DERIVATION OF EQUIVALENT KERNEL FOR GENERAL SPLINE SMOOTHING: A SYSTEMATIC APPROACH

Felix Abramovich * †

Department of Statistics & Operations Research, Raymond & Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv 69978, Israel

Vadim Grinshtein

Department of Mathematics, Raymond & Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv 69978, Israel

Running Title : EQUIVALENT KERNEL

Abstract

We consider first the spline smoothing nonparametric estimation with *variable* smoothing parameter and arbitrary design density function and show that the corresponding equivalent kernel can be approximated by the Green's function of a certain linear differential operator. Furthermore, we propose to use the standard (in applied mathematics and engineering) method for asymptotic solution of linear differential equations, known as the WKB-method, for *systematic* derivation of asymptotically equivalent kernel in this general case. The corresponding results for polynomial splines are a special case of the general solution. Then, we show how these ideas can be directly extended to the very general L-spline smoothing.

3

*The author responsible for correspondence

†Part of the work was done while the author was in the School of Mathematics, University of Bristol, UK

[3]*Key words and phrases.* Green's function, L-smoothing spline, nonparametric regression, variable smoothing parameter, WKB-method.

1. Introduction.

Consider the nonparametric regression problem of estimating the unknown response function $g(t)$ from noisy observations,

$$y_i = g(t_i) + \varepsilon_i, \qquad i = 1, ..., n$$

where $0 < t_1 < ... < t_n < 1$ and $\varepsilon_1, ..., \varepsilon_n$ are independent normal random variables with zero mean and variance σ^2 .

A widely known cubic smoothing spline estimator $\hat{g}(\cdot)$ is defined as a solution of the following minimization problem:

$$\hat{g}\ (t) = \arg \min_{f \in W_2^2} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y_i - f(t_i))^2 + k \int_0^1 (f''(t))^2 dt \right\}, \tag{1.1}$$

where $W_2^2[0,1]$ is the standard Sobolev space with the norm $\|f\|^2 = \int (f^2 + f'^2 + f''^2)$. The smoothing parameter k controls the tradeoff between the goodness-of-fit to the data, measured by the residual sum of squares, and smoothness of the estimate, expressed by the integral term. The solution of (1.1) is unique and is a natural cubic spline with knots t_i's. The explicit formula for $\hat{g}$ is given, for example, in Wahba (1978). The main results and extensive bibliographies on smoothing splines are summarized in Eubank (1988) and Wahba (1990).

However, the *global* value of smoothing parameter k does not adapt to the local behavior of $g(\cdot)$ in regions of high curvature where the cubic spline tends to oversmooth. This feature becomes especially problematic in interval estimation discussed in Section 3. Nychka (1988) showed that bias, usually the modest part of the mean squared error (MSE) of cubic spline estimators, increases significantly in these regions. The use of *variable* smoothing parameter allows one to be more flexible in controlling the tradeoff between bias and variance of the estimate. Reducing the penalty for lack of smoothness in regions of high curvature implies bias decreasing; where curvature is low, the estimate emphasizes smoothness and reduces the variance that dominates MSE. The idea of using splines with variable smoothing parameter and its estimation from the data are discussed in Abramovich and Steinberg (1996). Oehlert (1992) considered smoothing splines with variable smoothing parameter (relaxed boundary splines in his terminology) for decreasing the bias for cubic splines near the boundaries, which dominates the integrated mean squared error (IMSE) of the estimate. For this purpose he studied the specific case $k(t) \propto t^\alpha(1-t)^\alpha$ for some positive α.

To extend the definition (1.1) to the case of a variable smoothing parameter we define $\hat{g}(t)$ as follows:

$$\hat{g}\ (t) = \arg\min_{f \in W_2^2} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y_i - f(t_i))^2 + \int_0^1 k(t)(f''(t))^2 dt \right\}, \tag{1.2}$$

where the variable smoothing parameter $k(t) \in W_2^2[0,1]$ is strictly positive.

From the quadratic nature of (1.2), $\hat{g}(\cdot)$ is linear in the observations y_i's and, hence, may be expressed as

$$\hat{g}(t) = \frac{1}{n} \sum_{i=1}^{n} W_n(t, t_i) y_i$$

for a certain weight function $W_n(t, t_i)$. Thus, $\hat{g}(\cdot)$ is essentially a weighted moving average and, therefore, may be viewed as a kernel estimator with equivalent kernel $W_n(\cdot, \cdot)$.

The connection between spline smoothing and kernel estimation, originally based on different ideas, has deep foundation. It is well-known that for uniform design and for constant smoothing parameter k the equivalent kernel $W_n(\cdot, \cdot)$ asymptotically (for small k) may be well approximated by the Green's function for the differential operator $kD^4 + I$ with natural boundary conditions (see Speckman 1981, Cox 1983, Silverman 1984, Messer 1991, Nychka 1995). In the paper we extend these results to an arbitrary design density $p(t)$ and variable smoothing parameter $k(t)$ and show that in the general case the corresponding equivalent kernel is asymptotically connected to the Green's function of the differential operator $S_k = (1/p)D^2kD^2 + I$.

In order to obtain the Green's function of the operator S_k one has to find its fundamental system. It can be done by direct solving the corresponding homogeneous differential equation in the simplest case, where $k(t)$ and $p(t)$ are assumed to be constant (Speckman 1981, Cox 1983). However, the solution cannot be derived explicitly for general $k(t)$ and $p(t)$ and allows only *asymptotic* approximation. Silverman (1984) considered this problem for an arbitrary $p(t)$ (but constant k) but did not give any *constructive* method for finding the equivalent kernel. He "guessed" the asymptotic approximation without hinting at how it had been obtained.

In this paper we propose to use the WKB-method for asymptotic derivation of fundamental system of the operator S_k (and, hence, its Green's function) in the general case. The WKB-method is a standard method in applied mathematics and engineering for the asymptotic solution of linear differential equations (Coddington and Levinson 1955, Chapter 6, Bender and Orszag 1978, Chapter 10). Concerning the problem at hand, it allows the *systematic* derivation of asymptotically equivalent kernels for splines with arbitrary variable smoothing parameter and design density functions, yielding Silverman's result as a special case. Moreover, as we shall show, the WKB-method can be also used for finding asymptotically equivalent kernels for the very general L-splines, where the integral in (1.2) is of the form $\int (Lf)^2$ for some linear differential operator L.

The efficient computational algorithms for spline smoothing procedure are given in Wecker and Ansley (1983), Hutchinson and de Hoog (1986). The asymptotic properties of spline estimators are established, for example, in Speckman (1985), Nussbaum (1985). The equivalent kernel formulation developed in this paper is more of a conceptual interest, finding the relationship between two main nonparametric regression approaches and providing an intuition into what spline smoothing does to the data.

In the main Section 2 we establish the connection between spline smoothing with variable smoothing parameter and kernel estimation. Applying the WKB-method we derive the asymptotically equivalent kernel via the Green's function of the operator S_k and estimate the goodness of this asymptotic approximation for the original equivalent kernel $W_n(\cdot, \cdot)$. In Section 3 we apply the results obtained in Section 2 to statistical inference for spline estimates and, using the equivalence between Bayesian modeling and spline smoothing, we obtain asymptotic error bounds for the unknown response function. The last Section 4 gives a sketch for extension of the previous results to the very general L-spline smoothing. Remarkably, the corresponding asymptotically equivalent kernel depends only on the highest-order coefficient of the operator L and in this sense, for large samples, general smoothing splines behave similar to the "usual" smoothing spline with a corresponding variable smoothing parameter.

2. Derivation of asymptotically equivalent kernel.

We start this Section with intuitive explanation of the connection between the equivalent kernel for a smoothing spline with variable smoothing parameter and the Green's function of the operator S_k defined in the Introduction. Then we find the asymptotic approximation of the Green's function and finally Theorem 1, formulated at the end of the Section, justifies our preliminary considerations and rigorously summarizes all previous results.

2.1. Definition of the L_k-smoothing spline.

Consider the model

$$y(t_i) = g(t_i) + \varepsilon_i \quad i = 1, ..., n \tag{2.1}$$

where $0 < t_1 < ... < t_n < 1$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)' \sim \mathcal{N}(\mathbf{0}, \sigma^2 \ I_n)$.

We derive a smoothing spline with variable smoothing parameter $\hat{g}(t)$ by (1.2). Applying standard methods for solving (1.2) (see, for example, Kimeldorf and Wahba 1971, Wahba 1990) it can be shown that $\hat{g}(t)$ is an L-spline for the differential operator $L_k = k^{1/2}(t)D^2$ (L_k-spline), where D^2 is the second order differentiation operator, and is of the form

$$\hat{g}\ (t) = \alpha_0 + \alpha_1 t + \sum_{j=1}^{n} d_j Q(t_j, t)$$

where

$$Q(s,t) = \int_0^{min(s,t)} k(u)^{-1}(t-u)(s-u)du$$

Exact expressions for the coefficients α's and d's are given in Abramovich and Steinberg (1996) with minor changes due to their notation $k^2(t)$ for the smoothing parameter rather than $k(t)$ as we use here.

The estimate $\hat{g}(\cdot)$ satisfies the natural boundary conditions:

$$(k(t)\hat{g}\ ''(t))^{(p)}(0) = (k(t)\hat{g}\ ''(t))^{(p)}(1) = 0, \quad p = 0,\ 1 \tag{2.2}$$

Since $k(t)$ is strictly positive these conditions are equivalent to

$$\hat{g}\ ''(0) = \hat{g}\ ''(1) = \hat{g}\ '''(0) = \hat{g}\ '''(1) = 0. \tag{2.2a}$$

2.2. Equivalent kernel approach.

As we have already mentioned in the Introduction, from the quadratic nature of (1.2), $\hat{g}(t)$ is linear in the observations for fixed $k(t)$ and, hence,

$$\hat{g}\ (t) = \frac{1}{n}\sum_{i=1}^{n} W_n(t, t_i) y_i \tag{2.3}$$

where $W_n(\cdot,\cdot)$ is the equivalent kernel. Although $W_n(\cdot, s)$ is defined in (2.3) only for the data-points $s = t_i$ for further considerations we would like to treat it as a usual function of two variables. For this purpose we extend its definition for an arbitrary fixed $s \in [0,1]$, setting

$$W_n(t,s) = \arg\min_{f \in W_2^2} \left\{ \frac{1}{n}\sum_{i=1}^{n} f(t_i)^2 + \int_0^1 k(t)(f''(t))^2 dt - 2f(s) \right\} \tag{2.4}$$

A quadratic functional with positive second functional derivative always has a unique minimum over a closed convex set of functions, so the definition (2.4) is legitimate. For $s = t_i$, (2.4) is, up to the constant, the functional in (1.2) where $y_j = n\delta_{ij}, j = 1, ...n$ and one may easily verify that the extended definition (2.4) coincides with (2.3) for the data-points.

Now we derive an explicit asymptotic formula for $W_n(\cdot,\cdot)$. In this subsection we show the asymptotic connection between it and the Green's function for the differential operator $S_k = (1/p)D^2 kD^2 + I$.

Let F_n be the empirical distribution function of the design points, that is

$$F_n(t) = n^{-1} \times (\text{number of design points} \leq t)$$

and assume that the design points are distributed on $[0,1]$ with c.d.f. F, i.e. $\alpha(n) = \sup_{t\in[0,1]} |F_n(t) - F(t)| \to 0$ as n tends to infinity. Suppose that F is a differentiable function and let $p = F'$ be a design density. Suppose in addition that $p(t) > 0$ on $[0,1]$. Approximating the sum in (1.2) by the integral implies that asymptotically $\hat{g}(t)$ minimizes

$$\int_0^1 (y(t) - f(t))^2 p(t) dt + \int_0^1 k(t) f''(t)^2 dt$$

The Euler equation for this variational problem is

$$\frac{1}{p(t)}\{k(t)\hat{g}''(t)\}'' + \hat{g}(t) = y(t)$$

with boundary conditions (2.2).

One may see that this system is positive, so there exists a unique solution and

$$\hat{g}\ (t) = \int_0^1 G(t,s)y(s)ds \tag{2.5}$$

where $G(\cdot,\cdot)$ is the Green's function of the operator $S_k = (1/p)D^2kD^2 + I$ acting on the subspace of functions satisfying the boundary conditions (2.2).

Comparison of (2.5) with (2.3) indicates that for large samples the equivalent kernel $W_n(t,s)$ for spline smoothing with variable smoothing parameter $k(t)$ and design density $p(t)$ can be approximated by the kernel $W(t,s) = G(t,s)/p(s)$ (of course, this claim still must be rigorously formulated and proved!).

2.3. Asymptotical derivation of the Green's function.

The Green's function $G(\cdot,\cdot)$ in (2.5) cannot be obtained explicitly for general $k(t)$ and $p(t)$. However, we can approximate it by replacing the natural boundary conditions (2.2) by a requirement of vanishing $G(t,s)$ as t tends to $\pm\infty$. The solution of the latter problem, say $H(\cdot,\cdot)$, is essentially the Green's function for the operator $(1/p)D^2kD^2 + I$ acting on the subspace of functions satisfying homogeneous conditions at $\pm\infty$ and can be obtained asymptotically by the standard (in applied mathematics) method known as the WKB-method. The WKB-method, based on an asymptotic expansion with respect to a large parameter, was originally proposed by Jeffreys (1924) and in the works of Wentzel, Brillouin and Kramers (hence, it is called WKB or sometimes WKBJ-method) for an approximated solution of the Schrödinger equation and then extended to other linear differential equations (see, e.g. Coddington and Levinson 1955, Chapter 6, Bender and Orszag 1978, Chapter 10). Theorem 1 will establish the goodness of the WKB-approximation for $W_n(\cdot,\cdot)$ under very general assumptions given below.

We have finished now all the preliminary considerations and start the rigorous analysis.

First we formulate two assumptions on $k(\cdot)$ and $p(\cdot)$:

(i) Suppose that the smoothing parameter $k(t)$ tends to zero as n increases. We assume that $k(t)$ is of the form $k(t) = k_0(t)/h^4$, where $k_0(t)$ is a fixed known function independent of n, and h is parameter that depends on n in such a way that $h \to \infty$ as $n \to \infty$.

(ii) $k_0(\cdot)$ and the design density function $p(\cdot)$ lie in $C^3[0,1]$ and are bounded away from zero.

To construct the Green's function $H(\cdot,\cdot)$ we start from finding the fundamental solutions. Consider the homogeneous equation

$$\frac{1}{p(t)}\{k(t)f''(t)\}'' + f(t) = 0 \tag{2.6}$$

Recall that $k(t) = k_0(t)/h^4$ and, hence, (2.6) is equivalent to

$$\frac{1}{p(t)}\{k_0(t)f''(t)\}'' + h^4 f(t) = 0$$

We apply the WKB-method for solving the latter equation. Rewrite it in the form

$$f^{(\text{iv})}(t) + 2\frac{k_0'(t)}{k_0(t)}f'''(t) + \frac{k_0''(t)}{k_0(t)}f''(t) + h^4\frac{p(t)}{k_0(t)}f(t) = 0 \tag{2.7}$$

Following the idea of the WKB-method the asymptotic (for large h) solution of (2.7) is sought of the form

$$f(t) \sim \exp\left\{h\int_0^t \Psi(z)dz\right\}\sum_{k=0}^{\infty}\frac{C_k(t)}{h^k}$$

We prove in the Appendix that (2.7) satisfies the conditions of Theorem 6.3.1 of Coddington and Levinson (1955) and, hence, the solution of this form exists and yields the following fundamental system

$$f_j(t) = C_0(t)\exp\left\{h\eta_j\int_0^t\left(\frac{p}{k_0}\right)^{1/4}\right\}\{1 + O(1/h)\} \quad j = 1,...,4$$

as $h \to \infty$, where the η's are the four fourth roots of (-1). Substituting the latter expression into (2.7) and expanding the left-hand side by powers of h after straightforward calculus one obtains

$$f_j(t) = \frac{1}{k_0(t)^{1/8}p(t)^{3/8}} \exp\left\{h\eta_j \int_0^t (p/k_0)^{1/4}\right\}\{1 + O(1/h)\}$$

Construct the Green's function $H(t,s)$ as a linear combination of the fundamental solutions by standard methods:

$$H(t,s) = \begin{cases} \sum_{j=1}^{4} \theta_{1j}(s) f_j(t) & 0 \le t \le s \le 1 \\ \sum_{j=1}^{4} \theta_{2j}(s) f_j(t) & 0 \le s \le t \le 1 \end{cases}$$

The requirement that $H(t,s)$ vanishes as t tends to $\pm\infty$ implies that four coefficients of $f_j(\cdot)$ equal zero. The remaining four coefficients are defined from the continuity conditions of $H(t,s)$ and its derivatives at the point $t = s$ for every fixed s :

$$\begin{array}{ll} \sum_{j=1}^{2} \{\theta_{2j}(s) - \theta_{1j}(s)\} f_j^{(q)}(s) = 0 & q = 0, ..., 2 \\ \sum_{j=1}^{2} \{\theta_{2j}(s) - \theta_{1j}(s)\} \{k^2(t) f_j''(t)\}'|_{t=s} = p(s) & \end{array}$$

Solving the above system one finally obtains that $H(t,s) = H_0(t,s)(1 + O(1/h))$, where

$$H_0(t,s) = \frac{h}{2} \frac{p(s)^{5/8}}{\{k_0(t)k_0(s)\}^{1/8} p(t)^{3/8}} e^{-h\Phi_0(t,s)} \sin\{h\Phi_0(t,s) + \pi/4\} \tag{2.8}$$

and $\Phi_0(t,s) = \frac{1}{\sqrt{2}} \int_{\min(t,s)}^{\max(t,s)} (p/k_0)^{1/4}$.

2.4. The main theorem: proof and discussion.

The main part of $H(t,s)$ for large h, $H_0(t,s)$, is used now for approximating the original equivalent kernel $W_n(t,s)$. Define the function $W(t,s) = H_0(t,s)/p(s)$. The following Theorem derives an error in approximation $W_n(\cdot,\cdot)$ by $W(\cdot,\cdot)$:

Theorem 1 *Suppose that Assumptions* (i) and (ii) *are true. Then for all sufficiently large n*

$$|W(t,s)/h - W_n(t,s)/h| = O(h\alpha(n) + 1/h)$$

uniformly over all $t \in [0,1]$ and $s \in [\tau_1, \tau_2]$ for every τ_1 and τ_2, $0 < \tau_1 < \tau_2 < 1$.

Before starting the proof we recall that h is a function of n, $h \to \infty$ as $n \to \infty$ and $\alpha(n) = \sup_{t\in[0,1]} |F_n(t) - F(t)|$.

Proof. The idea of the proof is somewhat similar to that of Silverman's (1984) Theorem A. During the proof we use several lemmas proved in the Appendix. We denote for convenience $W(\cdot, s)$ and $W_n(\cdot, s)$ by $W_s(\cdot)$ and $W_{ns}(\cdot)$ respectively, and all corresponding derivatives will be taken with respect to t.

Let τ_1 and τ_2 be $0 < \tau_1 < \tau_2 < 1$. Fix $s \in [\tau_1, \tau_2]$ and n (and, therefore, h). Define the norm $\|u\|_n$ in W_2^2 by $\|u\|_n^2 = \int_0^1 ku''^2 + \int_0^1 u^2 dF_n$. Define a functional A_s in W_2^2 by

$$A_s(u) = \frac{1}{2}\|u\|_n^2 - u(s)$$

According to the definition (2.4), W_{ns} is a minimizer of A_s over W_2^2. From now on we drop the index s and denote the functional A_s simply by A.

It is easy to show that the functional derivative at v, $A'(v)(u) =< v, u >_n - u(s)$ for all v and u in W_2^2, where the inner product $< u, v >_n = \int ku''v'' + \int uv\, dF_n$ (e.g. Tapia and Thompson 1978). Since A is a quadratic functional its second functional derivative A'' is constant and

$$\begin{aligned} A''(w, u) &= A''(v)(w, u) = A'(v + w)(u) - A'(v)(u) =< v + w, u >_n - u(s) - < v, u >_n + u(s) \\ &= < w, u >_n \end{aligned}$$

W_{ns} is a minimizer of A and, therefore, $A'(W_{ns})$ is a zero functional. Hence, for all u in W_2^2

$$A'(W_s)(u) = \{A'(W_s) - A'(W_{ns})\}(u) = A''(W_s - W_{ns}, u) =< W_s - W_{ns}, u >_n$$

In particular,

$$A'(W_s)(W_s - W_{ns}) = \|W_s - W_{ns}\|_n^2 \tag{2.9}$$

As we have mentioned above,

$$A'(W_s)(u) =< W_s, u >_n - u(s) = \int kW_s''u'' + \int W_s u\, dF_n - u(s) \tag{2.10}$$

Lemma 1 *For sufficiently large h*

$$\int kW_s''u'' = [(kW_s'')u' - (kW_s'')'u]_0^1 - \int pW_s u + u(s)\{1 + O(1/h)\}$$

Substituting the result of Lemma 1 into (2.10) gives

$$A'(W_s)(u) = [(kW_s'')u' - (kW_s'')'u]_0^1 + \int W_s u \, d(F_n - F) + u(s)O(1/h) \tag{2.11}$$

The following two lemmas derive upper bounds for the absolute values of the first two terms in (2.11) for sufficiently large n, where the constants C_1, C_2 and δ used in the lemmas do not depend on s, n and h :

Lemma 2 *There exist $\delta > 0$ and C_1 such that*

$$\left| [(kW_s'')u' - (kW_s'')'u]_0^1 \right| \le C_1 \exp\{-\delta h\}(\sup|u| + \sup|u'|/h)$$

Lemma 3 *There exists C_2 such that*

$$\left| \int W_s u \, d(F_n - F) \right| \le C_2 h\alpha(n)(\sup|u| + \sup|u'|/h)$$

Applying Lemmas 2 and 3 to (2.11) yields

$$|A'(W_s)(u)| \le \gamma_n(\sup|u| + \sup|u'|/h) \tag{2.12}$$

where $\gamma_n = C_0(h\alpha(n) + \exp\{-\delta h\} + 1/h)$. For every δ we can choose h large enough so that $\exp(-\delta h) \ll 1/h$ and, therefore $\gamma_n \le C(h\alpha(n) + 1/h)$.

Setting $u = W_s - W_{ns}$, (2.9) and (2.12) imply

$$\frac{1}{h^4}\int_0^1 k_0 \Delta_n''^2 dt + \int_0^1 \Delta_n^2 dF_n(t) \le \gamma_n(\sup|\Delta_n| + \sup|\Delta_n'|/h), \tag{2.13}$$

where $\Delta_n(\cdot) = W_s(\cdot) - W_{ns}(\cdot)$.

Set $x = t \cdot h$ and define the functions $\tilde{\Delta}_n(x) = \Delta_n(x/h)$ and $\tilde{F}_n(x) = F_n(x/h)$. Then, from (2.13) we have

$$\frac{1}{h}\left(\int_0^h k_0 \tilde{\Delta}_n''^2 dx + \int_0^h \tilde{\Delta}_n^2 d\tilde{F}_n\right) \leq \gamma_n (\sup_{[0,h]} |\tilde{\Delta}_n(x)| + \sup_{[0,h]} |\tilde{\Delta}_n'(x)|), \tag{2.14}$$

where the derivatives of $\tilde{\Delta}_n$ and $\tilde{F}_n$ are with respect to x. By Cauchy-Schwarz inequality

$$\int_0^h \tilde{\Delta}_n^2 d\tilde{F}_n = \int_0^h \tilde{\Delta}_n^2 d\tilde{F}_n \;\cdot\; \int_0^h d\tilde{F}_n \geq \left(\int_0^h \tilde{\Delta}_n d\tilde{F}_n\right)^2$$

and thereby,

$$\int_0^h k_0 \tilde{\Delta}_n''^2 + \int_0^h \tilde{\Delta}_n^2 d\tilde{F}_n \geq \min(\inf k_0, 1)\left(\int_0^h \tilde{\Delta}_n''^2 + \left(\int_0^h \tilde{\Delta}_n d\tilde{F}_n\right)^2\right) \geq c_1 \|\tilde{\Delta}_n\|^2_{W_2^2[0,h]}$$

The last inequality directly follows from the classical theorem of equivalence of the norms in Sobolev spaces (e.g. Smirnov 1964, Theorem 114.3). From the Sobolev embedding theorem (e.g. the equation (157) of Smirnov, [114])

$$\sup |\tilde{\Delta}_n| + \sup |\tilde{\Delta}_n'| \leq c_2 \|\tilde{\Delta}_n\|_{W_2^2[0,h]}, \tag{2.15}$$

where the constants c_1 and c_2 do not depend on s, n and h.

Thus, (2.14) yields

$$\frac{1}{h}\|\tilde{\Delta}_n\|^2_{W_2^2[0,h]} \leq \frac{c_2}{c_1}\gamma_n \|\tilde{\Delta}_n\|_{W_2^2[0,h]}$$

Applying (2.15) again we finally get

$$\frac{1}{h}\sup|\Delta_n| = \frac{1}{h}\sup|\tilde{\Delta}_n| \leq \frac{c_2^2}{c_1}\gamma_n$$

or

$$\frac{1}{h}|W(t,s) - W_n(t,s)| \leq c_3(h\alpha(n) + 1/h)$$

The last result holds uniformly over all $t \in [0,1]$ and $s \in [\tau_1, \tau_2]$ for sufficiently large n and h.

□

Remark 1. All three components of γ_n in (2.12) have a very clear nature: the first term reflects the error caused by approximating the sum in (1.2) by the integral, the second component appears from replacing the natural boundary conditions (2.2) by the requirement that $H(\cdot, s)$ vanishes at $\pm\infty$, and the third term $1/h$ is implied from taking only the main part $H_0(t,s)$ of the true Green's function $H(t,s)$ in the definition of $W(t,s)$. Note that δ in the second component is completely determined by the lengths of the boundary intervals $[0,\tau_1]$ and $[\tau_2, 1]$ where the error bound obtained in Theorem 1 is not true. For every sufficiently large n, providing (2.13) uniformly over all $t \in [0,1]$ and $s \in [\tau_1, \tau_2]$, we may always choose τ_1 and τ_2 so that the corresponding δ satisfies $\exp(-h\delta) < 1/h$. Thus, we can say that the result of Theorem 1 holds uniformly over all t and over all s not "too close" to the boundaries.

It is important to note that asymptotic equivalence between the kernels $W_n(t,s)$ and $W(t,s)$ established in Theorem 1 does not claim asymptotic equivalence between the smoothing spline (1.2) itself and the corresponding kernel estimator with the kernel $W(t,s)$, but, nevertheless, gives intuition about what spline smoothing does to the data.

The obvious corollary from Theorem 1 confirms our preliminary considerations that $W(\cdot,\cdot)$ can be called an asymptotically equivalent kernel in the following sense:

Corollary 1 *Assume in addition to (i) and (ii) that $h \to \infty$ as $n \to \infty$ in such way that $h\alpha(n) \to 0$. Then $W_n(t,s)$, the equivalent kernel for spline smoothing with variable smoothing parameter $k(t) = k_0(t)/h^4$ and design density $p(t)$, is "asymptotically equivalent" to $W(t,s)$ in the sense that $|W_n(t,s)/h - W(t,s)/h| \to 0$ uniformly over all $t \in [0,1]$ and s not "too close" to the boundaries, where*

$$W(t,s) = \frac{h}{2}\{k_0(t)k_0(s)\}^{-1/8}\{p(t)p(s)\}^{-3/8}e^{-h\Phi_0(t,s)}\sin\{h\Phi_0(t,s) + \pi/4\} \tag{2.16}$$

and $\Phi_0(t,s)$ is defined in (2.8).

We finish this Section with several remarks.

Remark 2. It is interesting to compare $W(t,\cdot)$ from (2.16) with the weight function $K(t,\cdot)$ of the standard Priestley-Chao (1972) kernel estimator. The Priestley-Chao's $K(t,s)$ is of the form $K(t,s) = p(s)^{-1}b^{-1}K_0(|t-s|/b)$, where b is a bandwidth. The shape of $K(t,\cdot)$ is defined by $K_0(\cdot)$ and is the same for different t. The kernel $W(t,s)$, however, is not of the Priestley-Chao's convolutional type, so spline smoothing is nearly general weighted moving averaging of the data, where the shape of weight function varies with t.

Remark 3. The main results of this Section may be extended to the case where the second derivative $f''(\cdot)$ in (1.2) is replaced by $f^{(m)}(\cdot)$ for general m. As in the case $m = 2$ the derivation of $W(t,s)$ is based on the use of the WKB-method for finding a fundamental set of solutions of the $2m$-order homogeneous equation $(-1)^m(1/p)(k_0 f^{(m)})^{(m)} + h^{2m} f = 0$ and is of the form :

$$W(t,s) = \frac{h}{m}\{p(t)p(s)\}^{-(2m-1)/4m}\{k_0(t)k_0(s)\}^{-1/4m}\sum_{j=1}^{m/2} e^{-h\mathrm{Im}(\eta_j)\Phi_0(t,s)} \sin\{h\mathrm{Re}(\eta_j)\Phi_0(t,s) + \arg(\eta_j)\}$$

for even m and

$$\begin{aligned} W(t,s) &= \frac{h}{m}\{p(t)p(s)\}^{-(2m-1)/4m}\{k_0(t)k_0(s)\}^{-1/4m} \\ &\times \left[\frac{1}{2}e^{-h\Phi_0(t,s)} + \sum_{j=1}^{(m-1)/2} e^{-h\mathrm{Im}(\eta_j)\Phi_0(t,s)} \sin\{h\mathrm{Re}(\eta_j)\Phi_0(t,s) + \arg(\eta_j)\}\right] \end{aligned}$$

for odd m, where η's are the $2m$-th roots of (-1), $\eta_j = \exp\{i\pi(2j-1)/2m\}$ and $\Phi_0(t,s) = \int_{\min(t,s)}^{\max(t,s)}(p/k_0)^{1/2m}$. For the uniform design and *constant* smoothing parameter this result coincides with that of Messer and Goldstein (1993) for polynomial splines.

Remark 4. Consider $k_0(\cdot) \equiv 1$, that corresponds to the standard polynomial spline smoothing. We note that in this particular case the asymptotic approximation for equivalent kernel obtained here is more accurate than the well-known result of Silverman (1984) for cubic splines. His asymptotic is derived from (2.16) setting $p(\cdot) \equiv p(t)$ on the interval $[t,s]$ (cf. Corollary 1 here with Theorem A and (2.8) of Silverman) and, hence, implies rougher approximation. Thus, to obtain the results analogous to Theorem 1 of this Section he had to correct his approximation by an additional term (see Theorem B of Silverman).

Remark 5. The WKB-approximation of the solution of linear differential equations can be theoretically improved by adding higher terms of the WKB-series expansion.

3. Asymptotic error bounds.

In this Section we illustrate how the results of the previous Section may be applied for statistical inference for spline estimate and derive asymptotic error bounds for the unknown response function.

Wahba (1978, 1983), Wecker and Ansley (1983), Silverman (1985), Kohn and Ansley (1988), using different prior models for g, showed that smoothing splines have a natural interpretation as Bayes estimators of the unknown response function. Wahba's (1978, 1983) prior model for cubic splines assumes that $g''(t)$ is a homoscedastic "white noise". Abramovich and Steinberg (1996) showed that letting g'' be distributed *a priori* as a heteroscedastic "white noise" process, satisfying the stochastic differential equation $d^2X(t)/dt^2 = (\sigma/\sqrt{n})k(t)^{-1/2}\,dW(t)/dt$, where $W(t)$ is a Wiener process with $\text{Var}\{W(1)\} = 1$, yields the smoothing spline with variable smoothing parameter $k(t)$ (1.2) as the posterior mean of $g(t)$ given observations y_i's. Relaxing the requirement of prior homoscedasticity for the second derivative allows one to more closely express his prior belief about local properties of $g(t)$. The variable smoothing parameter $k(t)$ may serve as a vehicle for more accurate expression of information about the local non-linearity of the response function. The estimation of $k(t)$ from the data is considered in Abramovich and Steinberg (1996).

The Bayesian approach allows one to derive posterior pointwise probability intervals that provide natural error bounds for the true function (Wahba 1983, Silverman 1985, Nychka 1988, Ansley *et al.* 1993). Using the standard Bayesian methodology one can obtain the posterior variance of $g(t)$ at every point t. Wahba (1983) showed that for the design-points the posterior covariance $\text{Cov}((g(t_i), g(t_j))|\mathbf{y}) = (\sigma^2/n)W_n(t_i, t_j)$. This result may be extended to arbitrary points. Applying a technique similar to that used for the solution of the general spline smoothing problem (Kimeldorf and Wahba 1971, Wahba 1990), straightforward but tedious calculus provides explicitly the minimizer of (2.4), $W_n(t, s)$. After some matrix algebra it is possible to verify that $W_n(t, s)$ coincides with the posterior covariance function $\text{Cov}(\text{g(t)}, \text{g(s)}|\mathbf{y})$ derived in Gu and Wahba (1993) or in Abramovich and Steinberg (1996). The rigorous proof and explicit formulas are omitted here. Hence, the equivalent kernel $W_n(\cdot, \cdot)$ has a very clear statistical meaning and a $(1-\alpha)$-level Bayesian interval for $g(\cdot)$ is

$$\hat{g}(t) \pm z_{1-\alpha/2}\,\frac{\sigma}{\sqrt{n}}\sqrt{W_n(t,t)} \tag{3.1}$$

Applying the results of the previous Section and replacing $W_n(t,t)$ in (3.1) by $W(t,t)$ we can derive asymptotic $(1-\alpha)$-level Bayesian pointwise intervals.

Suppose that all the assumptions of Corollary 1 of Section 2 are true, that is $h \to \infty$ as $n \to \infty$ in such way that $h\alpha(n) \to 0$. Suppose in addition that h tends to infinity not faster than n, i.e. $h = O(n)$. For example, if design points are "regularly distributed with c.d.f. F", i.e. $t_i = F^{-1}((i-0.5)/n)$, then $\alpha(n) = O(1/n)$ and the additional assumption holds automatically. Under the assumption $h = O(n)$

$$\left|\sqrt{W_n/n} - \sqrt{W/n}\right|^2 \le \left|\sqrt{W_n/n} - \sqrt{W/n}\right| \left(\sqrt{W_n/n} + \sqrt{W/n}\right) \le |W_n/n - W/n| \le C\,|W_n/h - W/h|$$

for some constant C and for all t and s. Corollary 1 implies $|W_n/h - W/h| \to 0$ uniformly over all t and s not "too close" to the boundaries and, therefore, for large h we may replace asymptotically $\sqrt{W_n(t,t)/n}$ in (3.1) by $\sqrt{W(t,t)/n}$ and an asymptotic $(1-\alpha)$-level Bayesian interval for points not "too close" to the boundaries is

$$\hat{g}(t) \pm 2^{-3/4} z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} k(t)^{-1/8} p(t)^{-3/8} \tag{3.2}$$

For constant k the analogous results were obtained in Silverman (1985) from his asymptotic approximation of equivalent kernel discussed in the previous Section. However, since his kernel is not a symmetric function, it cannot be used for approximating posterior covariances.

The approximate intervals (3.2) lighten the role of smoothing parameter and design density for error bounds. The length of the error intervals (3.2) for $g(t)$ is reciprocal to $k(t)^{1/8}$ and $p(t)^{3/8}$. For the fixed design one may control the length of the pointwise intervals by controlling $k(t)$. In examples considered in Abramovich and Steinberg (1996) the appropriate choice of $k(t)$ (estimated from the data) in "almost linear" regions led to error intervals up to 30% narrower than the corresponding intervals for cubic splines with no significant drop in coverage probabilities. At the same time decreasing $k(t)$ in regions of rapid local changes gave the estimate "automatic" protection by broadening the error bounds and considerably improving coverage probabilities at those points. The (3.2) gives a clue for understanding this phenomenon.

4. Extension to general L-splines.

This Section gives a sketch of how the results of previous sections can be extended to the very general L-spline smoothing. As in Section 2 we first show that the corresponding equivalent kernel is asymptotically connected to the Green's function of a certain differential operator and then use the WKB-method for its asymptotical derivation in this very general case. Moreover, the WKB-method will prove that the asymptotically equivalent kernel depends only on the highest-order coefficient of the operator L and, hence, it is the same $W(t,s)$ from Remark 3 in Section 2 with corresponding variable smoothing parameter. The rigorous results analogous to those of Theorem 1 of Section 2 can be obtained in a similar way but involve much more tedious technical details and are omitted.

Define $\hat{g}(t)$ as the minimizer of:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + k\int_0^1 \{(Lf)(t)\}^2 dt, \tag{4.1}$$

where L is a general m-th order linear differential operator of the form $L = \sum_{j=1}^{m} a_j(t)D^j$ and k is a smoothing parameter (see Kimeldorf and Wahba, 1971; Wahba, 1978, 1985; Kohn and Ansley, 1983, 1988).

Let L^* be an adjoint operator to L with respect to a standard inner product in L_2. Then,

$$L^* f(t) = \sum_{j=1}^{m}(-1)^j D^j\{a_j(t)f(t)\} \tag{4.2}$$

It is well-known that the solution of (4.1) is a natural L-spline with the sets of knots $\{t_i\}$ satisfying $L^*L\hat{g} \equiv 0$ everywhere, except, maybe, the data points t_i's and providing $L\hat{g} \equiv 0$ in $[0, t_1]$ and $[t_n, 1]$. These boundary conditions yield $2m$ natural boundary conditions at the endpoints 0 and 1. The exact formula for $\hat{g}$ is given in Kimeldorf and Wahba (1971), Wahba (1990).

By the same arguments from Section 2, $\hat{g}\ (\cdot)$ may be expressed as

$$\hat{g}\ (t) = \frac{1}{n}\sum_{i=1}^{n} W_n(t, t_i) y_i, \tag{4.3}$$

where $W_n(\cdot,\cdot)$ is the corresponding equivalent kernel.

Now we find an asymptotic approximation to $W_n(\cdot,\cdot)$. Suppose again that the design points are distributed in $[0,1]$ with c.d.f. F and design density $p = F'$ which is strictly positive. Replacing the sum in (4.1) by the integral for large n, implies that asymptotically $\hat{g}(\cdot)$ minimizes

$$\int_0^1 (y(t) - f(t))^2 p(t) dt + k \int_0^1 \{(Lf)(t)\}^2 \, dt \tag{4.4}$$

Consider the differential operator $S = (k/p)L^* L + I$ acting on the subspace of functions satisfying natural boundary conditions, where the adjoint operator L^* is given by (4.2). Define in $L_2[0,1]$ the weighted inner product $[f_1, f_2] = \int f_1 f_2 p$. Then, S is obviously a positive self-adjoint operator with respect to this weighted inner product.

Consider the equation $Sf = y$. Since S is a positive self-adjoint operator, there exists a unique solution of the above equation

$$f(t) = \int G(t,s) y(s) ds, \tag{4.5}$$

where $G(\cdot,\cdot)$ is the Green's function of S. Moreover, the solution (4.5) is the minimizer of the functional $[Sf, f] - 2[f, y]$ over all f. Note that $[Sf, f] - 2[f, y] = \int (f^2 - 2fy)p + k \int (Lf)^2$ and up to the constant term $\int y^2 p$ coincides with (4.4). Thus, from (4.3) and (4.5) it follows that $W_n(t,s)$ can be asymptotically approximated by $G(t,s)/p(s)$. This result extends the analogous results of Abramovich (1993) for uniform design density and the results from Section 2 for L_k-splines.

Suppose that $k = 1/h^{2m}$ and let $h \to \infty$ as $n \to \infty$. First we approximate $G(t,s)$ for large h by the Green's function $H(t,s)$ replacing the natural boundary conditions by the requirement of vanishing as t tends to $\pm\infty$, and then use the WKB-method to find the main part $H_0(t,s)$ of $H(t,s)$.

We start from the homogeneous equation $Sf = (h^{-2m}/p)L^* Lf + f = 0$. After some calculus one can verify that

$$L^*Lf = (-1)^m\{a_m^2 f^{(2m)} + m(a_m^2)' f^{(2m-1)}\} + R,$$

where R contains all terms with derivatives of f less than $(2m-1)$-th order. The original homogeneous equation can be re-written now as

$$f^{(2m)} + m\frac{(a_m^2)'}{a_m^2} f^{(2m-1)} + R_1 + (-1)^m h^{2m}\frac{p}{a_m^2} f = 0, \tag{4.6}$$

where $R_1 = (-1)^m R/a_m^2$.

Following the WKB-method, we seek the asymptotic solution for (4.6) in the form

$$f(t) \sim \exp\left\{h\int_0^t \Psi(z)dz\right\}\sum_{k=0}^{\infty}\frac{C_k(t)}{h^k} \tag{4.7}$$

The slightly modified Proposition 1 from Appendix guarantees the conditions of Theorem 6.3.1 of Coddington and Levinson (1955) and, thus, the solution of this form exists.

Substituting (4.7) into (4.6) and expanding by powers of h after tedious straightforward calculus one has

$$\Psi^{2m}h^{2m} + \left(mC_0\frac{(a_m^2)'}{a_m^2}\Psi + (2m-1)mC_0\Psi' + 2mC_0'\Psi\right)\Psi^{2m-2}h^{2m-1} + R_2 + (-1)^m\frac{p}{a_m^2}h^{2m} = 0, \tag{4.8}$$

where R_2 contains all powers of h less than $2m-1$. As $h \to \infty$ only $C_0(\cdot)$ is relevant in (4.8) while all other $C_k(\cdot)$'s, $k > 0$ (and, therefore, R_2), contribute $O(1/h)$ to the resulting solution. From (4.8) we immediately have

$$\Psi = \eta_j\left(\frac{p}{a_m^2}\right)^{1/2m} \qquad C_0 = p^{-(2m-1)/4m}a_m^{-1/2m},$$

where η_j's are the $2m$-th complex roots of (-1) and, finally, the corresponding fundamental solutions are

$$f_j(t) = \frac{1}{p(t)^{(2m-1)/4m}a_m(t)^{1/2m}}\exp\left\{h\eta_j\int_0^t (p/a_m^2)^{1/2m}\right\}\{1 + O(1/h)\}, \quad j = 1, ..., 2m \tag{4.9}$$

It is important to note that the main part of the fundamental solutions (4.9) depends only on the highest-order coefficient a_m of the original operator L in (4.1). The Green's function $H(t,s)$ is a linear combination of the fundamental solutions and, hence, its main part $H_0(t,s)$ will also depend on t only through $a_m(t)$. Moreover, since S is a self-adjoint operator, $H_0(t,s)/p(s)$ is a symmetric function and, hence, $H_0(t,s)$ will also depend on s only through $a_m(s)$. Thus, the main part of the Green's function of the operator S is completely defined by the highest-order coefficient $a_m(\cdot)$ of the operator L. The equivalent kernel $W_n(t,s)$ in (4.3) can be asymptotically (for large h) approximated by $W(t,s) = H_0(t,s)/p(s)$, where $W(t,s)$ is exactly the same as in Remark 3 of Section 2 for the L_k-spline smoothing with $k_0(\cdot) = a_m^2(\cdot)$. These results may seem somewhat surprising for the "spline community" but are quite "natural" in asymptotical theory of differential equations. The WKB-method is key for the understanding of this phenomenon.

Summarizing the main results of this Section, we can conclude that for large samples and for small k, a)the asymptotically equivalent kernel for the general L-spine smoothing is the same as for the "usual" spline smoothing but with a variable smoothing parameter defined by the highest-order coefficient of the operator L; b)general spline smoothing is nearly weighted moving averaging of the data with the kernel $W(\cdot,\cdot)$ given in Remark 3 of Section 2. Unlike standard Priestley-Chao's (1972) convolutional kernel, the shape of the weight function at point t, $W(t,\cdot)$, varies with t.

Acknowledgements. The authors are most grateful to David Steinberg for stimulating and useful discussions and to Bernard Silverman for valuable remarks.

APPENDIX.

1. Derivation of the fundamental system of the homogeneous equation (2.7).

The following Proposition 1 provides the theoretical ground for applying the WKB-method for solving this equation.

Proposition 1 *Let η's be the fourth roots of (-1) and a function $\phi(t) \in C[0,1]$. There exists an angle Ω in the complex domain with vertex at the origin such that for every pair (i,j) the expression* $\mathrm{Re}\left\{h(\eta_i - \eta_j)(p(t)/k_0(t))^{1/4} + \phi(t)\right\}$ *has the same sign for all $t \in [0,1]$*

and for every sufficiently large $|h|$, where $h \in \Omega$.

Proof. It is sufficient to show that

$$\arg\left[h(\eta_i - \eta_j)\{p(t)/k_0(t)\}^{1/4} + \phi(t)\right] \neq \pm\frac{\pi}{2} \qquad (A.1)$$

for all sufficiently large $|h|$, $h \in \Omega$ and for all $t \in [0, 1]$.

Note that

$$\arg\left[h(\eta_i - \eta_j)\{p(t)/k_0(t)\}^{1/4}\right] = \arg(h) + \arg(\eta_i - \eta_j) + \arg(p/k_0)^{1/4} \qquad (A.2)$$

The second term in (A.2) is a constant for every pair (i, j), while the third one is zero since p/k_0 is a positive real function. Thus, we can always choose Ω such that

$$\left|\pm\frac{\pi}{2} - \arg[h(\eta_i - \eta_j)\{p(t)/k_0(t)\}^{1/4}]\right| \geq c > 0,$$

and, thereby, (A.1) always holds for sufficiently large $|h|$ due to the boundedness of $\phi(\cdot)$.
□

Now we may directly apply all corresponding results from Coddington and Levinson (1955, Section 6.3) to (2.7). Using their notations it is easy to show that

$$|A_0(t) - \lambda I| = \lambda^4 + \frac{p(t)}{k_0(t)}$$

where the matrix $A_0(t)$ is defined in Coddington and Levinson. The characteristic roots of the matrix $A_0(t)$ are $\lambda_j = (p(t)/(k_0(t))^{1/4}\eta_j$. Proposition 1 ensures that the hypothesis H of Coddington and Levinson (Section 6.3) holds and from their Theorem 6.3.1 for sufficiently large real h

$$f_j(t) = C_0(t)\exp\left\{h\eta_j \int_0^t \left(\frac{p}{k_0}\right)^{1/4}\right\}\{1 + O(1/h)\}$$

that justifies the form of solution we sought.

2. Proof of the lemmas from Section 2.

Here we prove the lemmas that were used in the proof of Theorem 1 in Section 2. Recall that all the derivatives are with respect to t.

Proof of Lemma 1. Recall that $W_s(t)\{1+O(1/h)\} = H(t,s)/p(s)$ where $H(t,s)$ is the Green's function for the operator with the differential expression $(1/p)D^2kD^2 + I$. Thus,

$$\frac{p(s)}{p(t)}\{k(t)W_s(t)''\}''\{1+O(1/h)\} + p(s)W_s(t)\{1+O(1/h)\} = \delta(t-s) \qquad (A.3)$$

Multiplying both parts of (A.3) by $u(t)p(t)/p(s)$ and integrating we have

$$\int (kW_s'')''u + \int pW_s u = u(s)\{1+O(1/h)\} \qquad (A.4)$$

Taking the first integral in (A.4) by parts twice completes the proof of the Lemma.

□

Proof of Lemma 2. Direct differentiation yields corresponding derivatives to be of the form

$$[kW_s'']_0^1 = \frac{1}{h^3}\left(e^{-h\{\Theta(1)-\Theta(s)\}}\sum_{j=0}^{2} x_j(1,s)h^j - e^{-h\{\Theta(s)-\Theta(0)\}}\sum_{j=0}^{2} x_j(0,s)h^j\right) \qquad (A.5a)$$

$$[(kW_s'')']_0^1 = \frac{1}{h^3}\left(e^{-h\{\Theta(1)-\Theta(s)\}}\sum_{j=0}^{3} z_j(1,s)h^j - e^{-h\{\Theta(s)-\Theta(0)\}}\sum_{j=0}^{3} z_j(0,s)h^j\right), \qquad (A.5b)$$

where $\Theta(s) = \int_0^s (p/k_0)^{1/4}$, x_j and z_j are certain functions uniformly bounded w.r.t. s and independent of n and h under Assumptions (i) and (ii) of Section 2.

Choose $\delta > 0$ such that

$$\delta \le \min(\tau_1, 1-\tau_2)\inf (p/k_0)^{\frac{1}{4}}$$

Then

$$\Theta(1) - \Theta(s) = \int_s^1 (p/k_0)^{\frac{1}{4}} \ge \delta \qquad \Theta(s) - \Theta(0) = \int_0^s (p/k_0)^{\frac{1}{4}} \ge \delta$$

So,

$$e^{-h(\Theta(1)-\Theta(s))} \le e^{-\delta h} \qquad e^{-h(\Theta(s)-\Theta(0))} \le e^{-\delta h}$$

that together with (A.5) gives

$$\left|[(kW_s'')u' - (kW_s'')'u]_0^1\right| \le \frac{e^{-\delta h}}{h^3}\left[\left(\sum_{j=1}^{3} q_{1j}(s)h^j\right)\sup|u| + \left(\sum_{j=1}^{2} q_{2j}(s)h^j\right)\sup|u'|\right],$$

where q_{1j} and q_{2j} are certain functions uniformly bounded w.r.t. s and independent of n and h. Thus,

$$\left|[(kW_s'')u' - (kW_s'')'u]_0^1\right| \le C_1 e^{-\delta h}(\sup|u| + \sup|u'|/h),$$

where δ and C_1 do not depend on s, n and h.

□

Proof of Lemma 3. Recall that $F_n(0) = F(0) = 0$ and $F_n(1) = F(1) = 1$. Thus, integrating by parts we have

$$\begin{aligned}\left|\int_0^1 W_s u\, d(F_n - F)\right| &= \left|\int_0^1 (W_s u)'(F_n - F)\, dt\right| \le \alpha(n)\left[\int_0^1 |W_s' u| + \int_0^1 |W_s u'|\right] \\ &\le \alpha(n)\left[\sup|u|\int_0^1 |W_s'| + \sup|u'|\int_0^1 |W_s|\right]\end{aligned}$$

Proposition 2 *There exist constants c_1 and c_2 independent of s and h such that*

$$\int_0^1 |W_s'| \le c_1 h \quad \text{and} \quad \int_0^1 |W_s| \le c_2$$

Proof of Proposition 2. From (2.8) the function $W_s(t)$ is of the form

$$W_s(t) = hU_1(t,s)\exp\left\{-h\int_{\min(s,t)}^{\max(s,t)} U_2(\rho)d\rho\right\}$$

and, hence,

$$W_s'(t) = h\exp\left\{-h\int_{\min(s,t)}^{\max(s,t)} U_2(\rho)d\rho\right\}\{P_1(t,s) + hP_2(t,s)\}$$

where $|P_1(t,s)|$ and $|P_2(t,s)|$ are bounded by some constant that does not depend on s and h due to the boundedness of all corresponding functions and their derivatives. Under

Assumption (ii), $U_2(\cdot) \geq \beta$ for some positive constant β. Then,

$$\exp\left\{-h\int_{\min(s,t)}^{\max(s,t)} U_2(\rho)d\rho\right\} \leq \exp\{-\beta h|t-s|\}$$

and, therefore,

$$|W_s'| \leq c_0 h^2 e^{-\beta h|t-s|}$$

So,

$$\int_0^1 |W_s'| \leq c_0 h^2 \int_0^1 e^{-\beta h|t-s|}dt = c_0 h^2 \left(\int_0^s + \int_s^1\right) = (c_0 h^2/\beta h)\left[e^{\beta h(t-s)}|_0^s - e^{\beta h(s-t)}|_s^1\right] \leq c_1 h$$

Repeating the same arguments for $|W_s|$ we get the second inequality of Proposition 2 and the rest of the proof of Lemma 3 follows immediately.

□

References

[1] Abramovich, F. (1993). The asymptotic mean squared error of L-smoothing splines. *Stat. Probab. Lett.*, **18**, 179-182.

[2] Abramovich, F. and Steinberg, D.M. (1996). Improved inference in nonparametric regression using L_k-smoothing splines. *J. Stat. Plann. Infer.* **49** 327-341.

[3] Ansley, C.F., Kohn, R. and Wong, C.-H. (1993). Nonparametric spline regression with prior information. *Biometrika* **80** 75-88.

[4] Bender, C.M. and Orszag, S.A. (1978). *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill.

[5] Coddington, E.A. and Levinson, N. (1955). *Theory of Ordinary Differential Equations.* McGraw-Hill.

[6] Cox, D.D. (1983). Asymptotics for M-type smoothing spline. *Ann. Statist.* **11** 530-551.

[7] Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression.* New York: Dekker.

[8] Gu, C. and Wahba, G. (1993). Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". *J. Comput. Graph. Stat.*, **2**, 97-117.

[9] Hutchinson, M. and de Hoog, F. (1986). Smoothing data with spline functions. *Numer. Math.* **47** 99-106.

[10] Jeffreys, H. (1924). On certain approximate solutions of linear differential equations of the second order. *London Math. Soc.* **23** 428-436.

[11] Kimeldorf, G.S. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82-95.

[12] Kohn, R. and Ansley, C.F. (1983). On the smoothness properties of the best linear unbiased estimate of a stochastic process observed with noise. *Ann. Statist.* **11** 1011-1017.

[13] Kohn, R. and Ansley, C.F. (1988). Equivalence between Bayesian smoothness priors and optimal smoothing for function estimation, in *Bayesian Analysis of Time Series and Dynamic Models* edited by J.C. Spall. Marcel Dekker, New York.

[14] Messer, K. (1991). A comparison of a spline estimate to its equivalent kernel estimate. *Ann. Statist.* **19** 817-829.

[15] Messer, K. and Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation. *Ann. Statist.* **21** 179-195.

[16] Nussbaum, M. (1985). Spline smoothing in regression and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984-997.

[17] Nychka, D.W. (1988). Bayesian confidence intervals for a smoothing spline. *J. Amer. Stat. Ass.* **83** 1134-1143.

[18] Nychka, G.W. (1995). Splines as local smoothers. *Ann. Statist.* **23** 1175-1197.

[19] Oehlert, G.W. (1992). Relaxed boundary smoothing splines. *Ann. Statist.* **20** 146-160.

[20] Priestley, M.B. and Chao, M.T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser.* **B 34** 385-392.

[21] Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12** 898-916.

[22] Silverman, B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Roy. Statist. Soc. Ser.* **B 47** 1-52 (with discussion).

[23] Smirnov, V.I. (1964). *A Course of Higher Mathematics*, Vol. **5**. Pergamon Press.

[24] Speckman, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. University of Oregon preprint.

[25] Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970-983.

[26] Tapia, R.A. and Thompson, J.R. (1978). *Nonparametric Probability Density Estimation.* Baltimore: Johns Hopkins University Press.

[27] Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser.* **B 40** 364-372.

[28] Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser.* **B 45** 133-150.

[29] Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Stat.* **13** 1378-1402.

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

[31] Wecker, W.E. and Ansley, C.F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *J. Amer. Stat. Ass.* **78** 81-89.