

## EQUIVALENT KERNELS FOR SMOOTHING SPLINES

P.P.B. EGGERMONT AND V.N. LARICCIA

To Ken Atkinson on the occasion of his 65th birthday

**ABSTRACT.** In the study of smoothing spline estimators, some convolution-kernel-like properties of the Green's function for an appropriate boundary value problem, depending on the design density, are needed. For the uniform density, the Green's function can be computed more or less explicitly. Then, integral equation methods are brought to bear to establish the kernel-like properties of said Green's function. We briefly survey how the Green's function arises in spline smoothing as the equivalent kernel, the reproducing kernel of a suitable Hilbert space, and as the Green's function for the Euler equations of a semi-continuous version of the spline smoothing problem.

**1. Introduction.** In this paper, we study the Green's function for the boundary value problem,

$$(1.1) \quad \begin{aligned} (-h^2)^m u^{(2m)} + wu &= v \quad \text{on } (0, 1), \\ u^{(k)}(0) = u^{(k)}(1) &= 0, \quad k = m, \dots, 2m - 1. \end{aligned}$$

Here,  $m$  is a positive integer,  $h$  is a positive parameter tending to 0, and  $w$  is a positive measurable function, which is bounded and bounded away from 0, i.e., there exist positive constants  $w_1$  and  $w_2$  such that

$$(1.2) \quad w_1 \leq w(t) \leq w_2, \quad \text{a.e. } t \in (0, 1).$$

Also,  $u^{(k)}$  denotes the  $k$ th derivative, for  $k = 1, 2, \dots$ . The above Green's function arises in the precise analysis of the *smoothing spline* estimator for the following, standard nonparametric regression problem. One observes the data  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ , which is interpreted as

$$(1.3) \quad Y_i = f_0(X_i) + D_i, \quad i = 1, 2, \dots, n.$$

---

2000 AMS *Mathematics Subject Classification.* Primary 34B27, 45A05, 62G08.  
*Key words and phrases.* Spline smoothing, random designs, equivalent kernels, reproducing kernels, Green's functions.

Received by the editors on August 16, 2005, and in revised form on December 14, 2005.

Copyright ©2006 Rocky Mountain Mathematics Consortium

Here,  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$  is the *random design*, i.e.,  $X_1, X_2, \dots, X_n$  are independent, identically distributed (iid) random variables with common probability density function (pdf)  $w(t)$  on a bounded interval, which we take to be  $(0, 1)$ . The noise  $\mathbf{D}_n = (D_1, D_2, \dots, D_n)$  are iid, conditional on the design  $\mathbf{X}_n$ , with

$$(1.4) \quad \mathbf{E}[\mathbf{D}_n \mid \mathbf{X}_n] = 0, \quad \mathbf{E}[\mathbf{D}_n^T \mathbf{D}_n \mid \mathbf{X}_n] = \sigma^2 I_{n \times n},$$

where  $\sigma^2$  is not known. In addition, one needs that, for some constant  $\kappa > 3$ ,

$$(1.5) \quad \mathbf{E}[|D_1|^\kappa \mid X_1] < \infty.$$

The goal is to estimate the function  $f_0$ , which is assumed to be smooth, i.e., for some integer  $m \geq 1$ ,

$$(1.6) \quad f_0 \in W^{m,2}(0, 1),$$

where  $W^{m,2}(0, 1)$  is the Sobolev space of order  $m$ . Specifically,

$$(1.7) \quad W^{m,2}(0, 1) = \left\{ f \in C^{m-1}[0, 1] \mid \begin{array}{l} f^{(m-1)} \text{ absolutely continuous} \\ f^{(m)} \in L^2(0, 1) \end{array} \right\}.$$

For general introductions to nonparametric regression and the various estimators, including smoothing splines, see, e.g., Eubank [17], Wahba [28] and Györfy et al. [18].

The smoothing spline estimator of  $f_0$  is then defined as the solution of the problem

$$(1.8) \quad \begin{aligned} \text{minimize} \quad & LS(f) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2 + h^{2m} \|f^{(m)}\|^2 \\ \text{such that} \quad & f \in W^{m,2}(0, 1). \end{aligned}$$

where  $\|\cdot\|$  denotes the  $L^2(0, 1)$  norm.

There are several ways in which the Green's function for (1.1) arises.

*Reproducing kernel Hilbert spaces.* First, one needs to worry about whether the problem (1.8) is properly formulated, in particular, whether

the point evaluations  $f \mapsto f(X_i)$  make sense. It is well known that  $W^{m,2}(0,1)$  embeds continuously into  $C[0,1]$ , so that there exists a constant  $c_m$  such that for all  $f \in W^{m,2}(0,1)$  and all  $x \in [0,1]$ ,

$$(1.9) \quad |f(x)| \leq c_m \left\{ \|f\|^2 + \|f^{(m)}\|^2 \right\}^{1/2},$$

see, e.g., Adams and Fournier [2]. Thus, the objective function in (1.8) is well defined, and establishing the existence and uniqueness of solutions of the problem (1.8) is straightforward. However, it makes sense to amend the result (1.9) in two ways. First, it seems reasonable to replace the  $L^2(0,1)$  norm,  $\|f\|^2$ , by the weighted  $L^2$  norm, defined by

$$(1.10) \quad \|f\|_{L^2(w)}^2 = \langle f, f \rangle_{L^2(w)},$$

where

$$(1.11) \quad \langle f, g \rangle_{L^2(w)} = \int_0^1 f(x) g(x) w(x) dx.$$

In other words, the design density is taken into account. Second, in view of the spline smoothing problem, it makes sense to weight  $\|f^{(m)}\|^2$  by the factor  $h^{2m}$ . Thus, we introduce the inner products

$$(1.12) \quad \langle f, g \rangle_{wmh} = \langle f, g \rangle_{L^2(w)} + h^{2m} \langle f^{(m)}, g^{(m)} \rangle_{L^2(0,1)},$$

and, as usual, define the norm by  $\|f\|_{wmh} = \{\langle f, f \rangle_{wmh}\}^{1/2}$ . Now, a scaling argument applied to (1.9) results in the inequality, for  $0 < h < 1$ , all  $x \in [0,1]$  and all  $f \in W^{m,2}(0,1)$ ,

$$(1.13) \quad |f(x)| \leq c_m h^{-1/2} \|f\|_{mh} \leq c_{mw} h^{-1/2} \|f\|_{wmh},$$

the last inequality because of (1.2), with  $c_{mw} = c_m w_1^{-1/2}$ . The inequality (1.13) says that, for each  $h$ , the space  $W^{m,2}(0,1)$  with the innerproduct  $\langle \cdot, \cdot \rangle_{wmh}$  is a reproducing kernel Hilbert space, so that there exists a function  $\mathfrak{R}_{wmh}(x, y)$ ,  $x, y \in [0,1]$ , such that  $\mathfrak{R}_{wmh}(x, \cdot) \in W^{m,2}(0,1)$  for each  $x \in [0,1]$ , and for all  $x \in [0,1]$  and all  $f \in W^{m,2}(0,1)$ ,

$$(1.14) \quad f(x) = \langle f, \mathfrak{R}_{wmh}(x, \cdot) \rangle_{wmh}.$$

Then, (1.13) implies the nifty bound

$$(1.15) \quad \|\mathfrak{R}_{wmh}(x, \cdot)\|_{wmh} \leq c_{mw} h^{-1/2},$$

with  $c_{mw}$  as in (1.13). For more on reproducing kernel Hilbert spaces, see Aronszajn [3]. Of course, the reproducing kernel  $\mathfrak{R}_{wmh}(x, y)$  is the Green's function for (1.1), see, e.g., Dolph and Woodbury [11].

The reproducing kernel gets used as follows. Taking the existence and uniqueness of the solution of (1.8) for granted, we denote the solution of (1.8) by  $f = f^{nh}$ . Then, since we are dealing with a quadratic minimization problem, one obtains the quadratic behavior of the objective function around its minimizer,

$$(1.16) \quad \frac{1}{n} \sum_{i=1}^n |\varepsilon(X_i)|^2 + h^{2m} \|\varepsilon^{(m)}\|^2 = LS(f_0) - LS(f^{nh}),$$

where  $\varepsilon \equiv f^{nh} - f_0$ . After some standard manipulations, as detailed in [14], one then arrives at

$$(1.17) \quad \frac{1}{n} \sum_{i=1}^n |\varepsilon(X_i)|^2 + h^{2m} \|\varepsilon^{(m)}\|^2 \leq S_n(\varepsilon) + h^{2m} \|f_0^{(m)}\| \|\varepsilon^{(m)}\|,$$

where, for  $f \in W^{m,2}(0, 1)$ ,

$$(1.18) \quad S_n(f) = \frac{1}{n} \sum_{i=1}^n D_i f(X_i).$$

Now, the problem is to bound  $S_n(\varepsilon)$  in terms of a suitable norm of  $\varepsilon$ . Since  $\varepsilon \in W^{m,2}(0, 1)$ , one obtains by way of the representation (1.14) that

$$(1.19) \quad S_n(\varepsilon) = \langle \varepsilon, \mathfrak{S}_{nh} \rangle_{wmh} \leq \|\varepsilon\|_{wmh} \|\mathfrak{S}_{nh}\|_{wmh},$$

in which  $\mathfrak{S}_{nh}(x) = S_n(\mathfrak{R}_{wmh}(\cdot, x))$ . In other words,

$$(1.20) \quad \mathfrak{S}_{nh}(x) = \frac{1}{n} \sum_{i=1}^n D_i \mathfrak{R}_{wmh}(X_i, x), \quad x \in [0, 1].$$

To get a feeling for the size of  $\mathfrak{S}_{nh}$ , using the properties (1.4) of the noise and the bound (1.15), it is straightforward to show that

$$(1.21) \quad \mathbf{E}[\|\mathfrak{S}_{nh}\|_{wmh}^2 \mid \mathbf{X}_n] \leq c_{mw}^2 (nh)^{-1}.$$

A moment's reflection shows that the rate  $(nh)^{-1}$  is sharp, since the "rate"  $h^{-1/2}$  in (1.15), respectively (1.13), is sharp.

Thus, the right-hand side of (1.17) may be (crudely) bounded by

$$\left\{ \|\mathfrak{S}_{nh}\|_{wmh} + h^{2m} \|f_0^{(m)}\| \right\} \cdot \|\varepsilon\|_{wmh},$$

and this would go a long way towards obtaining bounds on  $\|\varepsilon\|_{wmh}$ , if in the left-hand side of the inequality (1.17), one could replace the sum

$$(1.22) \quad \frac{1}{n} \sum_{i=1}^n |\varepsilon(X_i)|^2 \quad \text{by} \quad \|\varepsilon\|_{L^2(w)}^2.$$

For this, it suffices to obtain a bound like

$$(1.23) \quad \left| \frac{1}{n} \sum_{i=1}^n |\varepsilon(X_i)|^2 - \|\varepsilon\|_{L^2(w)}^2 \right| \leq \eta^{nh} \|\varepsilon\|_{wmh}^2,$$

where  $\eta^{nh} \rightarrow 0$  almost surely, provided  $h$  does not tend to 0 too fast. (Say  $nh^3$  bounded away from 0.) It is interesting that the reproducing kernel again enters into the picture. Starting with  $\varepsilon^2$  replaced by just  $f$ , and then, using the reproducing kernel Hilbert space trick  $f(X_i) = \langle f, \mathfrak{R}_{wmh}(X_i, \cdot) \rangle_{wmh}$  (and Fubini's theorem), one obtains for

$$(1.24) \quad \Delta_n(f) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n f(X_i) - \int_0^1 f(x) w(x) dx$$

the representation and bound

$$(1.25) \quad \begin{aligned} |\Delta_n(f)| &\leq |\langle f, w^{nh} - w_h \rangle_{wmh}| \\ &\leq \|f\|_{wmh} \|w^{nh} - w_h\|_{wmh}, \end{aligned}$$

where

$$(1.26) \quad \begin{aligned} w^{nh}(x) &= \frac{1}{n} \sum_{i=1}^n \mathfrak{R}_{wmh}(X_i, x), \\ w_h(x) &= \mathbf{E}[w^{nh}(x)] = \int_0^1 \mathfrak{R}_{wmh}(x, t) w(t) dt. \end{aligned}$$

Note that  $w^{nh}$  is an estimator of the design density. We are tempted to call it a reproducing kernel density estimator, in analogy with the standard kernel density estimator, in which  $\mathfrak{R}_{wmh}(X_i, x)$  is replaced by a convolution kernel  $K_h(X_i - x)$ , see the classic Devroye and Györfy [10], or the authors' favorite, [13].

Now, the task at hand is to bound  $\|w^{nh} - w_h\|_{wmh}$ , and obviously, this requires some properties on the reproducing kernel  $\mathfrak{R}_{wmh}$ . Finally, one needs to replace  $f$  by  $\varepsilon^2$ . See [14] for the full details.

*Remark 1.* The above elementary approach is a somewhat nonstandard way of dealing with the random sums  $S_n(\varepsilon)$ . The standard way is by considering a suitable closed set  $\mathcal{F} \subset W^{m,2}(0,1)$ , e.g., the unit ball, and studying the supremum of  $S_n(f)$  over  $f \in \mathcal{F}$ . Now, the “size” of  $\mathcal{F}$  as a subset of  $L^2(0,1)$  comes into play, where the “size” is measured in terms of the Kolmogorov (or metric) entropy of  $\mathcal{F}$ . We shall not address this further. See, e.g., Dudley [12].

*C-splines.* In the above, we outlined how the error in the smoothing spline is bounded by a suitable norm of  $\mathfrak{S}_{nh}$ . In fact, the error behaves exactly in this way. In [14], we show that under the conditions (1.2)–(1.6) that

$$(1.27) \quad f^{nh}(x) - \mathbf{E}[f^{nh}(x) \mid \mathbf{X}_n] = \mathfrak{S}_{nh}(x) + \delta^{nh}(x),$$

with

$$\|\delta^{nh}\|_\infty = o((nh)^{-1/2}) \quad \text{almost surely,}$$

provided  $n \rightarrow \infty$ ,  $h \rightarrow 0$ , with  $nh^3$  bounded away from 0. The way this comes about is as follows. Write

$$|f(X_i) - Y_i|^2 = |f(X_i) - f_0(X_i)|^2 - 2D_i(f(X_i) - f_0(X_i)) + |D_i|^2,$$

and as in (1.22), approximate the sum  $1/n \sum_{i=1}^n |f(X_i) - f_0(X_i)|^2$  by the corresponding integral. This leads to the minimization problem

$$(1.28) \quad \begin{array}{ll} \text{minimize} & CLS_n(f - f_0) + h^{2m} \|f^{(m)}\|^2 \\ \text{subject to} & f \in W^{m,2}(0,1), \end{array}$$

where

$$(1.29) \quad CLS(f) = \|f\|_{L^2(w)}^2 - \frac{2}{n} \sum_{i=1}^n D_i f(X_i).$$

We call the solution a  $C$ -spline estimator ( $C$  for continuous). This is slightly different from the “continuous” splines of Cox [7, 8]. The solution of (1.28) should be a pretty good approximation to the solution  $f^{nh}$  of (1.8), and indeed it is.

Now, by way of the Euler equations, one verifies that the solution of (1.28), denoted as  $\psi^{nh}(x)$ , may be written as

$$(1.30) \quad \psi^{nh}(x) = \mathbf{E}[\psi^{nh}(x)] + \frac{1}{n} \sum_{i=1}^n D_i \mathfrak{R}_{wmh}(X_i, x),$$

with

$$\mathbf{E}[\psi^{nh}(x)] = \int_0^1 \mathfrak{R}_{wmh}(x, y) f_0(y) w(y) dy, \quad x \in [0, 1].$$

This uses the fact that  $\mathfrak{R}_{wmh}$  is the Green’s function for the boundary value problem (1.1).

*Desirable properties of the Green’s function.* It is clear that for a detailed study of the random functions  $\mathfrak{S}_{nh}(x)$  and  $w^{nh}(x)$ , some information on the Green’s function  $\mathfrak{R}_{wmh}(x, y)$  is needed. In the probability literature, powerful results are available on random functions of the form

$$(1.31) \quad \varphi^{nh}(x) = \frac{1}{n} \sum_{i=1}^n D_i K_h(x - X_i), \quad x \in [0, 1],$$

where  $K_h(x) = h^{-1} K(h^{-1}x)$ , for some nice function  $K$ , in particular,

$$(1.32) \quad K \in L^1(\mathbf{R}) \cap BV(\mathbf{R}), \quad K(x) = 0 \quad \text{for } |x| > 1.$$

See, e.g., Deheuvels and Mason [9], and Einmahl and Mason [16], with precursors like Konakov and Piterbarg [20], and Härdle, Janssen and Serfling [19]. The functions  $K$  satisfying (1.32) are usually referred to as “kernels.” To distinguish them from reproducing kernels, we shall call them “convolution kernels.” Now, the Green’s function  $\mathfrak{R}_{wmh}(x, y)$  is not a convolution kernel, but we prove in this paper that it has properties quite analogous: There exist positive constants  $c, \gamma$  and  $\delta$  such that for all  $h, 0 < h < 1$ ,

$$(1.33) \quad \begin{aligned} \sup_{x \in [0,1]} \|\mathfrak{R}_{wmh}(x, \cdot)\|_\infty &\leq c h^{-1}, \\ \sup_{x \in [0,1]} \|\mathfrak{R}_{wmh}(x, \cdot)\|_1 &\leq c, \\ \sup_{x \in [0,1]} |\mathfrak{R}_{wmh}(x, \cdot)|_{BV(0,1)} &\leq c h^{-1}, \end{aligned}$$

and for all  $x, y \in [0, 1]$ ,

$$(1.34) \quad |\mathfrak{R}_{wmh}(x, y)| \leq \gamma h^{-1} \exp(-\delta h^{-1} |x - y|).$$

In (1.33),  $\|\cdot\|_p$  denotes the standard norm on  $L^p(0, 1)$ , for  $1 \leq p \leq \infty$ , and  $|\cdot|_{BV(A)}$  denotes the semi-norm on the space of functions (no equivalence classes) of bounded variation on  $A$  ( $[0, 1]$  or  $\mathbf{R}$ ); see, e.g., Ziemer [29]. Note that convolution kernels have these properties, except for the exponential decay (but obviously, a convolution kernel decays like an  $L^1$  function.) The properties (1.33) are useful for establishing rates of convergence of  $\mathfrak{S}_{nh}$  in various norms, e.g., the sup-norm. The exponential decay (1.34) is useful for showing that  $\sqrt{nh} \mathfrak{S}_{nh}$  converges to “white noise,” e.g., it implies that for  $x \neq y$ ,

$$(1.35) \quad nh \mathbf{E}[\mathfrak{S}_{nh}(x) \mathfrak{S}_{nh}(y) \mid \mathbf{X}_n] \longrightarrow 0, \quad h \rightarrow 0, \quad nh \rightarrow \infty.$$

An additional property of convolution kernels deals with measures of compactness of the sets  $\{K_h : a < h < b\}$  where  $0 < a < b < 1$ .

This comes about if one wishes to study the behavior of  $\varphi^{nh}$  as a function of  $h$ . This may take the form of inquiring about the almost sure boundedness of expressions like

$$(1.36) \quad \limsup_{n \rightarrow \infty} \sup_{h \in \mathcal{H}_n} D_n(h) / \sqrt{nh (\log(1/h) \vee \log \log n)}$$

where  $D_n(h) = \|\varphi^{nh}\|_\infty$  and  $\mathcal{H}_n$  is an interval, e.g.,  $\mathcal{H}_n = [n^{-1} \log n, 1/2]$ . The difficulty is that one cannot really deal with the supremum, other than by approximating it with a finite maximum. Ignoring the scaling in (1.36), one may consider

$$(1.37) \quad \sup_{h \in \mathcal{H}_n} D_n(h) \leq \max_{1 \leq i \leq N} D_n(h_i) + \sup_{h \in \mathcal{H}} \min_{1 \leq i \leq N} |D_n(h) - D_n(h_i)|,$$

where  $N$  and  $a = h_1 < h_2 < \dots < h_N = b$  must be chosen “appropriately” so as to balance the number of points  $N$  and the resulting approximation error (the second term on the right). This leads straight to the metric entropy of the aforementioned sets. See, e.g., Dudley [12] and Remark 1. Obviously, some information is required on the behavior of  $D_n(h)$  as a function of  $h$ . What is required are results like

$$\|K_h - K_\lambda\|_1 \leq c |1 - h/\lambda|$$

and

$$(1.38) \quad \|K_h - K_\lambda\|_\infty \leq c |h^{-1} - \lambda^{-1}|.$$

It is an exercise to show that  $K \in BV(\mathbf{R})$  and  $K$  having compact support imply (1.38). In Section 6, we formulate and prove the analogue of (1.38) for the reproducing kernels  $\mathfrak{R}_{wmh}$ .

*Equivalent kernels.* In the spline smoothing literature, the Green’s function goes under the name of “equivalent kernel,” see, e.g., Speckman [27], Cox [7, 8], Silverman [26], Messer [22], Messer and Goldstein [23], Nychka [24] and Chiang, Rice and Wu [6].

There are two aspects to the equivalent kernel set-up. One aspect concerns the convolution-kernel like properties of the reproducing kernel and the properties of the reproducing kernel estimator of the regression function. The other one deals with the accuracy of the reproducing

kernel estimator as an approximation to the original smoothing spline estimator.

Regarding the first problem, for the uniform design density, Cox [7] computes the Green's function for (1.1) with periodic boundary conditions by means of Fourier series, and then fixes the natural boundary conditions (for  $m = 2$ ). Messer and Goldstein [23] determine the Green's function for (1.1) on the line by means of Fourier transform methods, and then fix the natural boundary conditions on the finite interval. In Section 3, we give the details of this construction.

For "arbitrary," smooth design densities  $w$ , Nychka [24] for  $m = 1$  and Chiang et al. [6] and Abramovich and Grinshtein [1] for  $m = 2$ , use the venerable WKB method, although only the latter explicitly mention it. The WKB method applies to the boundary value problem,

$$(1.39) \quad \begin{aligned} &(-h^2)^m u^{(2m)} + wu = v, \quad \text{on the line,} \\ &u^{(k)}(x) \longrightarrow 0 \quad \text{for } x \longrightarrow \pm\infty, \quad \text{for } k = m, \dots, 2m-1, \end{aligned}$$

and deals with the asymptotic behavior of the solution as  $h \rightarrow 0$ . See, e.g., Mathews and Walker [21]. One drawback of this approach is that the boundary behavior of the Green's function is inaccessible, since the boundaries are pushed out to infinity. This implies that the approximations are only valid away from the boundary.

Regarding the error made when approximating the spline smoother by the equivalent kernel estimator, Nychka [24] and Chiang et al. [6], following Cox [7], employ an interesting operator equation method. First, since the spline smoother  $f^{nh}$  is linear in the data, there exist functions

$$r_{mh}(\cdot, X_i | \mathbf{X}_n) \in W^{m,2}(0,1), \quad i = 1, 2, \dots, n,$$

such that

$$(1.40) \quad f^{nh}(x) = \sum_{i=1}^n Y_i r_{mh}(x, X_i | \mathbf{X}_n), \quad x \in [0,1].$$

Now, introduce the operator  $F_n : W^{m,2}(0,1) \rightarrow W^{m,2}(0,1)$ ,

$$(1.41) \quad [F_n g](x) = \int_0^1 \mathfrak{R}_{wmh}(x,t) g(t) (dW_n(t) - dW(t)),$$

for  $x \in [0, 1]$ . Here,  $W(t)$  is the distribution function corresponding to the pdf  $w$ , and  $W_n(t)$  is the empirical distribution function for the design  $\mathbf{X}_n$ , i.e.,

$$(1.42) \quad W(t) = \int_0^t w(s) ds, \quad W_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq t).$$

Then, they show that  $\eta = f^{nh} - \psi^{nh}$  satisfies the operator equation

$$(1.43) \quad \eta + F_n \eta = -F_n \psi^{nh}.$$

Since  $\psi^{nh}$  is given explicitly in terms of  $\mathfrak{R}_{wmh}$ , see (1.30), it is now useful to study the Neumann series representation of  $(\mathcal{I} + F_n)^{-1} F_n \mathfrak{R}_{wmh}(\cdot, t)$  to get explicit approximations to the functions  $r_{mh}(t, X_i | \mathbf{X}_n)$  above. In fact, they obtain bounds of the form

$$(1.44) \quad |r_{mh}(x, X_i | \mathbf{X}_n) - \mathfrak{R}_{wmh}(x, X_i)| \leq c \delta_n \exp(-c_1 h^{-1} |x - X_i|),$$

for suitable positive constants  $c$  and  $c_1$  and

$$\delta_n = h^{-2} \|W_n - W\|_\infty.$$

(So the error is small and decays very fast as  $|x - X_i|$  increases.) Informally, (1.43) may be obtained from the Euler equations for the problems (1.8) and (1.28).

In this paper, only the properties of the Green's function are addressed. First, we more or less explicitly compute the Green's function for the uniform density case, using Fourier methods following Messer and Goldstein [23], including the precise treatment of the natural boundary conditions. Then, we show that the Green's function for "arbitrary" designs solves a Fredholm integral equation of the second kind, with the uniform Green's function as the kernel (more or less), and we take it from there.

The accuracy of the reproducing kernel estimator is treated in [14], by directly comparing the minimization problems (1.8) and (1.28).

*Remark 2.* In a different context, there is a huge literature dealing with the case  $m = 1$ . With periodic (as opposed to natural) boundary

conditions, it is usually referred to as Hill's equation. Here, for  $h = 1$  and  $w$  square integrable, the spectral properties are known in detail. See Pöschel [25] and the references therein. The relevance of this to the convolution-kernel like properties of the Green's function for  $h \rightarrow 0$  are not clear.

In the next section, we phrase the main theorem on the Green's function, and outline the proof. The details are provided in later sections.

*Notations.* For  $1 \leq p \leq \infty$ , we let  $\|\cdot\|_p$  denote the standard norm on  $L^p(0, 1)$ . The  $L^2(0, 1)$  norm is denoted simply as  $\|\cdot\|$ . We let  $\mathcal{I}$  denote the identity operator on  $L^p(0, 1)$  (for all  $p$ ). If  $T : L^p(0, 1) \rightarrow L^p(0, 1)$ , then the operator norm of  $T$  is again denoted as  $\|T\|_p$ , the case  $p = 2$  not being an exception this time. Also,  $|\cdot|_{BV(A)}$  denotes the semi-norm on the space of functions (no equivalence classes) of bounded variation on  $A$  ( $[0, 1]$  or  $\mathbf{R}$ ). If there should be no confusion about the set  $A$  in question, we write simply  $|\cdot|_{BV}$ . See, e.g., Ziemer [29].

**2. The main theorem.** In this section, we state the main theorem of the convolution-kernel like properties of the families of kernels  $\mathfrak{R}_{wmh}$  and  $\mathfrak{R}_{wmh}^{(m)}$ , where  $\mathfrak{R}_{wmh}^{(m)}(t, s)$  denotes the  $m$ th order derivative of  $\mathfrak{R}_{wmh}(t, s)$  with respect to  $s$  (or with respect to  $t$ , because of symmetry).

**Definition 1.** A family of piecewise continuous functions  $A_h$ ,  $0 < h < 1$ , defined on  $[0, 1] \times [0, 1]$ , is kernel-like if there exists a positive constant  $C_A$  such that for all  $h$ ,  $0 < h < 1$ , and all  $t \in [0, 1]$ ,

$$\begin{aligned} \|A_h(\cdot, t)\|_1 &\leq C_A, \\ \|A_h(\cdot, t)\|_\infty &\leq C_A h^{-1}, \\ |A_h(\cdot, t)|_{BV} &\leq C_A h^{-1}, \end{aligned}$$

and there exist positive constants  $\gamma$  and  $\kappa$  such that for all  $s, t \in [0, 1]$ ,

$$|A_h(s, t)| \leq \gamma h^{-1} \exp(-\kappa h^{-1}|s - t|).$$

**Main theorem.** *Under the assumption (1.2), the families*

$$h^\ell \mathfrak{R}_{wmh}^{(\ell)}, \quad 0 < h < 1, \quad \ell = 0, 1, \dots, m,$$

*are kernel-like in the sense of Definition 1.*

How the main theorem solves the problem of obtaining uniform error bounds for smoothing splines is demonstrated in [14].

We now outline the proof of the main theorem, filling in the details in later sections. The first observation is that the case of the uniform design density is straightforward albeit lengthy, relying as it does on more or less explicit computations using Fourier analysis. So, let

$$(2.1) \quad \mathcal{R}_{mh}(t, s), \quad 0 < h < 1, \quad \text{denote the Green's function for (1.1) with the uniform design density.}$$

Of course,  $\mathcal{R}_{mh}(t, s)$  is symmetric in  $t$  and  $s$ . The following theorem is proved in Section 3.

**Theorem 1** (Messer and Goldstein [23]). *The families*

$$h^\ell \mathcal{R}_{mh}^{(\ell)}(t, s), \quad 0 < h < 1, \quad \ell = 0, 1, \dots, m,$$

*are kernel-like in the sense of Definition 1.*

Next, integral equation methods are used to get to the  $\mathfrak{R}_{wmh}$ . Write the differential equation (1.1) as

$$(2.2) \quad (-\lambda^2)^m u^{(2m)} + u = \tilde{v} - Mu,$$

where  $w_{\text{low}} = w_1/2$ ;  $\lambda = h w_{\text{low}}^{-1/(2m)}$ ;  $M$  is the multiplication operator by the function  $M$ ,

$$(2.3) \quad M(t) = (w(t) - w_{\text{low}})/w_{\text{low}},$$

and  $\tilde{v} = v/w_{\text{low}}$ . Now, if  $v \in L^2(0, 1)$  and  $u$  solves the boundary value problem (1.1), then  $u$  solves the integral equation

$$(2.4) \quad u + \mathcal{T}_\lambda M u = w_{\text{low}}^{-1} \mathcal{T}_\lambda v,$$

where  $\mathcal{T}_\lambda : L^2(0, 1) \rightarrow L^2(0, 1)$  is defined as

$$(2.5) \quad [\mathcal{T}_\lambda u](t) = \int_0^1 \mathcal{R}_{m\lambda}(t, \tau) u(\tau) d\tau, \quad t \in [0, 1].$$

Of course, the point is that (2.4) is equivalent to the boundary value problem.

**Theorem 2.** *For each  $v \in L^2(0, 1)$ , the solution  $u \in L^2(0, 1)$  of the boundary value problem (1.1) exists and is unique, and is given by the unique solution in  $L^2(0, 1)$  of (2.4). Moreover,*

$$\sup_{0 < \lambda < 1} \|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_2 \leq 1 + \frac{w_2}{w_1}.$$

The main trick is now to infer the uniform invertibility of  $\mathcal{I} + \mathcal{T}_\lambda M$  on  $L^1(0, 1)$ , after which the rest is smooth sailing.

**Theorem 3.** *Under the assumption (1.2) on the design density, there exists a constant  $C_1$ , such that for all  $p$ ,  $1 \leq p \leq \infty$ ,*

$$\sup_{0 < \lambda < 1} \|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_p \leq C_1.$$

*Proof of the main theorem.* Fix  $s \in (0, 1)$ , and set  $u = \mathfrak{R}_{wmh}(\cdot, s)$ . Then,  $u$  is the solution to

$$(2.6) \quad u + \mathcal{T}_\lambda M u = w_{\text{low}}^{-1} \mathcal{R}_{m\lambda}(\cdot, s).$$

and by Theorem 3, we get  $\|u\|_1 \leq C_1 w_{\text{low}}^{-1} \|\mathcal{R}_{m\lambda}(\cdot, s)\|_1 \leq C_2$ , for a suitable constant  $C_2$ .

Now,  $|[\mathcal{T}_\lambda M u](t)| \leq \|\mathcal{R}_{m\lambda}(t, \cdot)\|_\infty \|Mu\|_1$ , and of course,

$$\|Mu\|_1 \leq w_{\text{low}}^{-1} \|w - w_{\text{low}}\|_\infty \|u\|_1.$$

So then, from (2.6),

$$|u(t)| \leq w_{\text{low}}^{-1} |\mathcal{R}_{m\lambda}(t, s)| + \|\mathcal{R}_{m\lambda}(t, \cdot)\|_\infty \|Mu\|_1,$$

and it follows with Theorem 1 that  $\|u\|_\infty \leq c_1 \lambda^{-1} \leq c h^{-1}$ . Moreover, this holds uniformly in  $s \in [0, 1]$ .

For the  $BV$ -property, note that, for all  $f$ ,

$$|\mathcal{T}_\lambda f|_{BV} \leq \sup_{s \in [0,1]} |\mathcal{R}_{m\lambda}(\cdot, s)|_{BV} \|f\|_1,$$

so that, again from (2.6),

$$|u|_{BV} \leq w_{\text{low}}^{-1} |\mathcal{R}_{m\lambda}(\cdot, s)|_{BV} + \sup_{s \in [0,1]} |\mathcal{R}_{m\lambda}(\cdot, s)|_{BV} \|Mu\|_1,$$

and the bound  $|u|_{BV} \leq c h^{-1}$  follows. Thus, the family of Green's functions  $\mathfrak{R}_{wmh}(t, s)$ ,  $0 < h < 1$ , are kernel-like.

Now, let  $1 \leq \ell \leq m$ . After  $\ell$  times differentiating both sides of (2.5), the above derivations may be repeated to show that the kernels

$$h^\ell \mathfrak{R}_{wmh}^{(\ell)}(t, s)$$

are kernel-like as well.

Apart from the exponential decay, the main theorem has been proved.  $\square$

In the remaining sections, Theorems 1, 2, 3 and the missing part of the main theorem regarding the exponential decay are proved.

**3. The Green's function for the uniform design.** In this section, we prove Theorem 1. There is little doubt that this is all very predictable: First, we determine a *fundamental* solution of the differential equation, ignoring the boundary conditions,

$$(-h^2)^m u^{(2m)} + w u = \delta_s,$$

with  $\delta_s$  the point mass at  $s$ , using Fourier methods. Then, all *homogeneous* solutions of the differential equation are computed, and finally, the correct linear combination of the homogeneous solutions is added to the fundamental solution so as to match the (natural) boundary conditions of (1.1). This follows Cox [7] for the case  $m = 2$  (but he

constructs the Green's function for (1.1) with *periodic* boundary conditions, and then matches the *natural* boundary conditions), and Messer and Goldstein [23] (who slightly fudge the natural boundary conditions, see the remark following (3.17)). When all is said and done, this leads to the following theorem.

**Theorem 4.** *Define the function  $\mathfrak{B}_{mh}(x)$ ,  $x \in \mathbf{R}$ , by its Fourier transform*

$$\widehat{\mathfrak{B}_{mh}}(\omega) = (1 + (2\pi h\omega)^{2m})^{-1}, \quad \omega \in \mathbf{R},$$

and let

$$\varphi_{\ell,h}(x) = \begin{cases} \exp(h^{-1}\varpi_{\ell}x) & \ell = 0, 1, \dots, m-1, \\ \exp(h^{-1}\varpi_{\ell}(x-1)) & \ell = m, m+1, \dots, 2m-1, \end{cases}$$

where

$$\varpi_{\ell} = \exp\left(\frac{2\ell + m + 1}{2m} \pi i\right), \quad \ell = 0, 1, \dots, 2m-1.$$

Then, for a suitable  $h_0 > 0$  and for all  $h < h_0$ , there exist functions  $a_{\ell,h}$  and positive constants  $c_m, \kappa_m$  such that for all  $x, y \in [0, 1]$ , the Green's function  $\mathcal{R}_{mh}$  may be represented as

$$\mathcal{R}_{mh}(x, y) = \mathfrak{B}_{mh}(x - y) + \sum_{\ell=0}^{2m-1} h^{-1} \varphi_{\ell,h}(x) a_{\ell,h}(y).$$

Moreover, for all  $y \in [0, 1]$ ,

$$\begin{aligned} \sup_{0 \leq \ell \leq m-1} |a_{\ell,h}(y)| &\leq c_m \exp(-h^{-1} \kappa_m y), \\ \sup_{m \leq \ell \leq 2m-1} |a_{\ell,h}(y)| &\leq c_m \exp(-h^{-1} \kappa_m (1 - y)), \end{aligned}$$

and  $\mathcal{R}_{mh}(x, y) = \mathcal{R}_{mh}(y, x)$  for all  $x, y \in [0, 1]$ .

Before proving Theorem 4, we show how it may be used to derive Theorem 1. This requires some information regarding the functions  $\mathfrak{B}_{mh}$  and  $\varphi_{\ell,h}$ , stated in the next lemma. Notationally,  $\mathfrak{B}_{mh}^{(k)}(t)$  denotes the  $k$ th order derivative of  $\mathfrak{B}_{mh}(t)$ .

**Lemma 1.** *Let  $m \geq 1$ . There exist positive constants  $\kappa_m$  and  $c_m$  such that for  $x \in [0, 1]$ ,  $t \in \mathbf{R}$  and  $k = 0, 1, \dots, 2m - 1$ ,*

- (a)  $|\varphi_{\ell,h}^{(k)}(x)| \leq c_m h^{-k} \exp(-h^{-1} \kappa_m x), \quad 0 \leq \ell \leq m - 1,$
- (b)  $|\varphi_{\ell,h}^{(k)}(x)| \leq c_m h^{-k} \exp(h^{-1} \kappa_m (x - 1)), \quad m \leq \ell \leq 2m - 1,$
- (c)  $|\mathfrak{B}_{mh}^{(k)}(t)| \leq c_m h^{-k-1} \exp(-h^{-1} \kappa_m |t|), \quad 0 \leq k \leq 2m - 1.$
- (d)  $\sup_{h>0} \sup_{\ell} \int_0^1 h^{k-1} |\varphi_{\ell,h}^{(k)}(y)| dy < \infty,$
- (e)  $\sup_{h>0} \sup_{x \in [0,1]} \int_0^1 h^k |\mathfrak{B}_{mh}^{(k)}(x - y)| dy < \infty.$

*Proof.* Only (c) needs some attention. First, define  $\mathfrak{B}_m$  by means of

$$\widehat{\mathfrak{B}}_m(h\omega) = \widehat{\mathfrak{B}}_{mh}(\omega), \quad \omega \in \mathbf{R}.$$

Now, observe that

$$1 + (2\pi\omega)^{2m} = \prod_{\ell=1}^{2m-1} \text{factor}(\omega, \ell),$$

where

$$\begin{aligned} \text{factor}(\omega, \ell) &= 2\pi\omega - \exp\left(\frac{(2\ell+1)\pi i}{2m}\right) \\ &= 2\pi\omega - (-i) \exp\left(\frac{(2\ell+m+1)\pi i}{2m}\right) \\ &= (-i) \left(2\pi i\omega - \exp\left(\frac{(2\ell+m+1)\pi i}{2m}\right)\right) \\ &= (-i)(2\pi i\omega - \varpi_\ell), \end{aligned}$$

with  $\varpi_\ell$  as in Theorem 4. Then, the partial fraction decomposition of

$$\widehat{\mathfrak{B}}_m^{(k)}(\omega) = (2\pi i \omega)^k \widehat{\mathfrak{B}}_m(\omega) = \frac{(2\pi i \omega)^k}{1 + (2\pi\omega)^{2m}}$$

may be written as

$$\widehat{\mathfrak{B}}_{mh}^{(k)}(\omega) = \sum_{\ell=0}^{2m-1} \alpha_{\ell,k} (2\pi i\omega - \varpi_{\ell})^{-1},$$

for suitable constants  $\alpha_{\ell,k}$ . Finally, observe that  $\{2\pi i\omega - \varpi_{\ell}\}^{-1}$  is the Fourier transform of

$$-\exp(x\varpi_{\ell})\mathbf{1}(x \leq 0) \quad \text{or} \quad \exp(x\varpi_{\ell})\mathbf{1}(x \geq 0)$$

depending on whether the real part of  $\varpi_{\ell}$  is positive or negative. Note that  $\operatorname{Re} \varpi_{\ell} \neq 0$  for all  $\ell$ .  $\square$

The properties of the families  $\mathcal{R}_{mh}(x, y)$ , and  $\mathcal{R}_{mh}^{(m)}$ ,  $h > 0$ , now follow.

*Proof of Theorem 1.* Lemma 1 (a), (b), (c) imply that for  $k = 0$  and  $k = m$ ,

$$\|h^k \mathcal{R}_{mh}^{(k)}(t, \cdot)\|_{\infty} \leq c h^{-1}.$$

Lemma 1 (d), (e), together with the bounds on the  $a_{\ell,h}$ , imply for  $k = 0, m$ , that

$$\|h^k \mathcal{R}_{mh}^{(k)}(t, \cdot)\|_1 \leq c h^{-1}, \quad |h^k \mathcal{R}_{mh}^{(k)}(t, \cdot)|_{BV} \leq c h^{-1},$$

where we used that

$$|\mathcal{R}_{mh}^{(k)}(x, \cdot)|_{BV} = \|\mathcal{R}_{mh}^{(k+1)}(x, \cdot)\|_1,$$

and that  $\mathcal{R}_{mh}(x, y)$  is symmetric.

The remaining properties follow likewise.  $\square$

*Proof of Theorem 4.* The first step is to determine a fundamental solution. Consider the boundary value problem on the line

$$(3.1) \quad \begin{aligned} &(-h^2)^m u^{(2m)} + u = v \quad \text{on} \quad (-\infty, \infty), \\ &u^{(k)}(x) \longrightarrow 0 \quad \text{as} \quad |x| \rightarrow \infty, \quad k = m, m+1, \dots, 2m-1, \end{aligned}$$

with  $v \in L^2(\mathbf{R})$ . The easiest way to solve this problem is by means of Fourier transforms. Letting

$$\widehat{u}(\omega) = \int_{\mathbf{R}} u(x) e^{-2\pi i \omega x} dx,$$

one obtains

$$\widehat{u}(\omega) = \frac{\widehat{v}(\omega)}{1 + (2\pi h \omega)^{2m}}, \quad \omega \in \mathbf{R},$$

and consequently,  $u$  is given as a convolution,  $u = \mathfrak{B}_{mh} * v$ , with

$$\widehat{\mathfrak{B}_{mh}}(\omega) = (1 + (2\pi h \omega)^{2m})^{-1}.$$

It follows that  $\mathfrak{B}_{mh}(x - y)$  is the Green's function for the boundary value problem (3.2), and a fundamental solution for (1.1). The required properties of  $\mathfrak{B}_{mh}$  follow from Theorem 1 and Lemma 1.

*All homogeneous solutions.* Consider the differential equation

$$(3.2) \quad (-h^2)^m u^{(2m)} + u = 0 \quad \text{on} \quad (0, 1).$$

The homogeneous solutions are of the form  $u(x) = \exp(i \lambda x)$  for suitable constants  $\lambda$ . Substituting this into the differential equation shows that  $\lambda$  must satisfy  $(h\lambda)^{2m} + 1 = 0$ , and one verifies that the solutions are given by  $\lambda = -i h^{-1} \varpi_\ell$ ,  $0 \leq \ell \leq 2m - 1$ . This gives the homogeneous solutions

$$u_\ell(x) = \exp(h^{-1} \varpi_\ell x), \quad \ell = 0, 1, \dots, 2m - 1.$$

It is useful to scale the  $u_\ell$  such that

$$\max_{x \in [0, 1]} |u_\ell(x)| = 1,$$

with the maximum occurring at either  $x = 0$  or  $x = 1$ . This leads to the  $2m$  homogeneous solutions  $\varphi_{\ell, h}$ ,  $\ell = 0, 1, \dots, 2m - 1$ , defined in Theorem 4.

Since these solutions are obviously linearly independent, they are a basis for the set of all homogeneous solutions of the differential equation (3.3).

*Taking care of the boundary conditions.* We now construct the Green’s function as a linear combination of the fundamental solution and the basic homogeneous solutions, in the form

$$(3.3) \quad \mathcal{R}_{mh}(x, y) = \mathfrak{B}_{mh}(x - y) + \sum_{\ell=0}^{2m-1} h^{-1} \varphi_{\ell,h}(x) a_{\ell,h}(y).$$

The coefficients  $a_{\ell,h}(y)$  are to be determined such that the boundary conditions of (1.1) are satisfied. This leads to the system of linear equations

$$(6.4) \quad \sum_{\ell=0}^{2m-1} \varpi_{\ell}^k \varphi_{\ell,h}(x) a_{\ell,h}(y) = \mathfrak{B}_m^{(k)}(h^{-1}(x - y)),$$

for  $x = 0, 1$ , and  $k = m, m + 1, \dots, 2m - 1$ .

We must show that the  $a_{\ell,h}$  exist, so that  $\mathcal{R}_{mh}(x, y)$  may indeed be represented by (3.4) and that the bounds of Theorem 4 apply.

*The bounds on the  $a_{\ell,h}$ .* Note that it is reasonable to partition the system (3.5) into two blocks of equations corresponding to the boundary conditions at  $x = 0$  and at  $x = 1$ . It turns out that, for  $h \rightarrow 0$ , this partitioning amounts to an asymptotic decoupling, and two  $m \times m$  systems of equations result with coefficient matrices independent of  $h$ . The existence of the solution, as well as the bounds on them, may then be read off.

To implement this, write (3.4) in matrix vector notation, and partition the unknown  $a_{\ell,h}$  into two blocks

$$(3.5) \quad b_0 = \begin{bmatrix} a_{0,h}(y) \\ a_{1,h}(y) \\ \vdots \\ a_{m-1,h}(y) \end{bmatrix}, \quad b_1 = \begin{bmatrix} a_{m,h}(y) \\ a_{m+1,h}(y) \\ \vdots \\ a_{2m-1,h}(y) \end{bmatrix},$$

and likewise for the right-hand sides,

$$(3.6) \quad \text{rhs}_0 = \begin{bmatrix} \mathfrak{B}_m^{(m)}(-y) \\ \mathfrak{B}_m^{(m+1)}(-y) \\ \vdots \\ \mathfrak{B}_m^{(2m-1)}(-y) \end{bmatrix}, \quad \text{rhs}_1 = \begin{bmatrix} \mathfrak{B}_m^{(m)}(1 - y) \\ \mathfrak{B}_m^{(m+1)}(1 - y) \\ \vdots \\ \mathfrak{B}_m^{(2m-1)}(1 - y) \end{bmatrix}.$$

The coefficient matrix is partitioned as

$$(3.7) \quad A = \begin{bmatrix} P & R \\ S & Q \end{bmatrix},$$

with

$$(3.8) \quad P = \begin{bmatrix} P_{m,0} & P_{m,1} & \cdots & P_{m,m-1} \\ P_{m+1,0} & P_{m+1,1} & \cdots & P_{m+1,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ P_{2m-1,0} & P_{2m-1,1} & \cdots & P_{2m-1,m-1} \end{bmatrix},$$

and similarly for the other matrices, and

$$(3.9) \quad \begin{aligned} P_{k,\ell} &= \varpi_\ell^k, & \ell &= 0, 1, \dots, m-1, \\ Q_{k,\ell} &= \varpi_\ell^k, & \ell &= m, m+1, \dots, 2m-1, \\ S_{k,\ell} &= \varpi_\ell^k \exp(-h^{-1}\varpi_\ell), & \ell &= m, m+1, \dots, 2m-1, \\ R_{k,\ell} &= \varpi_\ell^k \exp(-h^{-1}\varpi_\ell), & \ell &= 0, 1, \dots, m-1, \end{aligned}$$

and  $k = m, m+1, \dots, 2m-1$ . The system (3.4) then takes the form

$$(3.10) \quad \begin{bmatrix} P & R \\ S & Q \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = - \begin{bmatrix} \text{rhs}_0 \\ \text{rhs}_1 \end{bmatrix}.$$

A careful study of this system reveals that by Lemma 3.1,

$$(3.11) \quad \begin{aligned} \|\text{rhs}_0\|_\infty &\leq c \exp(-h^{-1}\kappa_m y), \\ \|\text{rhs}_1\|_\infty &\leq c \exp(-h^{-1}\kappa_m(1-y)), \end{aligned}$$

as well as

$$\begin{aligned} \|R\|_\infty &\leq mc_m \exp(-h^{-1}\kappa_m), \\ \|S\|_\infty &\leq mc_m \exp(-h^{-1}\kappa_m), \end{aligned}$$

uniformly in  $h$ . Here,  $\|\cdot\|_\infty$  denotes the max-norm on  $\mathbf{R}^m$ , as well as the induced matrix norm on  $\mathbf{R}^{m \times m}$ .

Now, the matrices  $P$  and  $Q$ , being Vandermonde matrices, see, e.g., Atkinson [4], are nonsingular (and they do not depend on  $h$ ). It follows that for some  $h_0 > 0$  and all  $h < h_0$ , the matrix

$$(3.13) \quad B \stackrel{\text{def}}{=} \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} A = \begin{bmatrix} I & P^{-1}R \\ Q^{-1}S & I \end{bmatrix}$$

is a small perturbation of the identity and so is invertible, with a bounded inverse, uniformly in  $h$  small enough. See, e.g., Atkinson [4, Section 7.4].

The new system of equations then reads as

$$(3.14) \quad \begin{bmatrix} I & P^{-1}R \\ Q^{-1}S & I \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = - \begin{bmatrix} P^{-1}\text{rhs}_0 \\ Q^{-1}\text{rhs}_1 \end{bmatrix}.$$

and the new right-hand sides satisfy the same bounds as before. It follows that

$$(3.15) \quad \sup_{h < h_0} \|b_0\|_\infty < \infty, \quad \sup_{h < h_0} \|b_1\|_\infty < \infty.$$

Moreover, from (3.14),

$$(3.16) \quad b_0 = -P^{-1}\text{rhs}_0 - P^{-1}Rb_1.$$

Now, the bound (3.15) on  $b_1$ , and the bound (3.12) on  $R$  imply that for all  $y \in [0, 1]$ ,

$$(3.17) \quad \begin{aligned} \|b_0\|_\infty &= \mathcal{O}(\exp(h^{-1}\kappa_m y)) + \mathcal{O}(\exp(h^{-1}\kappa_m)) \\ &= \mathcal{O}(\exp(h^{-1}\kappa_m y)). \end{aligned}$$

A similar derivation applies to  $b_1$ .  $\square$

*Remark 3.* Messer and Goldstein [23] fudge the natural boundary conditions slightly, by approximating the solution of (3.14) by the right-hand side. This introduces a negligible error.

**4. Convolution-like integral operators on  $L^p$  spaces.** In this section, we prove Theorems 2 and 3.

*Proof of Theorem 2.* One verifies that the boundary value problem (1.1) constitutes the Euler equations for the problem

$$\begin{aligned} \text{minimize} \quad & \|u\|_{L^2(w)}^2 - 2\langle u, v \rangle + h^{2m} \|u^{(m)}\|^2 \\ \text{subject to} \quad & u \in W^{m,2}(0, 1). \end{aligned}$$

Thus, for each  $v \in L^2(0, 1)$ , the (weak) solution of (1.1) exists and is unique. Moreover,  $u$  satisfies

$$\|u\|_{L^2(w)}^2 + h^{2m} \|u^{(m)}\|^2 = \langle u, v \rangle \leq \|u\|_{L^2(w)} \|v\|_{L^2(1/w)},$$

so that  $\|u\|_{L^2(w)} \leq \|v\|_{L^2(1/w)}$ , and then, by the assumption (1.2) on the design density,

$$(4.1) \quad \|u\| \leq w_1^{-1} \|v\|.$$

As far as the equivalence of (1.1) and (2.4) is concerned, obviously, if  $u$  solves (1.1), then it also is a solution of (2.4). For the converse, consider (2.4) with  $v \in L^2(0, 1)$ . The solution is unique: if  $u \in L^2(0, 1)$  and  $u + \mathcal{T}_\lambda M u = 0$ , then  $u = -\mathcal{T}_\lambda M u$ , so that  $u$  satisfies

$$(-\lambda^2)^m u^{(2m)} + u = -M u$$

together with the natural boundary conditions, but this implies that

$$(-h^2)^m u^{(2m)} + w u = 0$$

and consequently, see above,  $u = 0$ . Since  $\mathcal{T}_\lambda M : L^2(0, 1) \rightarrow L^2(0, 1)$  is a compact integral operator, the Fredholm alternative, see, e.g., Atkinson [5], now implies that the solution of (2.4) exists. Then, (4.1) implies that

$$\|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1} \mathcal{T}_\lambda\|_2 \leq w_{\text{low}}/w_1 = \frac{1}{2},$$

and so  $\|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1} \mathcal{T}_\lambda M\|_2 \leq (1/2)(w_2 - w_{\text{low}})/w_{\text{low}}$ , and finally, since  $(\mathcal{I} + \mathcal{T}_\lambda M)^{-1} = \mathcal{I} - (\mathcal{I} + \mathcal{T}_\lambda M)^{-1} \mathcal{T}_\lambda M$ , then

$$\|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_2 \leq 1 + (w_2 - w_{\text{low}})/(2w_{\text{low}}).$$

This is the bound of the theorem.  $\square$

*Proof of Theorem 3.* The goal is to apply Theorem 3.1 of Eggermont and Lubich [15], where the finiteness of  $\sup_{0 < \lambda < 1} \|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_\infty$  is deduced from the finiteness of  $\sup_{0 < \lambda < 1} \|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_2$ . To that end, we need to introduce classes of kernels on  $[0, 1] \times [0, 1]$ , denoted by  $\mathcal{F}(b, e)$  where  $b \in L^1(\mathbf{R})$  and  $e \in C(\mathbf{R})$ , with  $e(0) = 0$ , are given

functions. Now, we say that a function  $K$  defined on  $[0, 1] \times [0, 1]$  belongs to  $\mathcal{F}(b, e)$  if there exists an  $h$ , with  $0 < h < 1$ , such that

$$(4.2) \quad \begin{aligned} |K(t, s)| &\leq h^{-1}b(h^{-1}|t - s|), \\ \|K(t + \delta, \cdot) - K(t, \cdot)\|_1 &\leq e(h^{-1}\delta). \end{aligned}$$

Note that by Theorem 1, there exists a constant  $c$ , such that for all relevant  $t, s, \delta$  and  $h$ ,  $0 < h < 1$ ,

$$|\mathcal{R}_{m\lambda}(t, s)M(s)| \leq ch^{-1} \exp(-\kappa_m h^{-1}|t - s|),$$

as well as

$$\begin{aligned} \|\mathcal{R}_{m\lambda}(t + \delta, \cdot)M(\cdot) - \mathcal{R}_{m\lambda}(t, \cdot)M(\cdot)\|_1 &\leq c\delta\|\mathcal{R}'_{m\lambda}(\theta, \cdot)\|_1 \\ &\leq ch^{-1}|\delta|. \end{aligned}$$

Thus, the kernels  $\mathcal{R}_{m\lambda}(t, s)M(s)$  of the integral operators  $\mathcal{T}_\lambda M$  belong to a subset  $\mathcal{A}$  of the class  $\mathcal{F}(b, e)$ , with

$$b(t) = c \exp(-\kappa_m |t|) \quad \text{and} \quad e(t) = c|t|,$$

for a suitable constant  $c$ . Thus, by Theorem 3.1 of [15], now Theorem 2 implies that there exists a constant  $C_3$  such that

$$\sup_{0 < \lambda < 1} \|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_\infty \leq C_3.$$

Of course, since  $\mathcal{T}_\lambda$  and  $M$  are symmetric, then

$$\|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_\infty = \|(\mathcal{I} + M\mathcal{T}_\lambda)^{-1}\|_1,$$

and since  $(\mathcal{I} + \mathcal{T}_\lambda M)^{-1} = M^{-1}(\mathcal{I} + M\mathcal{T}_\lambda)^{-1}M$ , this gives

$$\begin{aligned} \|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1}\|_1 &\leq \|M\|_1 \|M^{-1}\|_1 \|(\mathcal{I} + M\mathcal{T}_\lambda)^{-1}\|_1 \\ &\leq 2(w_2 - w_{\text{low}})/w_{\text{low}} \cdot C_3 \\ &\leq C_4 < \infty. \end{aligned}$$

and we are done.  $\square$

**5. The decay of the Green's function.** The proof of the missing part of the main theorem regarding the exponential decay of the Green's function  $\mathfrak{R}_{umh}(t, s)$  rests on the following result. For  $\kappa > 0$ , define

$$(5.1) \quad a(t) = \lambda^{-1} \exp(-\kappa\lambda^{-1}|t|),$$

with  $\lambda$  as in (2.2). Also, recall Theorem 1, so that

$$(5.2) \quad |\mathcal{R}_{m\lambda}(t, s)| \leq c_m \lambda^{-1} \exp(-k_m \lambda^{-1}|t - s|).$$

**Lemma 2.** *For all  $0 < \kappa < k_m$ ,*

$$\int_0^1 \left| \frac{a(\tau - s)}{a(t - s)} - 1 \right| |\mathcal{R}_{m\lambda}(t, \tau)| d\tau \leq \frac{2c_m \kappa}{(k_m - \kappa)k_m}.$$

*Proof.* First, for  $|t| > |s|$ , we have

$$\begin{aligned} 0 &\leq \frac{a(s)}{a(t)} - 1 = \exp(\kappa\lambda^{-1}(|t| - |s|)) - 1 \\ &\leq \exp(\kappa\lambda^{-1}|t - s|) - 1. \end{aligned}$$

For  $|t| \leq |s|$ , one obtains likewise

$$\begin{aligned} 0 &\leq 1 - \frac{a(s)}{a(t)} \leq 1 - \exp(-\kappa\lambda^{-1}|t - s|) \\ &\leq \exp(\kappa\lambda^{-1}|t - s|) - 1. \end{aligned}$$

It follows that the integral in the lemma is bounded by

$$\begin{aligned} c_m \lambda^{-1} \int_0^1 \left( e^{\kappa\lambda^{-1}|t-\tau|} - 1 \right) e^{-k_m \lambda^{-1}|t-\tau|} d\tau \\ \leq c_m \int_{-\infty}^{\infty} \left( e^{\kappa|\tau|} - 1 \right) e^{-k_m|\tau|} d\tau. \end{aligned}$$

Now, multiply out the integrand and integrate.  $\square$

*Proof of the exponential decay of  $\mathfrak{R}_{wmh}$ .* Fix  $s \in [0, 1]$ . Then

$$v(t) = \mathfrak{R}_{wmh}(t, s)/a(t-s)$$

satisfies the equation, cf. (2.4),

$$(5.3) \quad v + T_{\lambda,a}v = b,$$

where  $b(t) = w_{\text{low}}^{-1} \mathcal{R}_{m\lambda}(t, s)/a(t-s)$ , and  $T_{\lambda,a}$  is defined by

$$[T_{\lambda,a}g](t) = \int_0^1 \frac{a(\tau-s)}{a(t-s)} \mathcal{R}_{m\lambda}(t, \tau) M(\tau) g(\tau) d\tau.$$

Note that  $b$  is bounded, uniformly in  $\lambda$ . Now, we may rewrite (5.3) as

$$v + \mathcal{T}_\lambda Mv + (T_{\lambda,a} - \mathcal{T}_\lambda)Mv = b$$

so that

$$(5.4) \quad v + \mathcal{E}v = (I + \mathcal{T}_\lambda M)^{-1}b,$$

where  $\mathcal{E} = (I + \mathcal{T}_\lambda M)^{-1}(T_{\lambda,a} - \mathcal{T}_\lambda)M$ . Now,

$$\|\mathcal{E}\|_\infty \leq \|(I + \mathcal{T}_\lambda M)^{-1}\|_\infty \|(T_{\lambda,a} - \mathcal{T}_\lambda)\|_\infty \|M\|_\infty \leq C\kappa$$

for a suitable constant  $C$ . This uses Theorem 3 with  $p = \infty$ , and Lemma 2 to bound  $\|(T_{\lambda,a} - \mathcal{T}_\lambda)\|_\infty$ . Now, choose  $\kappa \leq 1/(2C)$ . Then,  $\|\mathcal{E}\|_\infty \leq 1/2$ , so that the Banach contraction principle applied to (5.4) implies the inequality  $\|v\|_\infty \leq \text{const} \|b\|_\infty$ . Thus,  $v$  is bounded, uniformly in  $\lambda$ . But this implies the exponential decay of  $\mathfrak{R}_{wmh}(t, s)$ .  $\square$

**6. The dependence on  $h$ .** In this section, we study the dependence of the reproducing kernel  $\mathfrak{R}_{wmh}$  on the parameter  $h$ , analogous to the inequalities (1.38) for convolution kernels.

**Theorem 5.** *Under the conditions (1.2) on the design density, there exists a constant  $c$  such that for all  $h, \theta \in (0, 1)$ , and all  $p, 1 \leq p \leq \infty$ ,*

$$\sup_{s \in [0, 1]} \|\mathfrak{R}_{wmh}(\cdot, s) - \mathfrak{R}_{wm\theta}(\cdot, s)\|_p \leq ch^{-1+1/p} \left| 1 - \frac{h}{\theta} \right|.$$

*Proof.* It suffices to prove the cases  $p = 1$  and  $p = \infty$ . Let  $u = \mathfrak{R}_{wmh}(\cdot, s)$  and  $v = \mathfrak{R}_{wm\theta}(\cdot, s)$ . Let  $\lambda = hw_{\text{low}}^{-1/(2m)}$  and  $\eta = \theta w_{\text{low}}^{-1/(2m)}$ . Then, by Theorem 2, the functions  $u$  and  $v$  are the solutions to

$$(\mathcal{I} + \mathcal{T}_\lambda M)u = w_{\text{low}}^{-1} \mathcal{R}_{m\lambda}(\cdot, s)$$

and

$$(\mathcal{I} + \mathcal{T}_\eta M)v = w_{\text{low}}^{-1} \mathcal{R}_{m\eta}(\cdot, s).$$

It follows that

$$(6.1) \quad w_{\text{low}}(u - v) = \text{first} + \text{second},$$

with

$$(6.2) \quad \begin{aligned} \text{first} &= (\mathcal{I} + \mathcal{T}_\eta M)^{-1} \left\{ \mathcal{R}_{m\lambda}(\cdot, s) - \mathcal{R}_{m\eta}(\cdot, s) \right\}, \\ \text{second} &= \left\{ (\mathcal{I} + \mathcal{T}_\lambda M)^{-1} - (\mathcal{I} + \mathcal{T}_\eta M)^{-1} \right\} \mathcal{R}_{m\lambda}(\cdot, s). \end{aligned}$$

Everything is in place to bound the two terms. From the semi-explicit representation of Theorem 4, one obtains just as for convolution kernels that for all  $p$ ,  $1 \leq p \leq \infty$ ,

$$(6.3) \quad \left\| \mathcal{R}_{m\lambda}(\cdot, s) - \mathcal{R}_{m\eta}(\cdot, s) \right\|_p \leq c\lambda^{-1+1/p} \left| 1 - \frac{\lambda}{\eta} \right|.$$

and by Theorem 3, the same bound with a different constant applies to the first term.

Regarding the second term, observe that

$$(\mathcal{I} + \mathcal{T}_\lambda M)^{-1} - (\mathcal{I} + \mathcal{T}_\eta M)^{-1} = (\mathcal{I} + \mathcal{T}_\lambda M)^{-1} (\mathcal{T}_\eta - \mathcal{T}_\lambda) M (\mathcal{I} + \mathcal{T}_\eta M)^{-1},$$

so that, again with Theorem 3,

$$\|(\mathcal{I} + \mathcal{T}_\lambda M)^{-1} - (\mathcal{I} + \mathcal{T}_\eta M)^{-1}\|_p \leq c\|\mathcal{T}_\eta - \mathcal{T}_\lambda\|_p.$$

Now, since

$$\begin{aligned} \|\mathcal{T}_\lambda - \mathcal{T}_\eta\|_p &\leq c \sup_{x \in [0,1]} \|\mathcal{R}_{m\lambda}(x, \cdot) - \mathcal{R}_{m\eta}(x, \cdot)\|_1 \\ &\leq c_2 \left| 1 - \frac{\lambda}{\eta} \right|, \end{aligned}$$

then, with the bound from Theorem 4 valid for  $p = 1$ , and  $p = \infty$ ,

$$\|\mathcal{R}_{m\lambda}(x \cdot)\|_p \leq c\lambda^{-1+1/p},$$

the second term may be bounded as

$$|\text{second}| \leq c\lambda^{-1+1/p} \left|1 - \frac{\lambda}{\eta}\right|.$$

Finally, since  $\lambda/\eta = h/\theta$ , and  $h = w_{\text{low}}\lambda$ , the theorem follows.  $\square$

## REFERENCES

1. F. Abramovich and V. Grinshtein, *Derivation of equivalent kernels for general spline smoothing: a systematic approach*, Bernoulli **5** (1999), 359–379.
2. R.A. Adams and J.J.F. Fournier, *Sobolev spaces*, 2nd ed., Academic Press, Amsterdam, 2003.
3. N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.
4. K.E. Atkinson, *An introduction to numerical analysis*, John Wiley and Sons, New York, 1989.
5. ———, *The numerical solution of integral equations of the second kind*, Cambridge Univ. Press, Cambridge, 1997.
6. C. Chiang, J. Rice and C. Wu, *Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables*, J. Amer. Statist. Assoc. **96** (2001), 605–619.
7. D.D. Cox, *Asymptotics of M-type smoothing splines*, Ann. Statist. **11** (1984), 530–551.
8. ———, *Multivariate smoothing spline functions*, SIAM J. Numer. Anal. **21** (1984), 789–813.
9. P. Deheuvels and D.M. Mason, *General asymptotic confidence bands based on kernel-type function estimators*, Stat. Inference Stoch. Process. **7** (2004), 225–277.
10. L. Devroye and L. Györfi, *Density estimation: The  $L_1$ -view*, John Wiley and Sons, New York, 1985.
11. C.L. Dolph and M.A. Woodbury, *On the relation between Green's functions and covariances of certain stochastic processes and its application to unbiased linear prediction*, Trans. Amer. Math. Soc. **72** (1952), 519–550.
12. R.M. Dudley, *Real analysis and probability*, Cambridge Univ. Press, Cambridge, 2002.
13. P.P.B. Eggermont and V.N. LaRiccia, *Maximum penalized likelihood estimation*, Vol. I: *Density estimation*, Springer-Verlag, New York, 2001.
14. P.P.B. Eggermont and V.N. LaRiccia, *Uniform error bounds for smoothing splines*, manuscript, 2005.

15. P.P.B. Eggermont and Ch. Lubich, *Uniform error estimates of operational quadrature methods for nonlinear convolution equations on the half line*, Math. Comp. **56** (1991), 149–176.
16. U. Einmahl and D.M. Mason, *Uniform in bandwidth consistency of kernel-type function estimators*, Ann. Statist. **33** (2005), 1380–1403.
17. R.L. Eubank, *Spline smoothing and nonparametric regression*, Marcel Dekker, New York, 1999.
18. L. Györfi, M. Kohler, A. Krzyzak and A. Walk, *A distribution-free theory of nonparametric regression*, Springer-Verlag, New York, 2002.
19. W. Härdle, P. Janssen and R. Serfling, *Strong uniform consistency rates for estimators of conditional functionals*, Ann. Statist. **16** (1988), 1428–1449.
20. V.D. Konakov and V.I. Piterbarg, *On the convergence rate of maximal deviation distribution for kernel regression estimates*, J. Multivariate Anal. **15** (1984), 279–294.
21. J. Mathews and R.L. Walker, *Mathematical methods of physics*, Addison-Wesley, New York, 1979.
22. K. Messer, *A comparison of a spline estimate to its equivalent kernel estimate*, Ann. Statist. **19** (1991), 817–829.
23. K. Messer and L. Goldstein, *A new class of kernels for nonparametric curve estimation*, Ann. Statist. **21** (1993), 179–196.
24. D. Nychka, *Splines as local smoothers*, Ann. Statist. **23** (1995), 1175–1197.
25. J. Pöschel, *Hill's potentials in weighted Sobolev spaces and their spectral gaps*, manuscript, University of Stuttgart, 2004.
26. B.W. Silverman, *Spline smoothing: The equivalent variable kernel method*, Ann. Statist. **12** (1984), 898–916.
27. P.L. Speckman, *The asymptotic integrated mean squared error for smoothing noisy data by splines*, manuscript, University of Oregon, 1981.
28. G. Wahba, *Spline models for observational data*, SIAM, Philadelphia, 1990.
29. W.P. Ziemer, *Weakly differentiable functions*, Springer-Verlag, New York, 1989.

FOOD AND RESOURCE ECONOMICS, UNIVERSITY OF DELAWARE, NEWARK, DE  
19717-1303  
E-mail address: [eggermon@udel.edu](mailto:eggermon@udel.edu)

FOOD AND RESOURCE ECONOMICS, UNIVERSITY OF DELAWARE, NEWARK, DE  
19717-1303  
E-mail address: [lariccia@udel.edu](mailto:lariccia@udel.edu)