IRTG 1792 Discussion Paper 2020-010

Kernel Estimation: the Equivalent Spline Smoothing Method

Wolfgang K. Härdle [*]
Michael Nussbaum [*2]

[*] Humboldt-Universität zu Berlin, Germany
[*2] Weierstraß-Institut für Angewandte Analysis und Stochastik, Germany

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Kernel Estimation: the Equivalent Spline Smoothing Method *

Wolfgang Härdle
Humboldt-Universität zu Berlin
Wirtschaftswissenschaftliche Fakltät
Institut für Statistik und Ökonometrie
Spandauer Str. 1
D - 10178 Berlin
Germany

Michael Nussbaum
Weierstraß-Institut für Angewandte
Analysis und Stochastik
Mohrenstr. 39
D - 10117 Berlin
Germany

12. Juli 1994

* This paper was originally started as a "deutsch–deutsche Zusammenarbeit" financed by Sonderforschungsbereich 303, later the research on this paper was carried out within the Sonderforschungsbereich 373 at Humboldt-University Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

ABSTRACT

Among nonparametric smoothers, there is a well-known correspondence between kernel and Fourier series methods, pivoted by the Fourier transform of the kernel. This suggests a similar relationship between kernel and spline estimators. A known special case is the result of Silverman (1984) on the effective kernel for the classical Reinsch-Schoenberg smoothing spline in the nonparametric regression model. We present an extension by showing that a large class of kernel estimators have a spline equivalent, in the sense of identical asymptotic local behaviour of the weighting coefficients. This general class of spline smoothers includes also the minimax linear estimator over Sobolev ellipsoids. The analysis is carried out for piecewise linear splines and equidistant design.

Keywords: Kernel estimator, spline smoothing, filtering coefficients, differential operator, Green's function approximation, asymptotic minimax spline.

1. Introduction

It is part of the basic knowledge about smoothing methods that there is a correspondence between kernel and orthogonal series methods. Loosely speaking, and supposing a circular setting on the unit interval, we can say that a kernel estimator is equivalent to a tapered orthogonal series estimator, where the tapering coefficients are the Fourier coefficients of the kernel scaled with bandwidth parameter h. This is just a way of saying that convolution (which is what a kernel smoother does) is equivalent to multiplication of Fourier transforms. Such a relationship, which is elementary in the classical Fourier series context, can also be established between kernel and spline estimators. It is the purpose of the present paper to make this precise, and thus to contribute to a better understanding of smoothing methods in nonparametric estimation.

Our starting point is the result of Silverman (1984) who proved such a correspondence for the classical Reinsch-Schoenberg smoothing spline. Consider the nonparametric regression problem of estimating a curve m given observations

$$Y_i = m(x_i) + \varepsilon_i, \qquad i = 1, \dots n.$$

Assume that the design points $x_i \in [0, 1]$ are known and nonrandom, and the ε_i are random errors. The standard cubic spline smoother is defined to be the minimizer over functions g of

$$n^{-1} \sum_{i=1}^{n} (Y_i - g(x_i))^2 + \lambda \int (g''(x))^2 \, dx \tag{1.1}$$

where λ is a smoothing parameter. It was shown that this procedure is equivalent to using a certain kernel estimator, where in addition the bandwidth varies locally on $[0, 1]$ in dependence on the design density. It should be stressed that, although the theorem was proved in a statistical context, that result is *of purely analytic nature.* Indeed the smoothing philosophy can be developed in a deterministic framework, and the methods have been studied thoroughly. For other approximation-theoretic results on splines connected specifically with statistics see Utreras (1983) and Cox (1984a).

In our generalization we establish that, essentially, to each kernel estimator based on a kernel K there corresponds a certain spline estimator with 'effective kernel' K. This correspondence is analogous to the one between kernel and orthogonal series smoothers, and is based on the fact that there is a basis in the space of splines which is some way close to the classical Fourier basis. The Fourier transform of the Kernel K determines the shape of the spline smoother, and Silverman's (1984) result appears as a special case.

Let us introduce the following notations. By $(\cdot,\cdot)$ and $\|\cdot\|$ we denote the scalar product and norm in $L_2(0,1)$, respectively. For natural p, let D^p be the derivative of $f \in L_2(0,1)$ in the distributional sense, and let

$$W_2^p(0,1) \;= \{f \in L_2(0,1) \quad ; \quad \mathrm{D}^p f \in L_2(0,1)\}$$

be the Sobolev space of order p on the unit interval. For functions f and g we define the 'design inner product'

$$\langle f,g\rangle_n = n^{-1}\sum_{i=1}^n f(x_i)g(x_i)$$

and the differential bilinear form

$$(f,g)_p = (\mathrm{D}^p f, \mathrm{D}^p g).$$

The spline basis we have in mind is the Demmler-Reinsch basis, i. e. the n-tuple of functions $\psi_{in},\ i=1,\ldots,n$ in $W_2^p(0,1)$ which simultaneously diagonalize the bilinear forms $\langle\cdot,\cdot\rangle_n$ and $(\cdot,\cdot)_p$:

$$\langle\psi_{in},\psi_{jn}\rangle_n = \delta_{ij} \qquad , \qquad (\psi_{in},\psi_{jn})_p = \gamma_{in}\delta_{ij} \quad , \quad i,j=1,\ldots,n$$

and where $\gamma_{1n} \le \ldots \le \gamma_{nn}$ are minimal for all such n-tuples. It is well known that, for $p=2$, the minimizer of (1.1), $\tilde{g}$ say, is of the form

$$\tilde{g} = \sum_{i=1}^n c_i\psi_{in}\tilde{Y}_i \qquad , \qquad \tilde{Y}_i = \langle Y,\psi_{in}\rangle_n, \tag{1.2}$$

see Craven and Wahba (1979). To obtain the explicit form of the coefficients c_i, we have to minimize

$$\sum_{i=1}^{n}\{(1-c_i)^2\tilde{Y}_i^2 + \lambda\gamma_{in}c_i^2\tilde{Y}_i^2\}$$

which yields $c_i = (1+\lambda\gamma_{in})^{-1}$. For the spectral numbers γ_{in} asymptotic relations are known, see e.g. Speckman (1985), Nussbaum (1985). If the design is equidistant then

$$\gamma_{in} \;=\; (\pi i)^{2p}\;(1+o(1)), \quad i,\; n \;\to\; \infty. \tag{1.3}$$

Define $h = \lambda^{1/2p}$; then from (1.3) we infer

$$c_i \;\approx\; (1+(\pi i h)^{2p})^{-1}. \tag{1.4}$$

For $p = 2$ the function $\varphi(x) = (1+(2\pi x)^4)^{-1}$ is known as the 'Butterworth filter'; we have thus

$$c_i \;\approx\; \varphi(ih/2). \tag{1.5}$$

It turns out that Silverman's effective spline kernel function K_S is the inverse Fourier transform of the Butterworth filter:

$$\begin{aligned} K_S(t) &= \int_{-\infty}^{\infty} \exp\left(-2\pi \mathbf{i} t x\right)\; \varphi(x)dx \\ &= \frac{1}{2}\exp\left(-|u|/\sqrt{2}\right)\; \sin\left(|u|/\sqrt{2}+\pi/4\right). \end{aligned} \tag{1.6}$$

From the form of the coefficients (1.5) and the orthogonal expansion (1.2) we understand why K_S should be the 'effective kernel' of the classical spline smoother; we shall amplify on this below. Our recipe, to obtain an equivalent spline smoother for a kernel estimator with kernel K, is now obvious: take φ in (1.5) as the Fourier transform $\hat{K}$ of K. The correspondence will be made rigorous by a theorem on the local behaviour of the newly defined spline smoother. However we have been able to carry out this program on the rigorous level as yet only for piecewise linear splines. Hence Silverman's result, which refers to cubic splines, is not a special case, but rather its analog for the piecewise linear case. Still we believe this result to be instructive and pointing to the validity for splines of arbitrary degree.

A standard assumption in this context is that the nonrandom design points $x_1, \ldots, x_n$ behave regularly as $n \to \infty$, in the sense that the associated empirical distribution function L_n tends to a limit L which has a density ℓ. Then, according to Silverman (1984), the equivalent kernel estimator is one in which the bandwidth varies locally on $[0,1]$, in dependence on the limiting design density ℓ. For our result on the general class of spline smoothers, we confine ourselves to uniform ℓ; more specifically, an equidistant design will be assumed. It is easy to see that the local variability of the bandwidth of the equivalent kernel estimator in the case of nonuniform ℓ is a phenomenon which is independent of the kernel shape, and should hold in our general framework.

Using the terminology of time series analysis, the function φ in (1.5) may be termed a *filter*. It has been established that the Pinsker filter

$$\varphi(x) \;=\; (1 - |2\pi x|^p)_+$$

(cf. Pinsker (1980)) is connected with the minimax-among-linear estimator over Sobolev classes

$$W_2^p(Q) = \{f \in W_2^p(0,1) \quad ; \quad \|\mathrm{D}^p f\|^2 \le Q\}$$

when the loss is the squared norm deviation induced by the design inner product $\langle\cdot,\cdot\rangle_n$, see Speckman (1985). Also it is known that, for independent identically normally distributed ε_i, this spline estimator attains the best possible constant in the L_2-risk asymptotics, in a minimax sense over the Sobolev class; cf. Nussbaum (1985). In this setting the Butterworth filter, i. e. the classical spline smoother is not optimal, and this is one of the motivations for our extended class.

Dealing with the classical spline smoother, Cox (1983), (1984a) developed an effective framework for approximating it by the continuous analog, i. e. by a method-of-regularization operator. Our approach is inspired by these results; however, due to the particular simplicity of the selected special case, we are able to apply more direct methods. It should be noted that the conditions of Cox (1983), (1984a) exclude the piecewise linear case (a priori smoothness 1); thus our result seems to indicate a possible weakening of those regularity conditions.

Messer (1991) and Messer and Goldstein (1993) elaborate the result on the classical spline smoother, obtaining considerable analytic insight, but their analysis is still limited to Silverman's particular case. An important contribution to the general equivalence problem has been made by Thomas-Agnan (1991); we discuss this in the remarks at the end of the paper.

2. The Spline Kernel

To shed some more light on the equivalence which is the subject of this paper, we will follow Cox (1984a) in considering *the associated continuous smoothing problem.* In (1.1), put aside the randomness of the data Y_i for a moment, and assume that $Y_i = m(x_i),\ i = 1, \ldots, n$, where m is a continuous function on $[0, 1]$. Then as $n \to \infty$, the minimization criterion (1.1) will be close to

$$\int_0^1 (m(x) - g(x))^2\ \ell(x)dx + \lambda\ \|\mathrm{D}^p g\|^2 \tag{2.1}$$

(for $p = 2$; in the sequel p will be general). Similarly, the Demmler-Reinsch spline basis will tend to a limiting orthogonal system $\psi_i,\ i = 1, 2, \ldots$ in $L_2(0, 1)$ which may be characterized as follows. We have

$$\int_0^1 \psi_i \psi_j\ dL = \delta_{ij} \quad , \quad (\psi_i, \psi_j)_p = \gamma_i \delta_{ij} \quad , \quad i, j = 1, 2, \ldots$$

where $\gamma_1 \leq \gamma_2 \leq \ldots$, and the basis $\{\psi_i\}$ is extremal in the sense that the spectral values γ_i are minimal. The continuous analog of the smoothing operator

(1.2) then is

$$\tilde{g} = \sum_{i=1}^{\infty} c_i m_i \psi_i \qquad , \qquad m_i = (m, \psi_i) \tag{2.2}$$

where $c_i = (1+\lambda\gamma_i)^{-1}$. For our general class of smoothers, we put $c_i = \varphi(ih/2)$ for some filter φ (remind $\lambda = h^{2p}$). Thus the analysis of spline smoothing operators may be broken up into two parts:

- approximate the discrete problem by the continuous one, as $n \to \infty$, uniformly over a range of h
- study the continuous problem for smoothing parameter $h \to 0$.

Let us further examine the continuous problem, to see why a relationship like (1.6) should be expected between the filter function φ and the effective kernel K. For simplicity let us first assume that the limiting design density is uniform: $\ell \equiv 1$. It is well known that the basis functions ψ_i are eigenfunctions of the differential operator $(-\mathrm{D}^2)^p$ defined on functions in W_2^{2p} which satisfy natural (Neumann) boundary conditions:

$$(-\mathrm{D}^2)^p \psi_j \;=\; \gamma_j \psi_j \tag{2.3a}$$

$$\mathrm{D}^k \psi_j(0) = \mathrm{D}^k \psi_j(1) = 0, \;\; k = p, \ldots, 2p-1. \tag{2.3b}$$

The smoothing procedure (2.2) is an integral operator on $[0,1]$ with kernel

$$H(x,y) \;=\; \sum_{j=1}^{\infty} c_j \psi_j(x) \psi_j(y).$$

In the case $c_j = (1+\lambda\gamma_j)^{-1}$ which corresponds to the method of regularization criterion (2.1) H is the Green's function for the elliptic boundary value problem

$$(-\mathrm{D}^2)^p g + \lambda g \;=\; f \tag{2.4}$$

with boundary conditions (2.3b) on g. In our more general case $c_j = \varphi(jh/2)$ the function $H = H_h$ may be seen as a *generalized Green's function.* Silverman's result, if translated to the continuous smoothing case, says that the classical Green's function behaves locally like a kernel K_S:

$$h\ H_h(y+ht,y)\ \rightarrow\ K_S(t)\ \ \text{as}\ h\rightarrow 0 \tag{2.5}$$

for every $y\in[0,1]$.

This relationship may be very easily derived when we consider the *circular smoothing problem.* Suppose we seek the minimizer g of (2.1) subject to periodic boundary conditions on $\mathrm{D}^k g$. This will lead to the Green's function of the problem (2.4) with boundary conditions

$$\mathrm{D}^k g(0)=\mathrm{D}^k g(1),\ k=0,\ldots,2p-1 \tag{2.6}$$

which can also be expressed in terms of eigenfunctions. In the periodic case these are

$$\psi_0(x)=1,\quad \psi_j(x)=\sqrt{2}\cos{(2\pi jx)},\ j=1,2,\ldots$$
$$\psi_j(x)=\sqrt{2}\sin{(2\pi jx)},\ j=-1,-2,\ldots$$

with corresponding eigenvalues $(2\pi j)^{2p}$. Hence for the Green's function we have, with $\varphi(x)=(1+(2\pi x)^{2p})^{-1}$

$$\begin{aligned} H_h(x,y) &= 1+2\sum_{j=1}^{\infty}\varphi(jh)\{\cos{(2\pi jx)}\cos{(2\pi jy)}+\sin{(2\pi jx)}\sin{(2\pi jy)}\} \\ &= 1+2\sum_{j=1}^{\infty}\varphi(jh)\ \cos{(2\pi j(x-y))} \end{aligned}$$

since φ is symmetric about 0. Consequently we have

$$h\ H_h(y+ht,y)\ =\ h+2h\sum_{j=1}^{\infty}\varphi(jh)\ \cos{(2\pi jth)}$$
$$\approx 2\int_0^{\infty}\varphi(x)\cos{(2\pi xt)}dx = \int_{-\infty}^{\infty}\exp{(-2\pi \mathbf{i}xt)}\varphi(x)dx\ =\ K(t) \tag{2.7}$$

if K is the inverse Fourier transform of φ. This relationship will carry over to general φ provided the last set of displays remains true,which will be the case under appropriate smoothness and integrability conditions on φ. Thus in the periodic case we readily obtain our result on the local behaviour of the generalized Green's function

$$H_h(x,y) \;=\; \sum_{j=1}^{\infty} \varphi(jh/2)\; \psi_j(x)\psi_j(y). \tag{2.8}$$

However, to deal with the original spline smoothing problem we have to consider the nonperiodic case. Here the functions ψ_i are eigenfunctions of $(-\mathrm{D}^2)^p$ *under a different set of boundary conditions*, namely the Neumann set (2.3b). The heuristics then is clear: since we look at the local behaviour of the generalized Green's function in a neighborhood of a fixed point y in the interior of the interval, we can expect that the boundary conditions matter less and less as $h \to 0$, and the behaviour will be as in the periodic case. This interpretation is supported by the well known eigenvalue asymptotics in the Neumann case:

$$\gamma_j \;=\; (\pi j)^{2p}\; (1 + o(1)) \text{ as } j \to \infty$$

(see Agmon (1968), compare also the discrete analog (1.3)). This means that for large j the eigenvalues are close to those of the periodic problem (remind that those were $(2\pi j)^{2p}$, $j = \pm 1, \pm 2, \ldots$, with the same asymptotics under rearrangement). In (2.8), small values of j matter less as $h \to 0$, so if the eigenfunctions ψ_j have a similar tendency to approach those of the periodic problem we can expect the convergence (2.7). This is confirmed for the classical Green's function ($\varphi(x) = (1 + (2\pi x)^{2p})^{-1}$) by Silverman's result; we shall have to deal with the case of general φ fulfilling appropriate conditions.

We remark that Huber (1979) considered the discrete periodic smoothing problem in the case of an equidistant design $\{x_i\}$ on the unit interval, and obtained another approximation to the effective kernel of the procedure. It is shown to be equivalent to Silverman's result by Härdle (1989), chap. 3.4.

3. The continuous smoothing problem

We now proceed to derive the asymptotic relation (2.7) for the generalized Green's function (2.8) for the limiting continuous smoothing problem, in the nonperiodic case. Here the functions ψ_j figuring in (2.8) are the eigenfunctions in the problem (2.3) on the interval $[0,1]$. We are able to obtain the desired result as yet only in the case $p = 1$ and $\ell \equiv 1$ (uniform design density). The eigenfunctions in this case are

$$\psi_1(x) = 1, \quad \psi_j(x) = \sqrt{2}\cos(\pi(j-1)x), \; j = 2, 3, \ldots \qquad (3.1)$$

with corresponding eigenvalues $(\pi(j-1))^2$ (see Triebel (1972), theorem 23.3, p. 301).

Let us now fix appropriate conditions on the filter function φ and the kernel K. We shall use the following notations. By $L_q(a,b)$, $q = 1, 2$ we denote the L_q-space of complex-valued functions on an interval (a,b); when $(a,b) = \mathbb{R}$ we write L_q. Furthermore consider the Sobolev spaces $W_2^1(a,b)$ as defined in section 1; we write W_2^1 if $(a,b) = \mathbb{R}$. Integrals without limits extend over $\mathbb{R}$.

Now let K be a real-valued function on $\mathbb{R}$ with

$$K \in L_1 \;\; , \;\; \int K(x)dx \;=\; 1 \;\; , \;\; K(x) = K(-x). \qquad (3.2)$$

For any $g \in L_1$ let $\hat{g}$ be the Fourier transform of g:

$$\hat{g}(t) \;=\; \int \exp(2\pi \mathbf{i} t x) \; g(x) \; dx.$$

Define the filter function φ as $\varphi = \hat{K}$. Then we can state the following elementary result.

Proposition 3.1. *Let K be a kernel satisfying conditions (3.2). Then $\varphi = \hat{K}$ has properties*

(i) *φ is real and symmetric about* 0

(ii) $\varphi(0) = 1$, *φ is bounded and continuous.*

Furthermore, assume that $K \in L_2$ and understand the Fourier transform as defined on L_2. Then φ is also in L_2, and K is the inverse Fourier transform of φ:

$$K(u) \;=\; \hat{\varphi}(-u) \;=\; \hat{\varphi}(u).$$

At this point let us introduce tail and smoothness conditions on K. Define the set V_2^1 of complex valued functions on $\mathbb{R}$ as

$$V_2^1 \;=\; \{f \in L_2 \;\;,\;\; \int (1+|x|^2)|f(x)|^2 \; dx < \infty\}.$$

It is well known that $f \in W_2^1$ is equivalent to $\hat{f} \in V_2^1$, and $\widehat{\mathrm{D}K}(t) = 2\pi \mathbf{i} t \hat{f}(t)$. Our additional condition on K is

$$K \in W_2^1 \;\;,\;\; K' \in V_2^1. \tag{3.3}$$

Define the operator J by $(Jf)(x) = xf(x)$.

Proposition 3.2. *Let K be a kernel satisfying conditions (3.3). Then*

(iii) $\varphi \in V_2^1$

(iv) $J\varphi \in W_2^1$.

Lemma 3.1. *Let K be a kernel satisfying conditions (3.2), (3.3). Then for $\varphi = \hat{K}$ we have*

$$\sup_{h>0} \; h \; \sum_{j=1}^{\infty} (jh)^2 \; \varphi^2(jh) \;<\; \infty.$$

Proof. Define an interval $A_{jh} = ((j-1)h, jh)$. By standard imbedding theorems

$$(jh)^2 \varphi^2(jh) \;=\; (J\varphi(jh))^2 \;\leq\; C \; \{h^{-1}\|J\varphi\|^2(A_{jh}) + h\|(J\varphi)'\|^2(A_{jh})\}.$$

Now sum over j and use property (iv) of φ .

∎

We are now in a position to define our generalized Green's function: for any $x, y \in [0,1]$ and functions ψ_j from (3.1) we set

$$H_h(x,y) \;=\; 1 + 2\sum_{j=1}^{\infty} \varphi(jh/2)\;\cos(\pi j x)\cos(\pi j y). \tag{3.4}$$

Lemma 3.1 ensures convergence of the series uniformly over x, y. Putting

$$\cos(\pi j x) \;=\; \frac{1}{2}\left(\exp(\pi \mathbf{i} j x) + \exp(-\pi \mathbf{i} j x)\right)$$

and $x = y + th$, we obtain, using the symmetry of φ,

$$\begin{aligned} H_h(y+th, y) \;&=\; 1 + \frac{1}{2}\sum_{j=-\infty, j\neq 0}^{\infty} \varphi(jh/2)\;\big\{\exp(\pi \mathbf{i} j h t)\exp(2\pi \mathbf{i} j y) + \exp(\pi \mathbf{i} j h)\big\} \\ &=\; \frac{1}{2}\sum_{-\infty}^{\infty} \varphi(jh/2)\exp(\pi \mathbf{i} j h t)\;\big\{1 + \exp(2\pi \mathbf{i} j y)\big\}. \end{aligned}$$

Lemma 3.2. *For any t we have as $h \to 0$*

$$\frac{h}{2}\sum_{j=-\infty}^{\infty} \varphi(jh/2)\exp(\pi \mathbf{i} j h t) \;\to\; \int \exp(2\pi \mathbf{i} u t)\;\varphi(u)\;du.$$

Proof. Define

$$\varphi_t(x) \;=\; \varphi(x)\exp(2\pi \mathbf{i} x t).$$

For simplicity we substitute $h/2$ by h in the lemma. Consider intervals A_{jh} as in lemma 3.1. The difference of the two sides in the present lemma is

$$\begin{aligned} &\sum_{j=-\infty}^{\infty} \int_{A_{jh}} \left(\varphi_t(x) - \varphi_t(jh)\right)\,dx \\ \le\; &\sum_{j=-\infty}^{\infty} \int_{A_{jh}} \left|\varphi_t(x) - \varphi_t(jh)\right|\,dx. \end{aligned} \tag{3.5}$$

Consider first intervals A_{jh} which do intersect with $[-2,2]$. The corresponding sum of terms in (3.5) is $o(1)$, since φ_t is continuous. For the other intervals, the expression under the integral sign is bounded by

$$\Big(\int_{A_{jh}} (x\varphi_t'(x))^2\ dx\Big)^{1/2}\ \Big(\int_{A_{jh}} x^{-2} dx\Big)^{1/2}.$$

The Cauchy-Schwartz inequality then gives an upper bound for (3.5)

$$\Big(\int_{|x|>1} x^{-2} dx\Big)^{1/2}\ \|J(\varphi_t')\|\ h\ +\ o(1). \tag{3.6}$$

Now we have

$$(J\varphi)'\ =\ \varphi + J\varphi',$$

hence

$$\|J\varphi'\|\ \leq\ \|\varphi\| + \|(J\varphi)'\|.$$

Furthermore

$$\varphi_t'\ =\ (2\pi\mathbf{i}t\varphi + \varphi')\ \exp(2\pi\mathbf{i}t).$$

Consequently

$$\|J\varphi_t'\|\ \leq\ 2\pi t\|J\varphi\| + \|J\varphi'\| \leq\ 2\pi t\|J\varphi\| + \|\varphi\| + \|(J\varphi)'\|.$$

By proposition (3.2) all these terms are finite, hence (3.6) is $o(1)$.

$\square$

Lemma 3.3. *For any $\delta > 0$, we have as $h \to 0$*

$$\frac{h}{2}\sum_{j=-\infty}^{\infty} \varphi(jh/2)\exp(\pi\mathbf{i}jht)\ \exp(2\pi\mathbf{i}jy)\ =\ o(1)$$

uniformly over $y + ht \in (\delta, 1-\delta)$.

Proof. Let k be a natural, and observe that

$$\begin{aligned} &h\ \textstyle\sum_{|j|>k}\ \varphi(jh/2)\exp(2\pi\mathbf{i}j(y+ht/2)) \\ &\leq \Big(\textstyle\sum_{|j|>k} h^{-1}j^{-2}\Big)^{1/2}\ \Big(\textstyle\sum_{|j|>k} h(hj)^2\varphi^2(jh/2)\Big)^{1/2}. \end{aligned}$$

According to lemma 3.1 the second factor is bounded, uniformly over h and k. The first factor is

$$\left(h^{-1}\ O(k^{-1})\right)^{1/2}.$$

Suppose that $k \sim Mh^{-1}$; then for sufficiently large M the above term can be made less than $\varepsilon/2$. The remaining sum over terms $|hj| \leq M$ in the series is estimated as follows. This sum can be construed as being a series as in the assertion, with φ having support on $[-M, M]$ and being continuous there. Take a finite partition of $[-M, M]$ into intervals of equal length. Since φ can be approximated by corresponding step functions, uniformly on $[-M, M]$ if the partition becomes finer, it suffices to prove the lemma for each such step function. Each such step function is a linear combination of functions which are indicators of symmetric intervals $[-a, a]$, $a < M$. Hence it suffices to prove the lemma for each $\varphi = \chi_{(-a,a)}$, the indicator of some symmetric interval. In this case, for $r = [h^{-1}a]$ we have

$$h \sum\nolimits_{|j| \leq h^{-1}a} \exp(2\pi \mathbf{i} j(y + ht)) \;=\; h\ D_r(y + ht), \tag{3.7}$$

where $D_r(\cdot)$ is the Dirichlet kernel

$$D_r(x) \;=\; \frac{\sin(\pi(2r+1)x)}{\sin(\pi x)}.$$

Now for $x \in (\delta, 1-\delta)$ the numerator is bounded away from 0, hence $D_r(x)$ is uniformly bounded for $r \geq 1$, $x \in (\delta, 1-\delta)$. As $h \to 0$, (3.7) proves the lemma.

□

The final result on the generalized Green's function H_n can now be stated as follows. Observe beforehand that the convergence of lemma 3.2 holds uniformly over $|t| \leq C$, and also uniformly in h over any range $h \leq \overline{h}$ such that $\overline{h} \to 0$. The convergence of lemma 3.3 holds uniformly over $y + ht \in (\delta, 1-\delta)$ and $h \leq \overline{h}$.

Lemma 3.4. *We have for any $y \in (0,1)$, $t \in I\!R$*

$$h\ H_h(y+ht, y) \ \rightarrow \ K(t) \ \ as\ h \rightarrow 0,$$

and the convergence is uniform over $y \in (\delta, 1-\delta)$, $(\delta > 0)$, $|t| \le C$ and $h \in (0, \overline{h})$ where $\overline{h} \rightarrow 0$.

4. The spline basis

Having treated the limiting continuous smoothing problem for degree of differentiability $p = 1$ and uniform limiting design ($\ell \equiv 1$), we now look at the discrete analog, i. e. the problem with data observed at points $x_1, \ldots, x_n$. For this we assume that the regression design is of a particular uniformly spaced kind:

$$x_i \ = \ (i - 1/2)/n \ , \ i = 1, \ldots, n.$$

It is well known that the natural interpolation and smoothing splines for $p = 1$ are piecewise linear. For given $\{x_i\}$ as above and a function f defined on $[0,1]$, let $f^{(n)} = (f(x_1), \ldots, f(x_n))'$ be the trace of f on $\{x_i\}$. Let $S(f^{(n)})$ be the piecewise linear interpolant of f, uniquely defined on $[0,1]$ by the requirement to be constant on the marginal intervals $[0, 1/2n]$, $[1 - 1/2n, 1]$. The following fact is well known; see e.g. Laurent (1972), theorem 4.1.3.

Lemma 4.1. *For $f \in W_2^1(0,1)$, the function $S(f^{(n)})$ is in $W_2^1(0,1)$, and is the solution of*

$$\min\{\ \|\mathrm{D}g\|^2 \ ; \ g^{(n)} = f^{(n)} \ , \ g \in W_2^1(0,1)\ \}.$$

Let $\mathcal{S}_n = S(\mathrm{I\!R}^n)$ be the n-dimensional linear space of such piecewise linear spline functions. It is clear that there is a basis ψ_{jn}, $j = 1, \ldots, n$ in $\mathcal{S}_n$

which simultaneously diagonalizes the bilinear forms $\langle\cdot,\cdot\rangle_n$ and $(\cdot,\cdot)_1$. Lemma 4.1 implies that $\{\psi_{jn}\}$ coincides with the Demmler-Reinsch basis (for $p=1$) introduced in section 1. Obviously the standard smoothing spline for $p=1$, i. e. the minimizer over functions in $W_2^1(0,1)$ of

$$n^{-1}\sum_{i=1}^{n}(Y_i-g(x_i))^2+\lambda\int(\mathrm{D}g(x))^2\ dx \tag{4.1}$$

is in $\mathcal{S}_n$, and hence can be expressed in terms of the basis $\{\psi_{jn}\}$ according to (1.2). Here the filtering coefficients c_j are $c_j=(1+\lambda\gamma_{jn})^{-1}$; the interpolation spline $S(Y)$ is obtained for $c_j=1$ (no smoothing).

It turns out that in our particularly simple setting the functions ψ_{jn} are just the spline interpolants of the ψ_j from the limiting continuous problem, i. e. of the cosine functions given by (3.1).

Lemma 4.2. *The functions ψ_{jn} defined by*

$$\psi_{jn}\ =\ S(\psi_j^{(n)})\ ,\ j=1,\ldots,n$$

ψ_j being given by (3.1), satisfy

$$\langle\psi_{in},\psi_{jn}\rangle_n=\delta_{ij}\qquad,\qquad(\psi_{in},\psi_{jn})_1=\gamma_{jn}\delta_{ij}\quad,\quad i,j=1,\ldots,n$$

where

$$\gamma_{jn}\ =\ 4n^2\sin^2(\pi(j-1)/2n)\ ,\ j=1,\ldots,n\ .$$

Proof. Consider a set of points: $x_k=(k-1/2)/n,\ k=1,\ldots,2n$. Then for any natural $r,\ 1\le r\le 2n-1$, the set of points $\exp(\pi\mathbf{i}rx_k),\ k=1,\ldots,2n$ is evenly spaced on the unit circle in the complex plane. Hence

$$\sum_{k=1}^{2n}\exp(\pi\mathbf{i}rx_k)\ =\ 0. \tag{4.2}$$

Observe that each function $\cos(\pi r x)$, for $1 \le r \le 2n-1$ is symmetric on the interval $(0,2)$ with symmetry center 1. Hence

$$\sum_{k=1}^{n} \cos(\pi r x_k) = \frac{1}{2}\sum_{k=1}^{2n} \cos(\pi r x_k) = 0 \tag{4.3}$$

as a consequence of (4.2). Now we have for $i, j \ge 2$

$$\langle \psi_{in}, \psi_{jn} \rangle_n = n^{-1} \sum_{k=1}^{n} \{\cos(\pi(i-j)x_k) + \cos(\pi(i+j-2)x_k)\}.$$

This expression vanishes if $i \ne j$, according to (4.3), and equals 1 if $i = j$. The case where one of the ψ_{jn} is ψ_{1n}, i. e. identically 1, can be treated analogously. Thus the first orthogonality relation is proved. For the second, suppose first that either i or j is 1. Then, as $\mathrm{D}\psi_{1n} \equiv 0$ and $\gamma_{1n} = 0$, the claim about $(\cdot,\cdot)_1$ is clear. Suppose now that $i, j \ge 2$. Consider a set of points $z_k = k/n$, $k = 0, \ldots, 2n$. Analogously to (4.2) it can be shown that for $1 \le r \le 2n-1$

$$\sum_{k=1}^{2n} \exp(\pi \mathbf{i} r z_k) = 0. \tag{4.4}$$

Observe that each function $\sin(\pi r x)$, for $1 \le r \le 2n-1$ is antisymmetric on the interval $(0,2)$ about 1 and vanishes in 0 and 1. Hence for $1 \le i, j \le n$

$$\begin{aligned}\frac{2}{n}\sum_{i=1}^{n-1} \sin(\pi i z_k)\sin(\pi j z_k) &= n^{-1}\sum_{i=1}^{2n} \sin(\pi i z_k)\sin(\pi j z_k)\\ &= \delta_{ij}\end{aligned} \tag{4.5}$$

as a consequence of (4.4). Now

$$(\psi_{in}, \psi_{jn})_1 = n^{-1}\sum_{k=2}^{n}\big((\psi_{in}(x_k) - \psi_{in}(x_{k-1})\big)\ \big((\psi_{jn}(x_k) - \psi_{jn}(x_{k-1})\big)\ n^2.$$

Furthermore, writing $x_{k-1} = x_k - n^{-1}$, we obtain

$$\begin{aligned}(\psi_{in}(x_k) - \psi_{in}(x_{k-1})) &= \sqrt{2}\ \sin(\pi(i-1)(x_k - 1/2n))\ 2\sin(\pi(i-1)/2n)\\ &= 2\sqrt{2}\ \sin(\pi(i-1)z_{k-1})\ \sin(\pi(i-1)/2n).\end{aligned}$$

This yields in view of (4.5)

$$(\psi_{in}, \psi_{jn})_1 \;=\; \frac{2}{n} \sum_{k=1}^{n-1} \sin(\pi(i-1)z_k)\sin(\pi(j-1)z_k)\; 4n^2 \sin(\pi(i-1)/2n)\sin(\pi(j-1)/2n)$$

$$= \; \delta_{ij}\; 4n^2 \;\; \sin^2(\pi(j-1)/2n)$$

which proves the lemma.

□

Remark. The lemma describes the eigenvalues and eigenvectors of the $n \times n$ band matrix

$$\begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & & & & \\ & & & \ddots & & & \\ & & & & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}$$

as γ_{jn} and $\psi_j^{(n)}$, $j = 1, \ldots, n$. Note that

$$\gamma_{jn} \;=\; (\pi j)^2 \;(1 + o(1))$$

uniformly over $k_1(n) \le j \le k_2(n)$, for any $k_1(n) \to \infty$, $k_2(n) = o(n)$ as $n \to \infty$, which is a special case of (1.3).

Let us now describe the approximation property of the ψ_{jn} for the basis $\{\psi_j\}$.

Lemma 4.3. *We have*

$$\sup_{x\in[0,1]} |\psi_{jn}(x) - \psi_j(x)| \;\le\; n^{-1}\pi j \;,\; j = 1, \ldots, n.$$

Proof. Set $x_0 = 0$. We have for $x \in [x_{k-1}, x_k],\ k = 1, \ldots, n+1$

$$|\psi_{jn}(x) - \psi_j(x)| \ \leq\ \sup_{x \in [x_{k-1}, x_k]} |\psi_j'(x)| n^{-1} \ \leq\ n^{-1}\pi(j-1).$$

∎

This result can be immediately applied to describe the closeness of the generalized Green's function H_h and its discrete (spline) analog. Observe that given an observation vector Y, our spline estimator is the function of $x \in [0,1]$

$$\sum_{j=1}^{n} \varphi((j-1)h/2)\psi_{jn}(x)\langle Y, \psi_{jn}\rangle_n \ =\ n^{-1} \sum_{k,j=1}^{n} \varphi((j-1)h/2)\psi_{jn}(x)\psi_{jn}(x_k) Y_k. \tag{4.6}$$

Define for $x \in [0,1],\ k \in \{1, \ldots, n\}$

$$H_{hn}(x, x_k) \ =\ \sum_{j=1}^{n} \varphi((j-1)h/2)\psi_{jn}(x)\psi_{jn}(x_k). \tag{4.7}$$

Clearly this is the analog of the generalized Green's function (3.4).

Lemma 4.4. *Let $\underline{h}_n,\ \overline{h}_n$ be sequences: $\underline{h}_n \leq \overline{h}_n,\ \overline{h}_n \to 0,\ \underline{h}_n n \to \infty$ as $n \to \infty$. Then we have*

$$h\ |H_h(x, x_k) - H_{hn}(x, x_k)| \ \to\ 0$$

uniformly over $h \in [\underline{h}_n, \overline{h}_n],\ x \in [0,1],\ k = 1, \ldots, n$.

Proof. Since $\psi_j(x),\ \psi_{jn}(x)$ are uniformly bounded, we can use the method used in the proof of lemma (3.3) to show that in both H_h and H_{hn} we need only consider summation terms for $j \leq Mh^{-1}$ for some M. It then remains to show that

$$h\ \sum_{j \leq Mh^{-1}} \ \varphi((j-1)h/2)\ |\psi_j(x)\psi_j(x_k) - \psi_{jn}(x)\psi_{jn}(x_k)|$$

tends to zero uniformly. According to lemma (4.3), for $j \leq Mh^{-1}$

$$\sup_{x\in[0,1]} |\psi_{jn}(x) - \psi_j(x)| \;\leq\; n^{-1}\pi M h^{-1} \;\leq\; C\underline{h}_n^{-1} n^{-1} \;=\; o(1).$$

This proves the lemma.

□

Collecting the results of lemmas 3.4 and 4.4 we obtain the following result.

Theorem . *Suppose that in the regression model the design points x_i are $x_i = (i-1/2)/n,\ i = 1,\dots,n$. Let K be a kernel function satisfying conditions (3.2), (3.3), and let $\varphi = \hat{K}$ be its Fourier transform. Let $\psi_{jn},\ j = 1,\dots,n$ be the Demmler-Reinsch basis in the space $\mathcal{S}_n$ of piecewise linear splines with knots at x_i. Consider the spline estimator given by (4.6) for smoothing parameter h, and let H_{hn} be the corresponding weight function given by (4.7). Let $\underline{h}_n$, $\overline{h}_n$ be sequences: $\underline{h}_n \leq \overline{h}_n$, $\overline{h}_n \to 0$, $n\underline{h}_n \to \infty$. Then*

$$h\ H_{hn}(x_k + th, x_k) \;\to\; K(t)\ ,\quad n \to \infty$$

uniformly over $x_k \in (\delta, 1-\delta)$ $(\delta > 0)$, $h \in (\underline{h}_n, \overline{h}_n)$ and $|t| \leq C$.

5. Remarks

Having carried out our analysis for smoothness $p = 1$ (piecewise linear splines), it remains to include the classical spline smoother for $p = 1$ into this framework. Consider the minimizer of (4.1); as in (1.2)-(1.4) it can be seen that it corresponds to a filter function

$$\varphi(x) \;=\; (1 + (2\pi x)^2)^{-1}.$$

This filter function clearly satisfies conditions (i)-(iv) of section 3; hence its Fourier transform $K = \hat{\varphi}$ satisfies the condition of the theorem. We have

$$K(u) \;=\; \hat{\varphi}(u) \;=\; \frac{1}{2}\exp(-|u|)$$

so the double exponential density is the analog of Silverman's kernel K_S for $p = 1$. We conjecture that our main result can be generalized to arbitrary degree of smoothnes p and to a general limiting design density ℓ, provided the design tends to its limit sufficiently quickly. This is of course suggested by the results on the classical smoothing spline. We believe that more analytic results on the spectrum of differential operators and their approximation e. g. by Galerkin methods should be drawn upon for this. A useful reference is Chatelin (1983).

Let us stress again that so far our results did not involve stochastics, though they were obtained with a view to statistical smoothing. An interesting statistical result related to the subject of this paper was obtained by Cox (1984b). It was shown that the spline smoother applied to pure noise (i. e. to data ε_i) yields a random function on $[0, 1]$ which, when appropriately scaled, is close to a Gaussian process. This central limit theorem holds for general (nonnormal) noise distribution, and was used to show that the method of generalized cross-validation for choosing the smoothing parameter is asymptotically optimal. In turn, this study was motivated by a result of Speckman (1985) on the minimax linear spline, who established optimality of the bandwidth selector *under normality of the noise.* The normality assumption was removed by Cox (1984b), but the classical smoothing spline was substituted for the minimax linear one. Thus it appears a natural idea to generalize the limit theorem for spline estimators to our class. As the corresponding class of filters includes the Pinsker one, one should be able to infer optimality of the adaptive bandwidth choice for the minimax linear spline in the nongaussian case. This would complement a recently established lower asymptotic risk bound (see Golubev and Nussbaum, 1990), which showed that the minimax linear spline is a candidate for attainment also under nonnormal noise. That appears to be one way to confirm that this bound, which involves optimal rate *and constant*, is attainable adaptively by a spline estimator, without knowledge of the derivative bound Q and of the noise variance σ^2.

Thomas-Agnan (1991) defines a general class of spline-type smoothers, called α-splines, starting from the following observation. It is well known that in (1.1) the integral may be extended over $(0,1)$ or over the whole real line; in both cases the same spline minimizer results. If the whole line is used then (1.1) may be written in terms of the Fourier transform $\hat{g}$ of g

$$n^{-1}\sum_{i=1}^{n}(Y_i - g(x_i))^2 + \lambda \int \mid (2\pi t)^p \hat{g}(t) \mid^2 \, dt$$

Let α be a complex-valued function function defined on $\mathbb{R}$ fulfilling some regularity conditions; consider the minimizer g of

$$n^{-1}\sum_{i=1}^{n}(Y_i - g(x_i))^2 + \lambda \int \mid \alpha(t)(2\pi t)^p \hat{g}(t) \mid^2 \, dt$$

The solution is called an α-spline. For $\alpha \equiv 1$ and $p = 2$ one obtains the classical smoothing spline. The α-splines represent a large class of linear smoothers; in particular, they should be equivalent to kernel estimators. To see this heuristically, consider the corresponding continuous smoothing problem on the whole real line:

$$\int (m(x) - g(x))^2 dx + \lambda \int \mid \alpha(t)(2\pi t)^p \hat{g}(t) \mid^2 \, dt$$

Substituting the first integral by $\int (\hat{m}(t) - \hat{g}(t))^2 dt$ and arguing similarly to (1.2) we obtain a minimizer

$$\hat{g}(t) = (1 + \lambda\alpha(t)(2\pi t)^p)^{-1}\hat{m}(t) \qquad (5.1).$$

The Fourier transform expression for a general kernel smoother on the whole real line would be, using a filter function φ and bandwidth parameter h as before,

$$\hat{g}(t) = \varphi(ht/2)\hat{m}(t). \tag{5.2}$$

A choice $\alpha(t) = (2\pi ht)^{-p}\varphi^{-1}(ht/2) - 1, \lambda = h^p$ yields equality of (5.1) and (5.2). Though in the original concept α was assumed fixed, we see that a bandwidth-dependent choice of α makes the method sufficiently flexible to yield a spline-type optimization problem corresponding to the general kernel estimator. It is not essential in this connection that the α-splines are not necessarily polynomial splines. Thomas-Agnan (1991) discusses solution of the optimization problem via reproducing kernel Hilbert space methods. A rigorous proof of equivalence in the sense considered in this paper might be easier than for our estimator since Fourier transform methods are more directly at hand.

æ

References

Agmon, S. (1968). Asymptotic formulas with remainder estimates for eigenvalues of elliptic operators. *Arch. Rational Mech. Anal. , 28, 165–183.*

Chatelin, F. (1983). *Spectral approximation to linear operators. Academic Press, New York.*

Cox, D. D. (1983). Asymptotics for $M-$type smoothing splines. *Annals of Statistics, 11, 530–551.*

Cox, D. D. (1984a). Multivariate smoothing spline functions. *SIAM J. Numer. Anal. 21, 789-813*

Cox, D. D. (1984b). Gaussian approximation of smoothing splines. *manuscript*

Cox, D. D. (1988). Approximation of method of regularization estimators. *Annals of Statistics, 16, 694-712*

Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Annals of Statistics, 18, 758–778.*

Härdle, W. (1990). *Applied Nonparametric Regression. Cambridge University Press, Econometric Society Monograph Series 19.*

Huber, P.J. (1979). Robust smoothing. *in: Robustness in Statistics (E.Launer and G.Wilkinson, eds.) Academic Press, New York.*

Laurent, P. (1972). *Approximation et Optimisation. Hermann, Paris.*

Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Annals of Statistics, 13, 984–997.*

Messer, K. (1991). A comparison of a spline estimate to its "equivalent" kernel estimate. *Annals of Statistics, 19, 817–829.*

Messer, K. and Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation. *Annals of Statistics, 21, 179–195.*

Pinsker, M. S (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission, 16, No. 2, 52-68.*

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

Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Annals of Statistics, 13, 970–983.*

Thomas-Agnan, C. (1991). Spline functions and stochastic filtering. *Annals of Statistics, 19, 1512–1527.*

Triebel, H. (1972). *Höhere Analysis. Deutscher Verlag der Wissenschaften, Berlin.*

Utreras, F. (1983). Natural spline functions, their associated eigenvalue problem. *Numer. Math. 42, 107–118.*

IRTG 1792 Discussion Paper Series 2020

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Estimation and Determinants of Chinese Banks' Total Factor Efficiency: A New Vision Based on Unbalanced Development of Chinese Banks and Their Overall Risk" by Shiyi Chen, Wolfgang K. Härdle, Li Wang, January 2020.

002 "Service Data Analytics and Business Intelligence" by Desheng Dang Wu, Wolfgang Karl Härdle, January 2020.

003 "Structured climate financing: valuation of CDOs on inhomogeneous asset pools" by Natalie Packham, February 2020.

004 "Factorisable Multitask Quantile Regression" by Shih-Kang Chao, Wolfgang K. Härdle, Ming Yuan, February 2020.

005 "Targeting Cutsomers Under Response-Dependent Costs" by Johannes Haupt, Stefan Lessmann, March 2020.

006 "Forex exchange rate forecasting using deep recurrent neural networks" by Alexander Jakob Dautel, Wolfgang Karl Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2020.

007 "Deep Learning application for fraud detection in financial statements" by Patricia Craja, Alisa Kim, Stefan Lessmann, May 2020.

008 "Simultaneous Inference of the Partially Linear Model with a Multivariate Unknown Function" by Kun Ho Kim, Shih-Kang Chao, Wolfgang K. Härdle, May 2020.

009 "CRIX an Index for cryptocurrencies" by Simon Trimborn, Wolfgang Karl Härdle, May 2020.

010 "Kernel Estimation: the Equivalent Spline Smoothing Method" by Wolfgang K. Härdle, Michael Nussbaum, May 2020.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.