AN EQUIVALENT KERNEL METHOD FOR LEAST SQUARES SPLINE REGRESSION

by

S.Y. Huang and W.J. Studden
Wayne State University Purdue University

Technical Report #92-09

Department of Statistics
Purdue University

February 1992

*

AN EQUIVALENT KERNEL METHOD FOR LEAST SQUARES SPLINE REGRESSION

by

Su-Yun C. Huang and William J. Studden *

Department of Mathematics
Wayne State University
Detroit, MI 48202

Department of Statistics
Purdue University
W. Lafayette, IN 47907

AMS 1991 subject classification. Primary 62G07

Key words and phrases. Nonparametric regression, spline projection kernel, reproducing kernel Hilbert space.

* Research supported by NSF grant No. DMS-9101730

Abstract

This article is to study the asymptotically equivalent kernel method for least squares splines and its asymptotic behavior. This equivalent kernel is derived from an L_2 projection onto a certain spline space. The pointwise bias and variance of this equivalent kernel estimator display an interesting dependence on the distance from a knot. It is well known that the least squares spline has no dominant boundary effect (Agarwal and Studden, 1980). This equivalent kernel method gives some insight into the boundary behavior of the least squares spline.

1. Introduction

Consider the regression problem where we have observations Y_i following the model

$$Y_i = g(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{1.1}$$

Assume that the design points t_i are on a finite interval $[0,1]$ and ε_i are uncorrelated random errors with zero mean and common variance σ^2. The cubic smoothing spline estimator $\hat{g}$ of the regression curve is the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_i - g(t_i)\}^2 + \lambda \int_0^1 g''(t)^2 dt \tag{1.2}$$

over functions g in the Sobolev space $W^2 = \{g : g$ and g' are absolutely continuous on $[0,1]$ and $\int_0^1 g''(t)^2 dt < \infty\}$. The minimizer $\hat{g}$ of (1.2) is a natural cubic spline with knots at each design point t_i. This spline smoother can be written as (Silverman, 1984)

$$\hat{g}(s) = \frac{1}{n}\sum_{i=1}^{n} G(s, t_i) Y_i,$$

where $G(s,t)$ is certain weight function depending on the design points t_i and the smoothing parameter λ. Silverman (1984) showed that the weight function is approximately of a form corresponding to smoothing by a translation kernel function with bandwidth varying according to the local density $f(t)$ of the design points. The approximation is given by

$$G(s,t) \simeq \frac{1}{f(t)}\frac{1}{h(t)} K\left(\frac{s-t}{h(t)}\right)$$

with effective bandwidth $h(t) = \lambda^{1/4} f(t)^{-1/4}$ and K a certain kernel.

In the discussion to a paper by Silverman (1985), Drs. R.L. Parker and J.A. Rice inquired and comment on their use of a least squares spline (a spline with fewer knots than there are data points, fitted to the data by least squares) and a penalized version of it. The least squares spline can be written as

$$\hat{g}(s) = \frac{1}{n}\sum_{i=1}^{n} N'(s) M^{-1}(n) N(t_i) Y_i, \tag{1.3}$$

where $N'(s) = (N_1(s), \ldots, N_m(s))$ consists of the B-spline basis (see Schumaker (1981), § 4.3) and $M(n)$ depends on the design points t_i as follows. The (i,j)-th entry of $M(n)$ is

$$M_{ij}(n) = \frac{1}{n} \sum_{k=1}^{n} N_i(t_k) N_j(t_k), \qquad i,j = 1, \ldots, m.$$

Denote the weight function involved in the least squares spline smoothing as $K^{LSS}(s,t)$, i.e.

$$K^{LSS}(s,t) = N'(s) M^{-1}(n) N(t). \tag{1.4}$$

Parker and Rice also comment in the same paper that the asymptotically equivalent kernel of the least squares spline was unknown, and that the amount of smoothness, controlled by the number of knots used, could not be adjusted continuously.

The main purpose of this paper is to study the asymptotically equivalent kernel for least squares spline smoothing and the asymptotic behavior of this equivalent kernel. Our study of $K^{LSS}(s,t)$ will show that, under certain conditions, the weight function will be approximately of a form corresponding to smoothing by a spline projection kernel, which will be explicitly derived in section 2. This projection kernel, denoted by $K(s,t)$, is involved in the L_2 projection of functions onto certain linear space of splines. The L_2 projection of f, for Lebesque measure, can generally be written as

$$f_p(t) = \int K(s,t) f(s) ds.$$

We will be studying here spaces of cubic splines with a knot sequence becoming dense over the interval of interest. For a fixed t, the kernel $K(s,t)$ integrates to one and should approach a "delta function" at t as the knot sequence becomes dense. The kernel is not of a translation type, i.e. $K(s,t) \neq k(s-t)$. However, for fixed t, the kernel $K(s,t)$ can be written as a convex combination of translation type kernels. The weights, depending on t, are certain non-negative splines adding to one. For the interior of $[0,1]$ we will use a scaled version of the projection kernel onto the space of cubic splines on $(-\infty, \infty)$ with integer knots.

The dominant boundary effects on bias are well known for estimators like smoothing splines, interpolating splines and kernel estimators. See Gasser and Müller (1979), Rosenblatt (1976, 1981), Rice and Rosenblatt (1983), and Rice (1984). It is noteworthy that a least squares spline approximation has no such dominant boundary effects regardless of the boundary behavior of g (Agarwal and Studden, 1980). Gasser and Müller (1979) suggested modified boundary kernels to correct the boundary effects. We will show that the spline projection kernel for the interval $[0, 1]$ with equally spaced knots can be used as equivalent boundary kernels. This phenomenon gives insight into the boundary behavior of the least squares smoother and also explains well its absence of boundary effects.

The rest of the paper is organized as follows. The derivation of the spline projection kernels along with some properties is in section 2. Section 3 shows the asymptotically equivalent kernels for least squares spline smoothing. Section 4 deals with the asymptotic behavior of the spline projection kernel. Some discussions on boundary behavior, variable knots and design, and local behavior are in section 5. Most of the proofs are in the appendices.

2. The spline projection kernels

We will demonstrate how the cubic spline projection kernels on $(-\infty, \infty)$ are derived. The derivation of spline projection kernels of other orders is parallel to the cubic one. For simplicity of notations and formulas, we may choose the knot sequence to be $\{i\}_{i \in Z}$. Let

$$N'(t) = (\ldots, N_{-1}(t), N_0(t), N_1(t), \ldots) \tag{2.1}$$

be the normalized B-spline basis indexed in such a way that each $N_i(t)$ is centered at i. (See Schumaker (1981), § 4.3.) Let

$$S^2 = \{s(x) : \ \ s(x) = N'(x)\theta \ \text{ with } \ \theta = (\theta_i)_{i \in Z} \in \ell_2\}.$$

That is, S^2 is the $L_2(R)$ subspace of cubic splines with integer knots. Here $R = (-\infty, \infty)$. It is well known that the B-spline basis $N(t)$ constitutes an unconditional basis for S^2, in the sense that, there exist positive constants A and B such that

$$A\|\theta\|^2_{\ell_2} \ \leq \ \|N'(t)\theta\|^2_{L_2(R)} \ \leq \ B\|\theta\|^2_{\ell_2} \tag{2.2}$$

for all $\theta \in \ell_2$. Define

$$M = \int_{-\infty}^{\infty} N(t)N'(t)dt,$$

that is, the (i,j)-th entry of M is given by

$$M_{ij} = \int_{-\infty}^{\infty} N_i(t) N_j(t) dt.$$

This inner product matrix, whose figures are listed below, is a bi-infinite symmetric band matrix with bandwidth 7.

$$(2.3) \qquad M = \frac{1}{5040} \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & & 0 \\ \cdot & 2416 & 1191 & 120 & 1 & & \\ \cdot & 1191 & 2416 & 1191 & 120 & 1 & \\ \cdot & 120 & 1191 & 2416 & 1191 & 120 & \cdot \\ & 1 & 120 & 1191 & 2416 & 1191 & \cdot \\ & & 1 & 120 & 1191 & 2416 & \cdot \\ 0 & & & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Regarding M as linear operator in ℓ_2, the inequalities in (2.2) become

$$A\|\theta\|_{\ell_2}^2 \leq < \theta, M\theta > \leq B\|\theta\|_{\ell_2}^2,$$

where $<,>$ denotes inner product in ℓ_2. Therefore $M\theta = 0$ implies that $\theta = 0$. Hence M^{-1} exists in the range of M. The inverse of M is obtained by solving $Mx = e$, where $e = (e_i)_{i \in Z}$ with $e_i = \delta_{0i}$. That is, we solve the following difference equations.

$$x_{n-3} + 120x_{n-2} + 1191x_{n-1} + 2416x_n + 1191x_{n+1} + 120x_{n+2} + x_{n+3} = 5040\delta_{0n},$$

for $n = \ldots, -2, -1, 0, 1, 2, \ldots$. The result is given below.

Lemma 2.1. The solution to the above difference equations is

$$x_n = \sum_{\ell=1}^{3} c_\ell r_\ell^{|n|}.$$

The values (up to nine decimal places) of the constants c_ℓ and r_ℓ are listed below.

$c_1 =$ 6.0162 84002
$c_2 =$ -1.0558 20294
$c_3 =$ 0.0042 69179
$r_1 =$ -.5352 80431
$r_2 =$ -.1225 54615
$r_3 =$ -.0091 48695.

Proof. See Huang (1990).

Since each row of M is simply a shift of the previous row one entry to the right, Proposition 2.1 is an immediate result of Lemma 2.1.

Proposition 2.1. The (i,j)-th entry of M^{-1} is given by

$$M_{ij}^{-1} = \sum_{\ell=1}^{3} c_\ell r_\ell^{|i-j|}.$$

Theorem 2.1. The projection of $f \in L_2(R)$ onto S^2 can be written as

$$f_p(s) = \int_{-\infty}^{\infty} K(s,t) f(t) dt, \tag{2.4}$$

where $K(s,t) = N'(s) M^{-1} N(t)$.

Proof. Since $f_p \in S^2$, $f_p(t) = N'(t)\theta$ for some $\theta \in \ell_2$. The normal equations are

$$\int_{-\infty}^{\infty} N(t)N'(t)\theta dt = \int_{-\infty}^{\infty} N(t) f(t) dt.$$

Then, we have

$$\theta = M^{-1} \int_{-\infty}^{\infty} N(t) f(t) dt.$$

Therefore

$$\begin{aligned} f_p(s) =& N'(s) M^{-1} \int_{-\infty}^{\infty} N(t) f(t) dt \\ =& \int_{-\infty}^{\infty} K(s,t) f(t) dt. \end{aligned}$$ □

We name the kernel $K(s,t)$ in (2.4) the spline projection kernel. For a fixed t, the kernel $K(\cdot,t)$ is not of translation type and its shape depends on the position of t relative to knots. Plots of $K(\cdot,t)$ for various values of t are given in Figure 2.1.

Figure 2.1 here

Though $K(\cdot,t)$ is not of translation type, it can be written as a convex combination of translation type kernels as will be presented below. Define

$$H(t) = \sum_{i=-\infty}^{\infty} \left(\sum_{\ell=1}^{3} c_\ell r_\ell^{|i|} \right) N_i(t). \tag{2.5}$$

Then we have

$$\begin{aligned} K(\cdot,t) &= \sum_{i,j=-\infty}^{\infty} \sum_{\ell=1}^{3} c_\ell r_\ell^{|i-j|} N_i(\cdot) N_j(t) \\ &= \sum_{i,j=-\infty}^{\infty} \sum_{\ell=1}^{3} c_\ell r_\ell^{|i-j|} N_{i-j}(\cdot - j) N_j(t) \\ &= \sum_{j=-\infty}^{\infty} N_j(t) H(\cdot - j) \\ &= \sum_{j=[t]-1}^{[t]+2} N_j(t) H(\cdot - j). \end{aligned}$$

Let $w = t - [t]$. Then we have

$$K(\cdot,t) = \sum_{\ell=-1}^{2} N_\ell(w) H(\cdot - [t] - \ell).$$

Proposition 2.2. The spline projection kernel can be written as

$$K(\cdot,t) = \sum_{j=[t]-1}^{[t]+2} N_j(t) H(\cdot - j) \tag{2.6}$$

or

$$= \sum_{\ell=-1}^{2} N_\ell(w) H(\cdot - [t] - \ell),$$

where $w = t - [t]$. Note that $N_\ell(w) \geq 0$ and $\sum_{\ell=-1}^{2} N_\ell(w) \equiv 1$.

All the above discussions are based on splines with integer knots. Consider the knot sequence $\{ih\}_{i \in Z}$. The normalized B-spline basis becomes

$$N'\left(\frac{t}{h}\right) = \left(\ldots, N_{-1}\left(\frac{t}{h}\right), N_0\left(\frac{t}{h}\right), N_1\left(\frac{t}{h}\right), \ldots\right).$$

The inner product matrix then becomes

$$\int_{-\infty}^{\infty} N\left(\frac{t}{h}\right) N'\left(\frac{t}{h}\right) dt = hM.$$

And then the projection kernel is given by

$$K_h(s,t) = \frac{1}{h} K\left(\frac{s}{h}, \frac{t}{h}\right). \tag{2.7}$$

Following is a list of properties of the $L_2(R)$ spline projection.

Property 1. The $L_2(R)$ subspace of cubic splines with knots $\{ih\}_{i \in Z}$ given by

$$S_h^2 = \{s(t) : s(t) = N'\left(\frac{t}{h}\right)\theta, \qquad \theta \in \ell_2\}$$

is a reproducing kernel Hilbert space with reproducing kernel $K_h(s,t)$.

Property 2. The integral transform

$$f_p(s) = \int_{-\infty}^{\infty} K_h(s,t) f(t) dt$$

is well-defined for functions satisfying the following two conditions:

$$f \in L_1^{\text{local}} = \{f : \int_A |f| < \infty \text{ for all measurable } A \text{ with finite measure } \}$$

and

$$f(t) = O(|t|^\alpha) \text{ for some } \alpha \text{ as } |t| \to \infty.$$

This desirable property enables us to apply the "projection" to a larger class of functions than $L_2(R)$.

Property 3. Define

$$S_h^\infty = \{s(t) : s(t) = N'\left(\frac{t}{h}\right)\theta, \qquad \theta \in \ell_\infty\}.$$

The kernel reproduces S_h^∞ in the sense that

$$\int_{-\infty}^{\infty} K_h(s,t) f(t) dt = f(s) \text{ for } f \in S_h^\infty.$$

Property 4. The kernel $K_h(s,t)$ reproduces polynomials of order 4 (degree ≤ 3). Especially,

$$\int_{-\infty}^{\infty} K_h(s,t) dt \equiv 1.$$

Property 5. $H(t)$ is an order 4 kernel, i.e.

$$\int_{-\infty}^{\infty} H(t) t^m dt = \delta_{0m}, \qquad m = 0, 1, 2, 3$$

$$\int_{-\infty}^{\infty} H(t) t^4 dt \neq 0$$

Property 6. The kernel decays to zero at an exponential rate; i.e.

$$|K_h(s,t)| \le c\gamma^n,$$

where $n = |[\frac{s}{h}] - [\frac{t}{h}]|$, is the number of knots between s and t, and $c > 0$, $0 < \gamma < 1$ are some constants.

Property 7.

$$\int_{-\infty}^{\infty} K_h(s,t)t^4 dy = s^4 - B_4(w)h^4,$$

where $w = \frac{s}{h} - [\frac{s}{h}]$ and $B_4(w) = w^4 - 2w^3 + w^2 - \frac{1}{30}$ is the 4-th Bernoulli polynomial. The above equality says the error of approximating the polynomial t^4 by its "spline projection" behaves exactly as the 4-th Bernoulli polynomial in each cell between knots. The magnitude of the error is of order $O(h^4)$.

We refer to Huang and Studden (1990) for the proofs of the above properties.

3. Asymptotically equivalent kernel for least squares spline smoothing

In section 2, the $L_2(R)$ spline projection kernel was derived. Here we will introduce the $L_2[0,1]$ spline projection kernel. Notations K_h^R and K_h^I, $I = [0,1]$, will be used throughout to distinguish them. We will show that K_h^R is an asymptotically equivalent kernel for least squares spline smoothing in the interior of $[0,1]$. However this approximation deteriorates near the boundary. The asymptotically equivalent kernel near boundary will be shown to be K_h^I.

It is convenient first to establish some notations and to state the assumptions under which the main results will be proved. The main assumptions are:

(3.1) The design points $\{t_i\}_{i=1}^n$ are uniformly spaced over $I = [0,1]$.

(3.2) The knot sequence is taken to be $\{ih\}_{i\in Z}$ with $h = \frac{1}{k}$ and k a positive integer. Then $\{ih\}_{i=0}^k$ are knots in I.

(3.3) The smoothing parameter $h = h(n)$ depends on n in such a way that $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

Define the following notations and some results are immediate.

(3.4) Let $N_I'(t) = (N_{-1}(\frac{t}{h}), N_0(\frac{t}{h}), \ldots, N_{k+1}(\frac{t}{h}))$ for $t \in I = [0,1]$, where $N_i(\cdot)$ is as defined in (2.1). Then $N_I(t)$ constitutes a basis for the space of cubic splines in I with interior knots $\{ih\}_{i=1}^{k-1}$.

(3.5) The inner product matrix of N_I is $M_I = \int_0^1 N_I(t)N_I'(t)dt$. The dimension of M_I is $m \times m$ with $m = k+3$.

(3.6) The projection kernel of $L_2[0,1]$ onto the space of cubic splines with interior knots $\{ih\}_{i=1}^{k-1}$ is then given by

$$K_h^I(s,t) = N_I'(s)M_I^{-1}N_I(t).$$

(3.7) The weight function of the least squares spline with interior knots $\{ih\}_{i=1}^{k-1}$ is denoted by $K_h^{LSS}(s,t)$, i.e.

$$K_h^{LSS}(s,t) = N_I'(s)M_I^{-1}(n)N_I(t)$$

with

$$M_I(n) = \frac{1}{n}\sum_{k=1}^{n} N_I(t_k)N_I'(t_k).$$

The main results are stated below.

<u>Theorem 3.1.</u> For fixed $t \in (0,1)$, we have

$$|hK_h^{LSS}(t+hx,t) - K_1^R(w+x,w)| \;\le\; O\left(\frac{1}{nh} + |r_1|^{\frac{1}{2h}}\right)$$

as $n \to \infty$, for all x such that $t + hx \in [0,1]$, where $w = \frac{t}{h} - [\frac{t}{h}]$, the distance of t to the left nearest knot scaled by h, and r_1 is the constant given in Proposition 2.1.

Theorem 3.1 says that, for large n and small h and t not too close to the boundary, the approximation of the weight function corresponding to the observation at t is

$$\begin{aligned} K_h^{LSS}(t+hx,t) &\simeq \frac{1}{h}K_1^R\left(\frac{t}{h} - \left[\frac{t}{h}\right] + x,\; \frac{t}{h} - \left[\frac{t}{h}\right]\right) \\ &= \frac{1}{h}K_1^R\left(\frac{t}{h} + x,\; \frac{t}{h}\right). \end{aligned}$$

Setting $s = t + hx$, we have the approximation

(3.8) $$K_h^{LSS}(s,t) \simeq K_h^R(s,t).$$

The above approximation deteriorates as t approaches the boundary. The next theorem gives the asymptotically equivalent kernel near the boundary.

Theorem 3.2. For fixed $t \in [0,1]$, we have

$$|hK_h^{LSS}(t+xh,\ t) - hK_h^I(t+hx,\ t)| \ \leq\ O\left(\frac{1}{nh}\right)$$

as $n \to \infty$, for all x such that $t + xh \in [0,1]$.

Set $s = t + xh$, we have the approximation

$$K_h^{LSS}(s,t) \simeq K_h^I(s,t).$$

We will show in a later section that $K_h^I(\cdot, t)$, for $t \in [0,1]$, behaves like some boundary kernels proposed in Gasser and Müller (1979).

Proofs of Theorems 3.1 and 3.2 are in Appendix A.

In order to illustrate how well the approximation (3.8) works out, some explicit calculations were done. One hundred design points $\{\frac{i-1/2}{100}\}_{i=1}^{100}$ were used and 19 interior knots were placed at $\{\frac{i}{20}\}_{i=1}^{19}$. The weight function $K_h^{LSS}(s,t_i)$, together with approximation $K_h^R(s,t_i)$, for various values of $t_i = .025, .045, .095, .495$ are shown in Figure 3.1.

Figure 3.1 here.

4. Asymptotics

We will study the asymptotic behavior of the estimator given by

$$\hat{g}(s) = \frac{1}{n}\sum_{i=1}^{n} K_h^R(s,t_i)Y_i. \tag{4.1}$$

Suppose $g \in Lip^{4,\alpha}[0,1] = \{g \in C^4[0,1] : |g^{(4)}(x+\delta) - g^{(4)}(x)| \ \leq\ M\delta^\alpha$ for all $0 \leq x \leq x + \delta \leq 1\}$ for some $\alpha > 0$ and $h \to 0$, $nh \to \infty$ as $n \to \infty$. We have the following theorems.

Theorem 4.1. For a fixed $s \in (0,1)$, we have

$$E\hat{g}(s) - g(s) = -\frac{g^{(4)}(s)}{4!}B_4(w)h^4 + O\left(h^{4+\alpha} + \frac{1}{n}\right)$$

as $n \to \infty$, where $w = \frac{s}{h} - [\frac{s}{h}]$.

Theorem 4.2. For a fixed $s \in (0,1)$, we have

$$\text{Var } \hat{g}(s) = \frac{\sigma^2}{nh} K_1^R(w,w) + O\left(\frac{1}{n}\right)$$

as $n \to \infty$, where w is defined as above.

The pointwise bias and variance display an interesting dependence on the distance from a knot. The bias and variance plots are presented in Figure 4.1.

Figure 4.1 here.

Suppose g is periodic with period 1 (otherwise use the finite support kernel $K_h^I(s,t)$ for (4.1)). The integrated mean square error (IMSE) is given below.

Theorem 4.3. As $n \to \infty$, we have

$$\text{IMSE } = \frac{|B_8|}{8!} \|g^{(4)}\|_2^2 \, h^8 + \frac{\sigma^2}{nh} + O\left(h^{8+2\alpha} + \frac{1}{n}\right),$$

where B_8 is the 8-th Bernoulli number, which has value $-\frac{1}{30}$.

Proofs are in Appendix B.

5. Some remarks

1. Boundary kernels. In a kernel smoothing problem, the bias near the boundary is of a larger order of magnitude than in the interior, unless a periodicity assumption is made. Therefore the IMSE is dominated by the boundary effects. Similar problems arise from smoothing splines also, unless again a periodicity assumption is made or the boundary behavior is known and imposed. Gasser and Müller (1979) studied the asymptotics of kernel estimators near the boundary. They considered kernel smoothing using kernels with finite support $[-1,1]$. Suppose $s(n)$ is a sequence of points satisfying

$$s(n) = qh(n), \qquad q \in [0,1),$$

where h is the bandwidth. They introduced order d boundary kernels K_q. Here we let $d = 4$, since the cubic spline projection kernel is an order 4 kernel. The order 4 boundary kernels have the following properties.

I. Moment conditions.

a.

$$\int_0^1 \frac{1}{h} K_q \left(\frac{s-t}{h}\right) \left(\frac{s-t}{h}\right)^m dt = \delta_{0m}, \qquad m = 0, 1, 2, 3,$$

which can be converted to $\int_{-q}^{1} K_q(t) t^m dt = \delta_{0,m}, \qquad m = 0, 1, 2, 3.$

b. Uniformly bounded bias.

$$\int_0^1 \frac{1}{h} K_q \left(\frac{s-t}{h}\right) \left(\frac{s-t}{h}\right)^4 dt = \alpha_q, \tag{5.1}$$

$\alpha_q \not\equiv 0$ and α_q is uniformly bounded for $q \in [0, 1)$. Note that (5.1) can be converted to

$$\int_{-q}^{1} K_q(t) t^4 dt = \alpha_q.$$

II. Uniformly bounded variance.

$$\int_0^1 \left[\frac{1}{h} K_q \left(\frac{s-t}{2}\right)\right]^2 dt = \frac{1}{h} \int_{-q}^{1} K_q^2(t) dt.$$

$\int_{-q}^{1} K_q^2(t) dt$ has to be uniformly bounded for $q \in [0, 1)$.

III. The kernels depend continuously on q and

$$K_q \to K \quad \text{as} \quad q \to 1,$$

where K is the kernel for the interior. The idea of boundary kernels is to continuously modify the kernels toward the boundary to meet the moment conditions in such a way that the pointwise bias and variance are uniformly bounded.

The cubic spline projection kernels satisfy the moment condition

$$\int_0^1 K_h^I(s,t) \left(\frac{s-t}{h}\right)^m dt = \delta_{0,m}, \qquad m = 0, 1, 2, 3 \ ,$$

since it reproduces polynomials up to order 4 (degree ≤ 3). The 4-th moment

$$\int_0^1 K_h^I(s,t) \left(\frac{s-t}{h}\right)^4 dt$$

is uniformly bounded for all $s \in [0,1]$, since $K_h^I(s,t)$ is uniformly continuous in $I \times I$. The "uniformly bounded variance" condition is also met, as

$$\int_0^1 K_h^I(s,t)^2 dt = K_h^I(s,s)$$

by the reproducing property. $K_h^I(s,s)$ is uniformly bounded for all $s \in [0,1]$.

2. A locally weighted kernel estimate. We have, from equation (2.6)

$$\begin{aligned} K_h^R(s,t_i) =& \frac{1}{h} K_1^R\left(\frac{s}{h}, \frac{t_i}{h}\right) \\ =& \frac{1}{h} \sum_{\ell=-1}^{2} N_\ell(w) H\left(\frac{s}{h} - \left[\frac{t_i}{h}\right] - \ell\right) \\ =& \frac{1}{h} \sum_{\ell=-1}^{2} N_\ell(w) H\left(\frac{s - \xi_{i,\ell}}{h}\right), \end{aligned}$$

where $w = \frac{t_i}{h} - [\frac{t_i}{h}]$ and $\xi_{i,\ell} = ([\frac{t_i}{h}] + \ell)h$. The points $\{\xi_{i,\ell}\}_{\ell=-1}^2$ range over the 4 nearest knots from t_i. Along with the remark made previously, the least squares spline smoothing is asymptotically a locally weighted kernel (translation type) smoothing for data points not too close to the boundary, and a boundary kernel smoothing for data points near boundary.

3. Variable knots and design. Suppose the design points $\{t_i\}_{i=1}^n$ have a limiting density $f(t)$ and that the sequences of knots $T_k = \{0 = \xi_0, \ldots, \xi_k = 1\}$ are given by

$$\int_{\xi_{i-1}}^{\xi_i} p(t)dt = \frac{1}{k}, \qquad i = 1, \ldots, k, \tag{5.2}$$

where $p(t)$ is a positive continuous density function on $[0,1]$. By the mean-value theorem, we have from (5.2) that

$$p(\tau_{i-1})(\xi_i - \xi_{i-1}) = \frac{1}{k}$$

for some $\tau_i \in (\xi_{i-1}, \xi_i)$. Then $(\xi_i - \xi_{i-1}) = h/p(\tau_i)$, with $h = \frac{1}{k}$. As $n \to \infty$ (then $h \to 0$), the effective local bandwidth at the point t is $h(t) = h/p(t)$. The right kernel to use, corresponding to an observation at t_i, should be

$$\frac{1}{f(t_i)} \frac{1}{h(t_i)} K_1^R\left(\frac{s}{h(t_i)}, \frac{t_i}{h(t_i)}\right),$$

where the factor $\frac{1}{f(t_i)}$ can be replaced by $n(t_i - t_{i-1})$ if $f(t)$ is not known.

The choice of optimal design, number of knots and their positions are discussed in Agarwal and Studden (1980).

Appendix A.

This section is devoted to the proofs of Theorems 3.1 and 3.2 by bounding the L_∞ norms of some inverse matrices and by bounding some entrywise differences of matrices. The following lemmas will be needed.

Lemma A.1. Let

$$A_m = \frac{1}{5040}\begin{bmatrix} 2416 & 1191 & 120 & 1 & 0 & & & & 1 & 120 & 1191 \\ 1191 & 2416 & 1191 & 120 & 1 & \cdot & & & & 1 & 120 \\ 120 & 1191 & 2416 & 1191 & 120 & \cdot & \cdot & & & & 1 \\ 1 & 120 & 1191 & 2416 & 1191 & \cdot & \cdot & \cdot & & & \\ 0 & 1 & 120 & 1191 & 2416 & \cdot & \cdot & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & 1 \\ 1 & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & 120 \\ 120 & 1 & & & \cdot & \cdot & \cdot & \cdot & \cdot & & 1191 \\ 1191 & 120 & 1 & & & \cdot & 1 & 120 & 1191 & & 2416 \end{bmatrix}_{m\times m}$$

Then the (i,j)-th entry of A_m^{-1} is

$$(A_m^{-1})_{ij} = \sum_{\ell=1}^{3} c_\ell^{(m)}(r_\ell^{|i-j|} + r_\ell^{m-|i-j|}),$$

where r_ℓ are given in Proposition 2.1 and $c_\ell^{(m)}$ are some constants with limit $c_\ell = \lim_{m\to\infty} c_\ell^{(m)}$, which is also given in Proposition 2.1.

Proof. See Huang (1990).

The next lemma is an immediate consequence of Lemma A.1 and Proposition 2.1.

Lemma A.2. For fixed $j \in Z$, let $j' = [\frac{m}{2}] + j$, we have

$$|(A_m^{-1})_{ij'} - M_{ij'}^{-1}| = O(|r_1|^m)$$

for all $i = 1, 2, \ldots, m$ as $m \to \infty$.

Lemma A.3. For fixed $j \in Z$, let $j' = [\frac{m}{2}] + j$, we have

$$|(M_I^{-1})_{ij'} - (h^{-1}A_m^{-1})_{ij'}| \le O\left(\frac{|r_1|^{m/2}}{h}\right) \qquad \text{as } m \to \infty,$$

where M_I has dimension $m \times m$, $m = k+3$, and $h = \frac{1}{k}$ as indicated in (3.2) and (3.5).

Proof.

$$M_I^{-1} - h^{-1}A_m^{-1} = M_I^{-1}(hA_m - M_I)h^{-1}A_m^{-1}. \tag{A.1}$$

Let $D = hA_m - M_I$. It is easy to check the following statements.

$$|D_{\ell,q}| \leq h \text{ for } \ell, q \in \Lambda = \{1,2,3,4,m-3,m-2,m-1,m\},$$

and

$$D_{\ell,q} = 0 \text{ for } (\ell,q) \notin \Lambda \times \Lambda.$$

Therefore the (i,j')-th entry in (A.1) can be bounded by

$$\begin{aligned} & h \sum_{\ell,q\in\Lambda} |(M_I^{-1})_{i\ell}|\, |h^{-1}(A_m^{-1})_{qj'}| \\ & \leq \|M_I^{-1}\|_\infty \sum_{q\in\Lambda} |(A_m^{-1})_{qj'}| \\ & \leq \|M_I^{-1}\|_\infty\, O\left(|r_1|^{\frac{m}{2}}\right). \end{aligned}$$

The proof can then be completed by the next lemma.

Lemma A.4.

$$h\|M_I^{-1}\|_\infty \leq \rho$$

where ρ is some constant independent of h.

Proof. See deBoor (1976).

Proposition A.1. For fixed $t \in (0,1)$, we have

$$|K_h^I(s,t) - K_h^R(s,t)| \leq O\left(\frac{|r_1|^{m/2}}{h}\right)$$

as $m \to \infty$ (or equivalently $h \to 0$) for all $s \in [0,1]$.

Proof. $|K_h^I(s,t) - K_h^R(s,t)|$

$$= \sum_{i,j=-1}^{k+1} N_i\left(\frac{s}{h}\right) N_j\left(\frac{t}{h}\right) |(M_I^{-1})_{ij} - h^{-1}M_{ij}^{-1}| \tag{A.2}$$

Since t is a fixed point in the interior, we may apply Lemmas A.2 and A.3 and get

$$|(M_I^{-1})_{ij} - h^{-1}M_{ij}^{-1}| \leq O\left(\frac{|r_1|^{m/2}}{h}\right)$$

as $m \to \infty$ (or equivalently $k \to \infty$ or $h \to 0$), for all i and those j such that $N_j\left(\frac{t}{h}\right) \neq 0$. This completes the proof.

Lemma A.5.

$$h\|M_I^{-1}(n)\|_\infty \leq \rho^*,$$

where ρ^* is some constant independent of h and the data points $\{t_i\}_{i=1}^n$.

Proof. Let $D = M_I - M_I(n)$. We have

$$M_I^{-1}(n) = (M_I - D)^{-1} = M_I^{-1}(I - M_I^{-1}D)^{-1}.$$

Thus

$$(A.3) \qquad h\|M_I^{-1}(n)\|_\infty \;\leq\; h\|M_I^{-1}\|_\infty \, \|(I - M_I^{-1}D)^{-1}\|_\infty.$$

Let $B = M_I^{-1}D$. Since $\|B\|_\infty \;\leq\; \frac{\rho}{h}\|D\|_\infty = \frac{\rho}{h}O(\frac{1}{h}) = O\left(\frac{1}{nh^2}\right)$, we have from (A.3)

$$\begin{aligned} h\|M_I^{-1}(n)\|_\infty \;&\leq\; \rho \cdot \|\sum_{k=0}^{\infty} B^k\|_\infty \\ &\leq \frac{\rho}{1 - \|B\|_\infty} \\ &\leq 2\rho, \quad \text{when } \; n \; \text{ is large enough} \quad . \end{aligned}$$

Letting $\rho^* = 2\rho$ completes the proof.

Proposition A.2. For all $s, t \in [0, 1]$ we have

$$|K_h^{LSS}(s,t) - K_h^I(s,t)| \;\leq\; O\left(\frac{1}{nh^2}\right),$$

as $n \to \infty$.

Proof. We have

$$|K_h^{LSS}(s,t) - K_h^R(s,t)| \;\leq\; \sum_{i,j=-1}^{k+1} N_i\left(\frac{s}{h}\right) N_j\left(\frac{t}{h}\right) |(M_I^{-1}(n) - M_I^{-1})_{ij}|.$$

Proof can be completed by bounding $(M_I^{-1}(n) - M_I^{-1})_{ij}$, since $0 \le \sum_{i=-1}^{k+1} N_i(\frac{s}{h}) \le 1$, $0 \le \sum_{j=-1}^{k+1} N_j(\frac{t}{h}) \le 1$.

$$|(M_I^{-1}(n) - M_I^{-1})_{ij}| = |(M_I^{-1}(n)(M_I - M_I(n))M_I^{-1})_{ij}|$$

$$= \sum_{\ell,q=-1}^{k+1} |(M_I^{-1}(n))_{i\ell}(M_I - M_I(n))_{\ell q}(M_I^{-1})_{qj}| \tag{A.4}$$

Since $(M_I - M_I(n))_{\ell q} = O(\frac{1}{n})$ for all ℓ, q,

$$(A.4) \le \|M_I^{-1}(n)\|_\infty \; \|M_I^{-1}\|_\infty \; O\left(\frac{1}{n}\right)$$

$$\le O\left(\frac{1}{nh^2}\right),$$

and the proof is complete.

Theorems 3.1 and 3.2 are immediate from Propositions A.1 and A.2.

Appendix B.

Proof of Theorem 4.1.

$$(B.1) \qquad E\hat{g}(s) - g(s) = \frac{1}{n}\sum_{i=1}^{n} K_h^R(s,t_i)g(t_i) - g(s)$$

$$= \int_0^1 K_h^R(s,t)g(t)dt + O\left(\frac{1}{n}\right) - g(s)$$

Extend g to the whole real line in such a way that $g \in Lip^{4,\alpha}(R)$. Then

$$(B.1) = \int_{-\infty}^{\infty} K_h^R(s,t)g(t)dt - \int_{R\backslash[0,1]} K_h^R(s,t)g(t)dt - g(s) + O\left(\frac{1}{n}\right)$$

$$\stackrel{\text{def}}{=} I_1 - I_2 - g(s) + O\left(\frac{1}{n}\right)$$

$$(B.2) \quad I_1 - g(s) = \int_{-\infty}^{\infty} K_h^R(s,t)\left(g(s) + g'(s)(s-t) + \ldots + \frac{g^{(4)}(\xi)}{4!}(s-t)^4\right)dt - g(s),$$

where ξ is some point between s and t. Since K_h^R reproduces polynomials up to degree 3,

$$(B.2) = \int_{-\infty}^{\infty} K_h^R(s,t)\frac{g^{(4)}(\xi)}{4!}(s-t)^4 dt$$

$$= \int_{-\infty}^{\infty} K_h^R(s,t)\frac{g^{(4)}(s)}{4!}(s-t)^4 dt + \int_{-\infty}^{\infty} K_h^R(s,t)\left(\frac{g^{(4)}(\xi) - g^{(4)}(s)}{4!}\right)(s-t)^4 dt$$

$$(B.3) \qquad = -\frac{g^{(4)}(s)}{4!}B_4(w)h^4 + \int_{-\infty}^{\infty} K_h^R(s,t)\left(\frac{g^{(4)}(\xi) - g^{(4)}(s)}{4!}\right)(s-t)^4 dt,$$

by Property 7 in section 2. The proof can be completed by showing $I_2 = o\left(\frac{1}{n}\right)$ and the second term in (B.3) is $O(h^{4+\alpha})$. We have

$$|I_2| \le \|g\|_\infty \int_{R\backslash[0,1]} \left|\frac{1}{h}K_1^R\left(\frac{s}{h},\frac{t}{h}\right)\right| dt$$

$$= \|g\|_\infty \int_{R\backslash[0,\frac{1}{h}]} \left|K_1^R\left(\frac{s}{h},t'\right)\right| dt'.$$

Since $s \in (0,1)$ is fixed and $K_1^R\left(\frac{s}{h},t'\right)$ decays to zero at an exponential rate as $\left|\frac{s}{h}-t'\right| \to \infty$, it is clear $\int_{R\backslash[0,\frac{1}{h}]} \left|K_1^R\left(\frac{s}{h},t\right)\right| dt = o\left(\frac{1}{n}\right)$. We also have

$$\left|\int_{-\infty}^{\infty} K_h^R(s,t)\left(\frac{g^{(4)}(\xi) - g^{(4)}(s)}{4!}\right)(s-t)^4 dt\right|$$

$$= \left|\sum_{m=-\infty}^{\infty} \int_{\tau+(m-1)h}^{\tau+mh} K_h^R(s,t)\left(\frac{g^{(4)}(\xi) - g^{(4)}(s)}{4!}\right)(s-t)^4 dt\right|$$

where $\tau = [\frac{s}{h}]h$,

$$\leq \sum_{m=-\infty}^{\infty} \frac{cr^{|m-1|}}{h} \frac{|mh|^{\alpha}}{4!}(mh)^4 h, \quad \text{by Property 6}$$
$$= O(h^{4+\alpha}).$$

Proof of Theorem 4.2.

$$\begin{aligned} Var\ \hat{g}(s) =& \frac{\sigma^2}{n^2}\sum_{i=1}^{n} K_h^R(s,t_i)^2 \\ =& \frac{\sigma^2}{n}\int_0^1 K_h^R(s,t)^2 dt + O\left(\frac{1}{n}\right) \\ =& \frac{\sigma^2}{n}\int_{-\infty}^{\infty} K_h^R(s,t)^2 dt + \frac{\sigma^2}{n}\int_{R\setminus[0,1]} K_h^R(s,t)^2 dt + O\left(\frac{1}{n}\right) \\ =& \frac{\sigma^2}{n} K_h^R(s,s) + \frac{\sigma^2}{n}\int_{R\setminus[0,1]} K_h^R(s,t)^2 dt + O\left(\frac{1}{n}\right). \end{aligned} \tag{B.4}$$

Note that

$$\begin{aligned} K_h^R(s,s) =& \frac{1}{h} K_1^R\left(\frac{s}{h},\frac{s}{h}\right) \\ =& \frac{1}{h} K_1^R(w,w) \end{aligned}$$

where $w = \frac{s}{h} - [\frac{s}{h}]$. Proof can be completed by showing the second term in (B.4) is $o(\frac{1}{n})$.

$$\frac{\sigma^2}{n}\int_{R\setminus[0,1]} K_h^R(s,t)^2 dt = \frac{\sigma^2}{n}\int_{R\setminus[0,\frac{1}{h}]} \frac{1}{h} K_1^R\left(\frac{s}{h},t'\right)^2 dt'$$

Since $s \in (0,1)$ is fixed and $K_1^R(\frac{s}{h},t')$ approaches zero at an exponential rate as $|\frac{s}{h} - t'| \to \infty$, it is easy to see that $\int_{R\setminus[0,\frac{1}{h}]} \frac{1}{h} K_1^R(\frac{s}{h},t')^2 dt' = o(1)$ as $h \to 0$, which completes the proof.

To show Theorem 4.3 we will need the following lemmas.

Lemma B.1.

$$\int_0^1 \left(\frac{B_4(w)}{4!}\right)^2 dw = \frac{|B_8|}{8!}$$

Proof. See Ghizzetti and Ossicini (1970).

Lemma B.2.

$$\int_0^1 K_1^R(w,w)dw = 1$$

Proof.

$$\begin{aligned}
&\int_0^1 K_1^R(w,w)dw \\
&= \int_0^1 \sum_{\ell=-\infty}^{\infty} N_\ell(w)H(w-\ell)dw \\
&= \sum_{\ell=-\infty}^{\infty} \int_0^1 N_\ell(w)H(w-\ell)dw \\
&= \sum_{\ell=-\infty}^{\infty} \int_0^1 N_0(w)H(w-\ell)dw \\
&= \int_{-\infty}^{\infty} N_0(w)H(w)dw \\
&= \int_{-\infty}^{\infty} N_0(w) \sum_{i=-\infty}^{\infty} N_i(w) \left(\sum_{\ell=1}^{3} c_\ell r_\ell^{|i|} \right) dw \\
&= \sum_{i=-\infty}^{\infty} \int_{-\infty}^{\infty} N_0(w)N_i(w) \left(\sum_{\ell=1}^{3} c_\ell r_\ell^{|i|} \right) dw \\
&= \sum_{i=-\infty}^{\infty} M_{0i}(M^{-1})_{0i} = 1.
\end{aligned}$$

Note that we exchanged the order of integration and summation, which is justified because the sum is actually a finite sum.

Theorem 4.3 is an immediate consequence of Theorems 4.1 and 4.2 and Lemmas B.1 and B.2.

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**, 1307–1325.

2. deBoor, C. (1976). A bound on the L_∞-norm of L_2-approximation by splines in terms of a global mesh ratio. *Math. Comp.*, **30**, 765–771.

3. Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. Smoothing Techniques for curve estimation (Gasser and Rosenblatt, ed.). *Lecture notes in Mathematics*, **757**, Spring-Verlag.

4. Ghizzetti, A. and Ossicini, A. (1970). *Quadrature formulae.* Academic, New York.

5. Huang, S.Y. (1990). *Density estimation using spline projection kernels.* Thesis, Purdue University.

6. Huang, S.Y. and Studden, W.J. (1990). Density estimation using spline projection kernels. Technical report **#90-52**, Purdue University.

7. Parker, R.L. and Rice, J. (1985). Discussion of "Some aspects of the spline smoothing approach to non-parametric regression curve fitting" by Silverman. *J. Roy. Statist. Soc. B* (1985), **47**, 1–52.

8. Rice, J. (1984). Boundary modification for kernel regression. *Commun. Statist. Theor. Meth.*, **13**, 893–900.

9. Rice, J. and Rosenblatt, M. (1981). Integrated mean squared error of a smoothing spline. *J. Approx. Th.*, **33**, 353-369.

10. Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives, and deconvolution. *Ann. Statist.*, **11**, 141–156.

11. Rosenblatt, M. (1976). Asymptotics and representation of cubic splines. *J. Approx. Th.*, **17**, 332–343.

12. Schumaker, L.L. (1981). *Spline function: basic theory.* John Wiley and Sons.

13. Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.*, **12**, 898–916.

14. Silverman, B.W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. B*, **47**, 1–52.