations $\boldsymbol{y} = (y_1 \ \cdots \ y_n)'$, $\boldsymbol{b}_d = (b_{d,-p+1} \cdots \ b_{d,K_n})'$, $\boldsymbol{b} = (\boldsymbol{b}_1' \ \cdots \ \boldsymbol{b}_D')'$, $Z_d = (B_{-p+j}(x_{id}))_{ij}$, $Z = [Z_1 \ \cdots \ Z_D]$ and $\boldsymbol{\varepsilon} = (\varepsilon_1 \ \cdots \ \varepsilon_n)'$. The estimator $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ is defined as the minimizer of

$$\begin{aligned} L(\boldsymbol{b}) &= (\boldsymbol{y} - Z\boldsymbol{b})'(\boldsymbol{y} - Z\boldsymbol{b}) + \sum_{j=1}^{D} \lambda_{jn} \boldsymbol{b}_j' Q_m \boldsymbol{b}_j \\ &= (\boldsymbol{y} - Z\boldsymbol{b})'(\boldsymbol{y} - Z\boldsymbol{b}) + \boldsymbol{b}' Q_m(\lambda_n) \boldsymbol{b}, \end{aligned} \tag{3}$$

where $\lambda_{jn} (j = 1, \cdots, D)$ are the smoothing parameters, Q_m is the mth order difference matrix and $Q(\lambda_n) = \text{diag}[\lambda_{1n} Q_m \ \cdots \ \lambda_{Dn} Q_m]$. This estimation method is called the penalized least squares method and it has been frequently utilized in spline regression. For a fixed point $x_j \in (0,1)$, the estimator $\hat{f}_j(x_j)$ of $f_j(x_j)$ is

$$\hat{f}_j(x_j) = \sum_{k=-p+1}^{K_n} B_k(x_j)\hat{b}_{j,k}$$

and is called the penalized spline estimator of $f_j(x_j)$. The predictor of y at a fixed point $\boldsymbol{x} \in (0,1)^D$ is defined as

$$\hat{y} = \hat{f}_1(x_1) + \cdots + \hat{f}_D(x_D). \tag{4}$$

Since $E[f_j(X_j)] = 0$ is assumed for f_j, the estimator of each component f_j is usually centered. Hence, $\hat{f}_j(x_j)$ is rewritten as

$$\hat{f}_{j,c}(x_j) = \hat{f}_j(x_j) - \frac{1}{n}\sum_{i=1}^{n} \hat{f}_j(x_{ij}),$$

as discussed in Wang and Yang [13]. In this paper, however, we do not examine $\hat{f}_{j,c}$ because our interests are in asymptotics for $\hat{f}_j$ and $\hat{y}$, and asymptotic distributions of $\hat{f}_j(x_j)$ and $\hat{f}_{j,c}(x_j)$ become equivalent.

2.2. **The Ridge Corrected Penalized Spline Estimator.** In general, $\hat{\boldsymbol{b}} = (\hat{\boldsymbol{b}}_1' \ \cdots \ \hat{\boldsymbol{b}}_D')'$ is a solution of

$$\frac{\partial L(\boldsymbol{b})}{\partial \boldsymbol{b}} = \boldsymbol{0}. \tag{5}$$

However, this method has one defect: the $L(\boldsymbol{b})$ is not strictly convex as a function of $\boldsymbol{b}$ in general. Hence, the solution of (5) does not necessarily become the minimizer of (3). Actually, because each column sum of Z_j equal to 1, $Z'Z + Q_m(\lambda_n)$, the Hessian matrix of $L(\boldsymbol{b})$, has eigenvalue 0. Marx and Eilers [9] also noted this point as a typical problem of the additive spline regression. They studied a new method such that the loss function has strict convexity for obtaining the estimator of $\boldsymbol{b}$. Let $L_\gamma(\boldsymbol{b}) = L(\boldsymbol{b}) + \gamma \boldsymbol{b}'\boldsymbol{b}$, where $\gamma > 0$ is very small constant. Since $L_\gamma(\boldsymbol{b})$ is strictly convex, the solution $\hat{\boldsymbol{b}}_\gamma = (\hat{\boldsymbol{b}}_{1,\gamma}' \ \cdots \ \hat{\boldsymbol{b}}_{D,\gamma}')'$ of

$$\frac{\partial L_\gamma(\boldsymbol{b})}{\partial \boldsymbol{b}} = \boldsymbol{0}$$

can be obtained uniquely as

$$\hat{\boldsymbol{b}}_\gamma = (Z'Z + Q(\lambda_n) + \gamma I)^{-1} Z'\boldsymbol{y}.$$

The RCPS of $f_j(x_j)$ can be written as

$$\hat{f}_{j,\gamma}(x_j) = \boldsymbol{B}(x_j)'\hat{\boldsymbol{b}}_{j,\gamma},$$

where $\boldsymbol{B}(x_j) = (B_{-p+1}(x_j) \ \cdots \ B_{K_n}(x_j))'$. If $\gamma = 0$, $\hat{\boldsymbol{b}}_\gamma$ cannot be calculated because $Z'Z + Q(\lambda_n)$ is not invertible.

2.3. **The Backfitting Penalized Spline Estimator.** The merit and usage of the backfitting algorithm are clarified in Hastie and Tibshirani [7]. The ℓ-stage backfitting estimator $\boldsymbol{b}_j^{(\ell)}$ of $\boldsymbol{b}_j$ is defined as

$$\boldsymbol{b}_j^{(\ell)} = \Lambda_j^{-1} Z_j'(\boldsymbol{y} - Z_1\boldsymbol{b}_1^{(\ell)} - \cdots - Z_{j-1}\boldsymbol{b}_{j-1}^{(\ell)} - Z_{j+1}\boldsymbol{b}_{j+1}^{(\ell-1)} - \cdots - Z_D\boldsymbol{b}_D^{(\ell-1)}),$$

where $\boldsymbol{b}_j^{(0)}$'s are initial values. The ℓ-stage backfitting estimator $f_j^{(\ell)}(x_j)$ of $f_j(x_j)$ at $x_j \in (0,1)$ is defined as

$$f_j^{(\ell)}(x_j) = \sum_{k=-p+1}^{K_n} B_k(x_j) b_{j,k}^{(\ell)} = \boldsymbol{B}(x_j)'\boldsymbol{b}_j^{(\ell)}, \quad j = 1, \cdots, D.$$

For $D = 2$, the explicit form of $\boldsymbol{b}_j^{(\ell)}$ can be obtained (see Yoshida and Naito [15]). However, for general D, the exact form of $\boldsymbol{b}_j^{(\ell)}$ is too complicated to be written down. A mathematical property of the backfitting algorithm is that $\boldsymbol{b}^{(\infty)} = ((\boldsymbol{b}_1^{(\infty)})', \cdots, (\boldsymbol{b}_D^{(\infty)})')' \equiv \lim_{\ell\to\infty}((\boldsymbol{b}_1^{(\ell)})', \cdots, (\boldsymbol{b}_D^{(\ell)})')'$ satisfies

$$\left.\frac{\partial L(\boldsymbol{b})}{\partial \boldsymbol{b}}\right|_{\boldsymbol{b}=\boldsymbol{b}^{(\infty)}} = \boldsymbol{0}.$$

It is shown by Theorem 9 of Buja et al. [2] that $\boldsymbol{b}_j^{(\ell)}$ converges to $\boldsymbol{b}_j^{(\infty)}$, but those $\boldsymbol{b}_j^{(\infty)}(j = 1, ..., D)$ are depending on initial values $\boldsymbol{b}_j^{(0)}(j = 1, ..., D)$. This means that the convergence property of the backfitting estimator is guaranteed, but the uniqueness of the solutions is not trivial. We will study the asymptotic behavior of $\hat{f}_j(x_j) = f_j^{(\infty)}(x_j) = \boldsymbol{B}(x_j)'\boldsymbol{b}_j^{(\infty)}$, as well as the relationship between $\hat{\boldsymbol{b}}_{j,\gamma}$ and $\boldsymbol{b}_j^{(\infty)}$.

3. Asymptotic theory

We prepare some symbols and notations to be used hereafter. Define the $(K_n + p) \times (K_n + p)$ square matrix $G_k = (G_{k,ij})_{ij}$ with its (i,j)-component

$$G_{k,ij} = \int_0^1 B_{-p+i}(x)B_{-p+j}(x)q_k(x)dx$$

for $k = 1, \cdots, D$ and the $(K_n + p) \times (K_n + p)$ square matrix $\Sigma_k = (\Sigma_{k,ij})_{ij}$ having the (i,j)-component

$$\Sigma_{k,ij} = \int_{[0,1]^D} \sigma^2(\boldsymbol{x})B_{-p+i}(x_k)B_{-p+j}(x_k)q(\boldsymbol{x})d\boldsymbol{x}$$

for $k = 1, \cdots, D$. Let $\boldsymbol{b}_j^*$ be a best L_∞ approximation to the true function f_j. This means that $\boldsymbol{b}_j^*$ satisfies

$$\sup_{x\in(0,1)} \left| f_j(x) + b_{j,a}(x) - \boldsymbol{B}(x)'\boldsymbol{b}_j^* \right| = o(K_n^{-(p+1)}),$$

where

$$b_{j,a}(x) = -\frac{f_j^{(p+1)}(x)}{K_n^{p+1}(p+1)!} \sum_{k=1}^{K_n} I(\kappa_{k-1} \le x < \kappa_k) \mathrm{B}_{p+1}\left(\frac{x - \kappa_{k-1}}{K_n^{-1}}\right),$$

$I(a < x < b)$ is the indicator function of an interval (a, b) and $\mathrm{B}_p(x)$ is the pth Bernoulli polynomial, see Zhou et al. [17]. For a random sequence U_n, $E[U_n|\boldsymbol{X}_n]$ and $V[U_n|\boldsymbol{X}_n]$ designate the conditional expectation and the conditional variance of U_n given $(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n) = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$, respectively.

In spline smoothing, the smoothing parameter λ_{jn} is usually selected as $\lambda_{jn} \to \infty$ with $n \to \infty$ because a spline curve often yields overfitting for large n. In the following, we assume that $\lambda_{jn} = o(nK_n^{-1})$.

In this section, first we discuss the asymptotic distribution of the RCPS. Next we show that the difference of the BPS and RCPS asymptotically vanishes, by which we finally obtain the asymptotic distribution of the BPS.

3.1. **Asymptotic distribution of the RCPS.** We will show the asymptotic property of $\hat{f}_{j,\gamma}(x_j)(j = 1, \cdots, D)$. By using the result of the partitioned matrix of $Z'Z + Q_m(\lambda_n) + \gamma I$ and its asymptotic property, the form of $\hat{f}_{j,\gamma}(x_j)$ and its asymptotic expression can be clarified. As a result, we obtain the following Theorem.

Theorem 3.1. *Let $f_j \in C^{p+1}(j = 1, \cdots, D)$. Suppose that $K_n = o(n^{1/2})$ and $\lambda_{jn} = o(n/K_n)(j = 1, \cdots, D)$. Then, for $j, k = 1, \cdots, D$,*

$$\begin{aligned} E[\hat{f}_{j,\gamma}(x_j)|\boldsymbol{X}_n] - f_j(x_j) &= b_{j,\lambda}(x_j) + b_{j,\gamma}(x_j) + o_P(K_n^{-1}) + o_P(\lambda_{jn}K_n n^{-1}), \\ V[\hat{f}_{j,\gamma}(x_j)|\boldsymbol{X}_n] &= \frac{1}{n}\boldsymbol{B}(x_j)'G_j^{-1}\Sigma_j G_j^{-1}\boldsymbol{B}(x_j)(1 + o_P(1)) = O_P(K_n n^{-1}), \\ Cov(\hat{f}_{j,\gamma}(x_j), \hat{f}_{k,\gamma}(x_k)) &= O_P(n^{-1}), \quad j \neq k \end{aligned}$$

where

$$b_{j,\lambda}(x) = -\frac{\lambda_{jn}}{n}\boldsymbol{B}(x_j)'G_j^{-1}Q_m\boldsymbol{b}_j^* = O(\lambda_{jn}K_n n^{-1})$$

and

$$b_{j,\gamma}(x_j) = -\frac{\gamma}{n}\boldsymbol{B}(x_j)'G_j^{-1}\boldsymbol{b}_j^* = O(K_n n^{-1}).$$

In Theorem 3.1, the influence of γ appears only in $b_{j,\gamma}(x_j)$, which is in fact of negligible order. Furthermore, compared to Theorem 2 of Claeskens et al. [3], $\hat{f}_{j,\gamma}(x_j)$ is asymptotically equivalent to the penalized spline estimator based on the dataset $\{(y_i, x_{ij}) : i = 1, \cdots, n\}$ in the univariate regression model. By using Theorem 3.1 and Lyapunov's condition of the central limit theorem, we obtain the asymptotic joint distribution of $[\hat{f}_{1,\gamma}(x_1) \ \cdots \ \hat{f}_{D,\gamma}(x_D)]'$.

Theorem 3.2. *Suppose that there exists $\delta \geq 2$ such that $E[|\varepsilon_i|^{2+\delta}|X_i = x_i] < \infty$ and $f_j \in C^{p+1}$. Furthermore, K_n and λ_{jn} satisfy $K_n = o(n^{1/2})$, $n^{1/3} = o(K_n)$ and $\lambda_{jn} = o((nK_n^{-1})^{1/2})$. Then, for any fixed point $\boldsymbol{x} \in (0, 1)^D$, as $n \to \infty$,*

$$\sqrt{\frac{n}{K_n}}\begin{bmatrix} \hat{f}_{1,\gamma}(x_1) - f_1(x_1) \\ \vdots \\ \hat{f}_{D,\gamma}(x_D) - f_D(x_D) \end{bmatrix} \xrightarrow{d} N_D(\boldsymbol{0}, \Psi),$$

where $\Psi = \mathrm{diag}[\psi_1(x_1) \ \cdots \ \psi_D(x_D)]$ *and*

$$\psi_j(x_j) = \lim_{n\to\infty} \frac{1}{K_n} \boldsymbol{B}(x_j)' G_j^{-1} \Sigma_j G_j^{-1} \boldsymbol{B}(x_j), \quad j = 1, \cdots, D.$$

3.2. **Asymptotic distribution of the BPS.** It is easily confirmed that the BPS $\hat{f}_j(x_j)$ can be expressed as

$$\hat{f}_j(x_j) = f_j^{(1)}(x_j) + \lim_{\ell\to\infty} \sum_{k=1}^{\ell-1} \{f_j^{(k+1)}(x_j) - f_j^{(k)}(x_j)\}.$$

We show that for all $k \in \mathbb{N}$, $f_j^{(k+1)}(x_j) - f_j^{(k)}(x_j)$ asymptotically vanishes, from which we find that $\hat{f}_j(x_j)$ is asymptotically dominated by $f_j^{(1)}(x_j)$ for $j = 1, \cdots, D$. These properties are summarized in the following two Propositions.

Proposition 3.3. *Let* $f_j \in C^{p+1}$. *Suppose that* $K_n = o(n^{1/2})$ *and* $\lambda_{jn} = o(K_n n^{-1})$. *Then, as* $n \to \infty$,

$$f_j^{(k+1)}(x_j) - f_j^{(k)}(x_j) = O_P(K_n^{-k}) + O_P\left(K_n^{-(k-1)}(K_n n)^{-1/2}\right), \quad k = 1, 2, \cdots.$$

Proposition 3.4. *Under the same assumption as Proposition 3.3, as* $n \to \infty$,

$$\hat{f}_j(x_j) = f_j^{(1)}(x_j) + O_P(K_n^{-1}) + o_P\left(\sqrt{\frac{K_n}{n}}\right).$$

By its simple form of $f_j^{(1)}(x_j)$, it is easy to show that $f_j^{(1)}(x_j)$ is asymptotically equivalent to $\hat{f}_{j,\gamma}(x_j)$. Thus, Propositions 3.3 and 3.4 yield that the asymptotic equivalence between $\hat{f}_j(x_j)$ and $\hat{f}_{j,\gamma}(x_j)$. Consequently, Theorem 3.2 implies the asymptotic distribution of BPS summarized as follows:

Theorem 3.5. *Under the same assumption as Theorem 3.2, for any fixed point* $\boldsymbol{x} \in (0,1)^D$, *as* $n \to \infty$,

$$\sqrt{\frac{n}{K_n}} \begin{bmatrix} \hat{f}_1(x_1) - f_1(x_1) \\ \vdots \\ \hat{f}_D(x_D) - f_D(x_D) \end{bmatrix} \xrightarrow{d} N_D(\boldsymbol{0}, \Psi),$$

where Ψ *is that given in Theorem 3.2.*

From Theorem 3.5, for $i \neq j$, $\hat{f}_i(x_i)$ and $\hat{f}_j(x_j)$ are asymptotically independent. Asymptotic normality and the independence of $\hat{f}_i(x_i)$ and $\hat{f}_j(x_j)$ in kernel smoothing also hold, as shown in Wand [12]. Thus, the penalized spline estimator and the kernel estimator for the additive model have the same asymptotic property. Asymptotic normality of $\hat{y}$ in (4) can be shown as a direct consequence of Theorem 3.5. Though the BPS depends on the initial value, the effect of the initial value on the distribution of the BPS vanishes as $n \to \infty$, which means that the uniqueness of the BPS is asymptotically satisfied. Furthermore, Theorem 3.6 indicates that $\hat{\boldsymbol{b}} = \boldsymbol{b}^{(\infty)}$ minimizes $L(\boldsymbol{b})$.

Theorem 3.6. *Let $H(L)$ be the Hessian matrix of $L(\boldsymbol{b})$. Then, $H(L)$ is asymptotically positive definite.*

We now give the optimal order of K_n and λ_{jn} in the context of minimization of MSE of $\hat{f}_j(x_j)$. Note that K_n controls the trade-off between the bias and the variance of the estimator, both of which have been obtained by Proposition 3.4 and Theorem 3.1.

Corollary 3.7. *Under the same assumption as Proposition 3.3, it follows that*

$$\mathrm{MSE}(\hat{f}_j(x_j)) = E[\{\hat{f}_j(x_j) - f_j(x_j)\}^2 | \boldsymbol{X}_n] = O_P\left(\left\{\frac{K_n\lambda_{jn}}{n} + \frac{1}{K_n}\right\}^2\right) + O_P\left(\frac{K_n}{n}\right).$$

Furthermore taking $K_n = O(n^{1/3})$ and $\lambda_{jn} = O(n^{\nu}), \nu \leq 1/3$ leads to the rate of convergence, $\mathrm{MSE}(\hat{f}_j(x_j)) = O_P(n^{-2/3})$.

Asymptotic normality in Theorem 3.2 also holds even if the optimal orders $K_n = O(n^{1/3})$ and $\lambda_{jn} = O(n^{\nu}), \nu \leq 1/3$ in Corollary 3.7 are utilized, however the mean of the asymptotic distribution is not zero in such cases, that is, the bias terms given in Theorem 3.1 do not vanish. We see that the centered asymptotic distribution can be obtained by the assumptions for the orders of K_n and λ_{jn} as given in Theorem 3.2.

4. Discussion

In this paper, the asymptotic behavior of the penalized spline estimators in the additive models was investigated. The BPS and the RCPS have been shown to be asymptotically equivalent. For practical purposes, we compare the RCPS with the BPS from the view point of computation. We have to calculate the inverse of $M_D = Z'Z + Q_m(\lambda_n) + \gamma I$ in order to obtain the RCPS. Since the size of M_D is $\{D(K_n + p)\} \times \{D(K_n + p)\}$, the computation of M_D^{-1} is $O(\{D(K_n + p)\}^3)$. On the other hand, the ℓ-stage backfitting algorithm requires $O(\ell D\{K_n+p\})$ computations. Therefore when D is large, the BPS can be computed more quickly than the RCPS. Even for $D = 2$, the RCPS requires $O(8(K_n + p)^3)$ computations which is larger than the BPS. The same conclusion holds for a large sample size n, which is also detailed in Hastie and Tibshirani [7] as the advantage of the backfitting algorithm. So it might be better to utilize the BPS in additive penalized spline smoothing.

On the other hand, it is known that the BPS can be constructed by using a blend of backfitting and scoring algorithms also in the generalized additive models (GAM). However, since the backfitting algorithm is required to be implemented within each iteration of the scoring algorithm, the computation of the BPS finally becomes a heavy task. Therefore in GAM, the RCPS, the direct method without the backfitting cycle, might be better and its asymptotic theory should be investigated. Actually, Yoshida and Naito [16] showed that the RCPS has the asymptotic normality in GAM. Such theoretical research is a generalization of the works by this paper and Kauermann et al. [8].

Acknowledgements

Almost all of the work was done while the author was visiting the Graduate School of Science and Engineering, Shimane University. The author would like to thank Prof. Kanta Naito for his suggestions, encouragements and insightful comments.

References

[1] Aerts, M., Claeskens, G. and Wand, M.P. Some theory for penalized spline generalized additive models. *Journal of Statistical Planning and Inference*, **103** (2002), 455-470.
[2] Buja,A.Hastie,T. and Tibshirani,R. Linear smoothers and additive models (with discussion). *Ann. Statist.* **17** (1989),453-555.
[3] Claeskens,G., Krivobokova,T. and Opsomer,J.D. Asymptotic properties of penalized spline estimators. *Biometrika.* **96** (2009), 529-544.
[4] de Boor,C. *A Practical Guide to Splines.* Springer-Verlag, 2001.
[5] Eilers,P.H.C. and Marx,B.D. Flexible smoothing with B-splines and penalties(with Discussion). *Statist.Sci.* **11** (1996), 89-121.
[6] Hall,P. and Opsomer,J.D. Theory for penalized spline regression. *Biometrika.* **92** (2005), 105-118.
[7] Hastie,T. and Tibshirani,R. *Generalized Additive Models.* London: Chapman & Hall, 1990.
[8] Kauermann,G., Krivobokova,T., and Fahrmeir,L. Some asymptotic results on generalized penalized spline smoothing. *J. R. Statist. Soc.* B **71** (2009), 487-503.
[9] Marx,B,D. and Eilers,P.H.C. Direct generalized additive modeling with penalized liklihood. *Comp. Statist & Data Anal.* **28** (1998), 193-209.
[10] O'Sullivan,F. A statistical perspective on ill-posed inverse problems(with discussion). *Statist. Sci.* **1** (1986), 505–27.
[11] Ruppert,D., Wand,M.P. and Carroll,R.J. *Semiparametric Regression.* Cambridge University Press, 2003.
[12] Wand,M.P. A central limit theorem for local polynomial backfitting estimators. *J. Mult. Anal.* **70** (1999), 57–65.
[13] Wang,L. and Yang,L. Spline-backfitted kernel smoothing of nonlinear additive auto regression model. *Ann. Statist.* **35** (2007), 2474-2503.
[14] Wang,X., Shen,J. and Ruppert,D. On the asymptotics of penalized spline smoothing. *Ele. J. Statist.* **5** (2011), 1-17.
[15] Yoshida,T. and Naito,K. Asymptotics for penalized additive B-spline regression. *J. Japan. Statist. Soc.* **42** (2012), 81-107.
[16] Yoshida,T. and Naito,K. Asymptotics for penalized splines in generalized additive models. Revised. (2013).
[17] Zhou,S., Shen,X. and Wolfe,D.A. Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** (1998), 1760-1782.

Takuma Yoshida: Graduate School of Science and Engineering, Kagoshima University, Kagoshima, Kagoshima, 890-8580, Japan
E-mail address: yoshida@sci.kagoshima-u.ac.jp