Electronic Journal of Statistics
Vol. 13 (2019) 747–794
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1541

Asymptotic theory of penalized splines

Luo Xiao

Department of Statistics
North Carolina State University
Raleigh, NC, USA
e-mail: lxiao5@ncsu.edu

Abstract: The paper gives a unified study of the large sample asymptotic theory of penalized splines including the O-splines using B-splines and an integrated squared derivative penalty [22], the P-splines which use B-splines and a discrete difference penalty [13], and the T-splines which use truncated polynomials and a ridge penalty [24]. Extending existing results for O-splines [7], it is shown that, depending on the number of knots and appropriate smoothing parameters, the L_2 risk bounds of penalized spline estimators are rate-wise similar to either those of regression splines or to those of smoothing splines and could each attain the optimal minimax rate of convergence [32]. In addition, convergence rate of the L_∞ risk bound, and local asymptotic bias and variance are derived for all three types of penalized splines.

MSC 2010 subject classifications: 62G08, 62G20.
Keywords and phrases: Nonparametric regression, penalized splines, L_∞ convergence, L_2 convergence, local asymptotics, rate optimality.

Received September 2018.

Contents

1. Introduction

Penalized spline smoothing has become popular in the last two decades. The approach uses a flexible choice of bases and penalty and is often viewed as a bridge between regression splines and smoothing splines. Indeed, penalized splines exploit the mixed effect presentation of smoothing splines but are computationally much simpler because they use low rank bases; penalized splines inherit the computational simplicity of regression splines but relieve the overfitting of regression splines as they employ a smoothness penalty. Therefore, penalized splines have enjoyed widespread use in methodology development and applications. For example, penalized spline methods have been well developed in functional data analysis, e.g., [44, 15, 43, 40]. Another example is penalized splines methods for generalized additive models [38]. See [24] for a comprehensive introduction to penalized splines. The paper [25] gives a review of penalized splines in the 2000 decade and most recently, the paper [14] provides a review of P-splines [13], one popular type of penalized splines.

Theoretic understanding of penalized splines has been largely lagging behind. The flexibility of penalized splines in terms of usage of bases, placement of knots and choice of penalty actually dramatically increases the difficulty of the theoretic study of penalized splines. Indeed, the O-splines [22] use B-splines as bases and impose an integrated squared derivative penalty to control overfit, the P-splines [13] also use B-splines but directly impose a penalty on the associated coefficients, and the T-splines [24] use truncated polynomials and impose a ridge-type penalty on the associated coefficients. While it seems a consensus in the literature that all three types of penalized splines have similar practical performance, existing theoretic works usually focus on one type. The paper [7] is a seminal work and shows that, for the L_2 risk bound, depending on the number of knots and the penalty, O-splines behave similar to either regression splines or smoothing splines. Specifically, when the penalty is appropriately small, the number of knots for constructing the B-spline bases determines the convergence rate, leading to asymptotics rate-wise similar to those of regression splines; but when the penalty is large, the number of knots does not matter as long as it is sufficiently large and the penalty determines the convergence rate, leading to asymptotics rate-wise similar to these of smoothing splines. In particular, the optimal number of knots for the regression spline type asymptotics is rate-wise no bigger and can be much smaller than that required for the smoothing spline type asymptotics; thus, for penalized splines, the regression spline type asymptotics is referred to as the small number of knots scenario while the smoothing spline type asymptotics is referred to as the large number of knots scenario. It is interesting to study if such two-scenario asymptotics will also hold for P-splines and T-splines.

Table 1 gives a summary of existing theoretic works on penalized splines. As mentioned earlier, the paper [7] derives the L_2 risk bound of O-splines for both the small and the large numbers of knots. In addition, the work gives the local asymptotics of O-splines and T-splines for the small number of knots. The local asymptotics for the scenario of large number of knots is quite challenging to derive and equivalent kernel methods [30] have been adopted in the literature. Indeed, the papers [21], [37] and the unpublished [41] use equivalent kernel methods for P-splines, while the paper [27] studies O-splines. Note that all those works assume equally-spaced design points which is quite stringent. Another note is that, for the large number of knots scenario, P-splines have a slower convergence rate near the boundary [37]. Similar boundary behaviors of O-splines and T-splines have not been established yet. Early theoretic works of penalized splines also include [17] and [18]. For generalized additive models, the paper [45] studies the local asymptotics of a ridge-corrected P-spline estimator. For longitudinal data, the paper [5] studies the local asymptotics and the L_2 convergence rate of O-splines. Theoretic work on penalized splines in two dimensions include [42], [20] and [39].

Many theoretic gaps, e.g., the convergence rate of L_∞ risk bound, still remain and the paper intends to fill a number of the gaps. The checkmarks in Table 1 indicate the contributions of the paper. First, the L_2 convergence rates of P-splines and T-splines shall be established for both scenarios, extending [7]. Second, the L_∞ convergence rates of all three types of penalized splines shall be established and the rates are optimal for the small number of knots scenario. Third, the local asymptotic bias and variance of P-splines and T-splines will also be given.

TABLE 1

*Summary of theoretic works on penalized splines. The checkmarks indicate the major contributions of the paper. Note that the * around the ✓ means that the newly derived convergence rates for the bias might not be rate optimal. The details are provided in the theorems or remarks in parentheses.*

	Small number of knots		
	O-splines	P-splines	T-splines
L_2 (Rem 5.3)	[7]	✓(Rem 5.6)	✓(Rem 5.7)
L_∞ (Thm 6.1)	✓(Rem 6.5)	✓(Rem 6.5)	✓(Rem 6.5)
Local (Thm 7.1)	[7]	✓(Rem 7.1)	[7]
	Large number of knots		
	O-splines	P-splines	T-splines
L_2 (Rem 5.3)	[7]	✓(Rem 5.6)	✓(Rem 5.7)
L_∞ (Thm 6.1)	✓* (Rem 6.6)	[37]	✓* (Rem 6.6)
Local (Thm 7.1)	[27]	[21], [37], [41]	✓* (Rem 7.2)

In order to obtain the new theoretic results, two key observations are made and three key results are established. We first list the two observations.

- The three types of penalized splines differ essentially only in the penalty matrix being used and the three penalty matrices are inherently very similar to D, where D is a difference penalty matrix defined in Section 2; see also Section 4.
- The local variance of the three types of penalized splines can be studied via the bounds on the diagonals of $(I + \eta D)^{-1}$, where $\eta > 0$ is a scalar and D is the difference penalty matrix mentioned above.

Now we list the three key results.

- The decay rates of the eigenvalues of the three different penalty matrices are derived; see Propositions 4.1 and 4.2. These results show that the eigenvalues of the penalty matrices have similar decay rates.
- A property of penalized splines is established in Propositions 4.3, 4.4 and 4.5 for all three types of penalized splines.
- The rates of the local asymptotic variance of penalized splines are established for all three types of penalized splines; see Propositions 6.1, 6.2 and 6.3.

These three results ensure that a unified theoretic study is attainable. Finally, the paper provides several new theoretic results regarding the approximation accuracy of B-splines, which are useful for the theoretic derivations; see Section 3.

The rest of the paper is organized as follows. In Section 2, we introduce penalized splines. In Section 3, we consider the approximation accuracy of B-splines. In Section 4, we study the singular properties of penalized splines. In Section 5, we derive the L_2 convergence rate of penalized splines. In Section 6, we derive the L_∞ convergence rate of penalized splines. In Section 7, we derive the local asymptotic bias and variance of penalized splines. Proofs of theorems are provided in Section 8. Technical lemmas are given in Appendix B.3, the local variance of penalized splines is studied in Appendix B, and lower and upper risk bounds on the variance of penalized splines are derived in Appendix C.

We shall use the following notation convention. For a vector $\mathbf{a} = (a_k)$, $\|\mathbf{a}\|_2$ denotes its Euclidean norm and $\|\mathbf{a}\|_{\max} = \max_k |a_k|$. For a matrix $A = (A_{k\ell})$, $\|A\|_2$ is its operator norm, $\|A\|_{\max} = \max_{k\ell} |A_{k\ell}|$, $\|A\|_F$ is its Frobenius norm and $\|A\|_\infty = \max_k \sum_\ell |A_{k\ell}|$. For two square matrices A and B, $A \leq B$ means that $B - A$ is positive semidefinite. For a function $g(x)$ defined over an interval $\mathcal{T}$, we let $\|g\|$ be its supremum norm over $\mathcal{T}$, i.e., $\|g\| = \sup_{x\in\mathcal{T}} |g(x)|$, and we shall use $g^{(i)}(x)$ to denote its i^{th} derivative. We also use the notation $a \sim b$ to denote that $\lim_{n\to\infty} a/b = c$ for some constant $c > 0$.

2. Penalized splines

Consider the nonparametric regression problem

$$y_i = f(x_i) + e_i, \; i = 1, \dots, n,$$

where the n design points $x_i \in \mathcal{T} = [0, 1]$ can be either deterministic or random, y_i are the observed responses, and e_i are random errors. Let p be a fixed positive integer. It is assumed that $f \in \mathcal{C}^p(\mathcal{T})$, the class of functions with continuous p^{th} derivatives over $\mathcal{T}$. The aim is to estimate the unknown smooth function f by penalized splines. We introduce three types of penalized splines that are commonly used and then formulate a unified estimator that contains all of them.

2.1. O-splines

We first introduce splines [8]. A spline is a piecewise polynomial that is smoothly connected at its knots. More specifically, for a fixed integer m, denote by $\mathcal{S}(m, \underline{t})$ the set of spline functions with knots $\underline{t} = \{0 = t_0 < t_1 < \ldots < t_{K_0+1} = 1\}$. For $m = 1$, $\mathcal{S}(m, \underline{t})$ is the set of step functions with jumps at the knots and, for $m \geq 2$,

$$\begin{aligned}\mathcal{S}(m, \underline{t}) =&\{s \in C^{m-2}(\mathcal{T}) : s \text{ is a polynomial of degree} \\ &(m-1) \text{ on each interval } [t_i, t_{i+1}]\}.\end{aligned}$$

A basis for $\mathcal{S}(m, \underline{t})$ can be formed by B-splines, which are defined as

$$N_k^{[m]}(x) = (t_k - t_{k-m})[t_{k-m}, \ldots, t_k](t - x)_+^{m-1}, \quad 1 \leq k \leq K = K_0 + m,$$

where $(t - x)_+ = t - x$ if $t > x$ and 0 otherwise, $[t_{k-m}, \ldots, t_k](t - x)_+^{m-1}$ denotes the m^{th} order divided difference of $(t - x)_+^{m-1}$ as a function of t [8] and $t_{1-m} \leq \cdots \leq t_{-1} \leq t_0$, $t_{K_0+1} \leq t_{K_0+2} \leq \cdots \leq t_K$. The B-spline basis functions can also be recursively defined as $N_k^{[m]}(x) = \tilde{N}_{k-m}^{[m]}(x)$ with

$$\begin{aligned}\tilde{N}_k^{[1]}(x) &= \begin{cases} 1, & t_k \leq x < t_{k+1}, \\ 0, & \text{otherwise,}\end{cases} \\ \tilde{N}_k^{[m]}(x) &= \frac{x - t_k}{t_{k+m-1} - t_k}\tilde{N}_k^{[m-1]}(x) + \frac{t_{k+m} - x}{t_{k+m} - t_{k+1}}\tilde{N}_{k+1}^{[m-1]}(x),\end{aligned}$$

for $k = -(m-1), \ldots, K_0$. Here $0/0 = 0$. Then any spline function $s(x) \in \mathcal{S}(m, \underline{t})$ can be written as $\sum_{k=1}^K \theta_k N_k^{[m]}(x)$ with some scalars θ_k.

The O-spline estimator [22] is defined to be a spline function

$$\hat{f}_O \equiv \arg \min_{s \in \mathcal{S}(m, \underline{t})} \left[\frac{1}{n}\sum_{i=1}^n \{y_i - s(x_i)\}^2 + \lambda_O \int \left\{s^{(q)}(x)\right\}^2 dx\right], \tag{2.1}$$

where $q < m$ is a fixed integer, $s^{(q)}$ denotes the q^{th} derivative of s, and $\lambda_O \geq 0$ is a smoothing parameter. For O-splines, it is assumed that $t_{1-m} = \cdots = t_{-1} = t_0 = 0$ and $t_K = \cdots = t_{K_0+2} = t_{K_0+1} = 1$.

The derivative of a spline function is closely related to the difference operators. Let $\Delta_{K,1} \in \mathbb{R}^{(K-1)\times K}$ be the first order difference operator such that, for a vector $\theta \in \mathbb{R}^K$, $\Delta_{K,1}\theta = (\theta_2 - \theta_1, \ldots, \theta_K - \theta_{K-1})^T$. For $1 < q < K$, let

$\Delta_{K,q} \in \mathbb{R}^{(K-q)\times K}$ be the q^{th} order difference operator that is defined recursively as $\Delta_{K,q} = \Delta_{K-1,q-1}\Delta_{K,1}$. To simplify notation, denote $N_k^{[m]}(x)$ by $N_k(x)$ and let $N(x) = \{N_1(x), \dots, N_K(x)\}^T \in \mathbb{R}^K$. Similar to [46, equality (40)], we derive that

$$\frac{dN^{[m]}(x)}{dx} = \Delta_{K,1}^T W_K^{[m]} N^{[m-1]}(x),$$

where $W_K^{[m]} \in \mathbb{R}^{(K-1)\times(K-1)}$ is a diagonal matrix with the k^{th} diagonal element $(m-1)(t_k - t_{k-m+1})^{-1}$. Let $\tilde{\Delta}_{K,1,m} = W_K^{[m]}\Delta_{K,1} \in \mathbb{R}^{(K-1)\times K}$ be a weighted first order difference operator and define recursively $\tilde{\Delta}_{K,q,m} = \tilde{\Delta}_{K-1,q-1,m-1}\tilde{\Delta}_{K,1,m}$. To extend the definition to the case $q = m$ for T-splines to be introduced later, we let $W_K^{[1]} = h^{-1}I$. For simplicity, we suppress the dependence of the weighted operators $\tilde{\Delta}_{K,q,m}$ on $\underline{t}$. We obtain that

$$\frac{d^q N^{[m]}(x)}{dx^q} = \tilde{\Delta}_{K,q,m}^T N^{[m-q]}(x)$$

and hence

$$s^{(q)}(x) = \theta^T N^{(q)}(x) = \theta^T \tilde{\Delta}_{K,q,m}^T N^{[m-q]}(x). \tag{2.2}$$

Thus, $\int \left\{s^{(q)}(x)\right\}^2 dx = \theta^T P_q \theta$, where

$$O_q = \tilde{\Delta}_{K,q,m}^T G^{[m-q]} \tilde{\Delta}_{K,q,m} \tag{2.3}$$

with $G^{[m-q]} = \int N^{[m-q]}(x) N^{[m-q],T}(x) dx \in \mathbb{R}^{(K-q)\times(K-q)}$.

Let $Y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$ and $N = [N(x_1), \dots, N(x_n)]^T \in \mathbb{R}^{n\times K}$. It follows that the minimizer of (2.1) is $\hat{f}_O(x) = N^T(x)\hat{\theta}$ with

$$\hat{\theta} = \arg\min_{\theta} \left(\frac{1}{n} \|Y - N\theta\|_2^2 + \lambda_O \theta^T O_q \theta \right).$$

Therefore, $\hat{\theta} = \left(N^T N/n + \lambda_O O_q\right)^{-1} (N^T Y/n)$ and

$$\hat{f}_O(x) = N^T(x) \left(N^T N/n + \lambda_O O_q\right)^{-1} (N^T Y/n). \tag{2.4}$$

2.2. P-splines

The P-splines [13] imposes a penalty directly on the q^{th} order consecutive difference of the coefficient vector θ. Specifically, the P-spline estimator is also a spline function

$$\hat{f}_P \equiv \arg \min_{s\in\mathcal{S}(m,\underline{\tilde{t}})} \left[\frac{1}{n} \sum_{i=1}^n \{y_i - s(x_i)\}^2 + \lambda_P \theta^T D_{K,q} \theta \right],$$

where $D_{K,q} = \Delta_{K,q}^T \Delta_{K,q} \in \mathbb{R}^{K\times K}$, λ_P is also a smoothing parameter, and the set of spline functions $\mathcal{S}(m, \underline{\tilde{t}})$ is defined over equally-spaced knots, i.e., $\underline{\tilde{t}}$ contains knots with $t_i = i/(K_0+1), 1-m \le i \le K$. With slight abuse of notation, we still denote the corresponding B-spline basis functions by $N(x)$. The difference in the bases is minor for the theoretic study as we will discuss later. Then the P-spline estimator, denoted by $\hat{f}_P(x)$, takes the following form

$$\hat{f}_P(x) = N^T(x)\left(N^T N/n + \lambda_P D_{K,q}\right)^{-1}(N^T Y/n). \tag{2.5}$$

The difference penalty is effectively a smoothness penalty. Indeed, when $\theta^T D_{K,q}\theta = 0$, the resulting estimate reduces to a polynomial of degree $q-1$.

2.3. T-splines

Finally, we introduce the T-splines [24]. Let $\underline{t}$ be as in the O-splines and $F(x) = \{1, x, \cdots, x^{m-1}, (x-t_1)_+^{m-1}, \ldots, (x-t_{K_0})_+^{m-1}\} \in \mathbb{R}^K$. The T-splines is the estimator

$$\hat{f}_T \equiv \arg\min_{\theta \in \mathbb{R}^K} \left[\frac{1}{n}\sum_{i=1}^n \left\{y_i - F^T(x_i)\theta\right\}^2 + \lambda_T \theta^T \tilde{I}_{K,m}\theta\right],$$

where λ_T is a smoothing parameter and $\tilde{I}_{K,m} = \text{blockdiag}(\mathbf{0}_{m,m}, I_{K-m})$. We derive that

$$\hat{f}_T(x) = F^T(x)(F^T F/n + \lambda_T \tilde{I}_{K,m})^{-1}(F^T Y/n),$$

where $F = [F(x_1), \ldots, F(x_n)]^T \in \mathbb{R}^{n\times K}$. There exists an invertible transformation matrix $L_{K,m} \in \mathbb{R}^{K\times K}$ that depends only on $\underline{t}$ and such that $N(x) = L_{K,m}^T F(x)$, where $N(x)$ denote the B-spline bases for O-splines. Thus,

$$\hat{f}_T(x) = N^T(x)(N^T N/n + \lambda_T \tilde{D}_{K,m})^{-1}(N^T Y/n), \tag{2.6}$$

with $\tilde{D}_{K,m} = L_{K,m}^T \tilde{I}_{K,m} L_{K,m}$. We derive that (see Lemma A.8),

$$\tilde{D}_{K,m} = h^2\left\{\frac{1}{(m-1)!}\right\}^2 \tilde{\Delta}_{K,m,m}^T \tilde{\Delta}_{K,m,m}, \tag{2.7}$$

which can be thought of as an extension of the penalty matrix O_q for O-splines to the case $q = m$. Indeed, O_q and the O-splines can only be defined for $q < m$.

2.4. A unified penalized spline estimator

Comparing the three penalized splines estimators in (2.4), (2.5) and (2.6), we see that the main difference between them is the penalty matrix. In Section 4, we shall show that the penalty matrices (after adjusting the corresponding smoothing parameters) have eigenvalues of similar decay rates, which paves the way for a unified theoretic study of all three estimators. This motivates us to consider the following estimator

$$\hat{f}(x) = N^T(x)(N^T N/n + \lambda P_q)^{-1}(N^T Y/n), \tag{2.8}$$

where $P_q \in \mathbb{R}^{K\times K}$ is an arbitrary positive semi-definite matrix and two assumptions (Assumptions 3 and 4) on P_q will be made in Section 4. These two assumptions are satisfied by each type of penalized splines. We shall study the L_2 convergence rate, L_∞ convergence as well as local asymptotics of the unified penalized spline estimator $\hat{f}$ and then apply the theoretic results to the three types of penalized splines.

3. Spline approximation

In this section we establish some necessary results on the approximation accuracy of a smooth function by splines. We make the following assumption.

Assumption 1. *$K \geq n^\delta$ for some $\delta > 0$ and $K = o(n)$.*

We next specify some conditions on the placement of knots. Let $h_i = t_{i+1} - t_i$ and $h = \max_{0\leq i\leq K_0} h_i$. We assume that

$$\max_{0\leq i\leq K_0-1} |h_{i+1} - h_i| = o(K^{-1}) \quad \text{and} \quad \frac{h}{\min_{0\leq i\leq K_0} h_i} \leq M, \tag{3.1}$$

where M is a fixed constant. The same assumptions on the knots can be found in [46]. In many works, e.g., [3], the knots are assumed to be generated from a positive density, which will lead to (3.1). Therefore, (3.1) is slightly more general. Note that (3.1) implies that $h \sim K^{-1}$, i.e., h and K^{-1} are rate-wise equivalent.

If the design points $\underline{x} = (x_1, \ldots, x_n)$ are deterministic, we assume that

$$\|Q_n - Q\| = o(h), \tag{3.2}$$

where $Q_n(x) = n^{-1}\sum_{i=1}^n I(x_i \leq x)$ is the empirical cumulative distribution function and $Q(x)$ is a distribution with a positive and continuously differentiable density $\rho(x)$. Assumption (3.2) is also common; see, e.g., [46].

Denote by $B_k(x)$ the k^{th} Bernoulli polynomial function, i.e.,

$$B_k(x) = x^k + \binom{k}{1} B_1 x^{k-1} + \cdots + B_k, \quad x \in \mathcal{T},$$

where the B_ks are the Bernoulli numbers satisfying the following: for $k \geq 1$, $\int_{\mathcal{T}} B_k(x)dx = 0$ and for $k \geq 2$,

$$B_k^{(1)}(x) = kB_{k-1}(x), \quad B_k^{(j)}(0) = B_k^{(j)}(1),\, 0 \leq j \leq k-2.$$

As $f \in \mathcal{C}^p(\mathcal{T})$, define, whenever $p \geq m$, for $0 \leq k \leq K-1$,

$$b_f(x) = -f^{(m)}(t_k)\frac{h_k^m}{m!} B_m\left(\frac{x - t_k}{h_k}\right),\ t_k \leq x < t_{k+1}.$$

We use the notation $b_f^{(i)}(x)$ with values at the knots being the right derivatives.

Lemma 3.1. *If $f \in \mathcal{C}^p(\mathcal{T})$ with $p = m$, then there exists a spline $s_f \in \mathcal{S}(m, \underline{t})$ such that*

$$\left\| f^{(i)} - s_f^{(i)} + b_f^{(i)} \right\| = o\left(h^{m-i}\right), \ i = 0, 1, \ldots, m-2,$$

and for $i = m-1$,

$$\left\| f^{(m-1)} - s_f^{(m-1)} + b_f^{(m-1)} \right\| = O(h),$$

and

$$\left\| f^{(m-1)} - s_f^{(m-1)} + b_f^{(m-1)} \right\|_{L_2} = o(h).$$

Remark 3.1. *The lemma is adapted from Lemma 1 of [4] and can be easily extended to prove that: if $f \in \mathcal{C}^p(\mathcal{T})$ with $p < m$, then there exists a spline $s_f \in \mathcal{S}(m, \underline{t})$ such that,*

$$\left\| f^{(i)} - s_f^{(i)} \right\| = o\left(h^{p-i}\right), \ i = 0, 1, \ldots, \min(p, m-2).$$

and $\left\| f^{(p)} - s_f^{(p)} \right\|_{L_2} = o(1)$. To see this, let $g(x) = \sum_{k=0}^{p} f^{(k)}(0)x^k/k!$ which satisfies $\|f^{(i)} - g^{(i)}\| = o(h^{p-i})$ and apply Lemma 3.1 to g which gives $b_g(x) = 0$.

Remark 3.2. *The results in Lemma 3.1, Remark 3.1 and Lemma 3.2 for $i \geq 1$ are important for the theoretic study of O-splines which involves the derivative penalty of the spline estimators. To our best knowledge, such results have not been formally stated in the statistics literature.*

We denote the spline $s_f(x)$ in Lemma 3.1 by $\sum_k \beta_k N_k(x)$ and let $\beta = (\beta_1, \ldots, \beta_K) \in \mathbb{R}^K$. Lemma 3.2 below further characterizes the accuracy of s_f for approximating f.

Lemma 3.2. *Suppose that (3.1) holds and $f \in \mathcal{C}^{(p)}(\mathcal{T})$ with $p \leq m$. For $0 \leq i \leq \min(p, m-1)$ and $r \geq 1$,*

$$\max_k \left| \int_{\mathcal{T}} N_k^{[r]}(x) \left\{ f^{(i)}(x) - s_f^{(i)}(x) \right\} dQ(x) \right| = o\left(h^{p+1-i}\right).$$

Lemma 3.3. *Suppose that the assumptions in Lemma 3.2 hold. For the fixed design, if (3.2) holds, then*

$$\max_k \left| \int_{\mathcal{T}} N_k(x) \left\{ f(x) - s_f(x) \right\} dQ_n(x) \right| = o\left(h^{p+1}\right).$$

For the random design where the design points $\underline{x}$ are randomly sampled from $Q(x)$, if $K = o(n^{\delta^*})$ for some $\delta^* \in (0, 1/2)$, then

$$\max_k \left| \int_{\mathcal{T}} N_k(x) \left\{ f(x) - s_f(x) \right\} dQ_n(x) \right| = o\left(h^{p+1}\right), \ a.s.$$

Remark 3.3. *The fixed design case in Lemma 3.3 was proved in [1, Lemma 6.10].*

4. Properties of penalty of penalized splines

In this section we establish the properties of penalty of penalized splines. For smoothing splines, the eigenvalues of penalty play a fundamental role in the study of asymptotic properties; see, e.g., [34] and [31]. Given the similarity of penalized splines and smoothing splines, it seems not unreasonable to believe that such an approach can be extended for the theoretic study of penalized splines. Indeed, a number of theoretic studies of penalized splines used results on eigenvalues for smoothing splines albeit without a formal proof. Because smoothing splines use polynomial splines with some boundary conditions [36] while penalized splines do not satisfy those conditions; see, e.g., [37] for P-splines, a formal proof on the eigenvalues of penalty of penalized splines is needed but does not exist as far as we are aware of. We shall fill the gap and establish that the three penalty matrices of the respective penalized spline estimators introduced in Section 2 have eigenvalues of similar decay rates.

The assumptions on the knots and design points are summarized as below.

Assumption 2. *For the B-spline bases, assume that the interior knots are equally-spaced with $t_i = i/(K_0+1), 0 \le i \le K_0+1$. For the fixed design points,* (3.2) *holds.*

Remark 4.1. (3.1) *holds under Assumption 2 and $h = 1/(K_0+1)$.*

A square matrix $A = (a_{ij})$ is said r-banded if $a_{ij} = 0$ whenever $|i-j| > r/2$. If r is a finite number, we say that A has a finite band. For a symmetric matrix A, denote by $\lambda_k(A)$ its k^{th} smallest eigenvalue. We first derive the spectrum of the penalty matrix $D_{K,q}$ for P-splines. Note that $\lambda_k\left(D_{K,q}\right) = \lambda_k\left(O_q\right) = 0$ for $k \le q$ and $\lambda_k\left(\tilde{D}_{K,m}\right) = 0$ for $k \le m$.

Proposition 4.1. *Suppose that Assumption 1 holds. The matrix $D_{K,q}$ is $(2q)$-banded and there exists a constant $C_1 > 1$ that depends only on q and such that for $q+1 \le k \le K$,*

$$C_1^{-1}\left(\frac{k-q}{K}\right)^{2q} \le \lambda_k(D_{K,q}) \le C_1\left(\frac{k}{K}\right)^{2q}.$$

Proposition 4.2. *Suppose that Assumptions 1 and 2 hold. The matrices O_q and $\tilde{D}_{K,m}$ are both $(2m)$-banded. And there exists a constant $C_2 > 1$ that depends only on q and m and such that for $q+1 \le k \le K$,*

$$C_2^{-1}\left(\frac{k-q}{K}\right)^{2q} \le \lambda_k\left(h^{2q-1}O_q\right) \le C_2\left(\frac{k}{K}\right)^{2q}$$

and for $m+1 \le k \le K$,

$$C_2^{-1}\left(\frac{k-m}{K}\right)^{2m} \le \lambda_k\left(h^{2m-2}\tilde{D}_{K,m}\right) \le C_2\left(\frac{k}{K}\right)^{2m}.$$

Remark 4.2. *Propositions 4.1 and 4.2 show that the eigenvalues of $h^{2q-1}O_q$ and $D_{K,q}$ have the same decay rate and the eigenvalues of $h^{2m-2}\tilde{D}_{K,m}$ and $D_{K,m}$ have the same decay rate.*

Remark 4.3. *Results on the singular values of O_q without a rigorous proof has been used in the penalized spline literatures; see, e.g., [7] and [5]. To our best knowledge, our proof is the first one.*

Remark 4.4. *The results are comparable to the eigenvalues of smoothing splines [34] and our proofs are also similar to those in [34].*

Propositions 4.3, 4.4 and 4.5 give another useful property of the various penalized splines.

Proposition 4.3. *Suppose that Assumptions 1-2 hold. Assume that $f \in \mathcal{C}^p(\mathcal{T})$ with $q \le p \le m$ and $q < m$. Then*

$$\|O_q\beta\|_{\max} = O(K^{q-1}), \quad \beta^T O_q \beta = O(1),$$

where β is defined in Section 3 so that $s_f(x) = N^T(x)\beta$.

Proposition 4.4. *Suppose that Assumptions 1-2 hold. Assume that $f \in \mathcal{C}^m(\mathcal{T})$. Then*

$$\left\|h^{-1}\tilde{D}_{K,m}\beta\right\|_{\max} = O(K^{m-1}), \quad \beta^T\left(h^{-1}\tilde{D}_{K,m}\right)\beta = O(1).$$

Proposition 4.5. *Suppose that Assumptions 1-2 hold. Assume that $f \in \mathcal{C}^p(\mathcal{T})$ with $q \le p \le m$ and $q \le m$. Then*

$$\left\|h^{1-2q}D_{K,q}\beta\right\|_{\max} = O(K^{q-1}), \quad \beta^T\left(h^{1-2q}D_{K,q}\right)\beta = O(1).$$

5. L_2 convergence rate

The L_2 convergence rate of O-splines was first proved in [7] and similar proofs can be adopted to establish the L_2 convergence of the unified penalized spline estimator. Then the L_2 convergence of P-splines and T-splines can be established. However, it is worth noting that the adaption is not entirely trivial because O-splines only allows $q < m$, but T-splines uses essentially an m^{th} order penalty and we also allow $q = m$ for P-splines.

Assumptions 3 and 4 below summarize the key properties of penalty of all three types of penalized splines derived in Section 4.

Assumption 3. *Suppose that P_q is a symmetric and positive semi-definite square matrix with a finite band that depends only on q and m and satisfies: $\lambda_q(P_q) = 0$ and there exists a constant $C_3 > 1$ that depends only on q and m and such that for $q+1 \le k \le K$,*

$$C_3^{-1}\left(\frac{k-q}{K}\right)^{2q} \le \lambda_k\left(h^{2q-1}P_q\right) \le C_3\left(\frac{k}{K}\right)^{2q}.$$

Assumption 4. $\beta^T P_q \beta = O(1)$.

Let $G_n = N^T N/n$ and $H_n = G_n + \lambda P_q/n$. Define $G = \int N(x)N^T(x)\rho(x)dx \in \mathbb{R}^{K\times K}$ and $H = G + \lambda P_q/n$. Let $A = NG_n^{-\frac{1}{2}}$ and $\tilde{P}_q = G_n^{-\frac{1}{2}} P_q G_n^{-\frac{1}{2}}$. Let $f = \{f(x_1), \ldots, f(x_n)\}^T \in \mathbb{R}^n$ and $\hat{f} = \{\hat{f}(x_1), \ldots, \hat{f}(x_n)\}^T \in \mathbb{R}^n$. Then $\hat{f} = A(I_K + \lambda\tilde{P}_q)^{-1}A^T Y$. The singular values of $\tilde{P}_q$ plays an indispensable role in studying the L_2 convergence of $\hat{f}(x)$ [7]. Lemma 5.1 can be derived from Assumption 3 and A.1.

Lemma 5.1. *Suppose that Assumptions 1 - 3 hold. Then* $\lambda_k\left(\tilde{P}_q\right) = 0$ *for* $k \le q$ *and there exists a constant* $C_4 > 1$ *such that for* $q+1 \le k \le K$,

$$C_4^{-1} h^{-2q} \left(\frac{k-q}{K}\right)^{2q} \le \lambda_k\left(\tilde{P}_q\right) \le C_4 h^{-2q}\left(\frac{k}{K}\right)^{2q}.$$

Assumption 5. $\lambda = o(1)$.

Lemma 5.2. *Suppose that Assumptions 1 - 3 and 5 hold. Then*

$$\left\|\left(I + \lambda\tilde{P}_q\right)^{-1}\right\|_F^2 = O(h_e^{-1}),$$

where $\|\cdot\|_F$ *is the Frobenius norm and* $h_e = \max(h, \lambda^{\frac{1}{2q}})$.

Assumption 6. *The random errors* e_i *are independent from* x_i *and are i.i.d. with mean 0 and* $\mathbb{E}|e_i|^\tau < \infty$ *for some constant* $\tau > 2$.

We shall only give results for the fixed design. However, the extension to the random design is straightforward and the only additional assumption required is $K = O\left(n^{\delta^*}\right)$ for some $\delta^* \in (0, \frac{1}{2})$.

Theorem 5.1 (L_2 convergence)**.** *Suppose that Assumptions 1 - 6 hold. If* $f \in \mathcal{C}^p(\mathcal{T})$ *with* $q \le p \le m$, *then*

$$\mathbb{E}\left(\left\|\hat{f} - f\right\|_{L_2}^2\right) = O\left(K^{-2m}\right) + o(K^{-2p}) + O\left\{\min\left(\lambda^2 K^{2q}, \lambda\right)\right\} + O\left(\frac{1}{nh_e}\right),$$

where $h_e = \max(h, \lambda^{\frac{1}{2q}})$ *is defined in Lemma 5.2.*

Remark 5.1. *The first two terms correspond to the approximation bias of spline functions, the third term is the shrinkage bias due to the penalty, and the last term is the variance.*

Remark 5.2. *Assume in addition that* $e_i \sim N(0, \sigma^2)$, *then a concentration inequality on the term* $\int_{\mathcal{T}} \left\{\hat{f}(x) - \mathbb{E}\hat{f}(x)\right\}^2 \rho(x)dx$, *which measures the integrated variability of penalized splines, can be established; see Lemma C.1 in Appendix C.*

Remark 5.3. *As first observed in [7] for O-splines, depending on the number of knots* K *and the smoothing parameter* λ, *the convergence rates of penalized splines are similar to either those of regression splines or those of smoothing* *splines, giving rise to two-scenario asymptotics: the small number of knots scenario corresponding to regression spline asymptotics and the large number of knots scenario for the smoothing spline asymptotics.*

(a) *(Small number of knots scenario). Suppose that the conditions in Theorem 5.1 and that $f \in \mathcal{C}^m(\mathcal{T})$. If $\lambda K^{2q} = O(1)$, then*

$$\mathbb{E}\left(\left\|\hat{f} - f\right\|_{L_2}^2\right) = O\left(K^{-2m}\right) + O\left(\lambda^2 K^{2q}\right) + O\left(\frac{K}{n}\right),$$

and for $K \sim n^{\frac{1}{2m+1}}$, $\lambda = O\left(n^{-\frac{m+q}{2m+1}}\right)$, the estimator attains the optimal rate of convergence $n^{-\frac{2m}{2m+1}}$.

(b) *(Large number of knots scenario). Suppose that the conditions in Theorem 5.1 and that $f \in \mathcal{C}^q(\mathcal{T})$. There exists a sufficiently large constant C that does not depend on K or n such that, if $\lambda K^{2q} \geq C$, then*

$$\mathbb{E}\left(\left\|\hat{f} - f\right\|_{L_2}^2\right) = O\left(\frac{1}{n}\lambda^{-\frac{1}{2q}}\right) + O\left(\lambda\right) + O\left(K^{-2m}\right) + o\left(K^{-2q}\right),$$

and for $K \geq C^{\frac{1}{2q}}\lambda^{-\frac{1}{2q}}$, and $\lambda \sim n^{-\frac{2q}{2q+1}}$, the estimator attains the optimal rate of convergence $n^{-\frac{2q}{2q+1}}$.

Because $q \leq m$, the optimal number of knots K in (a) is rate-wise no bigger and can be much smaller than that in (b), explaining why the two-type asymptotics can be referred to as the small number of knots scenario and the large number of knots scenario.

Remark 5.4. *The same convergence rate can also be established for $AMSE(\hat{f}) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\hat{f}(x_i) - f(x_i)\right)^2$.*

Remark 5.5. *To apply Theorem 5.1 and Remark 5.3 to O-splines with $q < m$, just let $P_q = O_q$ and $\lambda = \lambda_O$.*

Remark 5.6. *To apply Theorem 5.1 and Remark 5.3 to P-splines with $q \leq m$, just let $P_q = h^{1-2q}D_{K,q}$ and $\lambda = \lambda_P h^{2q-1}$. As mentioned in Section 2, the bases for P-splines are different from those for O-splines. However, the difference does not matter for the theoretic study because Lemmas A.1, A.2 and A.3 in Appendix B.3 still hold for the P-spline bases, which can verified by checking the proofs in [46].*

Remark 5.7. *To apply Theorem 5.1 and Remark 5.3 to T-splines, just let $P_m = h^{-1}\tilde{D}_{K,m}$ and $\lambda = \lambda_T h$.*

6. L_∞ convergence rate

In this section, we shall establish the L_∞ convergence rate of penalized splines. Note that while bounds on the eigenvalues of P_q and $\tilde{P}_q$ are useful for deriving the L_2 convergence rate of penalized splines, they are not sufficient for studying the local and L_∞ convergence of penalized splines. For example, to study the local variance, proper bounds on the diagonals of H_n^{-1} and H_n^{-2} are required. Thus, we shall first derive the local asymptotic variances of the three types of penalized splines, in addition to those in Section 4. Then, we derive a unified convergence rate that applies to the three penalized splines.

6.1. Local asymptotic variance

Proposition 6.1. *Suppose that Assumptions 1, 2 and 5 hold. Let $\lambda = \lambda_O$ and $H_n = G_n + \lambda_O O_q$ with $q < m$. Then there exists a constant $C_5 > 1$ such that, for $r = 1$ and 2,*

$$C_5^{-1} h^{1-r} h_e^{-1} \le \min_k \left(H_n^{-r}\right)_{kk} \le \max_k \left(H_n^{-r}\right)_{kk} \le C_5 h^{1-r} h_e^{-1}.$$

Proposition 6.2. *Suppose that Assumptions 1, 2 and 5 hold. Let $\lambda = \lambda_T h$ and $H_n = G_n + \lambda_T \tilde{D}_{K,m}$. Then there exists a constant $C_6 > 1$ such that, for $r = 1$ and 2,*

$$C_6^{-1} h^{1-r} h_e^{-1} \le \min_k \left(H_n^{-r}\right)_{kk} \le \max_k \left(H_n^{-r}\right)_{kk} \le C_6 h^{1-r} h_e^{-1}.$$

Here $h_e = \max(h, \lambda^{\frac{1}{2q}})$ with $q = m$.

Proposition 6.3. *Suppose that Assumptions 1, 2 and 5 hold. Let $\lambda = \lambda_P h^{2q-1}$ and $H_n = G_n + \lambda_P D_{K,q}$ with $q \le m$. Then there exists a constant $C_7 > 1$ such that, for $r = 1$ and 2,*

$$C_7^{-1} h^{1-r} h_e^{-1} \le \min_k \left(H_n^{-r}\right)_{kk} \le \max_k \left(H_n^{-r}\right)_{kk} \le C_7 h^{1-r} h_e^{-1}.$$

6.2. Unified L_∞ convergence rate

The results in Section 6.1 can be summarized as follows for the unified penalized spline estimator.

Assumption 7. *There exists a constant $C_8 > 1$ such that, for $r = 1$ and 2,*

$$C_8^{-1} h^{1-r} h_e^{-1} \le \min_k \left(H_n^{-r}\right)_{kk} \le \max_k \left(H_n^{-r}\right)_{kk} \le C_8 h^{1-r} h_e^{-1}.$$

The results in Propositions 4.3, 4.4 and 4.5 can also be summarized as follows for the unified penalized spline estimator.

Assumption 8. $\|P_q \beta\|_{\max} = O\left(K^{q-1}\right)$.

Remark 6.1. *Assumption 8 is satisfied by O_q for $q < m$, $h^{-1}\tilde{D}_{K,m}$ for $q = m$ and $h^{1-2q} D_{K,q}$ for $q \le m$.*

We also need the following assumption, which is common for establishing uniform convergence rates [6].

Assumption 9. *Assume that*

$$K\left(\frac{n}{\log n}\right)^{\frac{2}{\tau}-1} = O(1),$$

where τ is in Assumption 6.

Theorem 6.1 (L_∞ convergence)**.** *Suppose that Assumptions 1 - 9 hold. If $f \in \mathcal{C}^p(\mathcal{T})$ with $q \le p \le m$, then*

$$\begin{aligned}\left\|\hat{f} - f\right\| =& O\left(K^{-m}\right) + o\left(K^{-p}\right) + O\left[\min\left\{\lambda K^q\left(1+\lambda K^{2q}\right)^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}}\right\}\right] \\ &+ O\left\{\left(\frac{\log n}{nh_e}\right)^{\frac{1}{2}}\right\}, a.s.\end{aligned}$$

Remark 6.2. *The terms on the right hand in the equation correspond to the L_∞ bound of the approximation bias of splines (first two terms), the L_∞ bound of the shrinkage bias due to the smoothness penalty (third term), and that of the variability (last term) of the penalized spline estimator.*

Remark 6.3. *The proof of the theorem can be adapted to show that the same rate holds for $\mathbb{E}\left\|\hat{f} - f\right\|$.*

Remark 6.4. *Assume in addition that $e_i \sim N(0, \sigma^2)$, then both lower and upper tail risk bound inequalities on the term $\sup_{x\in\mathcal{T}}\left|\hat{f}(x) - \mathbb{E}\hat{f}(x)\right|$ can be established; see Lemma C.2 and its following remark in Appendix C.*

Remark 6.5. *If $f \in \mathcal{C}^m(\mathcal{T})$ and $\lambda = o\left(K^{-2q}\right)$, then $h_e = h$, the shrinkage bias is negligible, and the L_∞ convergence rate is $O\left\{K^{-m} + \left(\frac{K\log n}{n}\right)^{\frac{1}{2}}\right\}$, the L_∞ rate for regression splines. In addition, when $K \sim \left(\frac{n}{\log n}\right)^{\frac{1}{2m+1}}$, the L_∞ convergence rate becomes $\left(\frac{\log n}{n}\right)^{-\frac{m}{2m+1}}$, the minimax optimal rate [6]. Similar to [6], such an optimal rate can be achieved when $\tau \ge 2 + \frac{1}{m}$ because of Assumption 9. Thus, λ does not matter as long as it is small and K as in regression splines serves as the smoothing parameter.*

Remark 6.6. *If λ is sufficiently large, then $h_e = \lambda^{\frac{1}{2q}}$, the approximation bias of spline functions is negligible, and the L_∞ convergence rate becomes $O\left[\min\left\{\lambda K^q(1+\lambda K^{2q})^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}}\right\} + \left(\frac{\log n}{n}\right)^{\frac{1}{2}}\lambda^{-\frac{1}{4q}}\right]$. When $\lambda \sim \left(\frac{\log n}{n}\right)^{\frac{2q}{2q+1}}$ and $K \sim \left(\frac{n}{\log n}\right)^{\frac{1}{2q+1}}$ and that λK^{2q} is sufficiently large, the L_∞ convergence rate is $\left(\frac{\log n}{n}\right)^{-\frac{q}{2q+1}}$, the minimax optimal rate [12]. Because of Assumption 9, such an optimal rate can be achieved when $\tau \ge 2 + \frac{1}{q}$.*

Remark 6.7. *Comparing the derived L_∞ rate with the optimal L_2 rate in part (b) of Remark 5.3, we see that when $\lambda K^{2q} \to \infty$, the derived L_∞ rate $O\left\{\left(\frac{\log n}{n}\right)^{\frac{1}{2}} \lambda^{-\frac{1}{4q}}\right\}$ for the asymptotic variance is optimal. However, the derived L_∞ rate $O\left[\min\left\{\lambda K^q(1+\lambda K^{2q})^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}}\right\}\right]$ for the shrinkage bias is suboptimal and we believe the optimal rate should be $O\left(\lambda^{\frac{1}{2}}\right)$.*

Remark 6.8. *To apply the unified L_∞ rate to the three types of penalized splines, we just follow the specifications in Remarks 5.5, 5.6 and 5.7, respectively.*

7. Local asymptotic bias and variance

Theorem 7.1 (Local asymptotics)**.** *Suppose that Assumptions 1 - 8 hold. If $f(x) \in \mathcal{C}^p(\mathcal{T})$ with $q \le p \le m$, then*

$$\left\|\mathbb{E}\hat{f} - f - 1_{\{p=m\}} b_f - b_\lambda\right\| = o\left[K^{-p} + \min\left\{\lambda K^q\left(1+\lambda K^{2q}\right)^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}}\right\}\right].$$

The shrinkage bias $b_\lambda(x)$ is $-N^T(x)H^{-1}(\lambda P_q)\beta$, where β is defined in Section 3. If additionally, $\lambda K^{2q} = O(1)$ or $\|Q_n - Q\| = o\left(h^{\frac{3}{2}}\right)$, then

$$\sup_x \left|\mathrm{var}\left\{\hat{f}(x)\right\} - \frac{\sigma^2}{n} N^T(x)H^{-1}GH^{-1}N(x)\right| = o\left(\frac{1}{nh_e}\right).$$

Remark 7.1. *If $\lambda K^{2q} = O(1)$,*

$$\left\|\mathbb{E}\hat{f} - f - 1_{\{p=m\}} b_f - b_\lambda\right\| = o\left(K^{-p} + \lambda K^q\right)$$

and

$$\sup_x \left|\mathrm{var}\left\{\hat{f}(x)\right\} - \frac{\sigma^2}{n} N^T(x)H^{-1}GH^{-1}N(x)\right| = o\left(\frac{1}{nh}\right).$$

The above results are the same as the local asymptotics of O-splines derived in [7] and also hold for P-splines and T-splines with specifications in Remarks 5.6 and 5.7, respectively. Suppose that $f \in \mathcal{C}^m(\mathcal{T})$ and $\lambda = o\left\{K^{-(m+q)}\right\}$, then the local asymptotics are the same as those for regression splines and hence are optimal.

Remark 7.2. *If λK^{2q} is sufficiently large, then the discussion is similar to Remark 6.6 and the derived rates may be suboptimal.*

8. Proofs

To simplify notation, we may use D for $D_{K,q}$, Δ for $\Delta_{K,q}$ and P for P_q in the proofs.

8.1. Proofs for Section 3

Proof of Lemma 3.2. We first consider the case $p = m$. Let $g(x) = f(x) - s_f(x) + b_f(x)$. By Lemma 3.1, $\left\|g^{(i)}\right\| = o(h^{m-i})$ if $i \leq m-2$. We derive that

$$\begin{aligned}&\left|\int_{\mathcal{T}} N_k^{[r]}(x)\left\{f^{(i)}(x) - s_f^{(i)}(x)\right\} dQ(x)\right| \\ &\leq \left|\int_{\mathcal{T}} N_k^{[r]}(x) g^{(i)}(x) dQ(x)\right| + \left|\int_{\mathcal{T}} N_k^{[r]}(x) b_f^{(i)}(x) dQ(x)\right|.\end{aligned} \tag{8.1}$$

Let $g_k(x) = N_k^{[r]}(x)\rho(x)$. By integration by parts,

$$\begin{aligned}\int_{\mathcal{T}} N_k^{[r]}(x) g^{(i)}(x) dQ(x) &= \left[g^{(i-1)}(x) g_k(x)\right]\Big|_{x=0}^{1} - \int_{\mathcal{T}} g^{(i-1)}(x) g_k^{(1)}(x) dx \\ &= o(h^{m+1-i}) - \int_{\mathcal{T}} g^{(i-1)}(x) g_k^{(1)}(x) dx.\end{aligned}$$

Note that $\left\|g^{(i-1)}\right\| = o(h^{m+1-i})$, $g_k^{(1)}(x) = \rho(x) dN_k^{[r]}(x)/dx + \rho^{(1)}(x) N_k^{[r]}(x)$ is non-zero for an interval of length $O(h)$, and is of order $O(h^{-1})$, uniformly for all k. Thus, $\int_{\mathcal{T}} g^{(i-1)}(x) g_k^{(1)}(x) dx = o(h^{m+1-i})$ and

$$\max_k \left|\int_{\mathcal{T}} N_k^{[r]}(x) g^{(i)}(x) dQ(x)\right| = o\left(h^{m+1-i}\right). \tag{8.2}$$

We now consider the second right hand term in (8.1) and shall prove that

$$\max_k \left|\int_{\mathcal{T}} N_k^{[r]}(x) b_f^{(i)}(x) dQ(x)\right| = o\left(h^{m+1-i}\right). \tag{8.3}$$

Note that $b_f^{(i)}(x) = -f^{(m)}(t_k)\frac{h_k^{m-i}}{(m-i)!} B_{(m-i)}\left(\frac{x-t_k}{h_k}\right)$, $t_k \leq x < t_{k+1}$ and we have defined $g_k(x) = N_k^{[r]}(x)\rho(x)$. It follows that

$$\begin{aligned}&\int_{\mathcal{T}} N_k^{[r]}(x) b_f^{(i)}(x) dQ(x) \\ &= -\sum_k f^{(m)}(t_k)\frac{h_k^{m-i}}{(m-i)!}\int_{t_k}^{t_{k+1}} g_k(x) B_{(m-i)}\left(\frac{x-t_k}{h_k}\right) dx \\ &= -\sum_k f^{(m)}(t_k)\frac{h_k^{m-i+1}}{(m-i)!}\int_0^1 g_k(t_k + h_k y) B_{(m-i)}(y) dy \\ &= -\sum_k f^{(m)}(t_k)\frac{h_k^{m-i+1}}{(m-i)!} w_k,\end{aligned}$$

where $w_k = \int_0^1 g_k(t_k + h_k y) B_{(m-i)}(y) dy$. By the definition of B-splines, w_k is non-zero only for a few ks. Moreover, using the fact that $\int_0^1 B_{(m-i)}(y) dy = 0$, we have

$$w_k = \int_0^1 \{g_k(t_k + h_k y) - g_k(t_k)\} B_{(m-i)}(y)dy = o(1)$$

uniformly with respect to k. It follows that (8.3) holds and together with (8.1) and (8.2), the proof for the case $p = m$ is complete.

Now we assume that $p < m$ and let $g(x) = f(x) - s_f(x)$ instead. By Remark 3.1, $\left\|g^{(i)}\right\|_{L_2} = o(h^{p-i})$. Let $g_k(x) = N_k^{[r]}(x)\rho(x)$. Then,

$$\max_k \left| \int_{\mathcal{T}} N_k^{[r]}(x) g^{(i)}(x) dQ(x) \right| \leq \max_k \|g_k\|_{L_2} \|g^{(i)}\|_{L_2} = O(h) o(h^{p-i}) = o(h^{p+1-i}).$$

The proof is now complete. □

Proof of Lemma 3.3. We first consider the fixed design. We derive that

$$\begin{aligned} &\max_k \left| \int_{\mathcal{T}} N_k(x) \{f(x) - s_f(x)\} dQ_n(x) \right| \\ \leq &\max_k \left| \int_{\mathcal{T}} N_k(x) \{f(x) - s_f(x)\} dQ(x) \right| \\ &\quad + \max_k \left| \int_{\mathcal{T}} N_k(x) \{f(x) - s_f(x)\} d(Q_n - Q)(x) \right|. \end{aligned}$$

By Lemma 3.1, $\max_k \left| \int_{\mathcal{T}} N_k(x) \{f(x) - s_f(x)\} dQ(x) \right| = o(h^{p+1})$. Hence, it suffices to show

$$\max_k \left| \int_{\mathcal{T}} N_k(x) \{f(x) - s_f(x)\} d(Q_n - Q)(x) \right| = o(h^{p+1}). \tag{8.4}$$

Let $g_k(x) = N_k(x)\{f(x) - s_f(x)\}$. By integration by parts,

$$\int_{\mathcal{T}} N_k(x) \{f(x) - s_f(x)\} d(Q_n - Q)(x) = -\int_{\mathcal{T}} (Q_n - Q)(x) g_k^{(1)}(x) dx.$$

Hence,

$$\left| \int_{\mathcal{T}} N_k(x) \{f(x) - s_f(x)\} d(Q_n - Q)(x) \right| = \left| \int_{\mathcal{T}} (Q_n - Q)(x) g_k^{(1)}(x) dx \right| \tag{8.5}$$

Note that

$$g_k^{(1)}(x) = N_k^{(1)}(x)\{f(x) - s_f(x)\} + N_k(x)\{f^{(1)}(x) - s_f^{(1)}(x)\}.$$

Since $\|f - s_f\| = O(h^p)$ and $\|f^{(1)} - s_f^{(1)}\| = O(h^{p-1})$,

$$\begin{aligned} \int_{\mathcal{T}} \left| g_k^{(1)}(x) \right| dx &\leq \int_{\mathcal{T}} \left| N_k^{(1)}(x) \right| |f(x) - s_f(x)| dx + \int_{\mathcal{T}} N_k(x) |f^{(1)}(x) - s_f^{(1)}(x)| dx \\ &= O(h^p). \end{aligned}$$

where the big O is uniform with respect to k and in the second to last equality we used the fact that $\|N_k^{(1)}\| = O(h^{-1})$ uniformly with respect to k and $N_k^{(1)}(\cdot)$ is non-zero for an interval of length $O(h)$ [26, Theorem 4.2]. It follows that

$$\left|\int_{\mathcal{T}} (Q_n - Q)(x) g_k^{(1)}(x) dx\right| \leq \|Q_n - Q\| \int_{\mathcal{T}} |g_k^{(1)}(x)| dx = o(h^{p+1}),$$

where we used the assumption $\|Q_n - Q\| = o(h)$ (equation (3.2)).

For the random design, by Serfling [29, Theorem 2.1.4b],

$$\|Q_n - Q\| = O\left\{n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}\right\} = o\left(K^{-1}\right)$$

almost surely since $K = o(n^{\delta^*})$ with $\delta^* < 1/2$. The proof is then similar to that for the fixed design. □

8.2. *Proofs for Section 4*

Proof of Proposition 4.1. Let $A_{K,q} = D_{K,1}^q - D_{K,q}$. Note that $A_{K,1} = 0$. By [2, pp. 289-290], for $q > 1$, $A_{K,q}$ are non-zero only for elements with indices $i, j \leq q$ and $K - i + 1 \leq q, K - j + 1 \leq q$. Moreover, it is easy to verify that the non-zero elements in $A_{K,q}$ depends only on q when K is sufficiently large. Thus, $D_{K,q}$ and $D_{K,1}^q$ should have similar singular eigenvalues (Lemma 8.2), and we could study the singular values of $D_{K,q}$ via those of $D_{K,1}^q$ (Lemma 8.3). We first present three Lemmas.

Lemma 8.1. *$A_{K,q} \geq 0$ and has at most $2q$ non-zero eigenvalues.*

Proof. The second claim is straightforward because $A_{K,q}$ has at most $2q$ rows with non-zero elements. So we focus on the first claim. Let $\check{D}_{K,1} = \Delta_{K,1}\Delta_{K,1}^T$. We first prove that $\check{D}_{K,1} \geq D_{K-1,1}$. Let $\mathbf{b} = (b_1, \ldots, b_{K-1})^T$. Then $\mathbf{b}^T \check{D}_{K,1}\mathbf{b} = b_1^2 + b_{K-1}^2 + \sum_{j=2}^{K-1}(b_j - b_{j-1})^2 \geq \sum_{j=2}^{K-1}(b_j - b_{j-1})^2 = \mathbf{b}^T D_{K-1,1}\mathbf{b}$. Note that $D_{K,q} = \Delta_{K,1}^T \Delta_{K-1,q-1}^T \Delta_{K-1,q-1} \Delta_{K,1} = \Delta_{K,1}^T D_{K-1,q-1} \Delta_{K,1}$ and $D_{K,1}^q = \Delta_{K,1}^T \check{D}_{K,1}^{q-1} \Delta_{K,1}$. Thus,

$$A_{K,q} = \Delta_{K,1}^T \left(\check{D}_{K,1}^{q-1} - D_{K-1,q-1}\right) \Delta_{K,1} \geq \Delta_{K,1}^T (D_{K-1,1}^{q-1} - D_{K-1,q-1}) \Delta_{K,1},$$

where the last term equals $\Delta_{K,1}^T A_{K-1,q-1} \Delta_{K,1}$. Therefore, the proof is complete by induction. □

Lemma 8.2. *For $2q + 1 \leq k \leq K$,*

$$\lambda_{k-2q}(D_{K,1}^q) \leq \lambda_k(D_{K,q}) \leq \lambda_k(D_{K,1}^q).$$

Proof of Lemma 8.2. The proof follows from Weyl's Theorem [28, pp. 117], Lemma 8.1 and that $\lambda_j(A_{K,q}) = 0$ for $j \leq K - 2q$. Specifically, the right hand side of the inequality follows from $D_{K,1}^q = D_{K,q} + A_{K,q}$ and $A_{p,q}$ is positive semidefinite by Lemma 8.1. For the left hand side, by Weyl's Theorem, $\lambda_{k-2q}(D_{K,1}^q) \leq \lambda_k(D_{K,q}) + \lambda_{K-2q}(A_{K,q}) = \lambda_k(D_{K,q})$ as $\lambda_{K-2q}(A_{K,q}) = 0$. □

The following lemma is adapted from Theorem 6.5.4 in [2].

Lemma 8.3. *The eigenvalues of* $D_{K,1}$ *are given by* $\nu_k = 4\left(\sin\frac{\pi(k-1)}{2K}\right)^2, k = 1,2,\ldots,K$ *and the corresponding eigenvectors are* $\mathbf{u}_1 = \sqrt{\frac{1}{K}}\{1,1,\ldots,1\}^T$ *and* $\mathbf{u}_k = \sqrt{\frac{2}{K}}\left\{\cos\frac{\pi(k-1)}{2K}, \cos\frac{3\pi(k-1)}{2K}, \ldots, \cos\frac{(2K-1)\pi(k-1)}{2K}\right\}^T$ *for* $k > 1$.

Now we prove Proposition 4.1. By Lemmas 8.2 and 8.3, for $k \geq 2q+2$,

$$4^q\left(\sin\frac{\pi(k-2q-1)}{2K}\right)^{2q} \leq \lambda_k(D_{K,q}) \leq 4^q\left(\sin\frac{\pi(k-1)}{2K}\right)^{2q}$$

It is easy to verify that $\frac{2}{\pi}x \leq \sin x \leq x$ for $x \in [0,\pi/2]$. Thus,

$$\left(\frac{2(k-2q-1)}{K}\right)^{2q} \leq \lambda_k(D_{K,q}) \leq \left(\frac{\pi(k-1)}{K}\right)^{2q}.$$

Since $(q+2)(k-2q-1) \geq k-q$ for $k \geq 2q+2$, we further obtain that for $k \geq 2q+2$,

$$\left(\frac{2}{q+2}\right)^{2q}\left(\frac{k-q}{K}\right)^{2q} \leq \lambda_k(D_{K,q}) \leq \pi^{2q}\left(\frac{k}{K}\right)^{2q}.$$

The proof is complete if for any $K > q$, $\lambda_{q+1}(D_{K,q}) \geq CK^{-2q}$ for some constant $C > 0$ that depends only on q. Note that $\lambda_{q+1}(D_{K,q}) = \lambda_1(\Delta_{K,q}\Delta_{K,q}^T)$. Hence, it suffices to show that

$$\Delta_{K,q}\Delta_{K,q}^T \geq CK^{-2q}I_{K-q}. \tag{8.6}$$

We shall prove (8.6) by induction on q. By Lemma 8.3, we have $\Delta_{K,1}\Delta_{K,1}^T \geq 4\left(\sin\frac{\pi}{2K}\right)^2 I_{K-1}$. Since $\sin\frac{\pi}{2K} \geq \frac{1}{K}$, for any $K > 1$,

$$\Delta_{K,1}\Delta_{K,1}^T \geq 4K^{-2}I_{K-1}. \tag{8.7}$$

Hence, (8.6) holds for $q = 1$. Note that

$$\Delta_{K,q}\Delta_{K,q}^T = \Delta_{K-1,q-1}\left(\Delta_{K,1}\Delta_{K,1}^T\right)\Delta_{K-1,q-1}^T \geq 4K^{-2}\Delta_{K-1,q-1}\Delta_{K-1,q-1}^T,$$

where in the last inequality (8.7) was used. Hence by an inductive proof, (8.6) holds for any q and the proof is complete. □

Proof of Proposition 4.2. We prove the inequalities for the eigenvalues of O_q as the proof for those of $\tilde{D}_{K,m}$ is similar. Note that the weight matrix $W_K^{[m]}$ has the k^{th} element $(m-1)/(t_k - t_{k-m+1})$ for $m \geq 2$ and $W_K^{[1]} = h^{-1}I$. By the assumption of equally-spaced interior knots in Assumption 2, $W_K^{[m]}$ is the same as $h^{-1}I_{K-1}$ except the first $m-2$ diagonal element. With slight abuse of notation, lnet $P_q = \tilde{\Delta}_{K,q,m}^T\tilde{\Delta}_{K,q,m}$. Then it can be shown that $h^{2q}P_q$ differs from $D_{K,q}$ in at most the first $k_{q,m}$ and the last $k_{q,m}$ rows, where $k_{q,m}$ is a finite constant that depends only on q and m. Then with a proof similar to that of Lemma 8.1, it can be shown that there exists a constant $c > 1$ such that $cD^q_{K,1} - h^{2q}P_q$ is positive semi-definite and that $D^q_{K,1} - h^{2q}P_q$ is non-zero only in the first and last few rows. Hence a proof similar to that of Proposition 4.1 proves the same inequalities for $h^{2q}P_q$. For example, inequality (8.6) can be similarly proved. Note that by Remark A.1, there exist constants $0 < c < \tilde{c}$ such that $chI \le G \le \tilde{c}hI$. Hence, $chP_q \le O_q \le \tilde{c}hP_q$, which implies that the eigenvalues of hP_q and O_q are rate-wise similar. Therefore, the inequalities for the eigenvalues of O_q are proved. □

Proof of Proposition 4.3. By the definition of O_q in (2.3),

$$O_q\beta = \tilde{\Delta}^T_{K,q,m} \int N^{[m-q]}(x) s_f^{(q)}(x)dx.$$

Denote $\int N^{[m-q]}(x)s^{(q)}(x)dx$ by $\tilde{\gamma} \in \mathbb{R}^{K-q}$. Then

$$\|O_q\beta\|_{\max} \le \|\tilde{\Delta}^T_{K,q,m}\|_\infty \|\tilde{\gamma}\|_{\max}$$

By definition, $\tilde{\Delta}_{K,q,m}$ is a sparse matrix with a finite band and each element is $O(h^{-q})$. Thus, $\|\tilde{\Delta}^T_{K,q,m}\|_\infty = O(h^{-q})$ and $\|O_q\beta\|_{\max} = O(K^{q-1})$ if

$$\|\tilde{\gamma}\|_{\max} = O(h). \tag{8.8}$$

By Lemma 3.2, we obtain that

$$\tilde{\gamma}_k = \int N_k^{[m-q]}(x) s_f^{(q)}(x)dx = \int N_k^{[m-q]}(x) f^{(q)}(x)dx + o(h^{p+1-q}) = O(h),$$

where $O(h)$ is uniform with respect to k and hence (8.8) holds.

Next, we derive that

$$\begin{aligned}\beta^T O_q \beta &= \int \left\{ s_f^{(q)}(x) \right\}^2 dx \\ &\le \int \left\{ f^{(q)}(x) \right\}^2 dx + \int \left\{ f^{(q)}(x) - s_f^{(q)}(x) \right\}^2 dx \\ &= O(1),\end{aligned}$$

where the last equality holds by Lemma 3.1 and Remark 3.1. The proof is complete. □

Proof of Proposition 4.4. By equation (2.2),

$$s_f^{(m-1)}(x) = \beta^T \tilde{\Delta}^T_{K,m-1,m} N^{[1]}(x).$$

Define $\gamma = (\gamma_1, \dots, \gamma_{K_0+1})^T = \tilde{\Delta}_{K,m-1,m}\beta$, then $s_f^{(m-1)}(x) = \gamma_k$ if $x \in [t_k, t_{k+1})$. By Lemma A.8,

$$\tilde{D}_{K,m} = \left\{ \frac{1}{(m-1)!} \right\}^2 \tilde{\Delta}^T_{K,m-1,m} D_{K-m+1,1} \tilde{\Delta}_{K,m-1,m}.$$

It follows that

$$\{(m-1)!\}^2\left(\tilde{D}_{K,m}\beta\right)=\tilde{\Delta}^T_{K,m-1,m}\Delta^T_{K-m+1,1}\left(\Delta_{K-m+1,1}\gamma\right)$$

and

$$\{(m-1)!\}^2\,\beta^T\tilde{D}_{K,m}\beta=\gamma^T D_{K-m+1,1}\gamma=\|\Delta_{K-m+1,1}\gamma\|_2^2=\sum_{k=2}^{K_0+1}(\gamma_k-\gamma_{k-1})^2.$$

Note that

$$\begin{aligned}&(\gamma_k-\gamma_{k-1})^2\\ =&\left\{s_f^{(m-1)}(t_k)-s_f^{(m-1)}(t_{k-1})\right\}^2\\ \le&\left\{s_f^{(m-1)}(t_k)-f^{(m-1)}(t_k)\right\}^2+\left\{s_f^{(m-1)}(t_{k-1})-f^{(m-1)}(t_{k-1})\right\}^2\\ &+\left\{f^{(m-1)}(t_k)-f^{(m-1)}(t_{k-1})\right\}^2\\ =&O(h^2),\end{aligned}$$

where the big O is uniform with respect to k and the last equalities follow by Lemma 3.1 and the fact that $f\in\mathcal{C}^m$. Since $\|\tilde{\Delta}^T_{K,m-1,m}\|_\infty=O(K^{m-1})$, $\|\Delta^T_{K-m+1,1}\|_\infty=O(1)$, we derive that

$$\begin{aligned}\{(m-1)!\}^2\,\|\tilde{D}_{K,m}\beta\|_{\max}&\le\|\tilde{\Delta}^T_{K,m-1,m}\|_\infty\|\Delta^T_{K-m+1,1}\|_\infty\|\Delta_{K-m+1,1}\gamma\|_{\max}\\ &=O(K^{m-1})O(1)O(h)\\ &=O(K^m).\end{aligned}$$

Thus, $\|h^{-1}\tilde{D}_{K,m}\beta\|_{\max}=O(K^{m-1})$.

Next, we derive that

$$\{(m-1)!\}^2\,\beta^T\tilde{D}_{K,m}\beta=\sum_{k=2}^{K_0+1}(\gamma_k-\gamma_{k-1})^2=O(h).$$

The proof is complete. □

Proof of Proposition 4.5. First note that Lemma 3.1 and Proposition 4.3 also hold when the boundary knots for O-splines are the same as those for P-splines. We first consider the case $q<m$. Then $\tilde{\Delta}_{K,q,m}=h^{-q}\Delta_{K,q}$ and

$$s_f^{(q)}(x)=N^{[m-q],T}(x)\tilde{\Delta}_{K,q,m}\beta=h^{-q}N^{[m-q],T}(x)\Delta_{K,q}\beta.$$

Hence,

$$\Delta_{K,q}\beta=h^q\left(G^{[m-q]}\right)^{-1}\int N^{[m-q]}(x)s_f^{(q)}(x)dx.$$

The proof in Proposition 4.3 (see equation (8.8)) shows that

$$\left\|\int N^{[m-q]}(x)s_f^{(q)}(x)dx\right\|_{\max} = O(h).$$

It follows that

$$\begin{aligned}\|D_{K,q}\beta\|_{\max} &\le \|\Delta_{K,q}^T\|_\infty \|\Delta_{K,q}\beta\|_{\max} \\ &\le h^q \left\|\left(G^{[m-q]}\right)^{-1}\right\|_\infty \left\|\int N^{[m-q]}(x)s_f^{(q)}(x)dx\right\|_{\max} \\ &= O(h^q).\end{aligned}$$

Therefore,

$$\|h^{1-2q}D_{K,q}\beta\|_{\max} = O(h^{1-q}).$$

Next, note that $O_q = h^{-2q}\Delta_{K,q}^T G^{[m-q]}\Delta_{K,q}$. Hence

$$O(1) = \beta^T O_q \beta = h^{-2q}\beta^T\Delta_{K,q}^T G^{[m-q]}\Delta_{K,q}\beta = h^{-2q}\gamma^T G^{[m-q]}\gamma,$$

where $\gamma = \Delta_{K,q}\beta$. By Remark A.1 after Lemma A.1, the eigenvalues of $G^{[m-q]}$ are of order h. Thus,

$$\beta^T D_{K,q}\beta = \|\gamma\|_2^2 = O(h^{2q-1})$$

and we have proved the cases with $q < m$.

Now we consider $q = m$. Let $\gamma = \Delta_{K,m}\beta$. Note that

$$s_f^{(m-1)}(x) = \beta^T\tilde{\Delta}_{K,m-1,m}^T N^{[1]}(x) = h^{1-m}\gamma^T N^{[1]}(x).$$

Hence, $s_f^{(m-1)}(x) = h^{1-m}\gamma_k$ if $x \in [t_k, t_{k+1})$. We next derive that

$$D_{K,m}\gamma = \Delta_{K,m}^T(\Delta_{K-m+1,1}\gamma)$$

and

$$\beta^T D_{K,m}\beta = \gamma^T D_{K-m+1,1}\gamma$$

and the rest of the proof is similar to that of Proposition 4.4. □

8.3. Proofs for Section 5

Proof of Lemma 5.1. Note that $\tilde{P}_q = G_n^{-\frac{1}{2}}P_q G_n^{-\frac{1}{2}}$. By Lemma A.1, it is easy to show that

$$\left(c_2^{-\frac{1}{2}} + o(1)\right)h^{-\frac{1}{2}} \le \lambda_{\min}\left(G_n^{-\frac{1}{2}}\right) \le \left(c_1^{-\frac{1}{2}} + o(1)\right)h^{-\frac{1}{2}}.$$

Since both $G_n^{-\frac{1}{2}}$ and P are non-negative and symmetric matrices, applying twice the inequalities 6.76 in [28, page 119] and Assumption 3 proves the lemma. □

Proof of Lemma 5.2. Let $\tilde{s}_k$ denote the k^{th} smallest eigenvalue of $\tilde{P}_q$, then $(1+\lambda s_k)^{-1}$ is the k^{th} largest eigenvalue of $(I+\lambda\tilde{P}_q)^{-1}$. Then

$$\left\|\left(I+\lambda\tilde{P}_q\right)^{-1}\right\|_F^2=\sum_{k=1}^{K}\left(\frac{1}{1+\lambda\tilde{s}_k}\right)^2.$$

By Lemma 5.1, $\tilde{s}_1=\ldots=\tilde{s}_q=0$ and $\tilde{s}_k\geq C_4h^{-2q}\left(\frac{k-q}{K}\right)^{2q}$ for $k\geq q+1$. It follows that

$$\sum_{k=1}^{K}\left(\frac{1}{1+\lambda\tilde{s}_k}\right)^2\leq q+\sum_{k=1}^{K-q}\left(\frac{1}{1+C_4\lambda h^{-2q}k^{2q}K^{-2q}}\right)^2$$

and the sum can be easily shown is $O(h_e^{-1})$. The proof is complete. □

Proof of Theorem 5.1. We first consider the bias $\mathbb{E}\hat{f}(x)-f(x)$ and derive that

$$\begin{aligned}\mathbb{E}\hat{f}(x)=&N^T(x)H_n^{-1}(N^Tf/n)\\=&N^T(x)\gamma-N^T(x)H_n^{-1}(\lambda P)\gamma,\end{aligned}$$

where $f=\{f(x_1),\ldots,f(x_n)\}^T\in\mathbb{R}^n$ and $\gamma=G_n^{-1}(N^Tf/n)$. Because

$$s_f(x)=N^T(x)\beta=N^T(x)G_n^{-1}(N^Ts_f/n),$$

where $s_f=\{s_f(x_1),\ldots,s_f(x_n)\}^T\in\mathbb{R}^n$, we further obtain

$$(\mathbb{E}\hat{f}-f)(x)=(s_f-f)(x)+N^T(x)G_n^{-1}\alpha-N^T(x)H_n^{-1}(\lambda P)\gamma, \tag{8.9}$$

where $\alpha=N^T(f-s_f)/n$. It follows that

$$\begin{aligned}&\frac{1}{3}\int\left\{\mathbb{E}\hat{f}(x)-f(x)\right\}^2\rho(x)dx\\&\leq\int\left\{s_f(x)-f(x)\right\}^2\rho(x)dx+\alpha^TG_n^{-1}GG_n^{-1}\alpha+\gamma^T(\lambda P)H_n^{-1}GH_n^{-1}(\lambda P)\gamma.\end{aligned}\tag{8.10}$$

We first derive $\|s_f-f\|$. Consider first $p=m$. By Lemma 3.1, $\|s_f-f-b_f\|=o(h^m)$. By definition, $\|b_f\|=O(h^m)$. Hence,

$$\|s_f-f\|=O(h^m). \tag{8.11}$$

Assume now $p<m$. By Remark 3.1, $\|s_f-f\|=o(h^p)$. Therefore, for general p with $p\leq m$, we obtain that

$$\|s_f-f\|=O(h^m)+o(h^p). \tag{8.12}$$

For the second right hand term in (8.9), note by (8.18) that $\|\alpha\|_{\max}=o(h^{p+1})$. Thus,

$$\alpha^TG_n^{-1}GG_n^{-1}\alpha\leq\|\alpha\|_2^2\|G_n^{-1}GG_n^{-1}\|_\infty=o(h^{2p}).$$

It follows from (8.10) that

$$\frac{1}{3}\int \left\{\mathbb{E}\hat{f}(x) - f(x)\right\}^2 \rho(x)dx = O\left(h^{2m}\right) + o\left(h^{2q}\right) + \gamma^T(\lambda P)H_n^{-1}GH_n^{-1}(\lambda P)\gamma. \tag{8.13}$$

Denote the last right hand term in (8.13) by ξ. Note that

$$\begin{aligned}\xi &= (1+o(1))\gamma^T(\lambda P)H_n^{-1}G_nH_n^{-1}(\lambda P)\gamma \\ &= (1+o(1))\gamma^T G_n^{\frac{1}{2}}\tilde{P}(I+\tilde{P})^{-2}\tilde{P}G_n^{\frac{1}{2}}\gamma,\end{aligned}$$

where $\tilde{P} = G_n^{-\frac{1}{2}}(\lambda P)G_n^{-\frac{1}{2}}$. Thus,

$$\xi = O(1)\gamma^T G_n^{\frac{1}{2}}\tilde{P}(I+\tilde{P})^{-2}\tilde{P}G_n^{\frac{1}{2}}\gamma. \tag{8.14}$$

Since $\tilde{P}(I+\tilde{P})^{-2}\tilde{P} \le \|\tilde{P}\|_2\tilde{P}$, we derive that

$$\begin{aligned}\xi &= O(1)\|\tilde{P}\|_2\gamma^T G_n^{\frac{1}{2}}\tilde{P}G_n^{\frac{1}{2}}\gamma \\ &= O(1)\|\tilde{P}\|_2\gamma^T(\lambda P)\gamma.\end{aligned}$$

Note that $\|\tilde{P}\|_2 \le \lambda\|G_n^{-\frac{1}{2}}\|_2^2\|P\|_2$, $\|G_n^{-\frac{1}{2}}\|_2 = O(h^{-\frac{1}{2}})$ and $\|P\|_2 = O(K^{2q-1})$ by Assumption 3. Therefore, $\|\tilde{P}\|_2 = O(\lambda K^{2q})$. In addition, by Lemma 8.4, $\gamma^T P\gamma = O(1)$. Thus,

$$\xi = O(\lambda^2 K^{2q}).$$

On the other hand, by (8.14) and the fact that $\tilde{P}(I+\tilde{P})^{-2}\tilde{P} \le \tilde{P}$, we obtain

$$\begin{aligned}\xi &= O(1)\gamma^T G_n^{\frac{1}{2}}\tilde{P}G_n^{\frac{1}{2}}\gamma \\ &= O(1)\gamma^T(\lambda P)\gamma \\ &= O(\lambda).\end{aligned}$$

Thus, we obtain that

$$\xi = O\left\{\min(\lambda^2K^{2q}, \lambda)\right\}. \tag{8.15}$$

Combining (8.13) and (8.15), we obtain

$$\frac{1}{3}\int \left\{\mathbb{E}\hat{f}(x) - f(x)\right\}^2 \rho(x)dx = O\left(h^{2m}\right) + o\left(h^{2p}\right) + O\left\{\min(\lambda^2K^{2q}, \lambda)\right\}, \tag{8.16}$$

which finishes the derivation of bias.

Next, we consider the variance and derive that

$$\int \mathrm{var}\{\hat{f}(x)\}\rho(x)dx = \frac{\sigma^2}{n}\mathrm{tr}\left(H_n^{-1}G_nH_n^{-1}G\right).$$

We derive that

$$\begin{aligned}\operatorname{tr}\left(H_n^{-1}G_nH_n^{-1}G\right) &= \{1+o(1)\}\operatorname{tr}\left(H_n^{-1}G_nH_n^{-1}G_n\right)\\ &= O(1)\operatorname{tr}\left\{\left(I+\tilde{P}\right)^{-2}\right\}\\ &= O(1)\left\|\left(I+\tilde{P}\right)^{-1}\right\|_F^2\\ &= O(h_e^{-1}),\end{aligned}$$

where the last equality follows by Lemma 5.2. It follows that

$$\int \operatorname{var}\{\hat{f}(x)\}\rho(x)dx = O\left(\frac{1}{nh_e}\right). \tag{8.17}$$

Thus, by combining (8.16) and (8.17), we obtain that

$$\begin{aligned}&\mathbb{E}\int\left\{\hat{f}(x)-f(x)\right\}^2\rho(x)dx\\ =&\int\left\{\mathbb{E}\hat{f}(x)-f(x)\right\}^2\rho(x)dx+\int\operatorname{var}\{\hat{f}(x)\}\rho(x)dx\\ =&O\left(h^{2m}\right)+o\left(h^{2p}\right)+O\left\{\min(\lambda^2K^{2q},\lambda)\right\}+O\left(\frac{1}{nh_e}\right).\end{aligned}$$

Because

$$\mathbb{E}\int\left\{\hat{f}(x)-f(x)\right\}^2dx \le \frac{1}{\inf_{x\in\mathcal{T}}\rho(x)}\mathbb{E}\int\left\{\hat{f}(x)-f(x)\right\}^2\rho(x)dx$$

and $\inf_{x\in\mathcal{T}}\rho(x)>0$, we obtain the desired bounds for $\mathbb{E}\|\hat{f}-f\|_{L_2}^2$. □

Lemma 8.4. *Suppose that Assumptions 1, 3-4 hold. Assume that* $f\in\mathcal{C}^p(\mathcal{T})$ *with* $q\le p\le m$. *Let* $\gamma=G_n^{-1}(N^Tf/n)$ *with* $f=\{f(x_1),\ldots,f(x_n)\}^T\in\mathbb{R}^n$. *Then*

$$\gamma^TP_q\gamma=O(1).$$

Proof. Note that $G_n^{-1}(N^Ts_f/n)=\beta$, where β is defined in Section 3. Hence

$$\gamma=\beta+G_n^{-1}\{N^T(f-s_f)/n\}=\beta+G_n^{-1}\alpha,$$

where $\alpha=N^T(f-s_f)/n$. Since P_q is positive semi-definite, it is easy to show that

$$\left(\gamma^TP_q\gamma\right)^{\frac{1}{2}}\le\left(\beta^TP_q\beta\right)^{\frac{1}{2}}+\left(\alpha^TG_n^{-1}P_qG_n^{-1}\alpha\right)^{\frac{1}{2}}.$$

By Assumption 4, $\beta^TP_q\beta=O(1)$. Thus, if suffices to show that $\alpha^TG_n^{-1}P_qG_n^{-1}\alpha=O(1)$. Note that

$$\|\alpha\|_{\max}=\|N^T(f-s_f)/n\|_{\max}=\left\|\int N(x)\{f(x)-s_f(x)\}dQ_n(x)\right\|_{\max}.$$

By Lemma 3.3,

$$\|\alpha\|_{\max} = o\left(h^{p+1}\right). \tag{8.18}$$

Thus,

$$\begin{aligned}\alpha^T G_n^{-1} P_q G_n^{-1} \alpha &\le \|\alpha\|_2^2 \|G_n^{-1} P_q G_n^{-1}\|_2 \\ &= o(h^{2p+2} h^{-1} h^{-2} h^{1-2q}) \\ &= o(1).\end{aligned}$$

And the proof is complete. □

8.4. Proofs for Section 6

Proofs of Propositions 6.1, 6.2 and 6.3. We first focus on $G_n + \lambda_O O_q$. Note that $O_q = \tilde{\Delta}_{K,q,m}^T G^{[m-q]} \tilde{\Delta}_{K,q,m}$ with $q < m$. By Lemma A.1, $c_1 hI \le G_n \le c_2 hI$, where c_1 and c_2 are constants in Lemma A.1. Similar to G_n, we may assume for the same two constants c_1 and c_2 that $c_1 hI \le G^{[m-q]} \le c_2 hI$. Then,

$$c_1 h \tilde{\Delta}_{K,q,m}^T \tilde{\Delta}_{K,q,m} \le O_q \le c_2 h \tilde{\Delta}_{K,q,m}^T \tilde{\Delta}_{K,q,m}.$$

Let $\bar{O}_q = \tilde{\Delta}_{K,q-1,m}^T \tilde{\Delta}_{K,q-1,m}$. It follows that

$$c_1 h \left(I + \lambda_O \bar{O}_q\right) \le G_n + \lambda_O O_q \le c_2 h \left(I + \lambda_O \bar{O}_q\right).$$

Thus, for $r = 1, 2$,

$$c_2^{-r} h^{-r} \left(I + \lambda_O \bar{O}_q\right)^{-r} \le (G_n + \lambda_O O_q)^{-r} \le c_1^{-r} h^{-r} \left(I + \lambda_O \bar{O}_q\right)^{-r}.$$

Hence, Proposition 6.1 holds if the minimum and maximum of the diagonal of $\left(I + \lambda_O \bar{O}_q\right)^{-r}$ are of order hh_e^{-1}. With a similar argument, Proposition 6.2 holds if the minimum and maximum of the diagonal of $\left(I + \lambda_T \tilde{D}_{K,m}\right)^{-r}$ are of order hh_e^{-1} with $h_e = \max(h, \lambda^{\frac{1}{2m}})$ and Proposition 6.3 holds if the minimum and maximum of the diagonal of $(I + \lambda_P D_{K,q})^{-r}$ are of order hh_e^{-1}. Note that $\bar{O}_m = \{(m-1)!\}^2 h^{-2} \tilde{D}_{K,m}$ by Lemma A.8. Because of Assumption 2, the matrix $h^{2q} \bar{O}_q$ is the same as $D_{K,q}$ except for the first and the last few rows and the matrix $h^{2(m-1)} \tilde{D}_{K,m}$ is the same as $D_{K,m}$ except for the first and the last few rows. As a result, an asymptotic study of $(I + \lambda_P D_{K,q})^{-r}$ for $r = 1, 2$ is the key and some minor technical modifications are sufficient to accommodate the differences in the first and last few rows. We first note that when $\lambda_P = O(1)$, the corresponding $h_e = O(1)$ and the diagonals of $(I + \lambda_P D_{K,q})^{-r}$ are necessarily $O(1)$, which proves the propositions. Therefore, we just need to focus on the case that if $\lambda_P \ge C$ for some sufficiently large constant C, there still exists a constant, say $\tilde{C} > 1$, such that

$$\tilde{C}^{-1} \lambda_P^{-\frac{1}{2q}} \le \min_k \left\{(I + \lambda_P D_{K,q})^{-r}\right\}_{kk} \le \max_k \left\{(I + \lambda_P D_{K,q})^{-r}\right\}_{kk} \le \tilde{C} \lambda_P^{-\frac{1}{2q}}.$$

The desired results are given in Theorem B.1 in Appendix B. □

Proof of Theorem 6.1. We consider the bias and variance of $\hat{f}(x)$ separately. By (8.9) and (8.12), we obtain that

$$\|\mathbb{E}\hat{f} - f\| \leq O(h^m) + o(h^p) + \left\|N^T(\cdot)G_n^{-1}\alpha\right\| + \left\|N^T(\cdot)H_n^{-1}(\lambda P)\gamma\right\|. \quad (8.19)$$

By (8.18), $\|\alpha\|_{\max} = o(h^{p+1})$. Hence,

$$\left\|G_n^{-1}\alpha\right\|_{\max} \leq \|G_n^{-1}\|_\infty \|\alpha\|_{\max} = O(h^{-1})\|\alpha\|_{\max} = o(h^p).$$

It follows that

$$\left\|N^T(\cdot)G_n^{-1}\alpha\right\| = o\left(h^p\right). \quad (8.20)$$

As for the last right hand term in (8.19), we have

$$\left\|N^T(\cdot)H_n^{-1}(\lambda P)\gamma\right\| \leq \left\|H_n^{-1}(\lambda P)\gamma\right\|_{\max}.$$

Note that $\gamma = \beta + G_n^{-1}\alpha$. Hence,

$$\left\|H_n^{-1}(\lambda P)\gamma\right\|_{\max} \leq \left\|H_n^{-1}(\lambda P)\beta\right\|_{\max} + \left\|H_n^{-1}(\lambda P)G_n^{-1}\alpha\right\|_{\max}.$$

Therefore,

$$\left\|N^T(\cdot)H_n^{-1}(\lambda P)\gamma\right\| \leq \left\|H_n^{-1}(\lambda P)\beta\right\|_{\max} + \left\|H_n^{-1}(\lambda P)G_n^{-1}\alpha\right\|_{\max}. \quad (8.21)$$

We derive the orders of each of the two right hand terms in (8.21) now. By Assumption 8 and Lemma A.4,

$$\left\|H_n^{-1}(\lambda P)\beta\right\|_{\max} \leq \lambda\|H_n^{-1}\|_\infty \|P\beta\|_{\max} = O\left\{\lambda K^q(1+\lambda K^{2q})^{\frac{3}{2}}\right\}.$$

On the other hand, we derive that

$$\begin{aligned}\left\|H_n^{-1}(\lambda P)\beta\right\|_{\max}^2 &\leq \beta^T(\lambda P)H_n^{-2}(\lambda P)\beta \\ &= O(h^{-1})\beta^T(\lambda P)H_n^{-1}G_nH_n^{-1}(\lambda P)\beta.\end{aligned}$$

Matrix algebra shows that $(\lambda P)H_n^{-1}G_nH_n^{-1}(\lambda P) \leq \lambda P$. Hence,

$$\left\|H_n^{-1}(\lambda P)\beta\right\|_{\max}^2 = O(h^{-1})\beta^T(\lambda P)\beta = O(\lambda K).$$

Thus, we have shown that

$$\left\|H_n^{-1}(\lambda P)\beta\right\|_{\max} = O\left[\min\left\{\lambda K^q(1+\lambda K^{2q})^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}}\right\}\right]. \quad (8.22)$$

Now we work on the second right hand term in (8.21). Note that by Lemma A.4, $\|H_n^{-1}\|_\infty = O\{K(1+\lambda K^{2q})^{\frac{3}{2}}\}$. We derive that

$$\begin{aligned}\left\|H_n^{-1}(\lambda P)G_n^{-1}\alpha\right\|_{\max} &\leq \lambda\|H_n^{-1}\|_\infty\|P\|_\infty\|G_n^{-1}\|_\infty\|\alpha\|_{\max} \\ &= O\left\{\lambda K(1+\lambda K^{2q})^{\frac{3}{2}}h^{1-2q}h^{-1}\right\}o(h^{p+1}).\end{aligned}$$

Since $p \geq q$, we obtain that

$$\left\| H_n^{-1}(\lambda P) G_n^{-1} \alpha \right\|_{\max} = o\left\{ \lambda K^q (1 + \lambda K^{2q})^{\frac{3}{2}} \right\}.$$

On the other hand,

$$\begin{aligned} \left\| H_n^{-1}(\lambda P) G_n^{-1} \alpha \right\|_{\max}^2 &\leq \alpha^T G_n^{-1} (\lambda P) H_n^{-2} (\lambda P) G_n^{-1} \alpha \\ &= O(h^{-1}) \alpha^T G_n^{-1} (\lambda P) H_n^{-1} G_n H_n^{-1} (\lambda P) G_n^{-1} \alpha \\ &= O(h^{-1}) \alpha^T G_n^{-1} (\lambda P) G_n^{-1} \alpha, \end{aligned}$$

The proof of Lemma 8.4 derives that $\alpha^T G_n^{-1} P G_n^{-1} \alpha = o(1)$. Thus,

$$\left\| H_n^{-1}(\lambda P) G_n^{-1} \alpha \right\|_{\max}^2 = o(\lambda K).$$

The above derivations prove that

$$\left\| H_n^{-1}(\lambda P) G_n^{-1} \alpha \right\|_{\max} = o\left[\min\left\{ \lambda K^q (1 + \lambda K^{2q})^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}} \right\} \right]. \tag{8.23}$$

Combining (8.21), (8.22) and (8.23), we have proved that

$$\left\| N^T(\cdot) H_n^{-1}(\lambda P) \gamma \right\| = O\left[\min\left\{ \lambda K^q (1 + \lambda K^{2q})^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}} \right\} \right]. \tag{8.24}$$

Now combining (8.19), (8.20) and (8.24), we have

$$\left\| \mathbb{E}\hat{f} - f \right\| = O\left(h^m\right) + o(h^p) + O\left[\min\left\{ \lambda K^q (1 + \lambda K^{2q})^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}} \right\} \right].$$

Now we consider $\hat{f}(x) - \mathbb{E}\hat{f}(x)$. Let

$$u(x) = \hat{f}(x) - \mathbb{E}\hat{f}(x) = \sum_{i=1}^n w_i(x) e_i,$$

where

$$w_i(x) = N^T(x) H_n^{-1} N(x_i)/n.$$

We shall use the following results: for any constant $r > 0$,

$$\max_i \|w_i\| = O\left(\frac{1}{nh_e}\right), \tag{8.25}$$

$$\sup_x \left[\sum_{i=1}^n \{w_i(x)\}^2 \right] = O\left(\frac{1}{nh_e}\right), \tag{8.26}$$

$$\sup_{x,z \in \mathcal{T}: |x-z| \leq n^{-r}} \max_i |w_i(x) - w_i(z)| = O\left\{ \frac{1}{n^r} \frac{1}{nh} \left(\frac{1}{hh_e}\right)^{\frac{1}{2}} \right\}. \tag{8.27}$$

The above equalities will be derived at the end of the proof. Define also

$$L_n = \left(\frac{nh_e}{\log n}\right)^{\frac{1}{2}}.$$

We derive that

$$u(x) = \sum_{i=1}^{n} w_i(x)e_i = u_1(x) + u_2(x), \tag{8.28}$$

where

$$u_1(x) = \sum_i w_i(x)e_i 1_{\{|e_i|>L_n\}},\ u_2(x) = \sum_i w_i(x)e_i 1_{\{|e_i|\le L_n\}}.$$

Note that

$$\left|e_i 1_{\{|e_i|>L_n\}}\right| \le L_n^{1-\tau}|e_i|^\tau,$$

where $\tau > 2$ is a constant in Assumption 9. Thus,

$$|u_1(x)| \le \sum_i |w_i(x)|L_n^{1-\tau}|e_i|^\tau \le \max_i |w_i(x)| \left(L_n^{1-\tau}n\right) \left(n^{-1}\sum_i |e_i|^\tau\right).$$

It follows by (8.25) and the strong law of large numbers that

$$\|u_1\| = O\left\{(nh_e)^{-1}(L_n^{1-\tau}n)\right\}, \quad \text{almost surely.}$$

By Assumption 6,

$$\|u_1\| = O(L_n^{-1}),\ a.s. \tag{8.29}$$

Let r be a sufficiently large constant such that

$$n^{-r}h^{-\frac{3}{2}}\left(\frac{n}{\log n}\right)^{\frac{1}{2}} = O(1). \tag{8.30}$$

Define $\chi(r) = \{n^{-r}, 2n^{-r}, \ldots, 1-n^{-r}, 1\}$. Then

$$\|u_2\| \le \sup_{x\in\chi(r)} |u_2(x)| + \sup_{x,z\in\mathcal{T}:|x-z|\le n^{-r}} |u_2(x) - u_2(z)|. \tag{8.31}$$

For the second term in (8.31), by Hölder's inequality,

$$\begin{aligned}|u_2(x) - u_2(z)| &\le \sum_i |w_i(x) - w_i(z)||e_i| \\ &\le \left\{\sum_i |w_i(x) - w_i(z)|^{\frac{1}{1-\frac{1}{\tau}}}\right\}^{1-\frac{1}{\tau}} \left(\sum_i |e_i|^\tau\right)^{\frac{1}{\tau}}.\end{aligned}$$

By derivation similar to that for (8.29), we obtain that

$$|u_2(x) - u_2(z)| \le \max_i |w_i(x) - w_i(z)|O(n),$$

where the big O is uniform with respect to x and z. By equality (8.27),

$$\sup_{x,z\in\mathcal{T}:|x-z|\le n^{-r}} |u_2(x) - u_2(z)| = O\left\{n^{-r}h^{-1}(hh_e)^{-\frac{1}{2}}\right\} = O(L_n^{-1}),\ a.s. \tag{8.32}$$

Note that the last equality above follows from (8.30). Next we focus on the first term in (8.31). Note that

$$\sum_i \mathbb{E}\left\{w_i^2(x)e_i^2 1_{\{|e_i|\le L_n\}}\right\} \le (\mathbb{E}|e_i|^\tau)^{\frac{2}{\tau}} \sum_i w_i^2(x) \le (\mathbb{E}|e_i|^\tau)^{\frac{2}{\tau}}\, \tilde{C}(nh_e)^{-1},$$

where $\tilde{C}$ is a constant and the last inequality follows from (8.26). In addition, by (8.25), there exists another constant $\bar{C} > 0$ such that

$$|w_i(x)e_i 1_{\{|e_i|\le L_n\}}| \le L_n\bar{C}(nh_e)^{-1}$$

uniformly with respect to x. By Bernstein's inequality for bounded random variables, for any constant $c > 0$,

$$\mathbb{P}\left\{\sup_{x\in\chi(r)} |u_2(x)| > cL_n^{-1}\right\} \le n^r \exp\left\{-\frac{c^2L_n^{-2}/2}{\check{C}(nh_e)^{-1} + c\bar{C}(nh_e)^{-1}/3}\right\} = n^{\gamma - c^*},$$

where $c^* = \frac{c^2/2}{\check{C}\sigma^2 + c\bar{C}/3}$ and $\check{C} = (\mathbb{E}|\epsilon_i|^\tau)^{\frac{2}{\tau}}\,\tilde{C}$. We can choose c sufficiently large so that the above inequality is summable. By the Borel-Cantelli lemma [11, pp.46],

$$\sup_{x\in\chi(r)} |u_2(x)| = O(L_n^{-1}),\ a.s.$$

The above equality together with (8.31) and (8.32) leads to

$$\|u_2\| = O(L_n^{-1}),\ a.s. \tag{8.33}$$

Combining (8.28), (8.29) and (8.33), we obtain that

$$\left\|\hat{f} - \mathbb{E}\hat{f}\right\| = O\left\{\left(\frac{\log n}{nh_e}\right)^{\frac{1}{2}}\right\},\ a.s.$$

and the proof is complete once we have established (8.25), (8.26) and (8.27).

For (8.25), we derive that

$$|w_i(x)| = n^{-1}\left|\sum_{k\ell} N_k(x)N_\ell(x_i)(H_n^{-1})_{k\ell}\right| \le n^{-1}\|H_n^{-1}\|_{\max}\sum_{k\ell} N_k(x)N_\ell(x_i),$$

where the latter inequality follows because $N_k(x) \ge 0, N_\ell(x_i) \ge 0$ and H_n^{-1} is symmetric and positive definite. Since $\sum_k N_k(x) = 1$ and $\sum_\ell N_\ell(x_i) = 1$, we obtain that $|w_i(x)| \le n^{-1}\|H_n^{-1}\|_{\max} = O\{(nh_e)^{-1}\}$ by Assumption 7 and (8.25) is proved. For (8.26), we derive that

$$\begin{aligned}
\sum_{i=1}^n \{w_i(x)\}^2 &= \frac{1}{n}N^T(x)H_n^{-1}G_nH_n^{-1}N(x) \\
&= O\left(\frac{1}{n}\right)\|H_n^{-1}G_nH_n^{-1}\|_{\max} \\
&= O\left(\frac{1}{n}\right)\|H_n^{-2}\|_{\max}\|G_n\|_\infty = O\left(\frac{1}{nh_e}\right),
\end{aligned}$$

where the last equality follows by Assumption 7 and Lemma A.1. Finally for (8.27), since $|N_k^{(1)}(x)| = O(h^{-1})$ uniformly for x and k, we derive that $|N_k(x) - N_k(z)| = O(h^{-1})|x-z|$ uniformly for x, z and k. Note that $N_k(x) - N_k(z) \neq 0$ only for a finite number of k. Thus, if $|x-z| \leq n^{-r}$,

$$\begin{aligned}|w_i(x) - w_i(z)|^2 &= n^{-2}\{N(x) - N(z)\}^T H_n^{-1} N(x_i) N^T(x_i) H_n^{-1}\{N(x) - N(z)\}\\ &= n^{-2} O(h^{-2} n^{-2r}) \|H_n^{-1} N(x_i) N^T(x_i) H_n^{-1}\|_2\\ &= n^{-2} O(h^{-2} n^{-2r}) N^T(x_i) H_n^{-2} N(x_i)\\ &= n^{-2} O(h^{-2} n^{-2r}) \|H_n^{-2}\|_{\max}\\ &= n^{-2} O(h^{-2} n^{-2r}) O(h^{-1} h_e^{-1}),\end{aligned}$$

and the big O is uniform with respect to i, x and z. Thus, $|w_i(x) - w_i(z)| = O\{(nh)^{-1} n^{-r} (hh_e)^{-\frac{1}{2}}\}$ uniformly for i, x and z and (8.27) is proved. □

8.5. Proofs for Section 7

Proof of Theorem 7.1. We first establish the asymptotic bias of $\hat{f}(x)$. With similar derivations for the bias in the proof of Theorem 5.1 (specifically, equalities (8.9), (8.11) and (8.12)) and the derivation of L_∞ rate of bias in the proof of Theorem 6.1 (specifically, equalities (8.21) and (8.23)), we obtain that

$$\begin{aligned}&\sup_x \left|\mathbb{E}\hat{f}(x) - (f + 1_{\{p=m\}} b_f)(x) - N^T(x) H_n^{-1}(\lambda P)\beta\right|\\ =&o\left[h^p + \min\left\{\lambda K^q (1 + \lambda K^{2q})^{\frac{3}{2}}, (\lambda K)^{\frac{1}{2}}\right\}\right].\end{aligned}$$

Next we focus on the asymptotic variance of $\hat{f}(x)$. First,

$$\operatorname{var}\{\hat{f}(x)\} = \frac{\sigma^2}{n} N^T(x) H_n^{-1} G_n H_n^{-1} N(x).$$

By Lemma A.6,

$$\begin{aligned}&\sup_x \left|\operatorname{var}\{\hat{f}(x)\} - \frac{\sigma^2}{n} N^T(x) H^{-1} G H^{-1} N(x)\right|\\ =&o\left(\frac{1}{n}\right) \sup_x N^T(x) H^{-1} G H^{-1} N(x).\end{aligned}$$

Again by Lemma A.6,

$$\begin{aligned}N^T(x) H^{-1} G H^{-1} N(x) &= (1 + o(1)) N^T(x) H_n^{-1} G_n H_n^{-1} N(x)\\ &= (1 + o(1)) h \|H_n^{-2}\|_{\max} = O(h_e^{-1})\end{aligned}$$

where the last equality holds by Assumption 7 and $h_e = \max\{h, \lambda^{1/(2q)}\}$. If $\lambda K^{2q} = O(1)$, $h \leq h_e = O(h)$, thus $N^T(x) H^{-1} G H^{-1} N(x) = O(h^{-1})$. Otherwise if $\lambda K^{2q} \geq C$ for a sufficiently large C, then $h_e = \lambda^{1/(2q)}$ and $N^T(x) H^{-1} G H^{-1} N(x) = \lambda^{-1/(2q)}$. The proof is now complete. □

Acknowledgements

The author wishes to thank the authors of [7] for their quick response regarding questions about their work and Professor David Ruppert for encouragement and help. The author would also like to thank the referees for their comments and suggestions, which lead to a much improved paper.

Appendix A: Technical lemmas

We assume that Assumptions 1 and 2 hold for Lemmas A.1-A.6. Lemmas A.1, A.2 and A.3 below list existing results concerning G_n from [46].

Lemma A.1. *There exist constants $c_2 > c_2 > 0$ such that*

$$c_1 h \leq \lambda_{\min}(G_n) \leq \lambda_{\max}(G_n) \leq c_2 h.$$

Moreover, $\|G_n\|_\infty = O(h)$.

Remark A.1. *The same inequalities holds for $G = \iint N(x)N^T(x)\rho(x)dx$ with a similar proof.*

Lemma A.2. *Denote the (i,j)th element of G_n^{-1} by α_{ij}. There exists a constant $c_3 > 0$ and $\gamma \in (0,1)$ such that, for large n, $|\alpha_{ij}| \leq c_3 h^{-1}\gamma^{|i-j|}$. In addition, $\|G_n^{-1}\|_\infty = O(h^{-1})$.*

Remark A.2. *The same inequalities holds for G^{-1}.*

Lemma A.3.

$$\begin{aligned}
\|G_n - G\|_{\max} &= O(\|Q_n - Q\|) = o(h), \\
\|G_n^{-1} - G^{-1}\|_{\max} &= O(h^{-2}\|Q_n - Q\|) = o(h^{-1}), \\
\|G_n^{-1} - G^{-1}\|_\infty &= O(h^{-2}\|Q_n - Q\|) = o(h^{-1}).
\end{aligned}$$

Lemma A.4. *Suppose that Assumption 3 holds. Then,*

$$\left\|H_n^{-1}\right\|_\infty = O\left\{K\left(1+\lambda K^{2q}\right)^{\frac{3}{2}}\right\}.$$

Proof. The lemma is proved by Theorem 2.4 in [10], which extends Theorem 2.2 in [9] by allowing a large conditional number in the matrix. Specifically, it can be shown that

$$\left|(H_n^{-1})_{k\ell}\right| = O\left\{K(1+\lambda K^{2q})\right\}\left(\frac{\sqrt{\mathrm{cond}(H_n)}-1}{\sqrt{\mathrm{cond}(H_n)}+1}\right)^{\frac{2|k-\ell|}{m^*}},$$

where m^* is the band of H_n and $\mathrm{cond}(H_n) = \lambda_{\max}(H_n)/\lambda_{\min}(H_n)$. Since $\lambda_{\max}(H_n)$ has the same order as $h(1+\lambda K^{2q})$ and $\lambda_{\min}(H_n)$ has the same order as h, straightforward calculation gives the desired result. □

Lemmas A.5 and A.6 below are useful for deriving the local asymptotic bias and variance of penalized splines.

Lemma A.5. *There exists an $\epsilon_1 > 0$ and $\epsilon_1 = o(1)$ such that the following inequalities hold:*

$$(1-\epsilon_1)G \leq G_n \leq (1+\epsilon_1)G,$$
$$(1-\epsilon_1)G^{-1} \leq G_n^{-1} \leq (1+\epsilon_1)G,$$
$$(1+\epsilon_1)^{-1}H^{-1} \leq H_n^{-1} \leq (1-\epsilon_1)^{-1}H^{-1}.$$

Proof. Since $G_n - G$ is a band matrix, $\|G_n - G\|_2 \leq \|G_n - G\|_\infty = O(\|G_n - G\|_{\max}) = o(h)$, where the latter equality holds by Lemma A.3. Note that Lemma A.1 shows that the eigenvalues of G are all of order h. Thus, there exists an $\epsilon_1 = o(1)$ such that $(1-\epsilon_1)G \leq G_n \leq (1+\epsilon_1)G$. It follows that $(1-\epsilon_1)H \leq H_n \leq (1+\epsilon_1)H$ because $H_n = G_n + P_q/n$, $H = G + P_q/n$ and P_q is positive semi-definite. Thus, $(1+\epsilon_1)^{-1}H^{-1} \leq H_n^{-1} \leq (1-\epsilon_1)^{-1}H^{-1}$. □

Lemma A.6. *Suppose that Assumptions 3 and 7 hold. If*

$$\left\|H_n^{-1}(G_n - G)^2 H_n^{-1}\right\|_\infty = o(1), \tag{A.1}$$

then there exists an $\epsilon_2 > 0$, which is $o(1)$ and independent from x and λ, such that:

$$1 - \epsilon_2 \leq \frac{N^T(x)H_n^{-1}G_nH_n^{-1}N(x)}{N^T(x)H^{-1}GH^{-1}N(x)} \leq 1 + \epsilon_2.$$

Remark A.3. *If $\lambda K^{2q} = O(1)$, then by Lemma A.4, $\|H_n^{-1}\|_\infty = O(h^{-1})$. Thus, the assumption (A.1) is satisfied without additional conditions. Otherwise, a sufficient condition for (A.1) is $\|Q_n - Q\| = o(h^{\frac{3}{2}})$ which leads to $\|G_n - G\|_\infty = o(h^{\frac{3}{2}})$. Indeed, one can derive that*

$$\begin{aligned}\left\|H_n^{-1}(G_n - G)^2 H_n^{-1}\right\|_\infty \leq& K\left\|H_n^{-1}(G_n - G)^2 H_n^{-1}\right\|_{\max} \\ \leq& K\left\|H_n^{-2}\right\|_{\max}\|G_n - G\|_\infty^2 \\ =& KO(h^{-2})o(h^3) = o(1).\end{aligned}$$

Proof. First note that the little o notation in the proof is independent from x and λ. By Lemma A.5, $-\epsilon_1 G \leq G_n - G \leq \epsilon_1 G$ with $\epsilon_1 = o(1)$. Thus, by Lemma A.7,

$$-\epsilon_1 H_n^{-1}GH_n^{-1} \leq H_n^{-1}(G_n - G)H_n^{-1} \leq \epsilon_1 H_n^{-1}GH_n^{-1}.$$

It follows that,

$$\left\|H_n^{-1}G_nH_n^{-1} - H_n^{-1}GH_n^{-1}\right\|_{\max} = o(1)\left\|H_n^{-1}GH_n^{-1}\right\|_{\max}. \tag{A.2}$$

It can be shown that

$$\begin{aligned}H_n^{-1}GH_n^{-1} - H^{-1}GH^{-1} \leq& \left(H_n^{-1} - H^{-1}\right)G\left(H_n^{-1} - H^{-1}\right) \\ =& H_n^{-1}(G_n - G)\left(H^{-1}GH^{-1}\right)(G_n - G)H_n^{-1}.\end{aligned}$$

Note that

$$\begin{aligned}&\left\|H_n^{-1}(G_n-G)\left(H^{-1}GH^{-1}\right)(G_n-G)H_n^{-1}\right\|_{\max}\\ \le&\left\|H_n^{-1}(G_n-G)^2H_n^{-1}\right\|_{\infty}\left\|H^{-1}GH^{-1}\right\|_{\max}.\end{aligned}$$

By the assumption $\left\|H_n^{-1}(G_n-G)^2H_n^{-1}\right\|_{\infty}=o(1)$, we derive that

$$\left\|H_n^{-1}(G_n-G)\left(H^{-1}GH^{-1}\right)(G_n-G)H_n^{-1}\right\|_{\max}=o(1)\left\|H^{-1}GH^{-1}\right\|_{\max}.$$

It follows that

$$\left\|H_n^{-1}GH_n^{-1}-H^{-1}GH^{-1}\right\|_{\max}=o(1)\left\|H^{-1}GH^{-1}\right\|_{\max}. \tag{A.3}$$

Combining (A.2) and (A.3), we derive that

$$\left\|H_n^{-1}G_nH_n^{-1}-H^{-1}GH^{-1}\right\|_{\max}=o(1)\left\|H^{-1}GH^{-1}\right\|_{\max}. \tag{A.4}$$

Therefore, the lemma is proved if we prove that

$$\left\|H^{-1}GH^{-1}\right\|_{\max}=O(1)\inf_x N^T(x)H^{-1}GH^{-1}N(x)$$

or equivalently,

$$\left\|H^{-1}GH^{-1}\right\|_{\max}=O(1)\min\operatorname{diag}\left(H^{-1}GH^{-1}\right).$$

Because of (A.4), it is also equivalent to prove that

$$\left\|H_n^{-1}G_nH_n^{-1}\right\|_{\max}=O(1)\min\operatorname{diag}\left(H_n^{-1}G_nH_n^{-1}\right). \tag{A.5}$$

Note that $\left\|H_n^{-1}G_nH_n^{-1}\right\|_{\max}\le\|H_n^{-2}\|_{\max}\|G_n\|_{\infty}=O(h_e^{-1})$ by Assumption 7. On the other hand, we have $H_n^{-1}G_nH_n^{-1}\ge(c_1+o(1))hH_n^{-2}$ by Lemma A.1 and here c_1 is a constant. Thus, again by Assumption 7, we obtain that $\min\operatorname{diag}\left(H_n^{-1}G_nH_n^{-1}\right)\ge Ch_e^{-1}$ for some constant $C>0$ and (A.5) is proved. The proof is complete. □

Lemma A.7. *Let A and B be two square and symmetric matrices of the same size. Assume C is another matrix with compatible dimension. If $A\le B$, then $C(A-B)C^T\le 0$.*

Proof. Let α be any vector of the size of A. Then $\alpha^TC(A-B)C^T\alpha=\gamma^T(A-B)\gamma\le 0$, where $\gamma=C^T\alpha$. □

Lemma A.8. *The penalty matrix $\tilde{D}_{K,m}$ for the T-splines satisfies*

$$\tilde{D}_{K,m}=h^2\left\{\frac{1}{(m-1)!}\right\}^2\tilde{\Delta}_{K,m,m}^T\tilde{\Delta}_{K,m,m}.$$

Proof. Define $\tilde{L}_{K,k,m} \in \mathbb{R}^{K\times K}$ for $k \leq m$ recursively as $-\tilde{L}^T_{K,1,m}$ is the left $K \times K$ matrix of $\Delta_{K+1,1}$ and for $2 \leq k \leq m-1$ with $m \geq 2$,

$$\tilde{L}^T_{K,k,m} = \tilde{L}^T_{K,1,m}\text{blockdiag}\left(\begin{pmatrix} m-1 \\ k-1 \end{pmatrix}, (k-1)^{-1}\, W^{[k]}_K \tilde{L}^T_{K-1,k-1,m}\right).$$

Then it can be shown that $L_{K,1} = \tilde{L}_{K,1,1}$ and for $m \geq 2$

$$L^T_{K,m} = L^T_{K,1}\text{blockdiag}\left(\begin{pmatrix} m-1 \\ m-1 \end{pmatrix}, (m-1)^{-1}\, W^{[m]}_K \tilde{L}^T_{K-1,m-1,m}\right).$$

It follows that

$$\tilde{D}_{K,1} = L^T_{K,1}\tilde{I}_{K,1}L_{K,1} = D_{K,1} = h^2\tilde{\Delta}^T_{K,1,1}\tilde{\Delta}_{K,1,1}$$

because $W^{[1]}_K$ is defined as $h^{-1}I$ and for $m \geq 2$

$$\begin{aligned}
&\tilde{D}_{K,m} \\
=&L^T_{K,m}\tilde{I}_{K,m}L_{K,m} \\
=&L^T_{K,1}\text{blockdiag}\left(0, \frac{1}{(m-1)^2}\, W^{[m]}_K \tilde{L}^T_{K-1,m-1,m}\tilde{I}_{K-1,m-1}\tilde{L}_{K-1,m-1,m}\, W^{[m]}_K\right)L_{K,1} \\
=&\frac{1}{(m-1)^2}\Delta^T_{K,1}\left(W^{[m]}_K \tilde{L}^T_{K-1,m-1,m}\tilde{I}_{K-1,m-1}\tilde{L}_{K-1,m-1,m}\, W^{[m]}_K\right)\Delta_{K,1} \\
=&\frac{1}{(m-1)^2}\tilde{\Delta}^T_{K,1,m}\left(\tilde{L}^T_{K-1,m-1,m}\tilde{I}_{K-1,m-1}\tilde{L}_{K-1,m-1,m}\right)\tilde{\Delta}_{K,1,m}.
\end{aligned}$$

Thus, using an inductive proof, we obtain the desired equality and the proof is complete. □

Appendix B: Local asymptotic variance of penalized splines

For $\eta > 0$, define $\Lambda = I + \eta D_{K,q}$. In this appendix, we study Λ^{-1} and derive its order of convergence in terms of η when η is large. The results can be used for studying the local asymptotic variance of P-splines. Note that essentially the same proof can be applied to $I + \eta\tilde{D}_{K,m}$ for T-splines and $I + \eta\tilde{\Delta}^T_{K,q-1,m}\tilde{\Delta}_{K,q-1,m}$ for O-splines.

With slight abuse of notation, we let $h_e = \eta^{\frac{1}{2q}}K^{-1}$. We assume that $K \to \infty$ and that there exists two constants δ_1 and $\delta_2 > 0$ such that $\eta > \delta_1 K^{\delta_2}$. In addition, we assume that $\eta = O(K^{2q})$ so that $h_e = O(1)$. We also use C and C_0 to denote constants that depend only on q and to simplify notation, they are allowed to vary from place to place.

The main result is Theorem B.1.

Theorem B.1. *Suppose that $\eta > C$ for a sufficiently large constant C. Then for $r = 1$ or 2,*

$$\min_k(\Lambda^{-r})_{kk} \simeq \max_k(\Lambda^{-r})_{kk} \simeq \eta^{-\frac{1}{2q}}.$$

To prove Theorem B.1, we shall follow [41] and invert Λ directly. Consider the equation

$$\eta(-1)^q(1-\rho)^{2q} + \rho^q = 0 \tag{B.1}$$

and let $\{\rho_\nu, \nu = 1, \dots, q\}$ be the q roots of (B.1) such that when η is large, the real parts of the first q roots are all positive and less than 1. Using a proof similar to that of Proposition 4.3 in [41], we derive that, when $\eta > C$ and C is a sufficiently large constant,

$$\rho_\nu = 1 - \psi_\nu \eta^{-\frac{1}{2q}} + \frac{1}{2}\psi_\nu^2 \eta^{-\frac{1}{q}} + O\left(\eta^{-\frac{3}{2q}}\right),\ 1 \le \nu \le q, \tag{B.2}$$

where $\psi_1, \dots, \psi_q$ are the roots of $x^{2q} + (-1)^q = 0$ with positive real parts. Notice that $|\rho_\nu| < 1$ if C is sufficiently large.

Notice that $(D)_{ij} = (D)_{(K-i),(K-j)}$, which implies that

$$(\Lambda^{-1})_{kk} = (\Lambda^{-1})_{(K-k),(K-k)}.$$

Hence, we may assume that $k \le K/2$ in this section. Define $S_k = \sum_{\nu=1}^q a_\nu U_k(\rho_\nu)$, where

$$U_k(\rho) = (\rho^{k-1}, \dots, \rho, 1, \rho, \dots, \rho^{K-k})^T \in \mathbb{R}^K$$

and $\mathbf{a} = (a_1, \dots, a_q)^T \in \mathbb{R}^q$ is the vector of coefficients to be determined soon. Fix k and assume that $k \ge q+1$. For $1 \le \nu \le q$, it can be shown that $U_k(\rho_\nu)$ is orthogonal to all columns of Λ except the first and last q columns and the j^{th} columns with $|k-j| < q$. Thus, for any $\mathbf{a}$, the same holds for S_k. Since S_k is linear in $\mathbf{a}$, there exists an $\mathbf{a}$ such that S_k is also orthogonal to the j^{th} columns with $k < j < k+q$ and $S_k^T \Lambda_k = 1$, where Λ_k is the k^{th} column of Λ. Note that it can be easily verified that S_k will be also orthogonal to Λ_j with $\max(q+1, k-q) \le j < k$. Therefore, S_k is orthogonal to all columns of Λ except the first and last q columns and the k^{th} column. By the derivations in Sections 4.2 in [41], we obtain that, $\mathbf{a}$ does not depend on k, and moreover, when $\eta > C$ for a sufficiently large constant C,

$$a_\nu = \frac{\psi_\nu}{2q}\eta^{-\frac{1}{2q}}\left\{1 + O\left(\eta^{-\frac{1}{q}}\right)\right\},\ 1 \le \nu \le q. \tag{B.3}$$

In particular, $\mathbf{a}$ satisfies the following equalities [41, pp. 11]:

$$\sum_{\nu=1}^q a_\nu\left(\rho_\nu^\ell - \rho_\nu^{-\ell}\right) = 0,\ \ell = 1, \dots, (q-1). \tag{B.4}$$

Note that both $U_1(\rho_\nu)$ and $U_K(\rho_\nu)$ are orthogonal to all columns of Λ except the first q and the last q columns. Hence, there exists $a_{k\nu}$ and $\tilde{a}_{k\nu}, \nu = 1, \dots, q$, such that, if we define $R_k = \sum_{\nu=1}^q a_{k\nu} U_1(\rho_\nu)$ and $T_k = \sum_{\nu=1}^q \tilde{a}_{k\nu} U_K(\rho_\nu)$, then

$$\Lambda(S_k + R_k + T_k) = \mathbf{e}_k, \tag{B.5}$$

where $\mathbf{e}_k \in \mathbb{R}^K$ is a unit vector with the k^{th} element 1 and others 0. Equation (B.5) implies that the k^{th} column of Λ^{-1} takes the form

$$(\Lambda^{-1})_k = S_k + R_k + T_k,\ q+1 \le k \le K/2. \tag{B.6}$$

Let $\mathbf{a}_k = (a_{k1}, \dots, a_{kq})^T \in \mathbb{R}^q$ and $\tilde{\mathbf{a}}_k = (\tilde{a}_{k1}, \dots, \tilde{a}_{kq})^T \in \mathbb{R}^q$.

Assume now $k \le q$. There exists $\mathbf{a}_k$ and $\tilde{\mathbf{a}}_k$ such that

$$\Lambda(R_k + T_k) = \mathbf{e}_k. \tag{B.7}$$

The existence of $\mathbf{a}_k$ and $\tilde{\mathbf{a}}_k$ follows from the fact that, for arbitrary $\mathbf{a}_k$ and $\tilde{\mathbf{a}}_k$, $R_k + T_k$ is orthogonal to all columns of Λ except the first and the last q columns, and that $\{U_1(\rho_1), \dots, U_1(\rho_q), U_K(\rho_1), \dots, U_K(\rho_q)\}$ are linearly independent.

B.1. Notation

Let $B_k = [U_k(\rho_1), U_k(\rho_2), \dots, U_k(\rho_q)] \in \mathbb{R}^{K\times q}$. Let

$$\mathbf{r}(x) = \{\exp(-\psi_1 x), \dots, \exp(-\psi_\nu x)\}^T \in \mathbb{R}^q.$$

Note that there exists a constant $\psi_0 > 0$ that depends only on q and satisfies $\exp(-\psi_\nu|x|) \le \exp(-\psi_0|x|)$ for all ν.

Define $\Psi_{q,1} \in \mathbb{R}^{q\times q}$ with the $(j,\nu)^{th}$ element ψ_ν^{q+j-1} and $\Psi_{q,2} \in \mathbb{R}^{q\times q}$ with the $(j,\nu)^{th}$ element $(-1)^{q+j}\psi_\nu^{q+j}$. Let Ω_1 be a $q\times q$ Vandermonde matrix with the $(j,\nu)^{th}$ element ρ_ν^{j-1} and Ω_2 be a $q\times q$ Vandermonde matrix with the $(j,\nu)^{th}$ element $\rho_\nu^{-(j-1)}$. Let Φ_1 be a $q\times q$ matrix with the $(j,\nu)^{th}$ element $(1-\rho_\nu)^{q+j-1}$ and Φ_2 be a $q\times q$ matrix with the $(j,\nu)^{th}$ element $(1-\rho_\nu^{-1})^{q+j-1}$.

Finally, for a matrix A, we shall use A_k to denote its k^{th} column.

B.2. Proof of theorem

We first describe a few lemmas and propositions, whose proofs are given in Section B.3.

Lemma B.1. *Assume that $\eta > C$ for a sufficiently large C. There exists a universal constant C_0 that depends only on q and satisfies*

$$\left|\rho_\nu^{K-k}\right| = O\left\{\exp\left(-C_0 h_e^{-1}\right)\right\},$$

for any $k \le K/2$ and $1 \le \nu \le q$.

Lemma B.2. *Fix $q+1 \le k \le K/2$. For any j with $j < k$,*

$$S_k^T(\Delta^T)_j = \sum_{\nu=1}^{q} a_\nu(-1)^q(1-\rho_\nu^{-1})^q\rho_\nu^{k-j}, \tag{B.8}$$

and if $j \ge k$,

$$S_k^T(\Delta^T)_j = \sum_{\nu=1}^{q} a_\nu(-1)^q(1-\rho_\nu)^q\rho_\nu^{j-k}. \tag{B.9}$$

We first derive $\mathbf{a}_k$ and $\tilde{\mathbf{a}}_k$ in the following two propositions.

Proposition B.1. *Suppose that $\eta > C$ for a sufficiently large constant C. Fix k with $q+1 \le k \le K/2$. Suppose that $\mathbf{a}_k$ and $\tilde{\mathbf{a}}_k$ satisfy (B.5). Then,*

$$\mathbf{a}_k = \frac{1}{2q}\eta^{-\frac{1}{2q}}\Psi_{q,1}^{-1}\Psi_{q,2}\mathbf{r}\left(\frac{k-1}{\eta^{\frac{1}{2q}}}\right) + O\left(\eta^{-\frac{1}{q}}\right)\exp\left(-\psi_0\frac{k-1}{h_e}\right)\mathbf{1}_q, \tag{B.10}$$

and there exists a universal constant C_0 that depends only on q and satisfies

$$\tilde{\mathbf{a}}_k = \exp\left(-C_0 h_e^{-1}\right) O(\mathbf{1}_q). \tag{B.11}$$

In particular, the big O notation is uniform with respect to $q+1 \le k \le K/2$.

Proposition B.2. *Suppose that $\eta > C$ for a sufficiently large constant C. Fix k with $k \le q$. Suppose that $\mathbf{a}_k$ and $\tilde{\mathbf{a}}_k$ satisfy (B.7). Then*

$$\mathbf{a}_k = \frac{(-1)^{q+1}}{q}\eta^{-\frac{1}{2q}}\Psi_{q,1}^{-1}\tilde{\mathbf{e}}_q + O\left(\eta^{-\frac{1}{q}}\right)\exp\left(-\psi_0\frac{k-1}{h_e}\right)\mathbf{1}_q,$$

where $\tilde{\mathbf{e}}_q \in \mathbb{R}^q$ is a unit vector with the q^{th} element 1, and there exists a universal constant C_0 that depends only on q and satisfies

$$\tilde{\mathbf{a}}_k = \exp\left(-C_0 h_e^{-1}\right) O(\mathbf{1}_q). \tag{B.12}$$

Proof of Theorem B.1. We shall first establish that

$$\|\Lambda^{-r}\|_{\max} = O\left(\eta^{-\frac{1}{2q}}\right). \tag{B.13}$$

Consider first that $q < k \le K/2$. By (B.6),

$$(\Lambda^{-1})_{k\ell} = \sum_{\nu=1}^{q}\left(a_\nu\rho_\nu^{|k-\ell|} + a_{k\nu}\rho_\nu^{\ell-1} + \tilde{a}_{k\nu}\rho_\nu^{K-\ell}\right). \tag{B.14}$$

Since $|\rho_\nu| < 1$ for large η,

$$(\Lambda^{-1})_{kk} \le \sum_{\nu=1}^{q}\left(|a_\nu| + |a_{k\nu}| + |\tilde{a}_{k\nu}|\right).$$

By Proposition B.1, $(\Lambda^{-1})_{kk} = O\left(\eta^{-\frac{1}{2q}}\right)$ uniformly for $q < k \le K/2$. Now for Λ^{-2}, we have

$$(\Lambda^{-2})_{kk} = \sum_{\ell=1}^{K}\left\{\sum_{\nu=1}^{q}\left(a_\nu\rho_\nu^{|k-\ell|} + a_{k\nu}\rho_\nu^{\ell-1} + \tilde{a}_{k\nu}\rho_\nu^{K-\ell}\right)\right\}^2.$$

We derive that

$$\left|\sum_{\ell=1}^{K} a_{\nu_1}a_{\nu_2}(\rho_{\nu_1}\rho_{\nu_2})^{|k-\ell|}\right| \le |a_{\nu_1}a_{\nu_2}|\frac{2}{1-|\rho_{\nu_1}\rho_{\nu_2}|},$$

uniformly for $q < k \leq K/2$ and because $1 - |\rho_{\nu_1}\rho_{\nu_2}| \geq \tilde{C}\eta^{-\frac{1}{2q}}$ for some constant $\tilde{C}$ and $a_\nu = O\left(\eta^{-\frac{1}{2q}}\right)$,

$$\left|\sum_{\ell=1}^{K} a_{\nu_1}a_{\nu_2}(\rho_{\nu_1}\rho_{\nu_2})^{|k-\ell|}\right| = O\left(\eta^{-\frac{1}{2q}}\right).$$

We can similarly show other terms in $(\Lambda^{-2})_{kk}$ are $O\left(\eta^{-\frac{1}{2q}}\right)$. Similar derivation can be done for $k \leq q$ by (B.7) and Proposition B.2. We have now established (B.13).

Next we show that there exists a constant $c > 0$ such that, for $r = 1$ or 2,

$$\min_k \left(\Lambda^{-r}\right)_{kk} \geq (c + o(1))\eta^{-\frac{1}{2q}}. \tag{B.15}$$

Note that $D_{K,q} \leq D_{K,1}^q$. Thus, $\Lambda \leq I + \eta D_{K,1}^q \leq \left(I + \eta^{\frac{1}{q}} D_{K,1}\right)^q$ as $D_{K,1}$ is positive semidefinite. Let $\Lambda_1 = I + \eta^{\frac{1}{q}} D_{K,1}$. Let $U = [u_1, \dots, u_K]$ be the eigenvectors of $D_{K,1}$ and $\nu = (\nu_1, \dots, \nu_K)^T$ be the vector of eigenvalues of $D_{K,1}$. Then

$$\Lambda^{-r} \geq U\left\{I + \eta^{\frac{1}{q}}\operatorname{diag}(\nu)\right\}^{-rq} U^T.$$

Since $\Lambda^{-1} \geq \Lambda^{-2}$ and hence $\left(\Lambda^{-1}\right)_{kk} \geq \left(\Lambda^{-2}\right)_{kk}$, it suffices to consider $\left(\Lambda^{-2}\right)_{kk}$. We derive that

$$\left(\Lambda^{-2}\right)_{kk} \geq \sum_{j=0}^{K-1} \frac{u_{jk}^2}{\left(1 + \eta^{\frac{1}{q}}\nu_j\right)^{2q}}.$$

By Lemma 8.3, $\nu_j = 4\left(\sin\frac{\pi j}{2K}\right)^2 \leq \frac{\pi^2 j^2}{K^2}$. Thus,

$$\left(\Lambda^{-2}\right)_{kk} \geq \sum_{j=0}^{K-1} \frac{u_{jk}^2}{\left\{1 + \left(\eta^{\frac{1}{2q}}\frac{\pi j}{K}\right)^2\right\}^{2q}} = O\left(\frac{1}{K}\right) + \frac{2}{K}\sum_{j=0}^{K-1} \frac{\left\{\cos\frac{(2k-1)\pi j}{2K}\right\}^2}{\left\{1 + \left(\eta^{\frac{1}{2q}}\frac{\pi j}{K}\right)^2\right\}^{2q}}.$$

By the Euler-Maclaurin formula,

$$\begin{aligned}
&\frac{1}{K}\sum_{j=0}^{K-1} \frac{\left\{\cos\frac{(2k-1)\pi j}{2K}\right\}^2}{\left\{1 + \left(\eta^{\frac{1}{2q}}\frac{\pi j}{K}\right)^2\right\}^{2q}} \\
&= \int_0^{K-1} \frac{\left\{\cos\frac{(2k-1)\pi x}{2K}\right\}^2}{K\left\{1 + \left(\eta^{\frac{1}{2q}}\frac{\pi x}{K}\right)^2\right\}^{2q}} dx + O\left(\frac{1}{K}\right) \\
&= \frac{1}{\eta^{\frac{1}{2q}}\pi}\int_0^{\eta^{\frac{1}{2q}}\pi} \frac{\left\{\cos\frac{(k-0.5)y}{\eta^{\frac{1}{2q}}}\right\}^2}{(1+y^2)^{2q}} dy + O\left(\frac{1}{K}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\eta^{\frac{1}{2q}}\pi}\int_0^\infty \frac{\left\{\cos\frac{(k-0.5)y}{\eta^{\frac{1}{2q}}}\right\}^2}{(1+y^2)^{2q}}dy + O\left(\frac{1}{K}\right) + o\left(\eta^{-\frac{1}{2q}}\right)\\
&= \frac{1}{2\eta^{\frac{1}{2q}}\pi}\int_0^\infty \frac{1+\cos\frac{(2k-1)y}{\eta^{\frac{1}{2q}}}}{(1+y^2)^{2q}}dy + O\left(\frac{1}{K}\right) + o\left(\eta^{-\frac{1}{2q}}\right).
\end{aligned}
$$

It can be shown that [16, pp. 430, equality 3.737.1], for any $k \geq 1$,

$$
\int_0^\infty \frac{\cos\frac{(2k-1)y}{\eta^{\frac{1}{2q}}}}{(1+y^2)^{2q}}dy > 0.
$$

Let

$$
c_q = \int_0^\infty \frac{1}{(1+y^2)^{2q}}dy > 0.
$$

Then,

$$
\frac{1}{K}\sum_{j=0}^{K-1}\frac{\left\{\cos\frac{(2k-1)\pi j}{2K}\right\}^2}{\left\{1+\left(\eta^{\frac{1}{2q}}\frac{\pi j}{K}\right)^2\right\}^{2q}} \geq \frac{c_q}{2\eta^{\frac{1}{2q}}\pi}(1+o(1)).
$$

It follows that

$$
\left(\Lambda^{-2}\right)_{kk} \geq \frac{c_q}{\eta^{\frac{1}{2q}}\pi}(1+o(1))
$$

for all k and we have proved (B.15).

The proof is complete by combining (B.13) and (B.15). □

B.3. Proofs of lemmas and propositions

Proof of Lemma B.1. Let $b_\nu = \eta^{1/(2q)}\log(\rho_\nu)$, then $b_\nu = \psi_\nu + O(\eta^{-1/q})$. Since $\exp(-\psi_\nu|x|) \leq \exp(-\psi_0|x|)$, we derive tat $|\exp(-b_\nu|x|)| \leq \exp(-0.5\psi_0|x|)$ for any x when η is sufficiently large. It follows that

$$
\begin{aligned}
\rho_\nu^{K-k} &= \exp\left(-b_\nu\eta^{\frac{1}{2q}}|K-k|\right)\\
&= O\left\{\exp\left(-0.5\psi_0\eta^{\frac{1}{2q}}|K-k|\right)\right\}.
\end{aligned}
$$

Since $k \leq K/2$, $|K-k|\eta^{\frac{1}{2q}} \geq 1/2h_e^{-1}$ and the proof is complete by letting $C_0 = \psi_0/4$. □

Proof of Lemma B.2. First we have

$$
S_k^T(\Delta^T)_j = \sum_{\nu=1}^q a_\nu\left\{\sum_{\ell=0}^q w_\ell\rho_\nu^{|k-j-\ell|}\right\} = \sum_{\ell=0}^q w_\ell\left\{\sum_{\nu=1}^q a_\nu\rho_\nu^{|k-j-\ell|}\right\},
$$

where $\mathbf{w} = (w_0, \dots, w_q)^T \in \mathbb{R}^{q+1}$ is the vector such that $\Delta\theta = \mathbf{w}^T\theta$. We derive that

$$\begin{aligned} & S_k^T(\Delta^T)_j - \sum_{\nu=1}^{q} a_\nu (-1)^q (1-\rho_\nu^{-1})^q \rho_\nu^{k-j} \\ &= \sum_{\nu=1}^{q} a_\nu \left\{ \sum_{\ell=0}^{q} w_\ell \rho_\nu^{|k-j-\ell|} \right\} - \sum_{\nu=1}^{q} a_\nu \left\{ \sum_{\ell=0}^{q} w_\ell \rho_\nu^{k-j-\ell} \right\} \\ &= \sum_{\ell=0}^{q} w_\ell r_\ell, \end{aligned}$$

where $r_\ell = \sum_{\nu=1}^{q} a_\nu \left(\rho_\nu^{|k-j-\ell|} - \rho_\nu^{k-j-\ell} \right)$. Thus, if $k-j-\ell \geq 0$, $r_\ell = 0$ and if $-q < k-j-\ell < 0$, by (B.4), r_ℓ is also 0. Therefore, (B.8) holds for $k-j > 0$.

Next for $k - j \leq 0$,

$$\sum_{\ell=0}^{q} w_\ell \rho_\nu^{|k-j-\ell|} = \sum_{\ell=0}^{q} w_\ell \rho_\nu^{-(k-j-\ell)} = \rho_\nu^{-(k-j)}(-1)^q(1-\rho_\nu)^q,$$

which proves (B.9). □

Proof of Proposition B.1. By equation (B.5),

$$(S_k + R_k + T_k)^T(\mathbf{e}_j + \eta D_j) = 0, \ \ j = 1, \dots, q, \ \text{ or } j = K-q+1, \dots, K, \tag{B.16}$$

where $\mathbf{e}_j$ is a length K vector with the j^{th} element 1 and other elements 0 and D_j is the j^{th} column of D. Note that $S_k^T\mathbf{e}_j = \sum_{\nu=1}^{q} a_\nu \rho_\nu^{k-j}$, $R_k^T\mathbf{e}_j = \sum_{\nu=1}^{q} a_{k\nu}\rho_\nu^{j-1}$ and $T_k^T\mathbf{e}_j = \sum_{\nu=1}^{q} \tilde{a}_{k\nu}\rho_\nu^{K-j}$. Then (B.16) for $1 \leq j \leq q$ can be rewritten into a matrix form:

$$\mathbf{0}_q = (\Omega_2 A^{k-1} + \eta\Sigma_1\tilde{\Sigma}_1 B_k)\mathbf{a} + (\Omega_1 + \eta\Sigma_1\tilde{\Sigma}_1 B_1)\mathbf{a}_k + (\Omega_2 A^{K-1} + \eta\Sigma_1\tilde{\Sigma}_1 B_K)\tilde{\mathbf{a}}_k, \tag{B.17}$$

where A is a $q \times q$ diagonal matrix with the ν^{th} diagonal element ρ_ν, Σ_1 is the top left $q \times q$ submatrix of Δ^T, and $\tilde{\Sigma}_1$ is the top $q \times K$ submatrix of Δ.

Similarly, we derive that

$$\begin{aligned} \mathbf{0}_q = {} & (\Omega_2 A^{K-q-k+1} + \eta\Sigma_2\tilde{\Sigma}_2 B_k)\mathbf{a} + (\Omega_1 A^{K-q} + \eta\Sigma_2\tilde{\Sigma}_2 B_1)\mathbf{a}_k \\ & + (\Omega_2 A^{q-1} + \eta\Sigma_2\tilde{\Sigma}_2 B_K)\tilde{\mathbf{a}}_k, \end{aligned} \tag{B.18}$$

where Σ_2 is the bottom right $q \times q$ submatrix of Δ^T and $\tilde{\Sigma}_2$ is the bottom $q \times K$ submatrix of Δ.

Let $A_1 = \Omega_2 A^{k-1} + \eta\Sigma_1\tilde{\Sigma}_1 B_k$, $A_2 = \Omega_1 + \eta\Sigma_1\tilde{\Sigma}_1 B_1$, $A_3 = \Omega_2 A^{K-1} + \eta\Sigma_1\tilde{\Sigma}_1 B_K$, $A_4 = \Omega_2 A^{K-q-k+1} + \eta\Sigma_2\tilde{\Sigma}_2 B_k$, $A_5 = \Omega_1 A^{K-q} + \eta\Sigma_2\tilde{\Sigma}_2 B_1$, and $A_6 = \Omega_2 A^{q-1} + \eta\Sigma_2\tilde{\Sigma}_2 B_K$. Then, we derive from (B.18) that

$$\tilde{\mathbf{a}}_k = -A_6^{-1}(A_4\mathbf{a} + A_5\mathbf{a}_k), \tag{B.19}$$

and then from (B.17) that

$$\mathbf{a}_k = -(I - A_2^{-1}A_3A_6^{-1}A_5)^{-1}A_2^{-1}(A_1 - A_3A_6^{-1}A_4)\mathbf{a}. \tag{B.20}$$

By Lemma B.1, because of A_3, each element of $A_2^{-1}A_3A_6^{-1}A_5$ and $A_3A_6^{-1}A_4$ as a function of h_e^{-1} decays to 0 exponentially fast and uniformly for $k \leq K/2$. Therefore,

$$\mathbf{a}_k = -A_2^{-1}A_1\mathbf{a} + o\left(\eta^{-\frac{1}{q}}\right)\exp\left(-Ch_e^{-1}\right)\mathbf{1}_q, \tag{B.21}$$

where the little o is uniform for $k \leq K/2$ and C is a constant.

Define $\mathbf{b}_k = -A_2^{-1}A_1\mathbf{a}$. Then,

$$(\Omega_1 + \eta\Sigma_1\tilde{\Sigma}_1B_1)\mathbf{b}_k = -(\Omega_2A^{k-1} + \eta\Sigma_1\tilde{\Sigma}_1B_k)\mathbf{a}.$$

A matrix perturbation analysis shows that

$$\mathbf{b}_k = -(\tilde{\Sigma}_1B_1)^{-1}(\tilde{\Sigma}_1B_k)\mathbf{a} + O\left(\eta^{-\frac{1}{q}}\right)\exp\left(-\psi_0\frac{k-1}{h_e}\right)\mathbf{1}_q, \tag{B.22}$$

where the big O is uniform for $k \leq K/2$.

Define $\mathbf{c}_k = -(\tilde{\Sigma}_1B_1)^{-1}(\tilde{\Sigma}_1B_k)\mathbf{a}$, which gives

$$\tilde{\Sigma}_1B_1\mathbf{c}_k = -\tilde{\Sigma}_1B_k\mathbf{a}. \tag{B.23}$$

It can be derived that $\tilde{\Sigma}_1B_1 = \Omega_1L_1$, where L_1 is a $q \times q$ diagonal matrix with the ν^{th} diagonal element $(-1)^q(1-\rho_\nu)^q$ and Φ_1 (as well as Φ_2 used later) is defined in Section B.1. Similarly, by Lemma B.2, we derive that $\tilde{\Sigma}_1B_k = \Omega_2L_2A^{k-1}$, where L_2 is a $q \times q$ diagonal matrix with the ν^{th} diagonal element $(-1)^q(1-\rho_\nu^{-1})^q$. Thus, (B.23) becomes

$$\Omega_1L_1\mathbf{c}_k = -\Omega_2L_2A^{k-1}\mathbf{a}. \tag{B.24}$$

By row transformations on Ω_1 and Ω_2, we obtain from (B.24) that

$$\Phi_1\mathbf{c}_k = -\Phi_2A^{k-1}\mathbf{a},$$

where Φ_1 and Φ_2 are defined in Section B.1. By (B.2) and (B.3), a simple matrix perturbation analysis gives

$$\mathbf{c}_k = \frac{1}{2q}\eta^{-\frac{1}{2q}}\Psi_{q,1}^{-1}\Psi_{q,2}\mathbf{r}\left(\frac{k-1}{\eta^{\frac{1}{2q}}}\right) + O\left(\eta^{-\frac{1}{q}}\right)\exp\left(-\psi_0\frac{k-1}{h_e}\right)\mathbf{1}_q, \tag{B.25}$$

where $\Psi_{q,1}$ and $\Psi_{q,2}$ are defined in Section B.1 and the big O is uniform for $k \leq K/2$. Combining (B.21), (B.22) and (B.25), we obtain (B.10). Then (B.11) follows from (B.10), (B.19) and Lemma B.1. □

Proof of Proposition B.2. Using the same notation in the proof of Proposition B.1, we first derive that

$$A_2\mathbf{a}_k + A_3\tilde{\mathbf{a}}_k \quad = \tilde{\mathbf{e}}_k,$$

$$A_5\mathbf{a}_k + A_6\tilde{\mathbf{a}}_k \quad = \mathbf{0}_q,$$

where $\tilde{\mathbf{e}}_k$ is a length q vector with the k^{th} element 1 and others 0. Then we get

$$\tilde{\mathbf{a}}_k = -A_6^{-1}A_5\mathbf{a}_k, \tag{B.26}$$

and

$$\mathbf{a}_k = (I - A_2^{-1}A_3A_6^{-1}A_5)^{-1}A_2^{-1}\tilde{\mathbf{e}}_k.$$

By Lemma B.1, we obtain that every element of $A_2^{-1}A_3A_6^{-1}A_5$ decays exponentially fast as a function of h_e^{-1}. Therefore,

$$\mathbf{a}_k = A_2^{-1}\tilde{\mathbf{e}}_k + O\left(\eta^{-\frac{1}{q}}\right). \tag{B.27}$$

Let $\mathbf{b}_k = A_2^{-1}\tilde{\mathbf{e}}_k$, i.e., $(\Omega_1 + \eta\Sigma_1\tilde{\Sigma}_1B_1)\mathbf{b}_k = \tilde{\mathbf{e}}_k$. Then another matrix perturbation analysis shows that

$$\mathbf{b}_k = (\tilde{\Sigma}_1B_1)^{-1}(\eta^{-1}\Sigma_1^{-1}\tilde{\mathbf{e}}_k) + O\left(\eta^{-\frac{1}{q}}\right). \tag{B.28}$$

Now let $\mathbf{c}_k = (\tilde{\Sigma}_1B_1)^{-1}(\eta^{-1}\Sigma_1^{-1}\tilde{\mathbf{e}}_k)$, which gives

$$\tilde{\Sigma}_1B_1\mathbf{c}_k = \eta^{-1}\Sigma_1^{-1}\tilde{\mathbf{e}}_k. \tag{B.29}$$

By the proof of Proposition B.1, $\tilde{\Sigma}_1B_1 = \Omega_1L_1$. Let R be the unique transformation matrix such that $R\Omega_1L_1 = \Phi_1$, we derive that

$$\mathbf{c}_k = \Psi_{q,1}^{-1}\Xi\left\{R(\Sigma_1)^{-1}\right\}\tilde{\mathbf{e}}_k + O\left(\eta^{-\frac{1}{q}}\right),$$

where Ξ is a $q\times q$ diagonal matrix with the j^{th} diagonal element $\eta^{-\frac{q-j+1}{2q}}$. This implies that

$$\mathbf{c}_k = \eta^{-\frac{1}{2q}}\Psi_{q,1}^{-1}\tilde{\mathbf{e}}_q\tilde{\mathbf{e}}_q^T\left\{R(\Sigma_1)^{-1}\right\}\tilde{\mathbf{e}}_k + O\left(\eta^{-\frac{1}{q}}\right).$$

It is easy to verify that $\tilde{\mathbf{e}}_q^TR(\Sigma_1)^{-1} = (-1)^{q+1}\mathbf{1}_q^T$. It follows that

$$\mathbf{c}_k = (-1)^{q+1}\eta^{-\frac{1}{2q}}\Psi_{q,1}^{-1}\tilde{\mathbf{e}}_q + O\left(\eta^{-\frac{1}{q}}\right), \tag{B.30}$$

and the proof is complete by combining (B.27), (B.28) and (B.30). □

Appendix C: Lower & upper bounds on the variance of penalized splines

Lemma C.1 (A concentration inequality on the variance of penalized splines)**.** *Suppose that Assumptions 1 - 6 hold. Assume further that $e_i \sim N(0,\sigma^2)$. Let $Z = \int_{\mathcal{T}}\{\hat{f}(x) - \mathbb{E}\hat{f}(x)\}^2\rho(x)dx$ and $\Sigma = \frac{1}{n^2}NH_n^{-1}GH_n^{-1}N^T$. Then there exists an absolute constant $c > 0$ such that, for every $t > 0$,*

$$\mathbb{P}\left\{\left|Z - \sigma^2\mathrm{tr}(\Sigma)\right| \geq t\sigma^2\mathrm{tr}(\Sigma)\right\} \leq 4\exp\left[-c\min\left(\frac{\pi^2}{4}\frac{t^2\{\mathrm{tr}(\Sigma)\}^2}{\|\Sigma\|_F^2}, \frac{\pi}{2}\frac{t\{\mathrm{tr}(\Sigma)\}}{\|\Sigma\|_2}\right)\right].$$

In particular, c can be chosen as $\frac{1}{64\pi}$.

Remark C.1. *It can be shown that* $\mathrm{tr}(\Sigma) \simeq (nh_e)^{-1}$*,* $\|\Sigma\|_F^2 \simeq (n^2 h_e)^{-1}$ *and* $\|\Sigma\|_2 = O(n^{-1})$*. Thus,* $\{\mathrm{tr}(\Sigma)\}^2/\|\Sigma\|_F^2 \geq c h_e^{-1}$ *and* $\mathrm{tr}(\Sigma)/\|\Sigma\|_2 \geq c h_e^{-1}$ *for some constant* $c > 0$*. Thus, the concentration inequality indeed shows that* Z *concentrates at* $\sigma^2 \mathrm{tr}(\Sigma)$*.*

Proof. It can be shown that $Z = \mathbf{e}^T \Sigma \mathbf{e}$, where $\mathbf{e} = (e_1, \ldots, e_n)^T$. Then $\mathbb{E}Z = \sigma^2 \mathrm{tr}(\Sigma)$. Let $C = \|e_1\|_{\psi_2}$, the sub-Gaussian norm of the random variable e_1 (see Definition 5.7 in [35]). By the Hanson-Wright inequality [23], there exists an absolute constant $c > 0$ such that, for every $t > 0$,

$$\mathbb{P}\left\{\left|Z - \sigma^2 \mathrm{tr}(\Sigma)\right| \geq t\right\} \leq 2\exp\left[-c\min\left(\frac{t^2}{C^4\|\Sigma\|_F^2}, \frac{t}{C^2\|\Sigma\|_2}\right)\right].$$

Note that for Gaussian random variables, $C = \sigma\sqrt{2/\pi}$. Hence, we obtain the desired inequality. By going through the proofs in [23] and [35], we find that c can be chosen as $\frac{1}{64\pi}$.

The proof is now complete. □

The following Lemma is adapted from [19] using the volume-of-tube formula [33].

Lemma C.2 (An equality on the supremum of variance of penalized splines)**.** *Suppose that Assumptions 1 - 6 hold. Assume further that* $e_i \sim N(0, \sigma^2)$*. Let* $Z(x) = \hat{f}(x) - \mathbb{E}\hat{f}(x)$*. Denote* $N^T(x) H_n^{-1} N^T/n$ *by* $\mathbf{g}(x)$ *and define* $\mathbf{v}(x) = \frac{\mathbf{g}(x)}{\|\mathbf{g}(x)\|_2}$*. Let* $c_0 = \int_{x\in\mathcal{T}} \left\|\frac{d\mathbf{v}(x)}{dx}\right\|_2 dx$*. Then, for* $t > 0$*,*

$$\mathbb{P}\left\{\sup_{x\in\mathcal{T}} \frac{|Z(x)|}{\sigma\|\mathbf{g}(x)\|_2} \geq t\right\} = \frac{c_0}{\pi}\exp\left(-\frac{t^2}{2}\right) + 2\left\{1 - \Phi(t)\right\} + o\left\{\exp\left(-\frac{t^2}{2}\right)\right\},$$

where Φ *is the cumulative distribution function of the standard normal and the little* o *is with respect to* t*.*

Remark C.2. *It can be shown that (see also the derivation in [19]) that* $c_0 \simeq K$*. Thus,* $\sup_{x\in\mathcal{T}} \frac{|Z(x)|}{\sigma\|\mathbf{g}(x)\|_2}$ *is of order* $\sqrt{\log K}$*. By Assumption 1, the order is equivalent to* $\sqrt{\log n}$*. Note that*

$$\frac{1}{\sigma \sup_x \|\mathbf{g}(x)\|_2}\|Z\| \leq \sup_{x\in\mathcal{T}} \frac{|Z(x)|}{\sigma\|\mathbf{g}(x)\|_2} \leq \frac{1}{\sigma \inf_x \|\mathbf{g}(x)\|_2}\|Z\|.$$

By Assumption 7, we can show that $\sup_x \|\mathbf{g}(x)\|_2 \simeq \inf_x \|\mathbf{g}(x)\|_2 \simeq (nh_e)^{-\frac{1}{2}}$*. Thus, up to multiplicative constants,* $\left(\frac{\log n}{nh_e}\right)^{\frac{1}{2}}$ *is both a lower and upper bound for* $\|Z\|$*.*

References

[1] Agarwal, G. G. and Studden, W. J. [1980], 'Asymptotic integrated mean square error using least squares and bias minimizing splines', *Ann. Statist.* **8**(6), 1307–1325. MR0594647

[2] Anderson, T. [1971], *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc, New York. MR0283939
[3] Barrow, D. L. and Smith, P. W. [1978] , 'Asymptotic properties of best $L_2[0,1]$ approximation by splines with variable knots', *Quarterly of Applied Mathematics* **36**(3), 293–304. MR0508773
[4] Barrow, D. and Smith, P. [1979], 'Efficient L_2 approximation by splines.', *Numerische Mathematik* **33**, 101–114. MR0545745
[5] Chen, H., Wang, Y., Paik, M. C. and Choi, H. A. [2013], 'A marginal approach to reduced-rank penalized spline smoothing with application to multilevel functional data', *Journal of the American Statistical Association* **108**(504), 1216–1229. MR3174699
[6] Chen, X. and Christensen, T. M. [2015], 'Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions', *Journal of Econometrics* **188**(2), 447 – 465. Heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao. MR3383220
[7] Claeskens, G., Krivobokova, T. and Opsomer, J. D. [2009], 'Asymptotic properties of penalized spline estimators', *Biometrika* **96**(3), 529–544. MR2538755
[8] de Boor, C. [1978], *A Practical Guide to Splines*, Springer, Berlin. MR0507062
[9] Demko, S. [1977], 'Inverses of band matrices and local convergence of spline projections', *SIAM Journal on Numerical Analysis* **14**(4), 616–619. MR0455281
[10] Demko, S., Moss, W. F. and Smith, P. W. [1984], 'Decay rates for inverses of band matrices', *Mathematics of Computation* **43**(168), 491–499. MR0758197
[11] Durret, R. [2005], *Probability: Theory and Examples, Third Edition*, Thomson. MR1609153
[12] Eggermont, P. P. B. and LaRiccia, V. N. [2006], *Uniform error bounds for smoothing splines*, Vol. Number 51 of *Lecture Notes–Monograph Series*, Institute of Mathematical Statistics, Beachwood, Ohio, USA, pp. 220–237. MR2387772
[13] Eilers, P. and Marx, B. [1996], 'Flexible smoothing with B-splines and penalties (with Discussion)', *Statist. Sci.* **11**, 89–121. MR1435485
[14] Eilers, P., Marx, B. and Durban, M. [2015], 'Twenty years of p-splines', *SORT* **39**(2), 1149–186. MR3467488
[15] Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B. and Reich, D. [2011], 'Penalized functional regression', *Journal of Computational and Graphical Statistics* **20**(4), 830–851. PMID: 22368438. MR2878950
[16] Gradshteyn, I. and Ryzhik, I. [2007] , *Table of Integrals, Series, and Products*, Academic Press, New York. MR0669666
[17] Hall, P. and Opsomer, J. D. [2005], 'Theory for penalised spline regression', *Biometrika* **92**(1), 105–118. MR2158613
[18] Kauermann, G., Krivobokova, T. and Fahrmeir, L. [2009], 'Some asymptotic results on generalized penalized spline smoothing', *Journal of the*

Royal Statistical Society: Series B (Statistical Methodology) **71**(2), 487–503. MR2649606

[19] Krivobokova, T., Kneib, T. and Claeskens, G. [2010], 'Simultaneous confidence bands for penalized spline estimators', *Journal of the American Statistical Association* **105**(490), 852–863. MR2724866

[20] Lai, M.-J. and Wang, L. [2013], 'Bivariate penalized splines for regression', *Statistica Sinica* **23**(3), 1399–1417. MR3114719

[21] Li, Y. and Ruppert, D. [2008], 'On the asymptotics of penalized splines', *Biometrika* **95**(2), 415–436. MR2521591

[22] O'Sullivan, F. [1986], 'A statistical perspective on ill-posed inverse problems', *Statist. Sci.* **1**(4), 502–518. MR0874480

[23] Rudelson, M. and Vershynin, R. [2013] , 'Hanson-wright inequality and subgaussian concentration', *Electron. Commun. Probab.* **18**, 9 pp. MR3125258

[24] Ruppert, D., Wand, M. and Carroll, R. [2003], *Semiparametric Regression*, Cambridge, New York. MR1998720

[25] Ruppert, D., Wand, M. and Carroll, R. J. [2009], 'Semiparametric regression during 2003?2007', *Electron. J. Statist.* **3**, 1193–1256. MR2566186

[26] Schumaker, L. [1981], *Spline Functions: Basic Theory*, Wiley-Interscience. MR0606200

[27] Schwarz, K. and Krivobokova, T. [2016], 'A unified framework for spline estimators', *Biometrika* **103**(1), 121. MR3465825

[28] Seber, G. [2007], *A Matrix Handbook for Statisticians*, Wiley-Interscience, New Jersey. MR2365265

[29] Serfling, R. [1980], *Approximation Theorems of Mathematical Statistics*, Wiley, New York. MR0595165

[30] Silverman, B. W. [1984], 'Spline smoothing: The equivalent variable kernel method', *Ann. Statist.* **12**(3), 898–916. MR0751281

[31] Speckman, P. [1985], 'Spline smoothing and optimal rates of convergence in nonparametric regression models', *Ann. Statist.* **13**(3), 970–983. MR0803752

[32] Stone, C. J. [1982], 'Optimal global rates of convergence for nonparametric regression', *Ann. Statist.* **10**(4), 1040–1053. MR0673642

[33] Sun, J. and Loader, C. R. [1994], 'Simultaneous confidence bands for linear regression and smoothing', *Ann. Statist.* **22**(3), 1328–1345. MR1311978

[34] Utreras, F. [1983], 'Natural spline functions, their associated eigenvalue problem', *Numerische Mathematik* **42**(1), 107–117. MR0716477

[35] Vershynin, R. [2012], *Introduction to the non-asymptotic analysis of random matrices*, Cambridge University Press, p. 210–268. MR2963170

[36] Wahba, G. [1990], *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics. MR1045442

[37] Wang, X., Shen, J. and Ruppert, D. [2011], 'On the asymptotics of penalized spline smoothing', *Electron. J. Statist.* **5**, 1–17. MR2763795

[38] Wood, S. [2006], 'Low-rank scale-invariant tensor product smooths for generalized additive mixed models', *Biometrics* **62**, 1025–1036. MR2297673

[39] Xiao, L. [2018], 'Asymptotics of bivariate penalised splines', *Journal of Nonparametric Statistics* **0**(0), 1–26.

[40] Xiao, L., Li, C., Checkley, W. and Crainiceanu, C. [2018], 'Fast covariance estimation for sparse functional data', *Statistics and Computing* **28**(3), 511–522. MR3761337

[41] Xiao, L., Li, Y., Apanasovich, T. and Ruppert, D. [2012], Local asymptotics of P-splines. Technical report. Available at https://arxiv.org/abs/1201.0708.

[42] Xiao, L., Li, Y. and Ruppert, D. [2013], 'Fast bivariate *P*-splines: the sandwich smoother', *J. R. Statist. Soc. B* **75**, 577–599. MR3065480

[43] Xiao, L., Zipunnikov, V., Ruppert, D. and Crainiceanu, C. [2016], 'Fast covariance estimation for high-dimensional functional data', *Statistics and Computing* **26**(1), 409–421. MR3439382

[44] Yao, F. and Lee, T. C. M. [2006], 'Penalized spline models for functional principal component analysis', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(1), 3–25. MR2212572

[45] Yoshida, T. and Naito, K. [2014], 'Asymptotics for penalised splines in generalised additive models', *Journal of Nonparametric Statistics* **26**(2), 269–289. MR3195983

[46] Zhou, S., Shen, X. and Wolfe, D. A. [1998], 'Local asymptotics for regression splines and confidence regions', *The Annals of Statistics* **26**(5), 1760–1782. MR1673277