The Annals of Statistics
1995, Vol. 23, No. 4, 1175–1197

SPLINES AS LOCAL SMOOTHERS[1]

BY DOUGLAS NYCHKA

North Carolina State University

A smoothing spline is a nonparametric curve estimate that is defined as the solution to a minimization problem. One problem with this representation is that it obscures the fact that a spline, like most other nonparametric estimates, is a local, weighted average of the observed data. This property has been used extensively to study the limiting properties of kernel estimates and it is advantageous to apply similar techniques to spline estimates. Although equivalent kernels have been identified for a smoothing spline, these functions are either not accurate enough for asymptotic approximations or are restricted to equally spaced points. This paper extends this previous work to understand a spline estimate's local properties. It is shown that the absolute value of the spline weight function decreases exponentially away from its center. This result is not asymptotic. The only requirement is that the empirical distribution of the observation points be sufficiently close to a continuous distribution with a strictly positive density function. These bounds are used to derive the asymptotic form for the bias and variance of a first order smoothing spline estimate. The arguments leading to this result can be easily extended to higher order splines.

1. Introduction. The basic model considered in nonparametric regression is

$$Z_j = \theta(t_j) + \varepsilon_j, \qquad 1 \le j \le n, \tag{1.1}$$

where $\mathbf{Z}$ is an observation vector depending on a smooth function θ, observation points $0 = t_1 \le t_2 \cdots \le t_n = 1$ and errors ε_j that are assumed to be independent, have mean zero and have common variance σ^2. A statistical problem posed by this model is to estimate θ without having to assume a parametric form for this function.

One class of estimators that has been studied extensively consists of weighted local averages of the observations where the weights are specified by a kernel function. For example, the kernel estimate attributed to Nadaraya (1964) and Watson (1964) has the form

$$\hat{\theta}(t) = \frac{1}{n}\sum_{j=1}^{n} w(t, t_j) Z_j, \tag{1.2}$$

Received February 1990; revised June 1994.
[1] Supported by NSF Grant DMS-87-15756.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62G15.
Key words and phrases. Nonparametric regression, adaptive smoothing, spline, equivalent kernel.

where

$$(1.3) \qquad w(t, t_j) = \frac{1}{h} K\left[\frac{t - t_j}{h}\right] \Big/ \sum_{k=1}^{n} \frac{1}{nh} K\left[\frac{t - t_k}{h}\right].$$

The kernel function K is assumed to be continuous, to be symmetric about zero and to integrate to 1. The bandwidth parameter h controls the relative weight given to observations as a function of their distance from t. It is easy to study the theoretical properties of kernel estimators because the estimate is an explicit function of the observations and the weight function w has a simple form. The formulation of the kernel estimate, however, constrains it to local operation on the data. This restriction may make it difficult to apply kernel methods to more complicated observational models than (1.1).

An alternative to kernel estimators are those based on maximizing a penalized likelihood. For example, under the assumption that the distribution of the errors may be approximated by a normal distribution, one has the log likelihood

$$(1.4) \qquad L(\mathbf{Z}|\theta, \sigma) = -\sum_{k=1}^{n} (Z_k - \theta(t_k))^2 / 2\sigma^2 - n \ln(\sqrt{2\pi\sigma}).$$

Furthermore, we will assume that

$$\theta \in W_2^m[0,1] = \{\theta : \theta, \dots, \theta^{(m-1)} \text{ absolutely continuous } \quad \theta^{(m)} \in L_2[0,1]\}$$

and will take as a roughness penalty $J(\theta) = \int_{[0,1]} (\theta^{(m)})^2 \, dt$. For $\delta > 0$ we have the penalized log likelihood

$$(1.5) \qquad L(\mathbf{Z}|\theta, \sigma) - \delta J(\theta).$$

A maximum penalized likelihood estimate is the function that maximizes (1.5) for all $\theta \in W_2^m[0,1]$. Thus to formulate an estimate for θ, all one needs to do is to specify the likelihood for the data in terms of the unknown function and subtract a roughness penalty. This approach has the advantage of being very flexible in adapting to more complicated models. See Nychka, Wahba, Pugh and Goldfarb (1984), O'Sullivan, Yandell and Raynor (1986), Silverman (1982), Nychka and Ruppert (1991) and Cole and Green (1992) for some examples of applications of splines to observational models that differ from (1.1).

The maximum penalized likelihood estimate just described actually takes the form of an mth order smoothing spline. Multiplying (1.5) by $-1/n$ and setting $\lambda = \delta/n$, one may define the maximum penalized likelihood estimate $\hat{\theta}_\lambda$ as the minimizer of

$$(1.6) \qquad \mathscr{L}(\xi) = \frac{1}{n} \sum_{k=1}^{n} (Z_k - \xi(t_k))^2 + \lambda \int_{[0,1]} (\xi^{(m)})^2 \, dt$$

over all $\xi \in W_2^m[0,1]$. Although the resulting estimate can be shown to be a linear function of the observations and, thus, can be written in the same form as (1.2), the spline weight function does not have a closed form.

The works of Silverman (1984), Messer (1991) and Messer and Goldstein (1993) identify kernels that will approximate the spline weight function. However, these results are limited. Although Silverman's approximation provides excellent intuition about how a spline estimate weights the data relative to fairly arbitrary distribution of the observation points, his analysis is not accurate enough for establishing asymptotic properties. Messer's Fourier analysis gives a high order approximation to $w(t, \tau)$ for all $m \geq 2$, but depends on $\{t_k\}$ being equally spaced. The results in this article fall in between the work of Silverman and Messer. Unequally spaced distributions of $\{t_k\}$ are considered and the analysis is accurate enough to establish convergence rates for the estimators. Although the assumptions on $\{t_k\}$ are not as restrictive as Messer's work, the case of a purely random design, covered by Silverman, is not addressed.

The author believes that the limitations of the results in this paper are largely technical and the methods can be extended in a straightforward manner to include random designs and higher order splines.

A conjecture for the pointwise bias and variance of a smoothing spline. One motivation for studying the spline weight function is the growing interest in locally adaptive smoothing [Staniswalis (1989), Härdle and Bowman (1988), Friedman and Silverman (1989), Schucany (1989), Eubank and Speckman (1993), Brockmann, Gasser and Herrmann (1993), Vieu (1991) and Filloon (1989). Although a single choice for the smoothing parameter λ may be adequate on the average for different values of t, it is reasonable to smooth less at sharp peaks and troughs in the function and smooth more where θ is linear. In order to implement a method for variable smoothing, it is important to understand how the mean squared error of a spline estimate varies from one point to another. Although work has been done [Cox (1984)] on the asymptotic properties of spline estimates, it is restricted to global measures of accuracy. Based on the proof techniques developed by Dennis Cox, Paul Speckman and others, it is easiest to analyze the convergence of $\hat{\theta}$ to θ in norms related to integrated squared error and the Sobolev norm suggested by the form of the roughness penalty. In fact, it seems difficult to understand the pointwise accuracy using previous work except in the case when $\{t_j\}$ are equally spaced [Messer and Goldstein (1993)]. This article presents a different approach for analyzing pointwise properties of spline estimates. Although a complete analysis is beyond the scope of a single paper, it is helpful to outline what general results might be expected.

The stating point for considering the asymptotic properties of a smoothing spline is to approximate the weight function in (1.2) by a Green's function to a particular $2m$th order differential equation. This function will be denoted by $G_\lambda(t, \tau)$ and will be referred to throughout this article as the equivalent kernel. The root $\rho = \lambda^{1/2m}$ will often be used in place of λ because it plays a role analogous to a bandwidth in a kernel estimator. In drawing this connection between G_λ and w, it should be noted that there is a natural extension of the spline weight function to values other than t_j in the second argument (see Section 3). So it makes sense to refer to $w(t,\tau)$ without requiring that τ be equal to t_j. Letting f denote the limiting density function for the independent variables $\{t_k\}$, then an exact form for G_λ will depend in a complicated manner on both f and λ. In addition, G_λ is not a convolution kernel and has a different shape depending on the distance of t and τ from the endpoints. However, suppose for the moment that a simple expression for G_λ is available. One might consider the approximations

$$\begin{aligned} E\left[\hat{\theta}(t)\right] &= E\left[\frac{1}{n}\sum_{j=1}^{n} w(t,t_j)Y_j\right] = \frac{1}{n}\sum_{j=1}^{n} w(t,t_j)\theta(t_j) \\ &\approx \int_0^1 w(t,\tau)\theta(\tau)f(\tau)\,dr \approx \int_0^1 G_\lambda(t,\tau)\theta(\tau)f(\tau)\,d\tau. \end{aligned} \tag{1.7}$$

In a similar manner, one is lead to

$$\operatorname{Var}\left[\hat{\theta}_\lambda(t)\right] \approx \frac{\sigma^2}{n}\int_0^1 [G_\lambda(t,\tau)]^2 f(\tau)\,d\tau. \tag{1.8}$$

In order to study (1.7) and (1.8), it turns out that it is not necessary to know the exact form for G_λ. Under suitable restrictions on the rate that λ converges to zero, if θ has $2m$ continuous derivatives, then it is reasonable to expect

$$E\left[\hat{\theta}_\lambda(t)\right] - \theta(t) \approx \frac{(-1)^{m-1}\lambda}{f(t)}\theta^{(2m)}(t) \tag{1.9}$$

and

$$\operatorname{Var}\left[\hat{\theta}_\lambda(t)\right] \approx \frac{\sigma^2 C_m}{nf(t)}\left(\frac{f(t)}{\lambda}\right)^{1/2m} \tag{1.10}$$

for t in the *interior* of $[0,1]$. Here C_m is a constant depending only on the order of the spline. Now set $\rho(t) = (\lambda/f(t))^{1/2m}$ and one obtains from (1.9) and (1.10),

$$E\left[\hat{\theta}_\lambda(t) - \theta(t)\right]^2 \approx \rho(t)^{4m}\left[\theta^{(2m)}(t)\right]^2 + \frac{\sigma^2 C_m}{n\rho(t)f(t)}.$$

In this form, $\rho(t)$ can be interpreted as a variable bandwidth and the accuracy of $\hat{\theta}$ is comparable to a $2m$th order kernel estimator. Based on the work of Fan (1992, 1993), the pointwise mean squared error is comparable to locally weighted regression estimators. If one wanted to achieve a constant bias or mean square error across t, one would have to consider not only the curvature of θ ($\theta^{(2m)}(t)$), but also the local density of the observations ($f(t)$). This discussion is only relevant to points t in the interior of $[0,1]$. The bias of a spline estimate at the boundary may exhibit slower convergence rates, depending on the derivatives of θ at the endpoints. This effect has been identified in Rice and Rosenblatt (1983) and is also well established for kernel estimators.

Although this overview may suggest the value of the following technical work, it should be emphasized that (1.7) and (1.8) are rigorously established in this paper only for the case $m = 1$. The extension to high order splines, however, will be a straightforward exercise.

Outline. The remainder of this article will be outlined. The next section formally states the main theorems in the paper. These are a general result on the local properties of w (Theorem 2.1) and a specific application to first order ($m = 1$) splines (Theorem 2.2). Section 3 reviews Cox's representation of the spline weight function as an infinite series and Section 4 proves Theorem 2.1.

The most difficult part of this analysis is characterizing a Green's function that corresponds to a $2m$th order differential equation. Section 5 discusses these functions and derives an explicit formula for the case $m = 1$ for uniform densities ($f = 1$). Section 5 also describes how to use the Green's function for uniform densities to approximate G_λ when f is not constant. This link is perhaps the most novel part of the article. Section 6 gives a proof for Theorem 2.2, and the last section briefly discusses how this work might be extended to higher order splines and random designs for $\{t_j\}$.

2. Main results.

2.1. *Definitions and assumptions.* Much of the theoretical work in nonparametric regression hinges on being able to approximate discrete sums such as (1.2) with integrals. In doing so, it is necessary to specify a limiting distribution for the independent variables. Let F_n denote the empirical distribution function for $\{t_j\}$, $1 \le j \le n$, let F be a distribution function with a continuous and strictly positive density function f on $[0, 1]$ and let

$$D_n = \sup_{t \in [0,1]} |F_n - F|.$$

To obtain bounds on w, the analysis in this article will require that D_n be sufficiently small relative to $\rho = \lambda^{1/2m}$ and to derive asymptotic approximations to the bias and variance of a smoothing spline, it will be necessary to assume almost sure convergence of D_n to zero.

Based on the choice for F, the weight function for an mth order spline estimate can be approximated by a Green's function to a particular $2m$th order differential equation.

DEFINITION OF EQUIVALENT KERNEL. $G_\lambda(t, \tau)$ is the Green's function associated with the differential equation

$$\lambda(-1)^m \frac{d^{2m}h(t)}{dt^{2m}} + f(t)h(t) = f(t)g(t) \tag{2.1}$$

for $t \in [0, 1]$ with boundary conditions $h^{(\nu)}(0) = h^{(\nu)}(1) = 0$ for $m \le \nu \le 2m - 1$.

For those readers not familiar with Green's functions, operationally speaking, if $h(t) = \int G_\lambda(t,\tau)g(\tau)f(\tau)\,d\tau$, then h will solve (2.1). Another useful way of characterizing G_λ is the following. If h has m continuous derivatives, then

$$\int_{[0,1]} G_\lambda(t,\tau)h(\tau)f(\tau)\,d\tau + \lambda\int_{[0,1]} \frac{\partial^m G_\lambda(t,\tau)}{\partial\tau^m}\frac{\partial^m h(\tau)}{\partial\tau^m}\,d\tau = h(t).$$

This version comes up in proving Cox's series representation for the weight function. One may note that G_λ is a reproducing kernel with respect to the inner product $\langle h_1, h_2\rangle = \int h_1 h_2 f\,d\tau + \lambda\int h_1^{(m)}h_2^{(m)}\,d\tau$ on $W_2^m[0,1]$. Thus from the elementary properties of reproducing kernels, G_λ is unique and symmetric in its arguments.

The results of the main theorem depend on G_λ satisfying the following condition:

ASSUMPTION A (Exponential envelope condition). Let $\rho = \lambda^{1/2m}$. There are positive constants $\alpha, \varepsilon, K < \infty$ such that for all $t, \tau \in [0,1]$,

$$|G_\lambda(t,\tau)| \le (K/\rho)\exp(-(\alpha-\varepsilon)|t-\tau|/\rho). \tag{2.2}$$

If $(\partial/\partial t)G_\lambda$ and $(\partial^2/\partial t\,\partial\tau)G_\lambda$ exist for all $t, \tau \in [0,1]$,

$$\left|\frac{\partial}{\partial t}G_\lambda(t,\tau)\right| \le \left(\frac{K}{\rho^2}\right)\exp\left(-(\alpha+\varepsilon)\frac{|t-\tau|}{\rho}\right), \tag{2.3}$$

$$\left|\frac{\partial^2}{\partial t\,\partial\tau}G_\lambda(t,\tau)\right| \le \left(\frac{K}{\rho^3}\right)\exp\left(-(\alpha+\varepsilon)\frac{|t-\tau|}{\rho}\right). \tag{2.4}$$

Otherwise, if $(\partial/\partial t)G_\lambda$ is not continuous when $t = \tau$, then

$$\frac{\partial}{\partial t}G_\lambda(t,\tau)|_{\tau=t^-} - \frac{\partial}{\partial t}G_\lambda(t,\tau)|_{\tau=t^+} = \frac{1}{\lambda}. \tag{2.5}$$

ASSUMPTION B. $\delta_n = 2K[1/\varepsilon + 1/\alpha]D_n/\rho < 1$.

Assumption A places certain restrictions on how G_λ and its derivatives must decrease as $|t-\tau|$ increases. The peculiar condition in (2.5) is needed for the situation when $m = 1$. Also, it should be noted that the separation of the exponent into a sum of α and ε is an artificial device to simplify the conclusions in Theorem 2.1. In this article, it showed that Assumption A holds for $m = 1$ and when f is strictly positive with a uniformly continuous derivative. Also, the work of Messer and Goldstein (1993) implies that this assumption will hold for $m > 1$ when f is constant. It is believed that the proof strategy for the $m > 1$ case can be extended to verify this assumption for more general densities (see Section 7), but a rigorous proof is not included. In general, it is conjectured that Assumption A will hold for all m and a wide class of design densities. Based on this conjecture, the main theorem has been phrased in terms of this more general assumption rather than specific results for the case when $m = 1$ or when f is constant.

2.2. *The main approximation theorem.*

THEOREM 2.1. *Let* $\rho = \lambda^{1/2m}$. *Under Assumptions* A *and* B,

$$|w(t,\tau)| < \frac{K}{(1-\delta_n)\rho}\exp\left(-\alpha\frac{|t-\tau|}{\rho}\right).$$

$$|w(t,\tau) - G_\lambda(t,\tau)| < \frac{\delta_n K}{(1-\delta_n)\rho}\exp\left(-\alpha\frac{|t-\tau|}{\rho}\right)$$

and

$$\left|\frac{\partial}{\partial t}w(t,\tau)\right| < \frac{K}{(1-\delta_n)\rho^2}\exp\left(-\alpha\frac{|t-\tau|}{\rho}\right),$$

uniformly over $t, \tau \in [0,1]$.

The proof for this theorem is given in Section 4. One interesting feature of this result is that this bound is not asymptotic and holds exactly for finite sample sizes. The only requirement is that the sizes of the equivalent bandwidth ρ and D_n must be balanced such that $\delta_n < 1$. In fact, F need not even be the "true" limiting distribution. All that is needed is for F to approximate the empirical distribution of $\{t_j\}$.

2.3. *The bias and variance of a first order* ($m = 1$) *smoothing spline.* One application of Theorem 2.1 is to derive an asymptotic form for the bias and variance of a spline estimate.

THEOREM 2.2. *Assume that* $\hat{\theta}_\lambda$ *is a first order* ($m = 1$) *smoothing spline estimate and the observation points are not equally spaced. Suppose that* $\theta \in C_2[0,1]$ *and satisfies the Hölder condition* $|\theta^{(2)}(t) - \theta^{(2)}(\tau)| \le M|t-\tau|^\beta$ *for some* $\beta > 0$ *and* $M < \infty$. *Assume that* f *has a uniformly continuous derivative and* $D_n \to 0$ *as* $n \to \infty$. *Choose* $\Delta > 0$,

$$\lambda_n \sim (D_n)^{2/3}\log(n) \quad \textit{and} \quad \Lambda_n \to 0 \quad \textit{as} \quad n \to \infty. \tag{2.6}$$

Then:

(i) $$E(\hat{\theta}_\lambda(t)) - \theta(t) = -(\lambda/f(t))\theta^{(2)}(t) + o(\lambda),$$

(ii) $$\operatorname{Var}(\hat{\theta}_\lambda(t)) = \frac{\sigma^2 f(t)^{1/2}}{8n\lambda^{1/2}f(t)}(1 + o(1)),$$

uniformly for $\lambda \in [\lambda_n, \Lambda_n]$ *and* $t \in [\Delta, 1-\Delta]$ *as* $n \to \infty$.

The proof of this theorem is given in Section 6.

3. Series representation of the spline weight function. One way of representing the minimizer of (1.6) is to examine the functional equations obtained from setting the Gateaux derivative equal to zero [see, for example, Nashed (1971)]. A necessary condition for $\hat{\theta}_\lambda$ to be a solution is that $(d/d\varepsilon)\mathscr{L}(\hat{\theta}_\lambda + \varepsilon h)|_{\varepsilon=0} = 0$ for all h in a dense subset of $W_2^m[0,1]$. This will be sufficient provided that there are at least $m + 1$ distinct values in $\{t_k\}$. Explicitly, this condition is

$$\frac{2}{n}\sum_{k=1}^{n} -\big(Z_k - \hat{\theta}_\lambda(t_k)\big)h(t_k) + 2\lambda\int_{[0,1]} \hat{\theta}_\lambda^{(m)}h^{(m)}\,dt = 0.$$

One way of characterizing $w(t, t_j)$ is to note that this function is actually the smoothing spline estimate applied to the "data": $Z_k = n$, for $j = k$, and $= 0$ otherwise. (This trick is the same as "smoothing" a vector of 0's with a 1 in the jth row to recover the weights used by a kernel estimate.) Thus we have

$$\frac{1}{n}\sum_{k=1}^{n} w(t_k, t_j)h(t_k) + \lambda\int_{[0,1]} \frac{\partial^m}{\partial t^m}w(t,t_j)\frac{d^m}{dt^m}h(t)\,dt - h(t_j) = 0$$

or

$$\int_{[0,1]} w(t,t_j)h(t)\,dF_n(t) + \lambda\int_{[0,1]} \frac{\partial^m}{\partial t^m}w(t,t_j)\frac{d^m}{dt^m}h(t)\,dt = h(t_j)$$

for $1 \le j \le n$,

and for all h in a dense set of $W_2^m[0,1]$. Note the similarity of this expression with the second characterization of G_λ.

A solution to this functional equation is given by Cox (1983, 1984) and is stated as the following lemma.

LEMMA 3.1. *Define the integral operator* $\mathscr{R}_{n\lambda}: C_0[0,1] \to C_0[0,1]$ *as*

$$\mathscr{R}_{n\lambda}(h)(t) = \int_{[0,1]} G_\lambda(u,t)h(u)\,d(F - F_n)(u)$$

and take $\mathscr{R}_{n\lambda}^{\nu}$ *as the* ν*th power of this operator*:

$$w(t,t_j) - G_\lambda(t,t_j) = \mathscr{R}_{n\lambda}\big(w(\cdot,t_j)\big)(t). \tag{3.1}$$

If $\|\mathscr{R}_{n\lambda}^{\nu}(G_\lambda(\cdot,t_j))(t)\| < \alpha^\nu$ for some $\alpha < 1$, *then*

$$w(t,t_j) = G_\lambda(t,t_j) + \sum_{\nu=1}^{\infty} \mathscr{R}_{n\lambda}^{\nu}\big(G_\lambda(\cdot,t_j)\big)(t). \tag{3.2}$$

The proof for this lemma is discussed in Cox (1983). However, a quick argument for (3.1) is to make the substitution $h(t) = G_\lambda(t, t_j)$ in the equation characterizing w and use the reproducing properties of G_λ.

The leading term in (3.2), $G_\lambda(t, t_j)$, does not depend on the empirical distribution of $\{t_j\}$. The other terms in the infinite sum can be interpreted as corrections to this asymptotic expression based on the difference between F and F_n. For example, the first order correction ($\nu = 1$) can be reexpressed as $\int G_\lambda(u,t)G_\lambda(u,t_j)d(F - F_n)(u)$. Most of the asymptotic theory for splines is based on showing that these correction terms are negligible. If $\{t_j\}$ are equally spaced, then these higher order corrections may be calculated using Fourier methods, and this is one way of interpreting the analysis in Messer (1991).

Based on the series representation in (3.2), it is easy to extend $w(t,\tau)$ to arbitrary values of τ. In Lemma 3.1 just replace the occurrences of t_j by τ. Also, since G_λ is symmetric (Section 2.1), from (3.2), $w(t,\tau)$ must also be a symmetric function in t and τ.

4. Proof of Theorem 2.1. The most important property used to prove Theorem 2.1 is Assumption A: Bounds on the equivalent kernel and its partial derivatives can be expressed in terms of a double exponential kernel. The basic idea behind the proof is elementary. Under the assumption that F_n is sufficiently close to F, an inductive argument will show that bounds on $\mathscr{R}^{\nu}_{n\lambda} G_\lambda(\cdot,\tau)$ may be inferred from bounds on $G_\lambda(t,\tau)$. These individual bounds form a convergent geometric series and thus can be summed to give a bound for w. In order to carry out the induction argument, two lemmas will be used: The first is the basic device for approximating sums by integrals and the second is a bound on the convolution of two double exponential kernels.

4.1. *Preliminary lemmas.*

LEMMA 4.1. *If h has an integrable first derivative, then*

$$\left|\int_{[0,1]} h\,dF_n - \int_{[0,1]} h\,dF\right| \le \sup|F_n - F| \int_{[0,1]} |h'|\,dF.$$

PROOF. For the integral $\int_{[0,1]} h d(F_n - F)$, integrate by parts and then apply Hölder's inequality. □

LEMMA 4.2. *Let $I(t,\tau,\rho) = \int_{[0,1]} \exp(-\alpha|t-u|/\rho - \alpha'|\tau - u|/\rho)\,du$, where $\alpha, \alpha' > 0$. Let $\alpha^* = \min(\alpha,\alpha')$ and $\varepsilon = |\alpha - \alpha'| > 0$:*

$$I(t,\tau,\rho) \le (1/\varepsilon + 1/\alpha^*)\rho\exp(-\alpha^*|t-\tau|/\rho).$$

PROOF. Extending the range of integration to all of $\mathbb{R}$,

$$I(t,\tau,\rho) < \int_{\mathbb{R}} \exp(-\alpha|t-u|/\rho - \alpha'|\tau-u|/\rho)\,du.$$

First consider the case when $t < \tau$ and $\alpha < \alpha'$. The integral above naturally breaks into three pieces, and letting $\Delta = |t - \tau|$, it follows that

$$\begin{aligned} I(t,\tau,\rho) &< \frac{\rho}{(\alpha+\alpha')}\exp\left(\frac{-\alpha'\Delta}{\rho}\right) + \frac{\rho}{\varepsilon}\left[1 - \exp\left(\frac{-\varepsilon\Delta}{\rho}\right)\right]\exp\left(\frac{-\alpha\Delta}{\rho}\right) \\ &\quad + \frac{\rho}{(\alpha+\alpha')}\exp\left(\frac{-\alpha\Delta}{\rho}\right) \\ &< \frac{2\rho}{(\alpha+\alpha')}\exp\left(\frac{-\alpha^*\Delta}{\rho}\right) + \frac{\rho}{\varepsilon}\left[1 - \exp\left(\frac{-\varepsilon\Delta}{\rho}\right)\right]\exp\left(\frac{-\alpha^*\Delta}{\rho}\right). \end{aligned}$$

Note that $1/(\alpha + \alpha') < 1/(2\alpha^*)$ and the bracketed term is always positive but less than 1. Therefore, collecting terms, the conclusion of the lemma must hold for this case. The remaining three cases can be established in a similar manner. □

LEMMA 4.3. *Under Assumptions* A *and* B, *for all* $\nu \geq 0$,

$$|\mathscr{R}^{\nu}_{n\lambda}(G_{\lambda}(\cdot,\tau))(t)| < (\delta_n)^{\nu}\left(\frac{K}{\rho}\right)\exp\left(-\alpha\frac{|t-\tau|}{\rho}\right), \tag{4.1}$$

$$\left|\frac{\partial}{\partial t}\mathscr{R}^{\nu}_{n\lambda}(G_{\lambda}(\cdot,\tau))(t)\right| \leq (\delta_n)^{\nu}\left(\frac{K}{\rho^2}\right)\exp\left(-\alpha\frac{|t-\tau|}{\rho}\right), \tag{4.2}$$

uniformly for $t, \tau \in [0,1]$.

PROOF. By Assumption A, (4.1) and (4.2) both hold when $\nu = 0$. Suppose that (4.1) and (4.2) hold for some $\nu = \mu$. Then it will be shown that these inequalities must also hold for $\nu = \mu + 1$. To simplify notation, let $g_{\tau}(u) = G_{\lambda}(u,\tau)$:

$$\mathscr{R}^{\mu+1}_{n\lambda}(g_{\tau})(t) = \int_{[0,1]} G_{\lambda}(u,t)\mathscr{R}^{\mu}_{n\lambda}(g_{\tau})(u)\,d(F(u) - F_n(u)).$$

Applying Lemma 4.1,

$$\begin{aligned}|\mathscr{R}^{\mu+1}_{n\lambda}(g_{\tau})(t)| &\leq \sup_{u\in[0,1]}|F(u) - F_n(u)|\int_{[0,1]}\left|\frac{\partial}{\partial u}[G_{\lambda}(u,t)\mathscr{R}^{\mu}_{n\lambda}(g_{\tau})(u)]\right|du\\ &< D_n\int_{[0,1]}\left|\frac{\partial}{\partial u}G_{\lambda}(u,t)\mathscr{R}^{\mu}_{n\lambda}(g_{\tau})(u)\right|\\ &\qquad+\left|G_{\lambda}(u,t)\frac{\partial}{\partial u}\mathscr{R}^{\mu}_{n\lambda}(g_{\tau})(u)\right|du.\end{aligned}$$

Now use the bounds on G_{λ} from Assumption A and the bounds on $\mathscr{R}^{\mu}$ from the induction hypothesis:

$$\leq D_n 2(\delta_n)^{\mu}(K^2/\rho^3)I(t,\tau,\rho).$$

Now apply Lemma 4.2 with $\alpha' = \alpha + \varepsilon$:

$$\leq D_n 2(\delta_n)^{\mu}(K^2/\rho^2)(1/\varepsilon + 1/\alpha)\exp(-\alpha|t-\tau|/\rho).$$

Collecting terms,

$$\leq (\delta_n)^{\mu+1}(K/\rho)\exp(-\alpha|t-\tau|/\rho).$$

Thus (4.1) must hold for $\nu = \mu + 1$.

If the mixed partial of G_{λ} exists for $t = \tau$, then a very similar argument is used to establish the second induction hypothesis. The proof will be completed by considering the case when the mixed partial for G_λ does not exist for $t = \tau$. Integrating by parts,

$$\begin{aligned}
&\frac{\partial}{\partial t}\mathscr{R}_{n\lambda}^{\mu+1}(g_\tau)(t)\\
&\quad = \frac{\partial}{\partial t}\int_{[0,1]}\left(\frac{\partial}{\partial u}G_\lambda(u,t)\right)h(u)\,du \qquad (4.3)\\
&\qquad + \frac{\partial}{\partial t}\int_{[0,1]} G_\lambda(u,t)\frac{\partial}{\partial u}\mathscr{R}_{n\lambda}^{\mu}(g_\tau)(u)(F(u) - F_n(u))\,du,
\end{aligned}$$

where $h(u) = (F(u) - F_n(u))R_{n\lambda}^{\mu}(g_\tau)(u)$. Considering the first term,

$$\begin{aligned}
&\frac{\partial}{\partial t}\int_{[0,1]}\left(\frac{\partial}{\partial u}G_\lambda(u,t)\right)h(u)\,du \qquad (4.4)\\
&\quad = \lim_{\delta\to 0}\frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\left(\frac{\partial}{\partial u}G_\lambda(u,t-\delta) - \frac{\partial}{\partial u}G_\lambda(u,t+\delta)\right)h(u)\,du
\end{aligned}$$

because, by assumption, $(\partial/\partial u)G_\lambda(u,t)$ is continuous for $t \neq \tau$. Based on (2.5), one can argue that the limit of (4.4) is $(1/\lambda)h(t)$. Now apply the induction hypothesis for $\mathscr{R}^{\mu}$ to bound this first term. The second part of (4.3) is handled by interchanging the order of the derivative and integral and bounding this expression based on Assumption A and the induction hypothesis. □

4.2. *Proof of Theorem* 2.1. From Lemma 4.3 and Assumption B the sum of (3.2) is bounded by a convergent geometric series and thus converges uniformly to $w(t,\tau)$. Also,

$$|w(t,\tau)| \le |G_\lambda(t,\tau)| + \sum_{\nu=1}^{\infty}|\mathscr{R}_{n\lambda}^{\nu}(g_\tau)(t)| \qquad (4.5)$$

and by Lemma 4.3

$$\begin{aligned}
&\le \left[1 + \sum_{\nu=1}^{\infty}\delta_n^{\nu}\right](K/\rho)\exp(-\alpha|t-\tau|/\rho)\\
&\le [1/(1-\delta_n)](K/\rho)\exp(-\alpha'|t-\tau|/\rho).
\end{aligned}$$

The second part of the theorem follows from

$$|w(t,\tau) - G_\lambda(t,\tau)| \le [\delta_n/(1-\delta_n)](K/\rho)\exp(-\alpha'|t-\tau|/\rho).$$

Finally, the third part follows in the same manner but uses the bounds for the derivative in Lemma 4.3. □

5. Exponential bounds on the Green's function. The main result in this article depends on Assumption A. To give a concrete example, this assumption is verified directly when $m = 1$ and f is a uniform density. For $m \ge 2$ and f uniform, the work of Messer and Goldstein on approximating G_λ

is applied to verify Assumption A. This section ends by extending the results for the uniform case when $m = 1$ to nonuniform design densities.

5.1. *Approximating kernel for a first order spline and uniform design density.* When the roughness penalty depends on the first derivative ($m = 1$), it is possible to derive a fairly simple expression for the Green's function G_λ. This integral kernel will be the solution of

$$-\lambda \frac{d^2}{dt^2} G_\lambda(t,\tau) + G_\lambda(t,\tau) = 0 \quad \text{for } t \neq \tau,$$

subject to the natural boundary conditions $(d/dt)G_\lambda(0,\tau) = (d/dt)G_\lambda(1,\tau) = 0$ and continuity conditions

$$G_\lambda(t,\tau)\big|_{\tau=t^-} - G_\lambda(t,\tau)\big|_{\tau=t^+} = 0$$

and

$$\frac{d}{dt} G_\lambda(t,\tau)\big|_{\tau=t^-} - \frac{d}{dt} G_\lambda(t,\tau)\big|_{\tau=t^+} = \frac{1}{\lambda}.$$

Also note that, by symmetry, $G_\lambda(t,\tau) = G_\lambda(t,\tau)$ and $G_\lambda(t,\tau) = G_\lambda(1-t, 1-\tau)$. Working with the homogeneous solution to this ordinary differential equation, it is possible to derive

$$G_\lambda(t,\tau) = \frac{1}{2\rho(1-\exp(-1/p))}\left[\exp\left(\frac{-(\tau-t)}{\rho}\right) + \exp\left(\frac{-(t+\tau)}{\rho}\right) + \exp\left(\frac{-(2-t-\tau)}{\rho}\right) + \exp\left(\frac{-(2-\tau+t)}{\rho}\right)\right]$$

for $t \le \tau$ and $\rho = \lambda^{1/2}$. The formula for G_λ when $t > \tau$ is obtained using the symmetry properties mentioned above. One can verify that G_λ given above satisfies parts (2.2) and (2.3) of Assumption A. The mixed partial does not exist when $t = \tau$, and so it is necessary to consider (2.5). However, this property follows directly from the continuity condition specified for the construction of this Green's function.

5.2. *Green's functions for $m \ge 2$ uniform density.* Calculating the Green's function for the case when $m \ge 2$ involves much more algebra. For example, when $m = 2$ there are eight linear equations of coefficients that need to be solved rather than four. Accurate approximations to G_λ in these cases have been developed by Messer and Goldstein (1993). In particular, Theorem 4.1 proved by these authors, together with their definition of the approximating function $K_b(x,t)$, will imply that the first two parts of Assumption A will hold. The bound for the mixed partial follows on noting that one can differentiate both sides of their second equation (23) on page 193 and still obtain the same type of exponential bounds in the bandwidth.

5.3. *An extension to nonuniform design points.* The Green's functions described above are limited to the case when the marginal distribution of $\{t_k\}$

converges to a uniform distribution. In general, the Green's function associated with the differential equation in (2.1) is much more difficult to derive. An alternative approach is to use the form of the Green's function for a uniform density to approximate these more general Green's functions. Restricting attention to the case when $m = 1$, we show that Assumption A will also hold for nonuniform densities provided that f' is uniformly continuous. Before stating the formal results, it is helpful to present the change of variables that is at the heart of this approximation.

Let G_λ denote the Green's function associated with the differential equation in (2.1) with $m = 1$ and let G_λ^U denote the Green's function when f is uniform ($f \equiv 1$).

Let

$$\kappa = \int_0^1 f(\tau)^{1/2}\, d\tau,$$

$$\Gamma(t) = (1/\kappa)\int_0^t f(\tau)^{1/2}\, d\tau \quad \text{and} \quad \gamma(t) = \Gamma(t)'.$$

(Note that Γ is a 1-1 transformation onto $[0,1]$.) Consider the kernel H such that

$$G_\lambda(t,\tau)f(\tau) = H(\Gamma(t),\Gamma(\tau))\gamma(\tau). \tag{5.1}$$

$H(u,v)$ will also be a Green's function for a particular second order differential equation. Due to the particular choice of transformation, H can be approximated (as $\lambda \to 0$) by G^U_{λ/κ^2}. To understand why this is reasonable, it is informative to derive the differential equation associated with H.

For any continuous function g, let

$$h(t) = \int_0^1 G_\lambda(t,\tau)g(\tau)f(\tau)\, d\tau$$

and

$$q(u) = \int_0^1 H_\lambda(u,v)g(\Gamma^{-1}(v))\, dv$$

By applying a change of variables to the second expression and using the chain rule for derivatives, one can verify that $h(t) = q(\Gamma(t))$. Recall that from the definition of G_λ, h must solve (2.1).

Using the correspondence between h and q given above, it also follows that q must solve

$$-(\lambda)(\gamma^2\circ\Gamma^{-1})\frac{d^2}{du^2}q + \lambda(\gamma'\circ\Gamma^{-1})\frac{dq}{du} + \phi q = \phi g\circ\Gamma^{-1} \tag{5.2}$$

subject to $q'(0) = q'(1) = 0$ and $\phi(u) = f(\Gamma^{-1}(u))$.

At this point it is not clear that any simplification has occurred from this change of variables. However, dividing through (5.2) by ϕ and regrouping the terms, one obtains

$$\left[\left(\frac{\lambda}{\kappa^2}\right)\frac{d^2}{du^2}q + q\right] + \lambda\psi\frac{dq}{du} = g \circ \Gamma^{-1}, \tag{5.3}$$

where now $\psi = (\gamma' \circ \Gamma^{-1})/f(\Gamma^{-1}(u))$ and $\kappa^2\gamma^2 = f$. Let $\mathscr{D}$ denote the operator for differentiation and let $\mathscr{L} = -(\lambda/\kappa^2)\mathscr{D}^2 + I$. With this notation the differential operator associated with the Green's function H can be expressed suggestively as $\mathscr{L} + \lambda\psi\mathscr{D}$. Thus, we see that this differential operator separates into the operator for the uniform density case plus an operator of lower order. For λ sufficiently small, one might expect the solution to (5.3) be approximated by the solution to $\mathscr{L}(q) = g \circ \Gamma^{-1}$. In other terms, the integral operator identified with $(\mathscr{L} + \lambda\psi\mathscr{D})^{-1}$ might be approximated by the integral operator associated with $\mathscr{L}^{-1}$ as $\lambda \to 0$. Note that we already have a link between G_λ and H_λ by (5.1), so if G^U_{λ/κ^2} can give an adequate approximation to H_λ, we are done.

The following theorem makes this heuristic discussion precise.

THEOREM 5.1. *Let f be a strictly positive density function with a uniformly continuous derivative and let G_λ be the Green's function associated with the differential equation* (2.1) *for $m = 1$. Let G^U_λ be the Green's function for $m = 1$ when f is constant. Let $\rho = \lambda^{1/2}$, $\xi = \sup|\psi|$ and $\omega = 2K[1/\alpha + 1/\varepsilon]\zeta\rho$, where K, ε and α are associated with the constants in Assumption* A *for G^U_{λ/κ^2}. If $\omega < 1$, then*:

(i) *G_λ will satisfy Assumption* A.

(ii) *There are constants $0 < C_1, C_2 < \infty$, such that*

$$\left|G_\lambda(u,v) - \frac{G^U_{\lambda/\kappa^2}(\Gamma(u),\Gamma(v))\gamma(v)}{f(v)}\right| < C_1 \exp\left(-C_2\frac{|u-v|}{\rho}\right)$$

uniformly for $u, v \in [0,1]$.

(iii) *For $\Delta > 0$ and $t \in [\Delta, 1-\Delta]$,*

$$\int_0^1 G_\lambda(t,\tau)^2 f(\tau)\,d\tau = \frac{1}{8\lambda^{1/2}f(t)^{1/2}}(1 + o(1))$$

uniformly as $\lambda^{1/2} \to 0$.

An outline of the proof of this theorem is in Appendix A.

6. Proof of the asymptotic form for the bias and variance. This section proves Theorem 2.2. Although the conclusion of this theorem is specific to first order splines, most of the proof does not depend on this restriction. For this reason, most of the discussion in this section considers smoothing splines of general order m.

First a lemma will be given that will be needed in the proof.

LEMMA 6.1. *Let G_λ be the Green's function defined in Section 2. Suppose that*

$$\left|\theta^{(2m)}(t) - \theta^{(2m)}(\tau)\right| \le M|t - \tau|^{\beta} \quad \text{for some } \beta > 0 \text{ and } M < \infty, \tag{6.1}$$

and furthermore θ satisfies $\theta^{(\nu)}(0) = \theta^{(\nu)}(1) = 0$ for $m \le \nu \le 2m - 1$. Then

$$\frac{d^{2m}}{dt^{2m}} \int_{[0,1]} G_\lambda(t, \tau)\theta(\tau)\, dF(\tau) \to (-1)^m \theta^{(2m)}(t) \quad \text{as } \lambda \to 0,$$

uniformly for $t \in [0, 1]$.

The proof is given in Appendix B.

PROOF OF THEOREM 2.2. (i) First consider the approximation to the pointwise bias of a smoothing spline:

$$\begin{aligned} E\hat{\theta}_\lambda(t) - \theta(t) &= \frac{1}{n}\sum_{j=1}^{n} w(t, t_j)\theta(t_j) - \theta(t) \\ &= \int_{[0,1]} w(t, \tau)\theta(\tau)\, dF_n(\tau) - \theta(t) \\ &= \int_{[0,1]} w(t, \tau)\theta(\tau)\, dF(\tau) - \theta(t) \\ &\quad + \int_{[0,1]} w(t, r)\theta(\tau) d(F_n - F)(\tau) \\ &= \int_{[0,1]} G_\lambda(t, \tau)\theta(\tau)\, dF(\tau) - \theta(t) \\ &\quad + \int_{[0,1]} [w(t, \tau) - G_\lambda(t, \tau)]\theta(\tau)\, dF(\tau) \\ &\quad + \int_{[0,1]} w(t, \tau)\theta(\tau) d(F_n - F)(\tau). \end{aligned} \tag{6.2}$$

It will be convenient to refer to the three terms of (6.2) as $b_\lambda(t) + R_1 + R_2$. The proof of part (i) of Theorem 2.2 will consist in showing that $b_\lambda(t)$ converges to the functional form stated in the theorem while R_1 and R_2 are $o(\lambda)$ uniformly over the specified ranges for λ and t.

For the first term in (4.5), assume that θ satisfies the boundary conditions in (2.1) and let $g(t) = \int_{[0,1]} G_\lambda(t,\tau)\theta(\tau)\,dF(\tau)$. By the definition of the Green's function,

$$(-1)^m \lambda \left(\frac{d^{2m}}{dt^{2m}} \right) g + fg = f\theta, \tag{6.3}$$

so

$$b_\lambda(t) = g(t) - \theta(t) = \frac{-\lambda(-1)^m}{f(t)} \left(\frac{d^{2m}}{dt^{2m}} \right) g.$$

From Lemma 6.1, $g^{2m} \to \theta^{(2m)}$ as $\lambda \to 0$ and it now follows that

$$b_\lambda(t) = \frac{(-1)^{m-1} \lambda \theta^{(2m)}(t)}{f(t)} (1 + o(1)) \tag{6.4}$$

uniformly for $t \in [0,1]$ as $\lambda \to 0$.

We now deal with the case when θ does not satisfy the boundary conditions. Let $\bar{\theta}$ denote a function that agrees with θ on $[\Delta/2, 1 - \Delta/2]$, but has been modified outside this interval so as to satisfy the natural boundary conditions:

$$g(t) = \int_{[0,1]} G_\lambda(t,\tau) \bar{\theta}(\tau)\, dF(\tau) + \int_{[0,1]} G_\lambda(t,\tau)\big(\theta(\tau) - \bar{\theta}(\tau)\big)\, dF(\tau).$$

Using the exponential bounds on G_λ and $\sup|\theta - \bar{\theta}| < \infty$, the second term will be $O(\exp(-\alpha\Delta/2\rho)/\rho) = o(\lambda)$ uniformly for $t \in [\Delta, 1 - \Delta]$ as $\lambda \to \infty$. Equation (6.4) now holds for this version of θ and for t in the subinterval $[\Delta, 1 - \Delta]$.

Now consider the first remainder term in (6.2). From the second assertion of Theorem 2.1,

$$\begin{aligned} |R_1| &< \sup_{[0,1]} |\theta(\tau) f(\tau)| \int_{[0,1]} |w(t,\tau) - G_\lambda(t,\tau)|\, d\tau \\ &< \sup_{[0,1]} |\theta(\tau) f(\tau)| \left[\frac{\delta_n}{1 - \delta_n} \right] \int_{[0,1]} \left(\frac{K}{\rho} \right) \exp\left(-\alpha \frac{|t - \tau|}{\rho} \right) d\tau \\ &= O(\delta_n) = O\left(\frac{D_n}{\rho} \right) \quad \text{as } n \to \infty. \end{aligned}$$

The last remainder term may be bounded using Lemma 4.1, Theorem 2.1 and the symmetry of w:

$$|R_2| < D_n \int_{[0,1]} \left| \frac{d}{d\tau} (w(t,\tau)\theta(\tau)) \right| d\tau = O\left(\frac{D_n}{\rho} \right) \quad \text{as } n \to \infty.$$

Finally, note that by the specification of λ_n in the hypothesis of the theorem,

$$\frac{(D_n / \lambda_n^{1/2m})}{b_{\lambda_n}(t)} = O\left(\frac{D_n}{\lambda_n^{(2m-1)/2m}} \right) = O\big(\log(n)^{-3/2}\big) \to 0 \quad \text{as } n \to \infty \tag{6.5}$$

This insures that the first term dominates the second and third terms in (6.2).

(ii) Expand the variance as

$$
\begin{aligned}
\operatorname{Var}\big(\hat{\theta}_{n\lambda}(t)\big) &= \sigma^2/n \int_{[0,1]} w(t,\tau)^2\, dF_n(\tau) \\
&= \sigma^2/n \Big\{ \int_{[0,1]} G_\lambda(t,\tau)^2 f(\tau)\, d\tau \\
&\qquad + \int_{[0,1]} \big[w(t,\tau)^2 - G_\lambda(t,\tau)^2 \big] f(\tau)\, d\tau \\
&\qquad\qquad + \int_{[0,1]} w(t,\tau)^2 d(F_n - F)(\tau) \Big\}.
\end{aligned} \tag{6.6}
$$

The first term in the above expression is asymptotically equivalent to $1/(8n\lambda^{1/2}f(t)^{1/2})$ when $m = 1$ by Theorem 5.1(iii). Now use the elementary identity $|a^2 - b^2| \le 2|a - b|\max(|a|, |b|)$ to bound the second term. With this bound one can use arguments similar to those for R_1 to show that this second term is $O(D_n/\lambda^{1/2m})$. The third term can be bounded using the same techniques to analyze R_2 and is also $O(D_n/\lambda^{1/2m})$. Thus, combining the asymptotic bounds for these three terms of (6.6), part (ii) now follows. □

7. Extensions to higher order splines, the variance of the spline estimate and random designs, and unequal weights. The arguments in this article have been structured so that the extension to higher order splines depends on Assumption A. This assumption in turn depends on establishing a version of Theorem 5.1 for $m > 1$, and a brief description will be given of this general theorem. For $m > 1$, the correct choice for Γ in Section 5 is to replace the exponent 1/2 by $1/2m$. Also, it will be necessary to assume that f has $2m - 1$ continuous derivatives along with some boundary conditions that guarantee that q will satisfy the natural boundary conditions of (2.1). Like the case for $m = 1$, a change of variables based on Γ will yield a differential operator with $(\lambda/\kappa^{2m})(-1)^m\mathscr{D}^{2m}$ as the first term plus a differential operator of lower order $(2m - 1)$. This second operator will be more complicated but can be handled using the same approach in the proof for the first order case. The proof will require that the Green's function for the uniform density satisfy more conditions than just Assumption A. One will need to use exponential bounds similar to those at (2.3) for the partials

$$
\left(\frac{\partial^\nu G_\lambda}{\partial t^\nu} \right), \qquad 1 \le \nu \le (2m - 1)
$$

and the bound in (2.5) for $(\partial^{2m}/\partial t^{(2m-1)}\partial\tau)(G_\lambda(t,\tau))$. However, these properties are guaranteed from the analysis in Messer and Goldstein (1993) and the continuity conditions used to construct G_λ^U. The proof of Theorem 5.1(iii) does not need to be altered because Messer and Goldstein's approximations to the uniform Green's function can serve in place of an explicit formula for G_λ^U.

The careful reader may have noticed a defect in the sharpness of Theorem 2.2 when $\{t_k\}$ are independent samples from a probability distribution (a random design). The limits for λ do not include the optimal rate of convergence for the mean squared error. For a random design $D_n = O(\log(n)/\sqrt{n})$ [Serfling (1980)] and so $\lambda_n \approx n^{-1/3}\log(n)^{4/3}$. However, to achieve the optimal convergence rate for mean squared error, one must have $\lambda \approx n^{-2/5}$. Therefore, the interval $[\lambda_n, \Lambda_n]$ will not contain a sequence of smoothing parameters that yield the optimal rate of convergence. The way around this problem is to sharpen the bound on the approximation error for the first order term in the expansion (3.2). (The bounds for the higher terms are already adequate.) The first order term should be analyzed using a uniform (with respect to λ, t and τ) strong law of large numbers rather than using integration by parts. With this different approach it is believed that one can obtain sharper bounds that will extend the conclusion to include the optimal convergence rates. It should be noticed that this problem is only peculiar to the bias for very smooth functions. The bounds for the variance approximation already include the random design case.

APPENDIX A

Outline of the proof of Theorem 5.1. (i) The first step is to derive bounds on the integral kernel H_λ. Let $\mathscr{M}_\psi$ denote the multiplication operator $\mathscr{M}_\psi h = h\psi$. Then one can associate the Green's function H_λ with the kernel of the integral operator $(\mathscr{L} + \lambda\mathscr{M}_\psi\mathscr{D})^{-1}$. To simplify notation, set $\mathscr{G} = \mathscr{L}^{-1}$ and $\mathscr{A} = \lambda\mathscr{M}_\psi\mathscr{D}\mathscr{G}$. Then at least in a formal sense,

$$\begin{aligned}(\mathscr{L} + \lambda\psi D)^{-1} &= \mathscr{G}(\mathscr{I} + \mathscr{A})^{-1}\\ &= \mathscr{G}\left(\mathscr{I} + \sum_{\nu=1}^{\infty}(\mathscr{A})^{\nu}\right).\end{aligned}$$

This expansion will be justified through the following analysis of the operators $\mathscr{A}^\nu$.

Let A^ν denote the integral kernel associated with the integral operator $\mathscr{A}^\nu$. It will be argued that for all $\nu \geq 1$ and $u, v \in [0,1]$,

$$|A^\nu(u,v)| < K\omega^\nu \exp(-\alpha|u-v|/\rho). \tag{A.1}$$

Clearly, if (A.1) holds for some $\omega < 1$, then the series expansion of $\mathscr{G}(I + \mathscr{A})^{-1}$ is valid in the sense that the integral kernel associated with the partial sum operator converges to a well-defined, integrable kernel. Moreover, in view of the bounds on G^U_{λ/κ^2}, integration can be interchanged with summation to give:

$$H_\lambda(u,v) = G^U_{\lambda/\kappa^2}(u,v) + \sum_{\nu=1}^{\infty}\int G^U_{\lambda/\kappa^2}(u,w)A^\nu(w,v)\,dw. \tag{A.2}$$

The inequality at (A.1) will be established by induction. Suppose that (A.1) holds for some $\nu = \mu$. Since $\mathscr{A}^{\mu+1} = \mathscr{A}^{\mu}(\lambda \mathscr{M}_{\psi} \mathscr{D})\mathscr{G}$, the integral kernel associated with $\mathscr{A}^{\mu+1}$ is

$$A^{\mu+1}(u,v) = \int_{[0,1]} A^{\mu}(u,w)\lambda\psi(w)\frac{\partial}{\partial w}G^{U}_{\lambda/\kappa^2}(w,v)\,dw$$

and so

$$|A^{\mu+1}(u,v)| < \int_{[0,1]}\left|A^{\mu}(u,w)\lambda\psi(w)\frac{\partial}{\partial w}G^{U}_{\lambda/\kappa^2}(w,v)\right|dw.$$

Substituting bounds for A^{μ}, G^{U}_{λ/κ^2} and ψ,

$$< \int_{[0,1]} \omega^{\mu}(\sup|\psi|)K^2 \exp(-\alpha|w-u|/\rho - (\alpha+\varepsilon)|w-v|/\rho)\,dw$$

where $\rho = \lambda^{1/2m}$.

Applying Lemma 4.2 and rearranging this expression,

$$< K\omega^{\mu}(\zeta K)[2\rho(1/\alpha + 1/\varepsilon)]\exp(-\alpha|u-v|/\rho).$$

Finally, by the definition of ω,

$$< K(\omega^{\mu+1})\exp(-\alpha|u-v|/\rho).$$

Clearly (A.1) holds when $\nu = 1$ and thus, by induction, (A.1) must hold for all ν. From the discussion above, (A.2) now follows. Using the bounds on G^{U}_{λ/κ^2} implied by Assumption A and Lemma 4.2,

$$\begin{aligned}
|H_\lambda(u,v)| &< \left(\frac{K}{\rho}\right)\exp\left(-(\alpha+\varepsilon)\frac{|u-v|}{\rho}\right) \\
&\quad + \sum_{\nu=1}^{\infty}\omega^{\nu}\int_{[0,1]}\frac{K^2}{\rho}\exp\left(-\alpha\frac{|w-u|}{\rho} - (\alpha+\varepsilon)\frac{|w-v|}{\rho}\right)dw \\
&< \left[\frac{K}{\rho} + \frac{\omega C_1}{(1-\omega)}\right]\exp\left(-\alpha\frac{|u-v|}{\rho}\right) \\
&< \frac{C_2}{\rho}\exp\left(-\alpha\frac{|u-v|}{\rho}\right),
\end{aligned} \tag{A.3}$$

where $C_1, C_2 < \infty$. Now recall the relationship between H_λ and G_λ at (5.1). By substituting $\Gamma(s)$ for u, $\Gamma(t)$ for v and multiplying both sides of the above inequality by $\gamma(t)/f(t)$, one obtains a bound on G_λ:

$$|G_\lambda(t,\tau)| \le \frac{C_2 f(\tau)}{\rho\gamma(\tau)}\exp\left(-\alpha\frac{|\Gamma(t)-\Gamma(\tau)|}{\rho}\right). \tag{A.4}$$

Using the fact that $|\Gamma(t) - \Gamma(\tau)| > (\inf\gamma)|t-\tau|$,

$$|G_\lambda(t,\tau)| \le \frac{C_3}{\rho}\exp\left(-C_4\frac{|t-\tau|}{\rho}\right), \tag{A.5}$$

for some $C_3 < \infty$ and $C_4 = \alpha \inf_{[0,1]} \gamma(u)$. Therefore, G_λ will satisfy the first bound in Assumption A. The other parts of Assumption A can be proved using similar inductive arguments applied to the partial derivatives of the kernel A.

(ii) From (A.2),

$$\text{(A.6)} \qquad \left|H_\lambda(u,v) - G^U_{\lambda/\kappa^2}(u,v)\right| < \sum_{\nu=1}^{\infty} \int \left|G^U_{\lambda/\kappa^2}(u,w) A^\nu(w,v)\right| dw,$$

and based on the same arguments leading to (A.3),

$$< C_1 \omega/(1-\omega) \exp\left(-\alpha \frac{|u-v|}{\rho}\right).$$

Now make the same substitutions detailed from part (i) to transform H_λ to G_λ.

(iii) To simplify notation, let $\bar{g}_t(\tau) = G^U_{\lambda/\kappa^2}(\Gamma(t), \Gamma(\tau))\gamma(\tau)/f(\tau)$ and $g_t(\tau) = G_\lambda(t,\tau)$:

$$\left|\int_0^1 g_t(\tau)^2 f(\tau)\, d\tau - \int_0^1 \bar{g}_t(\tau)^2 f(\tau)\, d\tau\right| \le \int_0^1 |g_t - \bar{g}_t|\ |g_t + \bar{g}_t| f(\tau)\, d\tau.$$

Both G_λ and G^U_λ satisfy Assumption A and a bound for $|g_t - \bar{g}_t|$ is given by part (ii). Substituting these double exponential bounds into the integral and applying Lemma 4.2, one can show that the integral is $O(1)$. Thus, it remains to approximate $\int_0^1 \bar{g}_t(\tau)^2 f(\tau)\, d\tau$:

$$\int_0^1 \bar{g}_t(\tau)^2 f(\tau)\, d\tau = \int_0^1 G^U_{\lambda/\kappa^2}(\Gamma(t), \Gamma(\tau))\gamma(\tau)^2/f(\tau)\, d\tau.$$

Now make the substitutions $v = \Gamma(\tau)$ and $u = \Gamma(t)$:

$$= \int_0^1 G^u_{\lambda/\kappa^2}(u,v)\,\omega(v)\, dv,$$

where $\omega(v) = \gamma(\Gamma^{-1}(v))/f(\Gamma^{-1}(v))$.

Examining the form for the uniform Green's function in Section 5.1, it is clear that for $u \in [\delta, (1-\delta)]$, $\delta > 0$,

$$G^u_{\lambda/\kappa^2} = \frac{\kappa}{2\rho} \exp\left(-\frac{|u-v|\kappa}{\rho}\right) + O\left(\frac{\exp(\kappa\delta/\rho)}{\rho}\right) \quad \text{as } \rho \to 0.$$

Based on this approximation,

$$\int_0^1 (\bar{g}_t(\tau))^2 f(\tau)\, d\tau = \int_0^1 \left(\frac{\kappa\omega(v)}{2\rho}\right) \frac{\kappa}{2\rho} \exp\left(\frac{-|u-v|\kappa}{\rho}\right) dv + o(\rho).$$

The integral is the expected value of $\kappa\omega(v)/2\rho$ with respect to a double exponential distribution. As $\rho \to 0$, probability is concentrated at u and so

$$= \kappa \frac{\omega(u)}{2\rho} + o(\rho).$$

Substituting in the values for ω and u, the theorem now follows. □

APPENDIX B

PROOF OF LEMMA 6.1. From the discussion in Section 2 of Cox (1988) there exists a basis $\{\phi_\nu\}$ for $W_2^m[0,1]$ such that

$$\int_{[0,1]} \phi_\nu \phi_\mu \, dF = \delta_{\nu\mu}$$

and

$$\int_{[0,1]} \phi_\nu^{(m)} \phi_\mu^{(m)} \, dt = \gamma_\nu \delta \nu\mu, \qquad 0 \le \gamma_1 \le \gamma_2 \cdots,$$

where $\delta_{\nu\mu}$ is Kronecker's delta. Moreover,

$$G_\lambda(t,\tau) = \sum_{\nu=1}^{\infty} \frac{\phi_\nu(t)\phi_\nu(\tau)}{1+\lambda\gamma_\nu} \tag{B.1}$$

and

$$\frac{\phi_\nu(t)}{(1+\lambda\gamma_\nu)} = \int_{[0,1]} G_\lambda(t,\tau)\phi_\nu(\tau)\, dF(\tau).$$

Thus, we see that ϕ_ν will satisfy the boundary conditions associated with the differential equation in (2.1). Also note that, by induction, $\phi_\nu \in C_\infty[0,1]$. Integrating parts and applying these boundary conditions,

$$\gamma_\nu \delta_{\nu\mu} = \int_{[0,1]} \phi_\nu^{(m)} \phi_\mu^{(m)} \, dt = (-1)^m \int_{[0,1\}} \phi_\nu \phi_\mu^{(2m)} \, dt.$$

Using the orthogonality relations given above and the fact that these functions are a basis for $W_2^m[0,1]$, it must follow that $\phi_\nu^{(2m)} = (-1)^m \gamma_\nu \phi_\nu$.

For $\theta \in W_2^m[0,1]$, we have the representation

$$\theta = \sum_{\nu=1}^{\infty} c_\nu \phi_\nu \quad \text{where } c_\nu = \int_{[0,1]} \theta\phi_\nu \, dF. \tag{B.2}$$

Moreover, because of the Lipschitz condition at (6.1) and the boundary conditions for θ, the partial sums of this series expansion will actually converge to θ in a norm that is stronger than the one for $C_{2m}[0,1]$. The reader is referred to Cox [(1988), Section 3] for a rigorous development of these norms and the related function spaces. Specifically, with respect to the scale of norms

$$\|\theta\|_p^2 = \sum_{\nu=1}^{\infty} c_\nu^2 (1+\gamma_\nu^p),$$

the partial sums in (B.2) are convergent for all $p < 2 + 1/2m + \beta/2m$. We will also use the fact that this interpolation norm dominates supremum norm for $p > 2 + 1/2m$.

Thus, it follows that

$$\text{(B.3)} \qquad \frac{d^{2m}}{dt^{2m}}\theta(t) = (-1)^m \sum_{\nu=1}^{\infty} c_\nu \gamma_\nu \phi_\nu(t)$$

and

$$\psi(t) = \frac{d^{2m}}{dt^{2m}}\left[(-1)^m \theta(t) - \int_{[0,1]} G_\lambda(t,\tau)\theta(\tau)\, dF(\tau)\right] = \sum_{\nu=1}^{\infty} c_\nu \frac{\lambda\gamma_\nu}{(1+\lambda\gamma_\nu)}\phi_\nu(t).$$

The proof will be completed by showing that ψ converges to zero uniformly as $\lambda \to 0$.

Now choose μ such $2 + 1/2m < \mu < 2 + 1/2m + \beta/2m$. It follows that

$$\text{(B.4)} \qquad \begin{aligned} \|\psi\|^2_{\mu-2} &= \sum_{\nu=1}^{\infty} c_\nu^2 \left[\frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}\right]^2 \left(1 + \gamma_\nu^{\mu-2}\right) < \lambda^2 \sum_{\nu=1}^{\infty} c_\nu^2\left(\gamma_\nu^2 + \gamma_\nu^{\mu}\right) \\ &< \lambda^2\left(\|\theta\|_2^2 + \|\theta\|_\mu^2\right) = O(\lambda^2). \end{aligned}$$

Since $\mu - 2 > 1/2m$, there is an $M < \infty$ that does not depend on ψ such that $\sup|\psi| < M\|\psi\|^2_{\mu-2}$ and thus by (B.4), $\sup|\psi| = O(\lambda^2)$ as $\lambda \to \infty$. □

Acknowledgments. The author would like to thank M. C. Jones, Karen Messer and Dennis Cox for their encouragement and insights.

REFERENCES

BROCKMANN, M., GASSER, T. and HERRMANN, E. (1993). Locally adaptive bandwidth choice for kernal regression estimators. *J. Amer. Statist. Assoc.* **88** 1302–1309.

COLE, T. J. and GREEN, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* **11** 1305–1319.

COX. D. D. (1983). Asymptotics for *M*-type smoothing splines. *Ann. Statist.* **11** 530–551.

COX, D. D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

COX, D. D. (1988). Approximation of method of regularization estimates. *Ann. Statist.* **16** 694–712.

EUBANK, R. and SPECKMAN, P. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301.

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216.

FRIEDMAN, J. and SILVERMAN, B. W. (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics* **31** 3–21.

FILLOON, T. (1989). Improved inference with smoothing splines through local cross-validation. Ph.D. dissertation, North Carolina State Univ., Raleigh.

HÄRDLE, W. and BOWMAN, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 102–110.

MESSER, K. (1991). A comparison of a spline estimate to its equivalent kernel estimate. *Ann. Statist.* **19** 817–829.

MESSER, K. and GOLDSTEIN, L. (1993). A new class of kernels for nonparametric curve estimation. *Ann. Statist.* **21** 179–195.

NADARAYA, F. A. (1964). Estimating regression. *Theory Probab. Appl.* **9** 141–142.

NASHED, M. Z. (1971). The role of differentials. In *Nonlinear Functional Analysis* (L. B. Rall, ed.). Academic Press, New York.

NYCHKA, D. and RUPPERT, D. (1995). Nonparametric transformations for both sides of a regression model. *J. Roy. Statist. Soc. Ser. B.* To appear.

NYCHKA, D., WAHBA, G., PUGH, T. and GOLDFARB, S. (1984). Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two-dimensional cross sections. *J. Amer. Statist. Assoc.* **79** 832–846.

O'SULLIVAN, F., YANDELL, B. and RAYNOR, W. J. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.* **81** 96–104.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: Regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

SCHUCANY, W. R. (1989). Locally optimal window widths for kernel density estimation with large samples. *Statist. Probab. Lett.* **7** 401–405.

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SILVERMAN, B. W. (1982). The estimation of a probability density function by maximum penalized likelihood method. *Ann. Statist.* **10** 795–810.

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

STANISWALIS, J. G. (1989). Local bandwidth selection for kernel estimates. *J. Amer. Statist. Assoc.* **84** 284–288.

VIEU, P. (1991). Nonparametric regression: Optimal local bandwidth choice. *J. Roy. Statist. Soc. Ser. B* **53** 453–464.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
BOX 8203
RALEIGH, NORTH CAROLINA 27695-8203