The Annals of Statistics
1992, Vol. 20, No. 1, 146–160

RELAXED BOUNDARY SMOOTHING SPLINES

BY GARY W. OEHLERT

University of Minnesota

Ordinary smoothing splines have an integrated mean squared error which is dominated by bias contributions at the boundaries. When the estimated function has additional derivatives, the boundary contribution to the bias affects the asymptotic rate of convergence unless the derivatives of the estimated function meet the natural boundary conditions. This paper introduces relaxed boundary smoothing splines and shows that they obtain the optimal asymptotic rate of convergence without conditions on the boundary derivatives of the estimated function.

1. Introduction and summary. Consider the nonparametric regression problem (x_i, y_i), $i = 1, 2, \dots, n$, $y_i = \mu(x_i) + \varepsilon_i$, where the ε_i's are uncorrelated, with zero mean and constant variance σ^2. We consider the case of uniformly spread data between 0 and 1: $x_i = (2i - 1)/2n$. Here the x values should be doubly indexed, because x_i depends on n, but this dependence will not be carried in the notation. Our goal is to estimate the regression function $\mu(x)$ and we will judge the quality of an estimator $\hat{\mu}(x)$ by integrated mean squared error (IMSE):

$$\text{IMSE}(\hat{\mu}) = E\int_0^1 (\hat{\mu}(t) - \mu(t))^2 \, dt.$$

As usual, the IMSE may be decomposed into an integrated squared bias term and an integrated variance term.

Smoothing splines are one popular method for such a nonparametric regression. An mth order smoothing spline $\mu_\lambda(x)$, $m \geq 2$, is defined to be that function with square integrable mth derivative which minimizes

$$\frac{1}{n}\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_0^1 (f^{(m)}(t))^2 \, dt,$$

where $f^{(m)}$ indicates the mth derivative of f. The solution to the minimization problem is well known: The minimizer is a polynomial in each of the intervals (x_i, x_{i+1}) with the polynomials constrained so that the function and its first $2m - 2$ derivatives are continuous. The polynomials are of degree $m - 1$ below x_1 and above x_n and degree $2m - 1$ otherwise.

The asymptotic properties of the IMSE for smoothing splines have been studied by a number of authors, including Wahba (1975), Craven and Wahba (1979), Speckman (1981), Rice and Rosenblatt (1983) and Cox (1988). Consider

Received July 1989; revised May 1991.
AMS 1980 *subject classifications*. Primary 62G07; secondary 62G20.
Key words and phrases. Splines, boundary effects, regularization.

cubic smoothing splines, where $m = 2$. If the function $\mu(x)$ has two derivatives, then the smoothing spline has an IMSE which is $O(n^{-4/5})$ provided that the data are spread regularly across $[0,1]$ (or differ from regularity by a sufficiently small amount) and that λ is reduced at the rate $n^{-4/5}$. Stone (1982) showed that this is the optimal rate of convergence for estimators of twice differentiable functions on compact intervals. If the function $\mu(x)$ has four derivatives, then the smoothing spline has an IMSE which converges to zero at a faster rate. However, Rice and Rosenblatt (1983) have shown that the rate at which the IMSE converges to zero depends on the second and third derivatives of μ at 0 and 1. If the second derivative is nonzero at either boundary, then the rate is $-5/6$; if the second derivatives are zero but either third derivative is nonzero the rate is $-7/8$; if the second and third derivatives are all zero at the boundaries, then the rate is $-8/9$, which is Stone's optimal rate for a four times differentiable function. Utreras (1988) has extended these results to the case $m > 2$.

The problem with the IMSE lies in the squared bias, not the variance. The smoothing spline estimator is a natural spline, that is, it is a polynomial of degree $m - 1$ near the boundaries, and it will be a good estimator of μ near the boundaries provided that μ also satisfies the natural boundary conditions. As μ fails to meet the natural boundary conditions to a greater and greater extent, the smoothing spline estimate of μ will become more and more biased near the boundaries since the spline estimate cannot adapt to the shape of μ near the boundaries.

How can this problem of boundary bias be corrected? Intuitively, the trade-off between bias and variance in a smoothing spline is controlled by the parameter λ. To reduce the bias, we reduce λ. A smaller λ will relax the spline, and should allow us to fit the boundary of μ more closely. However, we do not want to reduce λ globally, since that would increase the global variance. We only want to relax the spline near the boundaries. To do this we introduce a modified estimator which we call a relaxed boundary smoothing spline. This estimator is not a spline in the sense of being a piecewise polynomial; rather the name is intended to indicate the motivation and heritage of the estimator.

Let $w(t,\alpha)$ be a continuous weight function in the open interval $(0,1)$ given by

$$w(t,\alpha) = t^{\alpha}(1-t)^{\alpha}$$

for some $\alpha > 0$. An mth order relaxed boundary smoothing spline is the function g_λ which minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - g(x_i)\right)^2 + \lambda\int_0^1 w(t,m)\left(g^{(m)}(t)\right)^2\,dt,$$

where the minimization takes place over the weighted Sobolev space of functions $g(t)$ for which the integral converges.

This paper derives the rate of convergence (in the IMSE sense) for the relaxed boundary smoothing spline for $m \ge 4$. (Somewhat weaker results should hold for the case $m = 3$; see the Remarks.) We show that when μ has

m square integrable derivatives, then g_λ has an IMSE which converges at the optimal rate $n^{-2m/(2m+1)}$. Moreover, when μ has $2m$ square integrable derivatives, then g_λ has an IMSE which converges at the new optimal rate $n^{-4m/(4m+1)}$, with no further conditions on μ. Ordinary smoothing splines can only achieve this rate of convergence if μ satisfies the natural boundary conditions. Specifically, we prove:

THEOREM. *Let $x_1, x_2, \dots, x_n$ (the nth row of a triangular array of observation points) be such that $x_i = (2i - 1)/2n$. We observe data y_i satisfying $y_i = \mu(x_i) + \varepsilon_i$, where the ε_i are uncorrelated with zero mean and constant variance. Assume that the mean function $\mu(t)$ has $2m - 1$ absolutely continuous derivatives and square integrable $2m$th derivative. Then g_λ, the relaxed boundary smoothing spline estimate of μ has integrated mean square error of order $n^{-4m/(4m+1)}$ for $m \geq 4$, when λ is decreased as $n^{-2m/(4m+1)}$.*

The take-home idea from this paper is that using a relaxed boundary penalty function can lead to a better asymptotic IMSE than standard splines when the mean function has extra derivatives. Section 2 of the paper outlines a proof of the theorem. The last section discusses computation of the relaxed boundary smoothing spline and gives an example.

Other approaches are available for correcting the spline boundary bias problem. Eubank and Speckman (1989) describe a fairly general bias reduction technique for nonparametric regression. For spline smoothing, the technique amounts to regressing the spline residuals on polynomials and adding the predicted residuals to the spline fit. Heuristically, this approach is representing the expected value function as the sum of a polynomial and a function which meets the natural boundary conditions; see also Speckman (1988). A less desirable alternative would be to weight the IMSE criterion rather than the roughness penalty, for example, $E\int w(t)(\hat{\mu}(t) - \mu(t))^2\,dt$. Using $w(t, 2)$ from above and the results of Lemma 5 of Rice and Rosenblatt (1983), it is not hard to see that the integrated weighted squared bias for cubic smoothing splines will be reduced to the appropriate order without further boundary conditions. However, it seems much more appropriate to maintain our IMSE criterion function and change the class of estimators we use than to keep our estimators and change the criterion function.

2. Proof of theorem. The overall method of proof is that of Speckman (1981), who approximated the IMSE of a smoothing spline by the IMSE of a continuous analogue (Tikhonov regularization). Utreras (1988) derived the rate of convergence for the squared bias of the regularizer in spline smoothing. In fact, very little in their proofs is specific to the eigenfunctions and normed spaces of ordinary spline smoothing, and the elegant methods of Speckman and Utreras can be used without essential change by using norms in weighted Sobolev spaces instead of the usual Sobolev norms and by using appropriate eigenvalues and eigenfunctions in the continuous approximation. Thus, we will be able to defer to their proofs after establishing a few continuity (boundedness) conditions and a formula for the eigenvalues. We repeat the Utreras proof in some detail, however, since it is in the bias of the regularizer that relaxed boundary smoothing splines make their gains. Some technical lemmas are postponed until the end of the section.

Let $W(m, \alpha)$ be a weighted Sobdlev space of functions on the interval (0, 1). A function f is in $W(m, \alpha)$ if f and its first $m - 1$ derivatives are absolutely continuous and if $\int_0^1 w(t, \alpha)(f^{(k)}(t))^2\,dt$ is finite for $k = 0, 1, \ldots, m$. Note that $W(m, \alpha)$ is a richer space than the usual Sobolev space $W(m, 0)$, that is, $W(m, \alpha)$ contains more functions. Let $\|f\|_{m,\alpha}$ be the seminorm

$$\|f\|_{m,\alpha}^2 = \int_0^1 w(t, \alpha)\left(f^{(m)}(t)\right)^2 dt.$$

The usual weighted Sobolev norm on $W(m, \alpha)$ is $(\Sigma_{k=0}^{m}\|f\|_{k,\alpha})^{1/2}$ [Kufner (1980), page 18].

The continuous analogue of spline smoothing is called Tikhonov regularization. In regularization, a target function f is approximated by a regularizer function g and g is chosen to minimize a weighted sum of the norm of $f - g$ and a seminorm of g. For relaxed boundary splines, f is an element of $W(m, 0)$ and the Tikhonov regularizer of f is g_λ, the unique minimizer of

$$\min_{g \in W(m, m)} \left\|f - g\right\|_{0,0}^2 + \lambda\left\|g\right\|_{m,m}^2.$$

Utreras (1988) studied the asymptotic behavior of the Tikhonov regularizer for standard smoothing splines by looking at the differential eigenvalue problem

$$(-1)^m D^{2m}\psi_k = \chi_k\psi_k,$$
$$\psi_k^{(j)}(0) = \psi_k^{(j)}(1) = 0, \qquad j = m, \ldots, 2m - 1.$$

For relaxed boundary splines, we look at a related eigenvalue problem:

$$(-1)^m D^m\left(w(\cdot, m)D^m\psi_k(\cdot)\right) = \chi_k\psi_k,$$
$$D^j\left(w(\cdot, m)D^m\psi_k(t)\right)(0) = D^j\left(w(\cdot, m)D^m\psi_k(t)\right)(1) = 0,$$
$$j = 0, 1, \ldots, m - 1.$$

This is a singular, self-adjoint differential expression (the singularity is due to the zeros of the weight function at $t = 0, 1$) which is closely related to the Legendre differential equation. The eigenvalues can be expressed as $\chi_k = 4^{-m}(k - m + 1)(k - m + 2)\cdots(k + m)$ for $k = 0, 1, 2, \ldots,$ and the kth eigenfunction is exactly the kth (shifted) Legendre polynomial ψ_k; see Lemma 1. The ψ_k clearly form a complete orthonormal system for $W(0, 0)$. Furthermore, since $W(m, m)$ embeds in $W(2, 0)$ (see Lemma 2) which itself embeds in $W(0, 0)$, the ψ_k also span $W(m, m)$ as elements of $W(0, 0)$. In addition, $D^m\psi_k$ forms a complete orthogonal system for $W(0, m)$ (see Lemma 1).

Following Utreras (1988), we express the continuous regularization problem in terms of the eigenvalues and eigenfunctions. Any $f \in W(j, 0)$, $j \geq 0$, can be expanded in the Fourier series

$$f = \sum_{k \geq 0} f_k \psi_k,$$

where f_k is the kth Fourier coefficient

$$f_k = \int_0^1 f(x) \psi_k(x)\, dx.$$

Furthermore, for $k \geq m$ and $g \in W(m, m)$, we have

$$\begin{aligned} g_k &= \int_0^1 g(x) \psi_k(x)\, dx \\ &= \frac{(-1)^m}{\chi_k} \int_0^1 g(x) D^m [w(x, m) D^m \psi_k(x)]\, dx \\ &= \frac{1}{\chi_k} \int_0^1 D^m g(x) D^m \psi_k(x) w(x, m)\, dx. \end{aligned}$$

(Recall that $\chi_k = 0$ for $k < m$.) But since $D^m \psi_k$ is a complete orthogonal system in $W(0, m)$ and since $D^m g$ is in $W(0, m)$, we have that

$$\left\| D^m g \right\|_{0,m}^2 = \left\| g \right\|_{m,m}^2 = \sum_{k \geq 0} \chi_k g_k^2.$$

Substituting these Fourier expressions into the regularization problem, we want to minimize

$$\sum_{k \geq 0} (g_k - f_k)^2 + \lambda \sum_{k \geq 0} \chi_k g_k^2.$$

The solution is the function g with coefficients

$$g_k = \frac{f_k}{1 + \lambda \chi_k},$$

and the integrated squared difference between f and the regularizer g_λ can be written

$$\| f - g_\lambda \|_{0,0}^2 = \sum_{k \geq 0} \frac{\lambda^2 \chi_k^2}{(1 + \lambda \chi_k)^2} f_k^2.$$

Recall that

$$f_k = \int_0^1 \psi_k(x) f(x)\, dx = \frac{1}{\chi_k} \int_0^1 D^m \psi_k(x) D^m f(x) w(x, m)\, dx,$$

by integration by parts for $k \geq m$. If $f \in W(2m, 0)$, then all the derivatives of f up to order $2m - 1$ are uniformly bounded. In particular, this implies that $f \in W(2m, 0)$ satisfies the boundary conditions

$$D^j(w(\cdot, m) D^m f(t))(0) = D^j(w(\cdot, m) D^m f(t))(1) = 0, \qquad j = 0, 1, \ldots, m - 1.$$

Thus, we may continue to integrate by parts obtaining

$$f_k = \frac{(-1)^m}{\chi_k}\int_0^1 \psi_k(x) D^m[w(x)D^m f(x)]\,dx = \frac{(-1)^m}{\chi_k} f_k^*,$$

where $f^* = D^m[w(x)D^m f(x)]$. Note that $f \in W(2m,0)$ implies that $f^* \in W(0,0)$, since w and all its derivatives are bounded. Thus,

$$\sum_{k=m}^{\infty} (f_k^*)^2 \le \| f^* \|_{0,0}^2 < \infty.$$

Substituting the expression for Fourier coefficients into the formula for integrated squared error, we find that

$$\begin{aligned}\| f - g_\lambda \|_{0,0}^2 &= \sum_{k \ge 0} \frac{\lambda^2 \chi_k^2}{(1 + \lambda\chi_k)^2} f_k^2 \\ &= \lambda^2 \sum_{k \ge m} \frac{\chi_k^2 f_k^2}{(1 + \lambda\chi_k)^2} \\ &= \lambda^2 \sum_{k \ge m} \frac{f_k^{*2}}{(1 + \lambda\chi_k)^2} \\ &\le \lambda^2 \| f^* \|_{0,0}^2.\end{aligned}$$

Thus we have proved:

PROPOSITION 1. *Suppose $f \in W(2m,0)$ and let $f^* = D^m[wD^m f]$. Then*

$$\| f - g_\lambda \|_{0,0}^2 \le \lambda^2 \| f^* \|_{0,0}^2.$$

The advantage for relaxed boundary smoothing splines comes from the fact that the relaxed regularizer has integrated squared bias of order $O(\lambda^2)$ for all functions $f \in W(2m,0)$. Ordinary smoothing splines can only achieve this rate of convergence if the derivatives of the function f meet the natural boundary conditions, which permit the additional integration by parts.

To complete the proof of the theorem, we need to show that the integrated squared bias of the relaxed boundary smoothing spline can be approximated by the integrated squared bias of the regularizer to sufficient closeness and to derive an approximation for the integrated squared variance of the relaxed boundary smoothing spline. This is the content of the following two propositions.

PROPOSITION 2. *If $m \ge 4$, then the regularizer g_λ and relaxed boundary smoothing spline $g_{n,\lambda}$ satisfy*

$$\| f - g_{n,\lambda} \|_{0,0}^2 = \| f - g_\lambda \|_{0,0}^2 (1 + o(1)),$$

where $o(1)$ denotes a term independent of λ and n as $\lambda \to 0$, $n^2\lambda \to \infty$ and $n\|f - g_\lambda\|_{0,0}^2 \to \infty$.

PROPOSITION 3. *The integrated variance of the relaxed boundary smoothing spline $g_{n,\lambda}$ satisfies*

$$v_n(\lambda) = \frac{\int_0^\infty (1 + x^{2m})^{-2}\, dx}{n\lambda^{1/2m}}(1 + o(1)),$$

where $o(1)$ denotes a term independent of n and λ as $n \to \infty$, $\lambda \to 0$ and $n\lambda \to \infty$.

Propositions 2 and 3 are essentially Theorems 2.3 and 2.4 of Speckman (1981). The proofs in Speckman will work once we have established a framework of normed spaces for the relaxed boundary smoothing spline and the regularizer. Recall that we have an exact formula for the eigenvalues of the differential problem; these eigenvalues satisfy Speckman's Lemma 5.2.

Following Speckman, let $Y = W(0,0)$ have the norm $\|\cdot\|_{0,0}$ and let $X = W(0,0) + W(0,m)$ have the norm

$$|||(f,g)|||_{\lambda,m}^2 = \|f\|_{0,0}^2 + \lambda\|g\|_{0,m}^2$$

For $g \in W(m,m)$, let $Lg = (g, g^{(m)})$ and let

$$_{\lambda,m}\|g\| = |||Lg|||_{\lambda,m}^2 = \|g\|_{0,0}^2 + \lambda\|g^{(m)}\|_{0,m}^2.$$

The map L from $g \in W(m,m)$ to $(g, g^{(m)})$ in X is linear and injective. Furthermore, $W(m,m)$ continuously embeds in $W(2,0)$ (see Lemma 2) and thus in $W(0,0)$; hence the usual L_2 norm of g is bounded by the weighted Sobolev norm $\|g\|_{m,m}$. So, by the closed range theorem and the open mapping theorem, the image of $W(m,m)$ is closed in X. In particular, for any element $(f,g) \in X$, there is a unique $h \in W(m,m)$ with image under L closest to (f,g) in the norm $|||\cdot|||_{\lambda,m}$. Lh is the projection of (f,g) onto $LW(m,m)$. [In this notation, the Tikhonov regularizer is exactly the element $h_\lambda \in W(m,m)$ for which Lh_λ is closest to $(g,0) \in X$ in the $|||\cdot|||_{\lambda,m}$ norm.] Let $h_\lambda = P_\lambda g$. The projection from X onto $LW(m,m)$ is bounded, and the pseudoinverse of L is bounded since the $_{\lambda,m}\|\cdot\|$ norm is equivalent to the weighted Sobolev norm (see Lemma 3). Thus, the regularizer h_λ is obtained by a bounded linear transformation of g.

We also need a discrete analogue of the space X. Let $X_n = R^n + W(0,m)$, with norm

$$|||(y,g)|||_{n,\lambda,m}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 + \lambda\|g\|_{0,m}^2.$$

Let A_n be the map from $W(m,m)$ onto R^n defined by $A_n(g) = (g(x_1),\dots,g(x_n))$ and let L_n be a map from $W(m,m)$ to X_n defined by $L_n(g) = (A_n(g), g^{(m)})$. L_n is linear and for $n \geq m$, it is injective. We claim that L_n is bounded. We only need to show that $1/n\sum g(x_i)^2$ is bounded by a multiple of $\|g\|^2_{m,m}$. First, $W(m, m)$ is embedded in $W(2,0)$, so the L_2 norms of g and $g^{(2)}$ are both bounded by $\|g\|_{m,m}$. Next, standard quadrature formulae (e.g., Lemma 4.1 of Speckman) bound the absolute difference

$$\left|\frac{1}{n}\sum_{i=1}^{n} g(x_i)^2 - \int_0^1 g(t)^2\,dt\right| \le \frac{c}{n^2}\|g\|^2_{2,0}$$

for $g \in W(2,0)$. Thus, $1/n\Sigma g(x_i)^2$ must also be bounded by $\|g\|^2_{m,m}$. So, by the closed range theorem and open mapping theorem, $L_n W(m,m)$ is closed and $W(m,m)$ is complete with the norm

$${}_{n,\lambda,m}\|g\| = |||L_n g|||_{n,\lambda,m} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)^2 + \lambda\|g\|^2_{0,m}.$$

Let $f_{n,\lambda}$ be the projection of $a_n = ((y_1,\dots,y_n),0) \in X_n$ onto $L_n W(m,m)$; that is, $f_{n,\lambda}$ is the element in $W(m,m)$ which minimizes

$$|||a_n - L_n f|||_{n,\lambda,m}.$$

Again, standard quadrature formulae show that the pseudoinverse of L_n is bounded, so that

$$f_{n,\lambda} = P_{n,\lambda}\mathbf{y},$$

where $P_{n,\lambda}$ is a bounded linear transformation.

With these spaces, the proofs in Speckman may proceed, and the theorem is proved.

REMARK 1. The weighted Sobolev embeddings and equivalent norm results used in the proof of the theorem do not depend on the exact form of the weight function, but only on the rate at which the weight function goes to zero at the boundaries. Any positive weight function which goes to zero like t^m at the boundaries will work. Thus, we could use a weight function

$$w_\varepsilon(t,m) = \begin{cases} \left(\dfrac{t}{2\varepsilon}\right)^m\left(1-\dfrac{t}{2\varepsilon}\right)^m, & t \le \varepsilon,\\ 2^{-2m}, & \varepsilon \le t \le 1-\varepsilon,\\ \left(\dfrac{1-t}{2\varepsilon}\right)^m\left(1-\dfrac{1-t}{2\varepsilon}\right)^m, & t \ge 1-\varepsilon,\end{cases}$$

which should be more like an ordinary spline in the middle of the interval of estimation, and still have the embedding and equivalent norm results of Lemmas 2 and 4.

REMARK 2. The difficulty with using other weight functions comes when we try to find the eigenvalues for the differential system. Since the weight functions go to zero at the boundaries, the differential systems are singular, and standard asymptotic formulae for the eigenvalues of nonsingular systems [e.g., Naimark (1967)] do not apply. I have not been able to find analogous results for the singular case nor have I been able to extend the results for the nonsingular case to the singular case. If the eigenvalues are of the correct asymptotic order, the results will still hold.

REMARK 3. The same difficulty with eigenvalues arises when we take a nonuniform design for the observation points x_i. If the observation points follow the design density $p(t)$, then we must solve the eigensystem

$$(-1)^m D^m(w(\cdot, m) D^m \psi_k(\cdot)) = \chi_k \psi_k p(t),$$
$$D^j(w(\cdot, m) D^m \psi_k(t))(0) = D^j(w(\cdot, m) D^m \psi_k(t))(1) = 0,$$
$$j = 0, 1, \dots, m-1.$$

Again, if the eigenvalues have the correct asymptotic order, the results will go through.

REMARK 4. We proved the theorem under the assumption that $m \geq 4$, which excludes the standard cubic spline case of $m = 2$. The reason we cannot prove the theorem for m of 2 or 3 is that the quadrature approximation of $\int_0^1 g(t)^2\, dt$ by $\sum_{i=0}^{n} g(x_i)^2/n$ is only good to order n^{-1} (instead of n^{-2}) when $g \in W(1,0)$ and the spaces $W(2,2)$ and $W(3,3)$ embed in $W(1,0)$ but not in $W(2,0)$.

REMARK 5. We can get some results for $m = 3$ if we use the seminorm $\|\cdot\|_{3,2}$ as the roughness penalty instead of $\|\cdot\|_{3,3}$. The space $W(3,2)$ does embed in $W(2,0)$, so the quadrature approximation will be sufficiently accurate. However, the order of the weight function is not as large as the order of the derivative, so the boundary conditions

$$D^j(w(\cdot, m) D^m f(t))(0) = D^j(w(\cdot, m) D^m f(t))(1) = 0,$$
$$j = 0, 1, \dots, m-1,$$

will be met only if $D^m f(0)$ and $D^m f(1)$ are both 0. [It is curious that this weighted problem has restrictions on the mth order derivative rather than the $(2m-1)$th order derivative.] Furthermore, we again need to show that the eigenvalues have the correct asymptotic order.

REMARK 6. If the function f has only m derivatives, then we may show that the bias term is of order λ using the method of Section 6.3.2 of Eubank (1988), leading to $n^{-2m/(2m+1)}$ convergence. Also note that f having more than $2m$ derivatives will not produce a further decrease in the bias. Take, for example, the mth Legendre polynomial, which is infinitely differentiable but still has bias of order λ^2 and convergence of order $n^{-4m/(4m+1)}$.

REMARK 7. The weight used in the relaxed boundary smoothing spline is reminiscent of data tapers used in time series [see, e.g., Bloomfield (1976)]. While the motivations and details differ, both the weight function and data tapers have the effect of improving the rate at which certain Fourier coefficients approach zero.

We finish the section with four technical lemmas.

LEMMA 1. *The (shifted) Legendre polynomials ψ_k solve the differential eigenvalue problem*

$$(-1)^m D^m(w(\cdot, m)D^m\psi_k(\cdot)) = \chi_k\psi_k,$$

$$D^j(w(\cdot, m)D^m\psi_k(t))(0) = D^j(w(\cdot, m)D^m\psi_k(t))(1) = 0,$$

$$j = 0, 1, \dots, m - 1,$$

with eigenvalues $\chi_k = 4^{-m}\Gamma(m + k + 1)/\Gamma(m - k + 1)$.

PROOF. The shifted Legendre polynomials ψ_k are the orthogonal polynomials on the unit interval with respect to the weight function $w \equiv 1$, with the standardization that $\psi_k(1) = 1$. They differ from ordinary Legendre polynomials only in that ordinary Legendre are defined on the interval $(-1, 1)$; shifted Legendre polynomials can be obtained from ordinary Legendre polynomials by the change of variable $t = (x + 1)/2$. The differential problem is equivalent to the integral problem

$$\int_0^1 \psi_j(t)\psi_k(t)\, dt = \delta_{jk},$$

$$\int_0^1 \psi_j^{(m)}(t)\psi_k^{(m)}(t)t^m(1 - t)^m\, dt = \chi_k\delta_{jk},$$

with $0 = \chi_1 = \chi_2 = \cdots = \chi_m < \chi_{m+1} \leq \cdots$. Shifting back to the interval $(-1, 1)$ via $x = 2t - 1$, we get

$$\int_{-1}^1 \psi_j(x)\psi_k(x)\, dx = \delta_{jk},$$

$$\int_{-1}^1 \psi_j^{(m)}(x)\psi_k^{(m)}(x)(1 - x^2)^m\, dx = 4^m\chi_k\delta_{jk}.$$

The Legendre polynomials are orthogonal and thus meet the first of the two requirements. Furthermore, the mth derivative of the nth Legendre polynomial is $1 \cdot 3 \cdots (2m - 1)C_{n-m}^{(m+1/2)}(x)$, where $C_{n-m}^{(m+1/2)}(x)$ is an ultraspherical polynomial [see Abramowitz and Stegun (1965), Section 22.5]. These ultraspherical polynomials are a (complete) orthogonal system with weight function $(1 - x^2)^m$, exactly as required. The eigenvalues can be easily calculated given Table 22.2 of Abramowitz and Stegun (1965). □

LEMMA 2. *The weighted Sobolev spaces $W(m, m)$ continuously embed into the Sobolev space $W(2, 0)$ for $m \geq 4$.*

PROOF. First, $W(m, \omega)$ trivially embeds in $W(n, \omega)$ for $m \geq n$. Next, result 6.4 of Kufner (1980) shows that $W(m, \omega) \hookrightarrow W(m, \eta)$ for any $\eta \geq \omega$. Third, result 8.35 of Kufner (1980) shows that $W(k,\varepsilon) \hookrightarrow W(r, \varepsilon - 2(k-r))$ if $0 \le r < k$ and $\varepsilon > 2(k-r) - 1$. So, for $m = 2k \ge 4$, we have

$$W(2k,2k) \hookrightarrow W(k,0) \hookrightarrow W(2,0).$$

For $m = 2k + 1 \ge 5$, we have

$$W(2k+1, 2k+1) \hookrightarrow W(k+1,1) \hookrightarrow W(k+1,2) \hookrightarrow W(3,2) \hookrightarrow W(2,0). \quad \square$$

LEMMA 3. *The* ${}_{\lambda,m}\|\cdot\|$ *norm is equivalent to the usual weighted Sobolev norm on* $W(m,m)$.

PROOF. It suffices to show that the norms are equivalent on a dense subset of $W(m,m)$; we will show they are equivalent on $C^\infty([0,1])$, which is dense in $W(m,m)$ by Theorem 7.2 of Kufner (1980). Let $W_0(m,m)$ be the closure of $C_0^\infty(0,1)$ in the usual Sobolev norm, where $C_0^\infty(0,1)$ is the space of infinitely differentiable functions which are zero in a neighborhood of 0 and 1. Then any $f \in C^\infty([0,1])$ can be written uniquely as $f = f_0 + f_1$, where f_1 is the degree $2[(m-1)/2] + 1$ polynomial which agrees with f and its first $[(m-1)/2]$ derivatives at 0 and 1 and $f_0 \in W_0(m,m)$. The two norms are equivalent on the polynomial part and the two norms are equivalent on $W_0(m,m)$ by Proposition 9.2 of Kufner (1980). Since $W(m,m)$ is the direct sum of two subspaces and the norms are equivalent on both subspaces, the norms must be equivalent on the direct sum, $W(m,m)$. $\square$

3. Computations and examples. Kimeldorf and Wahba (1971) give an explicit solution to the smoothing problem in a reproducing kernel Hilbert space. To use their solution, we must first obtain the reproducing kernel. Decompose $W(m,m)$ into the direct sum of $\mathscr{P}$ (a space of polynomials of degree at most $m-1$) and $W_0(m,m)$ as in Lemma 3. The polynomials are written in the form

$$p(x) = \sum_{i=0}^{m-1} a_i x^i,$$

and the inner product of two polynomials is

$$\left\langle \sum_{i=0}^{m-1} a_i x^i, \sum_{i=0}^{m-1} b_i x^i \right\rangle = \sum_{i=0}^{m-1} a_i b_i,$$

so that the monomials x^i are orthogonal basis elements for $\mathscr{P}$. The reproducing kernel for $\mathscr{P}$ is

$$K_1(s,t) = \sum_{i=0}^{m-1} s^i t^i.$$

The space $W_0(m,m)$ has the inner product

$$\langle f,g \rangle = \int_0^1 t^m (1-t)^m f^{(m)} g^{(m)}\, dt,$$

and the (shifted) Legendre polynomials of degree m and higher (denoted ψ_i, $i \geq m$) form a complete orthogonal basis for $W_0(m, m)$ satisfying

$$D^m\left[t^m(1-t)^m D^m \psi_k\right] = \chi_k \psi_k,$$

where the eigenvalues χ_k are given in Lemma 1. Thus the reproducing kernel on $W_0(m, m)$ can be written

$$K_0(s,t) = \sum_{i=m}^{\infty} \frac{\psi_i(s)\psi_i(t)}{\chi_i}.$$

Now using Theorem 5.1 of Kimeldorf and Wahba, we have that the function $f \in W(m, m)$ which minimizes

$$\frac{1}{\lambda}\sum_{i=1}^{n}\left(f(x_i) - y_i\right)^2 + \int_0^1 t^m(1-t)^m\left(f^{(m)}(t)\right)^2 dt$$

may be written as

$$f(t) = w'(UM^{-1}U')^{-1}UM^{-1}\mathbf{y} + \eta' M^{-1}\left[I - U'(UM^{-1}U')^{-1}UM^{-1}\right]\mathbf{y},$$

where $w' = w'(t) = (1, t, t^2, \dots, t^{m-1})$, $\mathbf{y}' = (y_1, y_2, \dots, y_n)$,

$$U_{m\times n} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ & & \cdots & \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_n^{m-1} \end{bmatrix},$$

$$\eta = \left(\sum_{i=m}^{\infty} \frac{\psi_i(x_1)\psi_i(t)}{\chi_i}, \dots, \sum_{i=m}^{\infty} \frac{\psi_i(x_n)\psi_i(t)}{\chi_i}\right)'$$

and

$$M_{n\times n} = \left[\sum_{k=m}^{\infty} \frac{\psi_k(x_i)\psi_k(x_j)}{\chi_k}\right] + \lambda I.$$

The technique outlined above is not particularly efficient, but it is adequate for reasonably small sample sizes provided some care is exercised in evaluating terms involving M^{-1} when λ is small. Most problems can be avoided by noting that M can be written as $M = \lambda I + PP'$, so that

$$M^{-1} = \left(I - P(\lambda I + P'P)^{-1}P'\right)/\lambda.$$

Note also that this technique does not require that the observation points be uniformly spaced.

An approximate technique for uniformly spaced points can be derived from the analogous continuous regularization problem introduced during the analysis of the bias in Section 2. The solution of the continuous problem was a ridge regression on the (shifted) Legendre polynomials. As an approximation in the discrete case, we simply do a ridge regression on the (shifted) Legendre polynomials using the asymptotic eigenvalues (times the smoothing parameter λ) as the ridge constants. In the example below, the exact and approximate methods are indistinguishable.

In practice, we must use a finite number of Legendre polynomials in either the exact or approximate methods. The number of polynomials required to achieve accurate results will vary with m and λ. In the example below, computations with 15 and 30 polynomials differed in the fifth digit while computations with 25 and 30 polynomials differed in the seventh digit. Results below are based on computations with 30 polynomials.

Consider the function

$$f(x) = \cos(\pi x) + \frac{2}{27}\cos(2\pi x) - \frac{1}{729}\cos(3\pi x) - \frac{1}{864}\cos(4\pi x),$$

constructed in analogy with the example in Rice and Rosenblatt (1983) to have the property that

$$f^{(4)}(0) = \frac{1504}{729}\pi^4,$$

$$f^{(5)}(0) = f^{(6)}(0) = f^{(7)}(0) = f^{(4)}(1) = f^{(5)}(1) = f^{(6)}(1) = f^{(7)}(1) = 0.$$

We observe $y_i = f(x_i) + \varepsilon_i$ for $i = 1, \dots, 100$ and $x_i = (2i - 1)/200$. Assume that the variance of ε_i is 10^{-8}. The ordinary smoothing spline of order 4 will have an integrated mean squared error that is dominated asymptotically by bias at the boundary, since the function f does not satisfy the natural boundary conditions.

For this particular design and error variance, the minimum mean squared error at the 100 design points occurs when $\lambda \approx 10^{-12.4}$. The mean squared bias is 1.089×10^{-10} and the mean variance is 1.213×10^{-9} for a mean squared error of 1.322×10^{-9}. The bias and variance as functions of x are plotted as the solid lines in Figures 1 and 2. The relaxed boundary smoothing spline of order 4 has minimum mean squared error when $\lambda \approx 2 \times 10^{-5}$. The

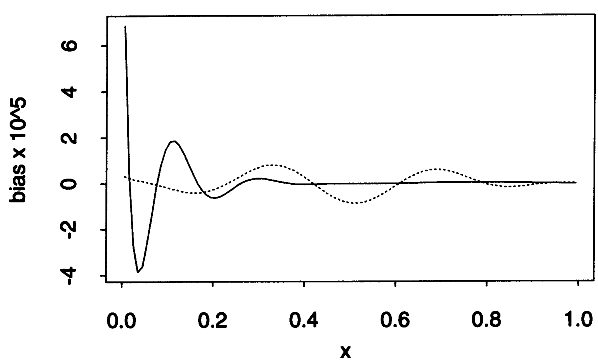

FIG. 1. *Bias in estimating* $f(x) = \cos(\pi x) + \frac{2}{27}\cos(2\pi x) - \frac{1}{729}\cos(3\pi x) - \frac{1}{864}\cos(4\pi x)$. *Solid line*: *smoothing spline*, $m = 4$, $\lambda = 10^{-12.4}$. *Dashed line*: *relaxed boundary smoothing spline*, $m = 4$, $\lambda = 2 \times 10^{-5}$.

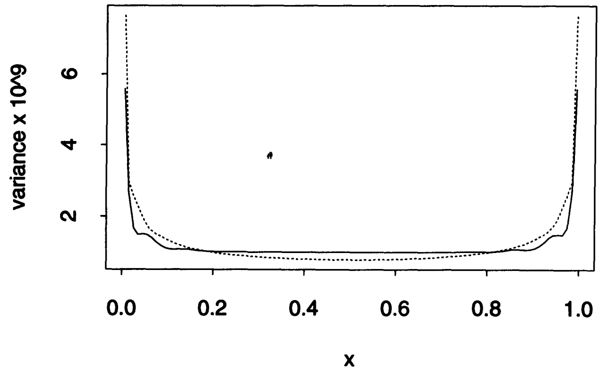

FIG. 2. *Variance in estimating* $f(x) = \cos(\pi x) + \frac{2}{27}\cos(2\pi x) - \frac{1}{729}\cos(3\pi x) - \frac{1}{864}\cos(4\pi x)$ *when* $\sigma^2 = 10^{-8}$. *Solid line: smoothing spline,* $m = 4$, $\lambda = 10^{-12.4}$. *Dashed line: relaxed boundary smoothing spline,* $m = 4$, $\lambda = 2 \times 10^{-5}$.

mean squared bias is 1.767×10^{-11} and the mean variance is 1.247×10^{-9} for a mean squared error of 1.264×10^{-9}. The bias and variance as functions of x are plotted as the dashed lines in Figures 1 and 2.

While the bias does not dominate the variance for the ordinary spline in this design, the ordinary spline bias itself is dominated by boundary contributions. This is clearly not the case for the relaxed boundary spline, where the bias is spread over the entire interval. On the variance side, the relaxed boundary spline has variances which are less than the ordinary spline in the middle of the interval and greater near the endpoints. This is intuitively what would be expected from the shape of the weight function.

Acknowledgments. I would like to thank two anonymous referees and an Associate Editor for helpful comments on a previous version of this paper.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1965). *Handbook of Mathematical Functions.* Dover, New York.

BLOOMFIELD, P. (1976). *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

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

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31** 377–403.

EUBANK, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* Dekker, New York.

EUBANK, R. L. and SPECKMAN, P. L. (1989). A bias reduction theorem with applications in nonparametric regression. *Scand. J. Statist.* To Appear.

KIMELDORF, G. S. and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95.

KUFNER, A. (1980). *Weighted Sobolev Spaces.* Wiley, New York.

NAIMARK, M. A. (1967). *Linear Differential Operators, I.* Ungar, New York.

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

SPECKMAN, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. Unpublished manuscript.

SPECKMAN, P. (1988). Regression analysis for partially linear models. *J. Roy. Statist. Soc. Ser. B* **50** 413–436.

STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

UTRERAS, F. (1988). Boundary effects on convergence rates for Tikhonov regularization. *J. Approx. Theory* **54** 235–249.

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

DEPARTMENT OF APPLIED STATISTICS
UNIVERSITY OF MINNESOTA
1994 BUFORD AVENUE
SAINT PAUL, MINNESOTA 55108