The Annals of Statistics
1991, Vol. 19, No. 2, 817–829

A COMPARISON OF A SPLINE ESTIMATE TO ITS EQUIVALENT KERNEL ESTIMATE

BY K. MESSER[1]

California State University, Fullerton

It has been observed that to a smoothing spline operator there corresponds an equivalent kernel operator; these two operators have been compared in a variety of norms [Cox (1984), Silverman (1984)]. In this paper, we refine the existing bounds for the particular case of the spline estimator considered in Rice and Rosenblatt (1983) and its corresponding equivalent kernel estimator. We obtain detailed asymptotic expressions for the bias and covariance functions of the two estimates and provide rates of convergence. Direct comparison then shows that the two estimates are similar. They may differ somewhat in their higher order boundary behavior.

1. Introduction. Consider the problem of estimating the function $f(x)$ from noisy observations $y_i = f(x_i) + \varepsilon_i$, $i = 0, \dots, n$, where the errors ε_i are i.i.d. random variables with finite variance σ^2 and the design points x_i are equispaced in $[0, 1]$. Two competing approaches are given by spline and kernel smoothing.

A cubic spline estimate $\tilde{f}(x)$ of $f(x)$ is given as the minimizer over functions g of

$$\frac{1}{n} \sum_{i=0}^{n} \left(y_i - g(x_i)\right)^2 + \lambda \int_0^1 g^{(2)}(x)^2 \, dx, \tag{1.1}$$

where $g^{(j)}$ represents the jth derivative of g. The smoothing parameter λ is chosen by the user and must approach zero at an appropriate rate; it will trade the bias against the variance of the resulting estimate.

The idea of a smoothing spline dates to Whittaker (1923), the modern formulation to Schoenberg (1964) and Reinsche (1967) and the statistical analysis to Wahba (1975). For more recent work, see Rice and Rosenblatt (1983), Cox (1984) and Silverman (1984, 1985). A current bibliography may be found in Eubank (1988).

Of the many possible forms of a kernel estimate $\tilde{f}(x)$ of $f(x)$, we shall discuss

$$\tilde{f}(x) = \frac{1}{nb} \sum_{i=0}^{n} K_b(x, x_i) y_i,$$

suitable for the case of equally spaced observations, $x_i = i/n$. Here $K_b(x, t)$ is

Received August 1986; revised May 1990.
[1]Research partially supported by NSF Grant DMS-86-02018.
AMS 1980 *subject classifications*. Primary 62G05, 62G20; secondary 62J02.
Key words and phrases. Nonparametric regression, spline, kernel, boundary bias.

often a translation kernel, that is, $K_b(x,t) = k((x-t)/b)$ for some univariate function $k(t)$. More generally, we shall let $K_b(x,t) = k(x/b, t/b)$ for a bivariate function $k(x,t)$. The bandwidth b plays the same role as the smoothing parameter λ. Kernel estimates have been discussed by many authors; we refer to Rosenblatt (1971).

In this paper we shall discuss the relation between kernel and spline smoothing. A connection between the two has been established theoretically by Silverman (1984) and related approximations to a Green's function have been established by Speckman (1981) and Cox (1984). The connection between kernel and spline smoothing has also been recognized practically and exploited to provide approximate spline estimates for large scale problems [for example, see Parker and Shure (1982)].

To see how a connection between kernel and spline estimators might arise, consider the minimization problem (1.1). It is well known that the solution $\hat{f}$ is linear in the observations y_i and hence may be expressed as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=0}^{n} G_{n,\lambda}(x, x_i) y_i \tag{1.2}$$

for a weight function $G_{n,\lambda}(x, x_i)$. It has been observed that, for x_i away from the boundary, there is a fixed kernel function $k(t)$, given in (2.5), such that

$$G_{n,\lambda}(x, x_i) \approx \lambda^{-1/4} k\big((x - x_i)\lambda^{-1/4}\big)$$

and that the kernel approximation of the weight function is excellent in practice [Silverman (1984)]. In other words, for x in the interior of $[0,1]$, computationally the spline behaves much as a kernel smoother with simple translation kernel k and bandwidth $\lambda^{1/4}$. Silverman (1984) has extended this relation to x near the boundary by introducing a variable kernel $\lambda^{-1/4}K(x,t,\lambda)$, as well as to the more general case of data distributed according to density $\rho(x)$. In this latter case, which is beyond the scope of this paper, the proper bandwidth is $\lambda^{1/4}\rho^{-1/4}(x)$.

To gain more insight into the equivalent variable kernel, we contrast it with the Green's function approximation first introduced by Speckman (1981) and Cox (1984). Consider the mean of the spline estimate, $E[\hat{f}(x)]$. It follows from (1.2) that $E[\hat{f}(x)]$ solves problem (1.1) with $f(x_i)$ replacing y_i. If we now approximate the sum in (1.1) by an integral, we are led to

$$\min_g \int (f(x) - g(x))^2\,dx + \lambda \int g^{(2)}(x)^2\,dx.$$

The Euler equations for this are

$$\begin{gathered} \lambda g^{(4)}(x) + g(x) = f(x), \\ g^{(2)}(0) = g^{(2)}(1) = g^{(3)}(0) = g^{(3)}(1) = 0. \end{gathered} \tag{1.3}$$

The preceding system is self-adjoint, with solution $h(x)$ given by integration against a unique Green's function $G_\lambda(x, t)$:

$$h(x) = \int G_\lambda(x, t) f(t)\, dt.$$

It may be seen that $G_\lambda(x, t)$ is approximately of the form $\lambda^{-1/4}k((x - x_i)\lambda^{-1/4})$ for x in the interior of $[0, 1]$. However, in general, construction of $G_\lambda(x, t)$ requires the solution of a system of linear equations for each value of λ and the system becomes ill-conditioned as $\lambda \to 0$. For details, see Speckman (1981) or Coddington and Levinson (1955). In particular, $G_\lambda(x, t)$ does not scale, as does a variable kernel.

From both a theoretical and a practical point of view, then, a natural question is how close a spline estimate is to a particular associated kernel estimate. Silverman (1984) has established bounds for the difference between the weight function $G_{n,\lambda}(x, x_i)$ of the spline estimate and an appropriate kernel K, in sup norm, for a general smooth design. In this general setting, however, the bound is not sharp enough for the present purposes. For example, consider the case of uniformly spaced observations and $f \in C^4[0, 1]$, the space of functions on $[0, 1]$ with four continuous derivatives, and suppose that $n^2\lambda \to \infty$ as $n \to \infty$, $\lambda \to 0$. Then, exploiting the symmetry of the weight function, it follows from Silverman (1984), Proposition 2 that

$$\lim \left| G_{n,\lambda}(x, x_i) - \lambda^{-1/4}K\big((x - x_i)\lambda^{-1/4}\big) \right| < c$$

uniformly in x, x_i for x, x_i away from the boundary of $[0, 1]$. [With a more complicated variable kernel, as given later, the bound is extended to all x_i; see Theorem B of Silverman (1984).] Such a bound does not guarantee, for example, that the difference between the bias functions of the two estimates approaches 0.

The present paper attempts to address the question, what is the practical difference between a spline estimate and its associated kernel estimate, for moderate n? It is well known, and follows from Theorems 1 and 2, that the bias of each of the two estimates is of order $O(\lambda)$ in the interior of $[0, 1]$, and if $f^{(2)}(0) \neq 0$, of order $O(\lambda^{1/2})$ near 0. Hence we should like a bound on the approximation error between the two estimates to within the previous orders.

We shall obtain such bounds in a particular setting, at the cost of some generality. We compare the bias and covariance functions of a particular cubic spline estimate, with uniformly spaced observations, to those of an equivalent kernel estimate. We shall see, for reasonable choices of λ, that both estimates converge to their respective asymptotic forms at a rate which is $O(n^{-1})$. We may then compare directly the asymptotic bias of the two estimates, and see that, to first order, they are the same. Both estimates exhibit clearly the boundary bias phenomenon described in Rice and Rosenblatt (1983); the shape of the boundary bias may differ somewhat in its higher order terms between the two. This difference may be an artifact of the particular spline estimate studied, which has been modified somewhat at the boundary. The covariance functions of the two estimates exhibit a similar relation. We shall give the kernel explicitly. In a related paper [Messer and Goldstein (1989)], we show that it is nearly the previously discussed Green's function.

Our approach is to exploit a discrete Fourier analysis introduced in Rice and Rosenblatt (1983). To that end, we replace (1.1) in the definition of the smoothing spline by

$$\frac{1}{n}\sum_{i=0}^{n}{}^{*}(y_i - g(x_i))^2 + \int_0^1 g^{(2)}(x)^2\,dx, \tag{1.4}$$

where now $\sum_{i=0}^{n}{}^{*}a_i = \frac{1}{2}(a_0 + a_n) + \sum_{i=1}^{n-1}a_i$. Note that we have averaged the two boundary observations, y_0 and y_n. This will introduce a perturbation in the higher order terms of the asymptotic expansion for the spline estimate, but leaves the qualitative behavior unchanged. The analysis does not readily extend to the case of nonuniformly spaced data; we are able to obtain these more precise results only at the cost of considerable generality. Similar results for higher order splines may be obtained analogously.

In summary, the present results support the empirically observed excellent fit between certain kernel and spline operators. Further empirical work is called for; for some data sets the equivalent kernel estimate may be computationally more convenient than the corresponding spline estimate. This will be pursued elsewhere.

The remainder of the paper is organized as follows. In Section 2 the approximating kernel is presented and discussed. Section 3 contains the main assumptions and results. Section 4 contains proofs of theorems and Section 5 contains concluding remarks.

2. The kernel estimator. Here we present the approximating kernel and define the approximating estimator. Note that the kernel depends on the bandwidth b. Let

$$k(t) = \frac{1}{2\sqrt{2}}e^{-|t|/\sqrt{2}}\left\{\cos\left(\frac{t}{\sqrt{2}}\right) + \sin\left(\frac{|t|}{\sqrt{2}}\right)\right\}. \tag{2.5}$$

Let $K_b'(x,t)$ for $x, t \in R$ be given by

$$\begin{aligned} K_b'(x,t) = k\left(\frac{x-t}{b}\right) &+ 2\sqrt{2}\left\{k^{(2)}\left(\frac{x}{b}\right)k^{(2)}\left(\frac{t}{b}\right) + k^{(2)}\left(\frac{1-x}{b}\right)k^{(2)}\left(\frac{1-t}{b}\right)\right.\\ &\left. + k^{(3)}\left(\frac{x}{b}\right)k^{(3)}\left(\frac{t}{b}\right) + k^{(3)}\left(\frac{1-x}{b}\right)k^{(3)}\left(\frac{1-t}{b}\right)\right\}. \end{aligned} \tag{2.6}$$

Notice that smoothing with a smoothing spline corresponds to smoothing with a variable kernel estimator; that is, the shape of the kernel $K_b'(x,t)$ as a function of t will depend on the point x at which $f(x)$ is to be estimated for x near the boundary. Such modified kernel estimates have been investigated by Rice (1984) and Gasser and Muller (1979) in the context of boundary bias reduction.

As a final complication, the identification (1.4) of the boundary observations for the spline induces a wrapping of the equivalent kernel around the interval [0, 1]. Hence we take as equivalent kernel

$$K_b(x,t) = K_b'(x,t) + K_b'(-1+x,t) + K_b'(1+x,t). \tag{2.7}$$

This wrapping arises from the first few terms of the Poisson summation formula. [Cf. Feller (1966), page 630; for example, see the remarks following (4.12)].

Finally, we define the equivalent kernel estimate $\tilde{f}$, with kernel K_b and bandwidth b, as

$$\tilde{f}(x) = \frac{1}{nb}\sum_{i=0}^{n}{}^{*}K_b(x,x_i)y_i, \tag{2.8}$$

where Σ^* is as in (1.4). The appropriate bandwidth is $b = \lambda^{1/4}$.

Here we have modified the usual estimate at the endpoints, as for the spline. The modification here is not necessary; it amounts to using an extended trapezoidal quadrature rule and will improve the rate of convergence of the bias of the estimate to its asymptotic form for x near the boundary.

Remark 1. The derivatives $k^{(2)}(t)$ and $k^{(3)}(t)$ decay exponentially away from the boundary. Hence in the interior of [0, 1], $K_b(x,t)$ behaves as the translation kernel $k((x-t)/b)$.

Remark 2. It can be shown that for $b = \lambda^{1/4}$,

$$\sup_{x,t}|K_b(x,t) - G_\lambda(x,t)| = O(b^{-1}e^{-1/(\sqrt{2}b)}); \tag{2.9}$$

see Messer and Goldstein (1989).

Remark 3. Comparison of $K_b(x,t)$ with the equivalent kernel in Silverman (1984), Section 5, shows asymptotic agreement at the above rate for $K_b(x,0)$ as well as for x, t in the interior of [0, 1]. The two are not in exact agreement as can be seen by expanding the latter kernel using trigonometric identities.

3. Assumptions and results. We first introduce some notation. Then we state the assumptions under which the results of this paper will be proved. Finally, we state and discuss the theorems of the paper.

We will find the following notation useful. Let $\|g\| = (\int_0^1 g(x)^2\,dx)^{1/2}$ denote the L^2 norm on [0, 1]. Let the space of functions $H^4[0,1]$ consist of all functions h on [0, 1] such that $h^{(3)}$ exists and is absolutely continuous with $\|h^{(4)}\| < \infty$. Let $\Delta h = h(1) - h(0)$ and let E be the expectation operator.

We make the following assumptions:

Assumption 1. The unknown function f satisfies $f \in H^4[0,1]$.

ASSUMPTION 2. The observation points x_i satisfy $x_i = i/n$, $i = 0, \dots, n$.

ASSUMPTION 3. The errors ε_i are independent identically distributed mean 0 random variables with finite variance σ^2.

ASSUMPTION 4. The kernel estimate $\tilde{f}$ is given by (2.8), where $K_b(x, t)$ is as in (2.7).

ASSUMPTION 5. The spline estimate $\hat{f}$ is defined to be the minimizer over functions $g \in H^4[0, 1]$ of (1.4).

ASSUMPTION 6. The bandwidth $b = \lambda^{1/4}$ and λ depends on n in such a way that $\lambda \to 0$, $\lambda n \to \infty$ as $n \to \infty$.

The following theorem bounds the approximation error for the integral approximation to the bias of the kernel estimate.

THEOREM 3.1. *Let*

$$\tilde{B}(x) = \frac{1}{b} \int_0^1 f(t) K_b(x, t)\, dt - f(x).$$

Then

$$\left|(E\tilde{f}(x) - f(x)) - \tilde{B}(x)\right| \le \left(\frac{1}{24\sqrt{2}} \frac{1}{n^2} \frac{1}{b^3} \max_t |f(t)| \right)(1 + O(b)).$$

PROOF. This follows from the standard error bound for an extended trapezoidal rule of numerical integration [e.g., see Apostol (1969), page 604] and from the simple bound $|(\partial^2/\partial t^2)K(x, t)| \le 1/2\sqrt{2}$. $\square$

REMARK 4. A better rate of convergence can be obtained by using a higher order Newton–Coates-type quadrature rule in the definition of $\tilde{f}$. The present bound may be large for small λ, as has been pointed out by a referee. For example, in Rice and Rosenblatt [(1983), Figure 1], $n = 50$, $\lambda = 10^{-6}$ and the bound is about $.373 \max |f|$. For this example, the asymptotically optimal value of λ in terms of integrated mean squared error is about $(.6\sigma^2/n)^{2/3}$. A fourth order quadrature rule might be more appropriate in this setting; we have used the trapezoidal rule for simplicity.

REMARK 5. From (2.9), we see that the kernel will inherit properties from the Green's function. Integrating by parts, we may then exploit known properties of the Green's function [viz., $G_\lambda(x, t)$ satisfies (1.3) as a function of t for fixed $x \neq t$ and has the appropriate discontinuity in the third derivative for

$x = t$; see Speckman (1981), Messer and Goldstein (1989)] to establish

$$\tilde{B}(x) = -b^4 \int b^{-1} K_b(x,t) f^{(4)}(t)\, dt + b^3 K_b(x,t) f^{(3)}(t)\Big|_{t=0}^{t=1} \tag{3.10}$$
$$- b^3 \frac{\partial K_b(x,t)}{\partial t} f^{(2)}(t)\Big|_{t=0}^{t=1} + O\big(b e^{-1/(\sqrt{2}b)}\big).$$

The two boundary terms on the right give rise to the boundary bias of the kernel estimate. The integral term is easily seen to be of order $O(b^4)$. The first boundary term is of order $O(b^3)$, unless $f^{(3)}$ satisfies the appropriate boundary conditions. The second boundary term is of order $O(b^2)$, since the differentiation introduces another factor of b^{-1}. Notice that these boundary terms drop out if f satisfies $f^{(i)}(j) = 0$ for $i = 2, 3$ and $j = 0, 1$. These are the principle boundary bias terms of the spline estimate as well.

The next theorem bounds the difference between $\tilde{B}(x)$ and the bias of the spline estimate. The second statement of the theorem compares the bias of the spline to $\tilde{B}(x)$ in the special case where the boundary bias terms drop out.

THEOREM 3.2. *Let $\tilde{B}(x)$ be as in Theorem* 1. *Then if $n\lambda \to \infty$,*

$$E\hat{f}(x) - f(x)$$
$$= \big\{\tilde{B}(x) - D_n(x)\big(-\lambda^{1/2} \Delta f^{(2)} + \sqrt{2}\lambda^{3/4}\big(f^{(3)}(0) + f^{(3)}(1)\big) + O(\lambda)\big)\big\}$$
$$\times (1 + O(1/n)),$$

where $D_n(x)$ is as in (4.18), $\|D_n\| = O(1/\sqrt{n})$ *and the bounds $O(\cdot)$ are uniform in x.*

We turn to the covariance of the smoothing spline. Let K_b^* be the kernel as in (2.7), but now with $k(\cdot)$ replaced by $k^*(\cdot) = k(\cdot) + D_n(\cdot)$ in (2.6) and D_n as in (4.18). Let $\tilde{f}^*$ be the corresponding kernel estimate. We have:

THEOREM 3.3.

$$\operatorname{cov}\big(\hat{f}(s), \hat{f}(t)\big) = \operatorname{cov}\big(\tilde{f}^*(s), \tilde{f}^*(t)\big)(1 + o(1/n)),$$

where the term $o(1/n)$ is independent of s, t, and f.

4. Proofs of theorems. The proofs of Theorems 3.2 and 3.3 rely on a Fourier analysis of the minimization problem defining the smoothing spline, as in Rice and Rosenblatt (1983). Proposition 4.1 may be essentially found in that paper.

We first establish some notation.

We shall denote the kth Fourier coefficient of g by g_k, where $g_k = \int_0^1 g(x) e^{2\pi i k x}\, dx$. We take the Fourier transform $\hat{g}$ of g to be $\int_{-\infty}^{\infty} g(x) e^{-itx}\, dx$ and the corresponding inverse transform to be $(1/2\pi)\int_{-\infty}^{\infty} \hat{g}(x) e^{itx}\, dx$. By $f * g(x)$, we mean $\int_{-\infty}^{\infty} f(t) g(x - t)\, dt$, where f may be understood to be 0 for t outside of $[0, 1]$.

Let $a_k = (2\pi ik)^{-1}$, $b_k = (2\pi ik)^{-2}$ and $r_k = (2\pi ik)^{-4}$. Then if $f \in H^4$, $f_k = \Delta fa_k - \Delta f^{(1)}b_k + \Delta f^{(2)}a_k b_k - \Delta f^{(3)}r_k + f_k^{(4)}r_k$.

Given a sequence c_k define $c_k^{[n]}$ as

$$c_k^{[n]} = \lim_{v \to \infty} \sum_{-v}^{v} c_{k+sn}$$

for $k = 0, \dots, n-1$.

We may use the Parseval relation to recast (1.4) defining the smoothing spline in terms of the Fourier coefficients of f and g. We obtain the equivalent problem

$$\min_{g \in H^4} \left| \frac{\hat{y}_0}{\sqrt{n}} - \hat{g}_0^{[n]} \right|^2 + \sum_{k=1}^{n-1} \left| \frac{\hat{y}_k}{\sqrt{n}} - \Delta g a_k^{[n]} + \Delta g^{(1)} b_k^{[n]} - \left(g_k^{(2)} b_k\right)^{[n]} \right|^2$$

$$+ \lambda \left\{ \left(\Delta g^{(1)}\right)^2 + \sum_{k \neq 0} \sum_{s} \left| g_{k+sn}^{(2)} \right|^2 + \sum_{s \neq 0} \left| g_{sn}^{(2)} \right|^2 \right\},$$

which may be minimized term by term for fixed k. Here,

$$\hat{y}_k = \frac{1}{\sqrt{n}} \sum_{j=0}^{n} {}^{*} y_j \exp(2\pi ijk/n).$$

Note that $E[\hat{y}_k] = \sqrt{n} f_k^{[n]}$, $0 \le k < n$.

After some algebra, we obtain the following expressions. The reader is referred to Rice and Rosenblatt [(1983), Lemma 2] and expression (3.30) and (3.31) for details. Note that our $r_k^{[n]}$ is their r_k and our $f_k^{(2)}$ is their m_k.

PROPOSITION 4.1. *Let $\phi(x) = E[\hat{f}](x)$. Then*

$$\phi(x) - f(x) = \Delta(\phi - f) A_3(x) - \Delta(\phi^{(1)} - f^{(1)}) A_2(x) - \Delta f^{(2)} A_1(x) + \Delta f^{(3)} A_0(x) + I(x),$$

where the functions $A_i(x)$ and $I(x)$ are as follows:

$$A_3(x) = \sum_s \sum_{k=1}^{n-1} \left\{ a_{k+sn} - \frac{r_{k+sn}}{\lambda + r_k^{[n]}} a_k^{[n]} \right\} e^{-2\pi i(k+sn)x},$$

$$A_2(x) = \sum_s \sum_{k=1}^{n-1} \left\{ b_{k+sn} - \frac{r_{k+sn}}{\lambda + r_k^{[n]}} b_k^{[n]} \right\} e^{-2\pi i(k+sn)x},$$

(4.11)

$$A_1(x) = -\sum_s \sum_{k=1}^{n-1} \left\{ (a_k b_k)^{[n]} \frac{r_{k+sn}}{\lambda + r_k^{[n]}} - a_{k+sn} b_{k+sn} \right\} e^{-2\pi i(k+sn)x},$$

$$A_0(x) = -\sum_s \sum_{k=1}^{n-1} \left\{ r_k^{[n]} \frac{r_{k+sn}}{\lambda + r_k^{[n]}} - r_{k+sn}^{[n]} \right\} e^{-2\pi i(k+sn)x},$$

$$I(x) = \sum_s \sum_{k=1}^{n-1} \left\{ \left(f^{(4)} r_k\right)^{[n]} \frac{r_{k+sn}}{\lambda + r_k^{[n]}} - f_{k+sn}^{(4)} r_{k+sn} \right\} e^{-2\pi i(k+sn)x}. \tag{4.12}$$

Further,

$$\Delta\phi - \Delta f = \left\{ \sum_{k=1}^{n-1} \frac{\left(f_k^{(2)} b_k\right)^{[n]} \bar{a}_k^{[n]}}{\lambda + r_k^{[n]}} \right\} \left\{ \sum_{k=1}^{n-1} \frac{|a_k^{[n]}|^2}{\lambda + r_k^{[n]}} \right\}^{-1} \tag{4.13}$$

and

$$\begin{aligned} &\Delta\phi^{(1)} - \Delta f^{(1)} \left\{ 1 - \left\{ 1 + \sum_{k=1}^{n-1} \frac{|b_k^{[n]}|^2}{\lambda + r_k^{[n]}} \right\}^{-1} \right\} \\ &\quad = - \left\{ \sum_{k=1}^{n-1} \frac{\left(f_k^{(2)} b_k\right)^{[n]} b_k^{[n]}}{\lambda + r_k^{[n]}} \right\} \left\{ 1 + \sum_{k=1}^{n-1} \frac{|b_k^{[n]}|^2}{\lambda + r_k^{[n]}} \right\}^{-1}. \end{aligned} \tag{4.14}$$

We now derive Theorem 3.2. The function $I(x)$ in Proposition 4.1 will give the principle contribution to the bias in the interior of $[0, 1]$; the remaining terms will be boundary bias. We first show:

PROPOSITION 4.2.

$$I(x) = -\lambda\left(\lambda^{-1/4} k_{\lambda^{1/4}} * f^{(4)}(x)\right) + \varepsilon(x),$$

where $|\varepsilon(x)| < \lambda^{-1} n^{-3} \| f^{(4)} \| (1 + o(1))$.

We may interchange the order of summation in (4.12) to obtain

$$I(x) = -\lambda \sum_{k=1}^{n-1} \frac{\left(f_k^{(4)} b_k^2\right)^{[n]}}{\lambda + r_k^{[n]}} e^{-2\pi i k x}. \tag{4.15}$$

Note that

$$\left(f_k^{(4)} b_k^2\right)^{[n]} = \frac{f_k^{(4)}}{(2\pi i k)^4} + \sum_{s \neq 0} f_{k+sn}^{(4)} (2\pi i (k + sn))^{-4}.$$

For $k < n/2$, we may bound the right-hand sum in absolute value by $\| f^{(4)} \| \varepsilon_k$, where

$$0 \le \varepsilon_k < (2\pi n)^{-4} (\zeta(4) + 2^{-4}).$$

Here $\zeta(t)$ is the Reimann zeta function. For $k > n/2$, $k = n - j$, we have similarly

$$\left| \left(f_k^{(4)} b_k^2\right)^{[n]} - f_{-j}^{(4)} \right| \le \| f^{(4)} \| \varepsilon_k.$$

In a like manner we obtain

$$r_k^{[n]} = \begin{cases} (2\pi i k)^{-4} + \varepsilon_k, & k \le n/2, \\ (2\pi i(-j))^{-4} + \varepsilon_k, & k > n/2. \end{cases}$$

Substituting into expression (4.15), we have

$$I(x) = -\lambda \sum_{-n/2}^{n/2} \frac{f_k^{(4)}}{\lambda (2\pi i k)^4 + 1} e^{-2\pi i k x} (1 + O(n^{-4})) + \delta(x), \tag{4.16}$$

where the term $O(\cdot)$ is uniform in k and where

$$|\delta(x)| < \| f^{(4)} \| \sum_{-n/2}^{n/2} \frac{\varepsilon_k (2\pi k)^4}{\lambda(2\pi k)^4 + 1}.$$

The previous expression will contribute the main approximation error in the interior of $[0,1]$; it may be easily bounded by $c\| f^{(4)} \| \lambda^{-1} n^{-3}$, where c is a constant less than 1.

It remains to evaluate the principle term in (4.16). The function $\lambda^{-1/4} k_{\lambda^{1/4}}$ has inverse Fourier transform $(t^4\lambda + 1)^{-1}$ as may be verified by a contour integration. Thus the principle term in (4.16) is the partial Fourier series expansion of the function

$$(4.17) \qquad -\lambda \sum_{-\infty}^{\infty} \lambda^{-1/4} f^{(4)} * k_{\lambda^{1/4}}(j - x),$$

the differentiability of which is established by a dominated convergence argument. It follows that (4.16) converges to (4.17) by the Poisson summation formula. We need only consider the terms $j = 0$ and $j = \pm 1$, as the sum of the remaining terms decays as $\lambda^{1/4} e^{-1/(\sqrt{2}\lambda^{1/4})} \| f^{(4)} \|$ uniformly in x. The difference between the truncated sum (4.16) and (4.17) may be bounded in absolute value by $\| f^{(4)} \| n^{-4} \lambda^{1/2}$.

Next we examine the first four terms of Proposition 4.1; we shall see these are the boundary bias functions of the smoothing spline. We have, by a similar analysis of (4.13):

PROPOSITION 4.3.

$$\Delta(E\hat{f} - f) = \left\{ -\lambda^{1/2} \Delta f^{(2)} + 2\sqrt{2}\lambda^{3/4} \tfrac{1}{2}\left(f^{(3)}(0) + f^{(3)}(1) \right) - r_{0\lambda} \right\} \times \left(1 + O(n^{-1}\lambda^{1/8})\right) + \varepsilon,$$

where

$$r_{0,\lambda} = -2\sqrt{2}\lambda^{3/4} \int_0^1 f^{(4)}_{\text{odd}}(t) k^{(3)}(t\lambda^{-1/4})\, dt \left(1 + O\left(e^{-1/(\sqrt{2}\lambda^{1/4})}\right)\right),$$

and where $|\varepsilon| \le \| f^{(4)} \| n^{-7} \lambda^{-5/4}(1 + o(1))$.

Similarly, from (4.14) we have:

PROPOSITION 4.4.

$$\Delta(E\hat{f}^{(1)} - f^{(1)}) = \left\{ -2\sqrt{2}\lambda^{1/4} \tfrac{1}{2}\left(f^{(2)}(0) + f^{(2)}(1) \right) + \lambda^{1/2} \Delta f^{(3)} - r_{1\lambda} \right\} \times \left(1 + O(n^{-1}\lambda^{1/4})\right) + \varepsilon,$$

where

$$r_{1,\lambda} = 2\sqrt{2}\lambda^{1/2} \int_0^1 f^{(4)}_{\text{even}}(t) k^{(2)}(t\lambda^{-1/4})\, dt \left(1 + O\left(e^{-1/(\sqrt{2}\lambda^{1/4})}\right)\right),$$

and where $|\varepsilon| < (\Delta f^{(2)} + \| f^{(4)} \|) n^{-3/2}$ *if* $\lambda < 1$.

The expressions for $r_{0,\lambda}$ and $r_{1,\lambda}$ arise from a Poisson summation argument similar to that for (4.17). Here f_{even} and f_{odd} refer to the even and odd parts of f about the point $\frac{1}{2}$:

$$f_{\text{even}} = \tfrac{1}{2}(f(x) + f(1-x)),$$
$$f_{\text{odd}} = \tfrac{1}{2}(f(x) - f(1-x)).$$

In order to compare the boundary bias of the spline to the kernel estimate, we need an explicit representation of the functions $A_i(x)$. Theorem 3.2 will follow from Propositions 4.1–4.4 and the following:

PROPOSITION 4.5.

$$A_3(x) = k^{(3)}(-x\lambda^{-1/4}) + k^{(3)}((1-x)\lambda^{-1/4}) + D_n(x) + \varepsilon_3(x),$$
$$A_2(x) = \lambda^{1/4}\big(k^{(2)}(-x\lambda^{-1/4}) + k^{(2)}((1-x)\lambda^{-1/4})\big) + \varepsilon_2(x),$$
$$A_1(x) = \lambda^{1/2}\big(k^{(1)}(-x\lambda^{-1/4}) + k^{(1)}((1-x)\lambda^{-1/4})\big) + \varepsilon_1(x),$$
$$A_0(x) = \lambda^{3/4}\big(k(-x\lambda^{-1/4}) + k((1-x)\lambda^{-1/4})\big) + \varepsilon_0(x),$$

where $|\varepsilon_i(x)| = O(n^{-2})$ *uniformly in* x *and where* $\|D_n\| = O(1/\sqrt{n})$.

It remains to demonstrate Proposition 4.5. A Poisson summation argument shows that

$$\big\|A_j - \lambda^{(3-j)/4}\big(k^{(j)}(-x\lambda^{-1/4}) + k^{(j)}((1-x)\lambda^{-1/4})\big)\big\| = o(\lambda^{(7-2j)/8});$$

see Rice and Rosenblatt [(1983), Lemma 3]. Proposition 4.5 requires us to refine this bound. Let

$$B_j(x) = \lambda \sum_{k\neq 0} \frac{(2\pi ik)^j}{\lambda(2\pi ik)^4 + 1} e^{-2\pi ikx}.$$

By a Poisson summation argument, we see that

$$B_j(x) = \lambda^{(3-j)/4}\big(k^{(j)}(-x\lambda^{-1/4}) + k^{(j)}((1-x)\lambda^{-1/4})\big) + O(e^{-1/(\sqrt{2}\lambda^{1/4})}).$$

A short computation shows that

$$a_k^{[n]} = \frac{1}{2\pi ij(k)} + \frac{\operatorname{sgn}(j(k))kC(|k/n|)}{2in^2}$$

for $k = 1,\dots,n-1$, where

$$j(k) = \begin{cases} k, & k < n/2, \\ k-n, & k > n/2, \end{cases}$$

and where $C(t) = (\frac{1}{2}t)(\coth(\pi t) - 1/(\pi t))$. Substituting into expression (4.11) and switching the order of summation, we obtain

$$A_3(x) = B_3(x) + \left\{\frac{-1}{2(\pi n)^2}\sum_{-n/2}^{n/2}\frac{(2\pi ik)C(k/n)}{\lambda(2\pi k)^4+1}e^{-2\pi ikx} + \lambda\sum_{|k|>n/2}\frac{(2\pi ik)^3}{\lambda(2\pi k)^4+1}e^{-2\pi ikx}\right\} + \varepsilon(x), \tag{4.18}$$

with $|\varepsilon(x)| = O(1/n^2)$ uniformly in x. We shall denote the previous term in brackets by $D_n(x)$; a short computation gives $\|D_n(x)\| = O(1/\sqrt{n})$. Similarly, we obtain

$$A_2(x) = B_2(x) - \frac{1}{(2\pi n)^2} \sum_{-n/2}^{n/2} \frac{C(k/n)}{\lambda(2\pi ik)^4 + 1} e^{-2\pi ikx} + \varepsilon(x),$$

where $|\varepsilon(x)| = O(1/n^2)$. The second term may be bounded in absolute value by $1/(2^4 3\sqrt{2}\, n)$ uniformly in x. The expansions for $A_1(x)$ and $A_0(x)$ follow similarly.

This establishes Proposition 4.5; Theorem 3.2 will follow by substituting the expressions for the $A_i(x)$ into Proposition 4.1, expanding the expressions for the $k^{(i)}(x)$ and collecting terms.

We turn to the proof of Theorem 3.3.

Notice that $\operatorname{cov}(\hat{y}_k, \hat{y}_j) = (\delta_{jk} - 1/(2n))\sigma^2$. We may write $\operatorname{cov}(\hat{f}(s), \hat{f}(t)) = E[h(s)h(t)]$, where

$$h(s) = (\Delta \hat{f} - E\Delta \hat{f})A_3(s) + (\Delta \hat{f}^{(1)} - E\Delta \hat{f}^{(1)})A_2(s) + \sum_{k=0}^{n-1} \left(\frac{\hat{y}_k}{\sqrt{n}} - f_k^{[n]} \right) \left(\frac{1 + \varepsilon_k k^4}{\lambda(2\pi ik)^4 + 1} \right) e^{-2\pi iks},$$

and where ε_k is uniformly $O(n^{-4})$ by an analysis similar to that following (4.15). The theorem follows from the approximations

$$\operatorname{var}(\Delta \hat{f}) = \frac{\sigma^2}{n} \lambda^{-1/4} C_1 \left(1 + O\left(\frac{1}{n}\right)\right),$$

$$\operatorname{var}(\Delta \hat{f}^{(1)}) = \frac{\sigma^2}{n} \lambda^{-3/4} C_2 \left(1 + O\left(\frac{1}{n}\right)\right),$$

$$\operatorname{cov}(\Delta \hat{f}, \Delta \hat{f}^{(1)}) = \frac{\sigma^2}{n} \lambda^{-1/2} C_3 \left(1 + O\left(\frac{1}{n}\right)\right),$$

$$\operatorname{cov}\left(\Delta \hat{f}, \frac{\hat{y}_k}{\sqrt{n}}\right) = \frac{\sigma^2}{n} \lambda^{3/4} \sqrt{2}\, \frac{\bar{a}_k^{[n]}}{\lambda + r_k^{[n]}} \left(1 + O\left(\frac{1}{n}\right)\right),$$

$$\operatorname{cov}\left(\Delta \hat{f}^{(1)}, \frac{\hat{y}_k}{\sqrt{n}}\right) = \frac{\sigma^2}{n} (\lambda^{1/4} - \lambda^{1/2}) \frac{b_k^{[n]}}{\lambda + r_k^{[n]}} \left(1 + O\left(\frac{1}{n}\right)\right),$$

where

$$C_1 = \left(\int \frac{x^6}{(x^4+1)^2}\, dx \right) \left(\int \frac{x^2}{(x^4+1)^2}\, dx \right)^{-2},$$

$$C_2 = \left(\int \frac{x^4}{(x^4+1)^2}\, dx \right) \left(\int \frac{1}{(x^4+1)^2}\, dx \right)^{-2},$$

$$C_3 = 2 \int \frac{x}{(x^4+1)^2}\, dx,$$

and collecting terms.

5. Concluding remarks. We briefly mention that the results of this paper may be used to develop a correction for the boundary bias of a spline estimate. Further possible applications may be to investigate the behavior of $\hat{\lambda}$, where $\hat{\lambda}$ is an estimate of the optimal λ and to give a confidence band for a smoothing spline.

Acknowledgments. This work is derived from the author's dissertation written under the supervision of Murray Rosenblatt, whose guidance is gratefully acknowledged. Thanks are also due to an Associate Editor.

REFERENCES

APOLSTOL, T. (1969). *Calculus, Volume II*, 2nd ed. Xerox College Publishing, Waltham, Mass.

CODDINGTON, E. and LEVINSON, N. (1955). *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.

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

EUBANK, R. (1988). *Spline Smoothing and Non-parametric Regression*. Dekker, New York.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications, Volume II*, 2nd ed. Wiley, New York.

GASSER, T. and MULLER, H. G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.) 23–68. Springer, New York.

GASSER, T. and MULLER, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 197–211.

MESSER, K. and GOLDSTEIN, L. (1989). A new class of kernels for nonparametric curve estimation. *Ann. Statist.* To appear.

PARKER, R. and SHURE, L. (1982). Efficient modeling of the earth's magnetic field with harmonic splines. *Geophys. Res. Lett.* **2** 812–815.

REINSCHE, C. (1967). Smoothing by spline functions. *Numer. Math.* **10** 177–183.

RICE, J. (1984). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* **13** 893–900.

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

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1841.

SCHOENBERG, I. J. (1964). Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. USA* **52** 947–950.

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

SILVERMAN, B. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B* **47** 1–22.

SPECKMAN, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. *Numer. Math.* To appear.

WAHBA, G. (1975). Smoothing noisy data with spline functions. *Numer. Math.* **24** 383–393.

WHITTAKER E. T. (1923). On a new method of graduation. *Proc. Edinburgh Math. Soc. (2)* **41** 63–75.

DEPARTMENT OF MATHEMATICS
CALIFORNIA STATE UNIVERSITY, FULLERTON
FULLERTON, CALIFORNIA 92634-4080