JOURNAL OF APPROXIMATION THEORY 33, 353–369 (1981)

Integrated Mean Squared Error of a Smoothing Spline

JOHN RICE* AND MURRAY ROSENBLATT*

Department of Mathematics, University of California, San Diego, La Jolla, California 92093

Communicated by Oved Shisha and Samuel Karlin

Received September 15, 1980

IN MEMORY OF ARTHUR SARD

1. INTRODUCTION

We consider a global measure (the integrated mean squared error) of the approximation of a function by a cubic smoothing spline, and show that the rate at which this measure converges to zero is determined by the boundary behavior of the function.

We assume we have observations

$$\begin{aligned} x_i &= g(t_i) + \varepsilon_i, \\ t_i &= i/n, \qquad i = 0,\dots, n-1, \end{aligned} \tag{1}$$

where $g(t)$ is an unknown function on $[0, 1]$ and where the ε_i are random errors satisfying

$$\begin{aligned} E(\varepsilon_i) &= 0, \\ E(\varepsilon_i \varepsilon_j) &= \sigma^2 \delta_{ij}, \\ \sigma^2 &> 0. \end{aligned} \tag{2}$$

A smoothing spline estimate of g is a function $f(t; \lambda, n)$ minimizing

$$\frac{1}{n} \sum_{i=0}^{n-1} [f(t_i) - x_i]^2 + \frac{\lambda}{(2\pi)^2} \int_0^1 [f''(t)]^2 \, dt. \tag{3}$$

(The factor $(2\pi)^2$ is merely included for convenience and makes no essential difference.) Smoothing splines were proposed in [5, 7, 9] and some of their properties have been discussed in [2, 8].

* The research was supported in part by NSF Grant MCS-7901800 and ONR Contract N00014-75-C-0428.

0021-9045/81/120353-17$02.00/0

We will examine the behavior of f as an approximate of g as $n \to \infty$ and $\lambda = \lambda(n) \to 0$ by considering a global measure of the approximation—the integrated mean square error (IMSE)

$$E \int_0^1 |f(t; \lambda, n) - g(t)|^2 \, dt \tag{4}$$

as $n \to \infty$. We show that the rate at which this measure converges to zero not only depends on the differentiability of g on $[0, 1]$ but on the differentiability of g periodically extended, and thus on its boundary behavior. These results are established via Fourier analysis.

It is well known that the function minimizing Eq. (3) is a natural cubic spline with knots at the points i/n, $i = 0, \ldots, n-1$.

In this paper $\lambda = \lambda(n)$ is a positive quantity.

2. Statement and Discussion of Results

It is useful to decompose the IMSE into the integrated squared bias and the integrated variance

$$E \int_0^1 [f(t) - g(t)]^2 \, dt = \int_0^1 [Ef(t) - g(t)]^2 \, dt + \int_0^1 \sigma^2[f(t)] \, dt, \tag{5}$$

where $\sigma^2[f(t)] = E[f(t) - Ef(t)]^2$.

Initially we will assume that $f, g \in C^2[0, 1]$ with

$$f(0) = f(1), \quad g(0) = g(1) \qquad \text{and} \qquad f'(0) = f'(1), \quad g'(0) = g'(1). \tag{6}$$

Under this condition we have

THEOREM 1. *In order that* $\int_0^1 \sigma^2[f(t)] \, dt \to 0$ *as* $n \to \infty$ *and* $\lambda(n) \to 0$ *it is necessary and sufficient that* $\lambda n^4 \to \infty$. *In this case*

$$\int_0^1 \sigma^2[f(t)] \, dt \cong \frac{\sigma^2}{n\lambda^{1/4}} \int_{-\infty}^{\infty} \frac{dx}{(1 + x^4)^2} + o(n^{-1}\lambda^{-1/4}). \tag{7}$$

We will frequently use the Fourier series of g:

$$g(t) = \sum_{k=-\infty}^{\infty} a_k \exp(2\pi i k t). \tag{8}$$

THEOREM 2. *If $\lambda \to 0$ and $n^4\lambda \to \infty$ as $n \to \infty$ under assumption* (6) *and* $|a_k|^2 = O(|k|^{-5-\epsilon})$ *for some $\varepsilon > 0$ we have*

$$\int_0^1 |Ef(t) - g(t)|^2 \, dt = \sum_{|j| \leqslant n/2} |a_j|^2 \frac{\lambda^2 j^8}{(1 + \lambda j^4)^2} + O(n^{-5-\epsilon}\lambda^{-1/4} + n^{-4}). \tag{9}$$

If

$$|a_k|^2 \cong |k|^{-5-\epsilon}, \qquad 0 < \varepsilon < 4, \tag{10}$$

this expression is

$$\cong \lambda^{1+\epsilon/4}. \tag{11}$$

If

$$\sum |a_k|^2 k^8 < \infty \tag{12}$$

then the expression is

$$\cong \lambda^2. \tag{13}$$

COROLLARY. *Under the conditions of Theorem* 1, *if* $|a_k|^2 \cong k^{-5-\epsilon}$, $0 < \varepsilon < 4$, *the optimal rate of decay of the IMSE is*

$$n^{-(4+\epsilon)/(5+\epsilon)}. \tag{14}$$

If $\sum |a_k|^2 k^8 < \infty$ *the optimal rate of decay is*

$$n^{-8/9}. \tag{15}$$

The latter case of the corollary has also been proved in [8].

The assumption $g(0) = g(1)$, $g'(0) = g'(1)$ can be relaxed and it appears that the rate of convergence is slower yet if these conditions do not hold. The proofs are substantially more complicated and will appear elsewhere.

These results show that the optimal rate at which the IMSE tends to zero is determined by the rate at which the Fourier coefficients of g decay, which depends on the smoothness of g periodically extended. Thus, even though g may be very smooth as a function on $[0, 1]$, if derivatives at 0 and 1 do not match, the rate of decay may be relatively slow.

This seems an undesirable property of the smoothing spline procedure. If the boundary behavior were known, one could modify the procedure by adding an appropriate polynomial to the data. If the boundary behavior is unknown, it may be possible to estimate it from the data and modify the procedure accordingly, but the properties of such a modified procedure are not immediately clear. In some ways the importance of the boundary behavior is not too surprising; the Gibb's phenomenon is well known for classical Fourier series, and in the case of interpolating splines the boundary behavior can determine a global measure of approximation [6]. In the statistical literature, Hall [3] has pointed out the importance of boundary behavior for orthogonal series estimates of a probability density function, and Gasser and Müller [4] note the effect of the boundary on a non-parametric estimate of a regression function. It is noteworthy that a least-squares spline approximation (a spline with fewer knots than there are data points, fitted to the data by least squares) has the property that $\mathrm{IMSE} \cong n^{-8/9}$ if $g \in C^4$ regardless of the boundary behavior [1].

Our results are global and do not indicate in which regions the bias is large. It is plausible that the integrated squared bias is dominated by the integral over a small region near the boundary, and that perhaps the rate of convergence of the IMSE is faster on an interval $[\varepsilon, 1-\varepsilon]$. Cross-validation [2] has been suggested as a procedure for determining a value of λ to minimize IMSE. Since the boundary behavior has a crucial influence on the minimal IMSE, it is likely that it plays an important role in the outcome of cross-validation, perhaps overwhelming the influence of the interior. This is suggested by the results in Rosenblatt [6], where it is shown that the integrated squared error of a "natural spline" interpolator converges to zero at a slower rate than the squared error away from the boundary. This is due to a skin effect near the boundary which dominates the integrated squared error asymptotically. This can be remedied as indicated in the paper by appropriately estimating the derivative at the boundary and incorporating the estimate in the interpolator.

We have carried out our analysis for a cubic smoothing spline. Similar results hold for a more general smoothing spline. We would also expect comparable, or even a more drastic, influence of boundary behavior in the multidimensional case, and that similar phenomena occur in other regularization problems.

We note that the following generalization of the earlier theorems and corollary hold. The details of the proofs are not given since they are quite analogous to the derivations given for the cubic smoothing spline.

THEOREM 3. *Let us consider a smoothing spline estimate f of g (see Eqs.* (1) *and* (2)) *that minimizes*

$$\frac{1}{n}\sum_{i=0}^{n-1} |f(t_i)-x_i|^2 + \frac{\lambda}{(2\pi)^j}\int_0^1 |f^{(j)}(t)|^2\,dt. \tag{16}$$

Assume that $f, g \in C^{(j)}[0,1]$ with

$$f^{(k)}(0)=f^{(k)}(1), \qquad g^{(k)}(0)=g^{(k)}(1), \qquad k=0,1,\ldots,j-1. \tag{17}$$

In order that $\int_0^1 \sigma^2[f(t)]\,dt \to 0$ *as* $n \to \infty$ *and* $\lambda(n) \to 0$ *it is necessary and sufficient that* $\lambda(n)\,n^{2j} \to \infty$. *Then*

$$\int_0^1 \sigma^2[f(t)]\,dt \cong \frac{\sigma^2}{n\lambda^{1/(2j)}} \int_{-\infty}^{\infty} \frac{dx}{(1+x^{2j})^2} + o(n^{-1}\lambda^{1/(2j)}). \tag{18}$$

THEOREM 4. *If* $\lambda \to 0$, $n^{2j}\lambda \to \infty$ *under assumption* (17) *and* $|a_k|^2 = O(|k|^{-2j-1-\epsilon})$ *for some* $\varepsilon > 0$, *then*

$$\int_0^1 [Ef(t) - g(t)]^2\,dt = \sum_{|k| \leqslant n/2} |a_k|^2 \frac{\lambda^2 k^{4j}}{(1+\lambda k^{2j})^2} + O(n^{-2j-1-\epsilon}\lambda^{-1/(2j)} + n^{-2j}) \tag{19}$$

as $n \to \infty$. *If* $|a_k|^2 \cong |k|^{-2j-1-\epsilon}$, $0 < \varepsilon < 2j$, *this expression is*

$$\cong \lambda^{1+\epsilon/(2j)}. \tag{20}$$

If $\sum |a_k|^2 k^{4j} < \infty$ *then the expression is*

$$\cong \lambda^2. \tag{21}$$

COROLLARY. *Under the conditions of Theorem* 4, *if* $|a_k|^2 \cong k^{-2j-1-\epsilon}$, $0 < \varepsilon < 2j$, *the optimal rate of decay of the IMSE is*

$$n^{-(2j+\epsilon)/(2j+1+\epsilon)} \tag{22}$$

while if $\sum |a_k|^2 k^{4j} < \infty$ *the optimal rate of decay is*

$$n^{-4j/(4j+1)}. \tag{23}$$

3. NUMERICAL EXAMPLES

Since the results of our theorems are asymptotic in nature, we thought it interesting to compute some numerical examples. For the first example we consider the function $g(t) = te^{-2t}$, $0 \leqslant t \leqslant 1$. To ascertain the effect of the boundary conditions we also compute a modified estimate in which a cubic polynomial is added to the data and then subtracted from a smoothing spline fit to the modified data. The polynomial $p(t)$ is chosen so that $\tilde{g}(t) = g(t) + p(t)$ satisfies $\tilde{g}^{(k)}(1) - \tilde{g}^{(k)}(0) = 0$, $k = 0, 1, 2$.

The results for $n = 20$ and 100 are summarized in Table I, which shows for several values of λ the integrated squared bias (B^2) of the smoothing spline, the integrated squared bias (B^2 mod) of the modified estimate, V/σ^2

TABLE I

$\lambda/(2\pi)^2$	B^2	B^2 (mod.)	V/σ^2	σ (opt.)
		$(n = 20)$		
10^{-1}	1.2×10^{-3}	8.2×10^{-5}	9.5×10^{-2}	1.1×10^{-1}
10^{-2}	8.7×10^{-4}	6.0×10^{-5}	1.0×10^{-1}	9.3×10^{-2}
10^{-3}	1.8×10^{-4}	1.3×10^{-5}	1.2×10^{-1}	3.9×10^{-2}
10^{-4}	1.6×10^{-5}	1.2×10^{-6}	1.7×10^{-1}	1.0×10^{-2}
10^{-5}	1.0×10^{-6}	8.3×10^{-8}	2.6×10^{-1}	2.5×10^{-3}
		$(n = 100)$		
10^{-1}	1.3×10^{-3}	8.9×10^{-5}	6.0×10^{-3}	4.7×10^{-1}
10^{-2}	9.0×10^{-4}	6.0×10^{-5}	1.0×10^{-2}	3.0×10^{-1}
10^{-3}	2.1×10^{-4}	1.3×10^{-5}	1.8×10^{-2}	1.1×10^{-1}
10^{-4}	2.2×10^{-5}	1.2×10^{-6}	3.3×10^{-2}	2.6×10^{-2}
10^{-5}	1.8×10^{-6}	1.0×10^{-7}	5.9×10^{-2}	5.5×10^{-3}

(V is the integrated variance), and the value of σ for which the corresponding value of λ is optimal for the smoothing spline. We note that $g(t)$ varies from 0 to 1.8×10^{-1}; thus $\sigma = 10^{-1}$ would give additive noise that greatly masks the underlying function, whereas $\sigma = 10^{-2}$ would be 10% noise at the maximum value of g. From Table I it is seen that the modification reduces B^2 by about a factor of 10 for all given values of λ. Figures 1 and 2 show the local bias for $n = 20$, $\lambda/(2\pi)^2 = 10^{-5}$ and Figs. 3 and 4 show the bias for $n = 100$, $\lambda/(2\pi)^2 = 10^{-5}$. There is little qualitative difference between the bias for $n = 20$ and $n = 100$. The major contribution to B^2 for the smoothing

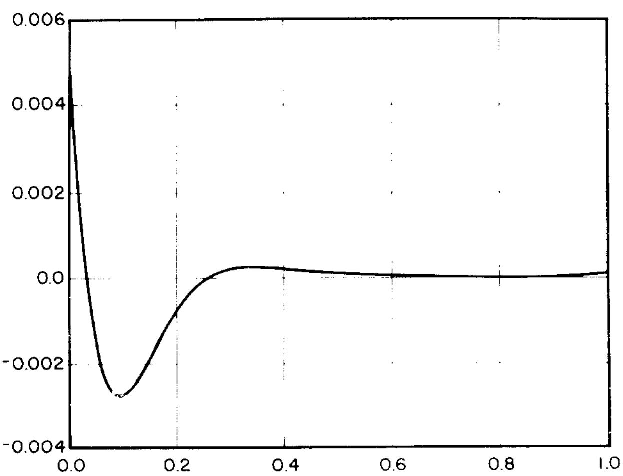


FIG. 1. Bias of smoothing splines. $g(t) = te^{-2t}$, $n = 20$, $\lambda/(2\pi)^2 = 10^{-5}$.

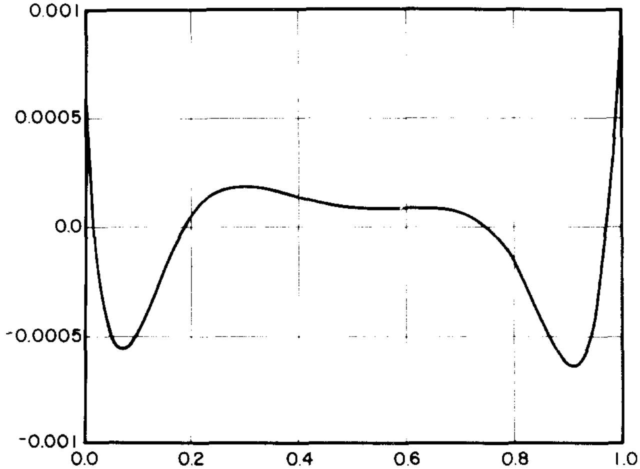


FIG. 2. Bias of modified smoothing spline. $g(t) = te^{-2t}$, $n = 20$, $\lambda/(2\pi)^2 = 10^{-5}$.

spline comes from the boundary region near $t = 0$, where the first and second derivatives are large. It appears that the width of this region is proportional to $\lambda^{1/4}$, which may correspond to something like a bandwidth for the smoothing procedure. Figures 5 and 6 show the bias and modified bias for $n = 20$, $\lambda/(2\pi)^2 = 10^{-3}$. This corresponds to smoothing with a wider bandwidth and the bias is spread more throughout the interval. The results for $n = 100$ are very similar.

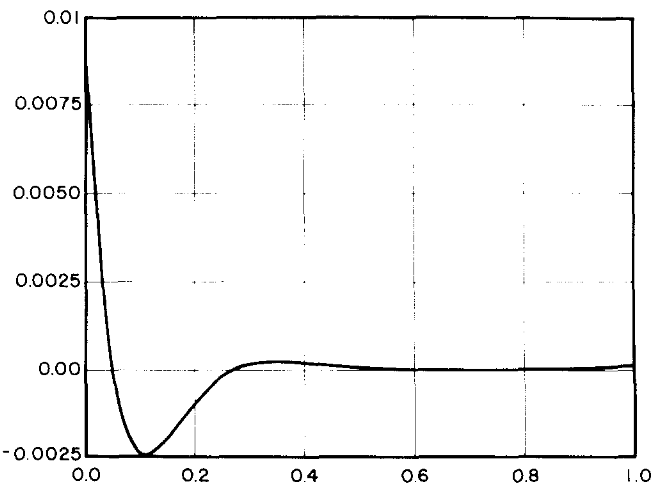


FIG. 3. Bias of smoothing spline. $g(t) = te^{-2t}$, $n = 100$, $\lambda/(2\pi)^2 = 10^{-5}$.

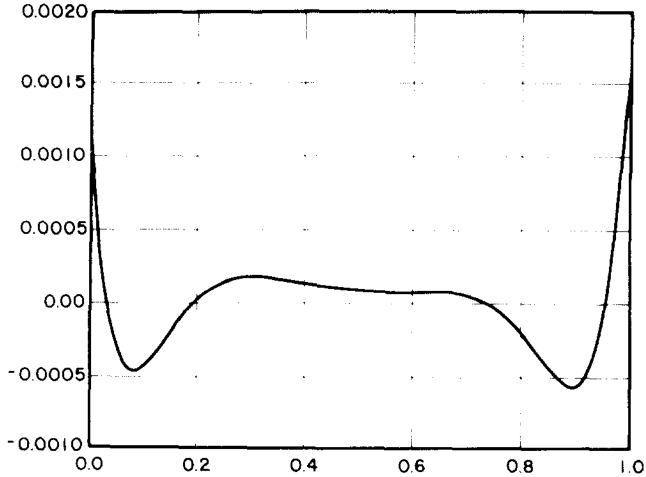


FIG. 4. Bias of modified smoothing spline. $g(t) = te^{-2t}$, $n = 100$, $\lambda/(2\pi)^2 = 10^{-5}$.

For a second example we chose the function $g(t) = (t + 0.1)^{1/2}(1.1 - t)^{1/4}$, $0 \leqslant t \leqslant 1$. The maximum value of g is approximately 1. Table II summarizes the results for various values of λ. In this case the modification appears to be more effective for smaller values of λ. Figures 7 and 8 show the bias for $n = 100$, $\lambda/(2\pi)^2 = 10^{-3}$. Again the principal contributions to B^2 for the smoothing spline come from regions near the boundaries. This effect becomes more pronounced as λ decreases.

Although the smoothing spline is in a sense a third order method, it is

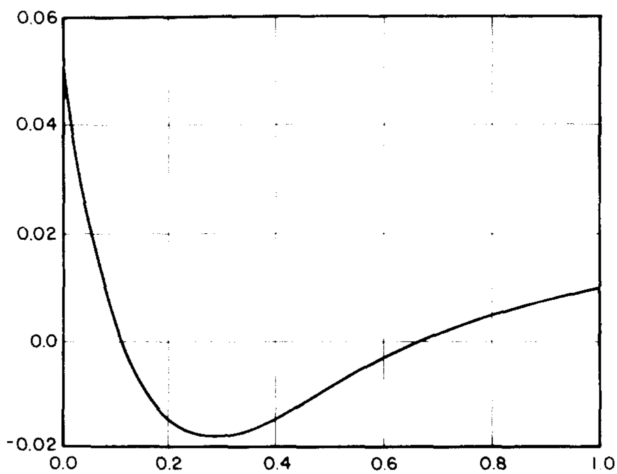


FIG. 5. Bias of smoothing spline. $g(t) = te^{-2t}$, $n = 20$, $\lambda/(2\pi)^2 = 10^{-3}$.

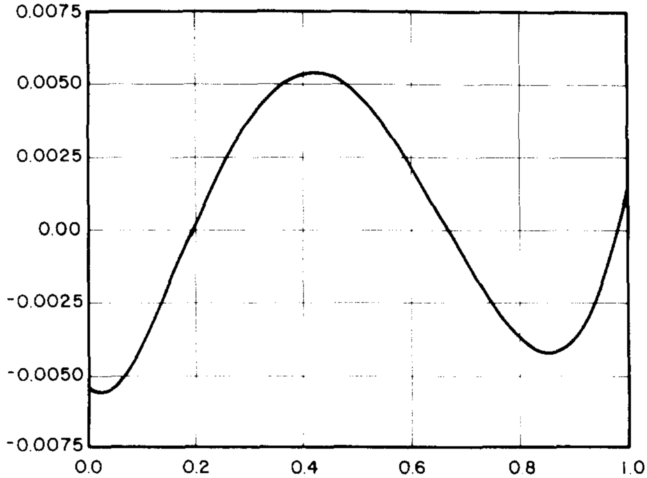


FIG. 6. Bias of modified smoothing spline. $g(t) = te^{-2t}$, $n = 20$, $\lambda/(2\pi)^2 = 10^{-3}$.

biased for cubic polynomials. Figures 9 and 10 show the bias in approximating the function $g(t) = (t - \frac{1}{4})(t - \frac{1}{2})(t - \frac{3}{4})$ for $\lambda/(2\pi)^2 = 10^{-3}$ and 10^{-5}. Again we see that the boundary regions make the major contribution to B^2 for small λ, whereas the bias is spread throughout the interval for larger λ.

TABLE II

$\lambda/(2\pi)^2$	B^2	B^2 (mod.)	V/σ	σ (opt.)
		($n = 20$)		
10^{-2}	2.7×10^{-3}	1.1×10^{-3}	1.0×10^{-1}	1.6×10^{-1}
10^{-3}	3.9×10^{-4}	2.0×10^{-4}	1.2×10^{-1}	5.7×10^{-1}
10^{-4}	4.0×10^{-5}	1.3×10^{-5}	1.7×10^{-1}	1.5×10^{-2}
10^{-5}	4.5×10^{-6}	1.6×10^{-6}	2.6×10^{-1}	4.2×10^{-3}
		($n = 100$)		
10^{-2}	2.8×10^{-3}	1.1×10^{-3}	1.0×10^{-2}	5.3×10^{-1}
10^{-3}	5.1×10^{-4}	2.1×10^{-4}	1.8×10^{-2}	1.7×10^{-1}
10^{-4}	5.6×10^{-5}	1.1×10^{-5}	3.3×10^{-2}	4.1×10^{-2}
10^{-5}	7.0×10^{-6}	1.6×10^{-6}	5.9×10^{-2}	1.0×10^{-2}

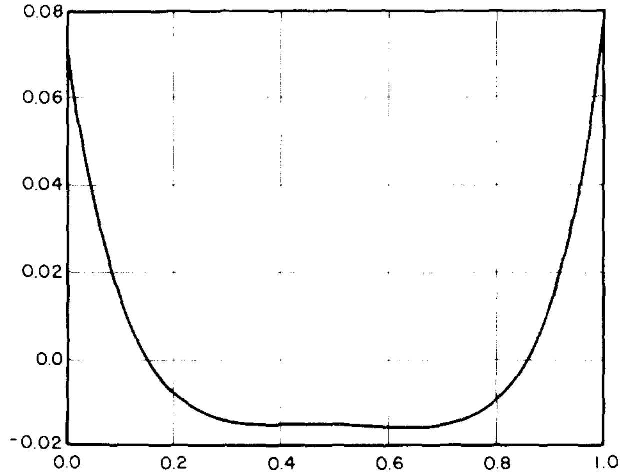


FIG. 7. Bias of smoothing spline. $g(t) = (t + 0.1)^{1/2}(1.1 - t)^{1/4}$, $n = 100$, $\lambda/(2\pi)^2 = 10$

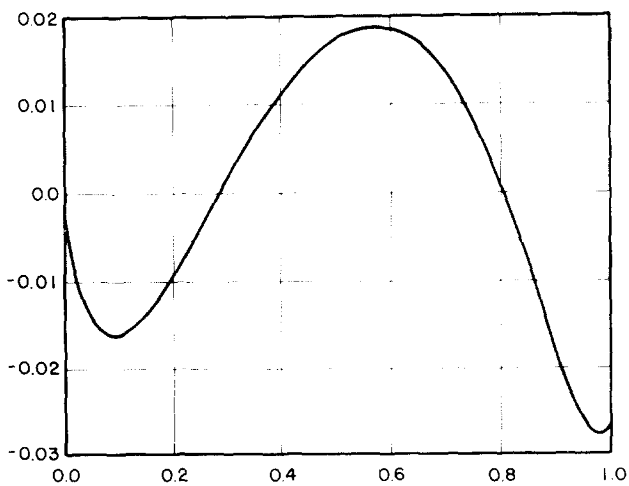
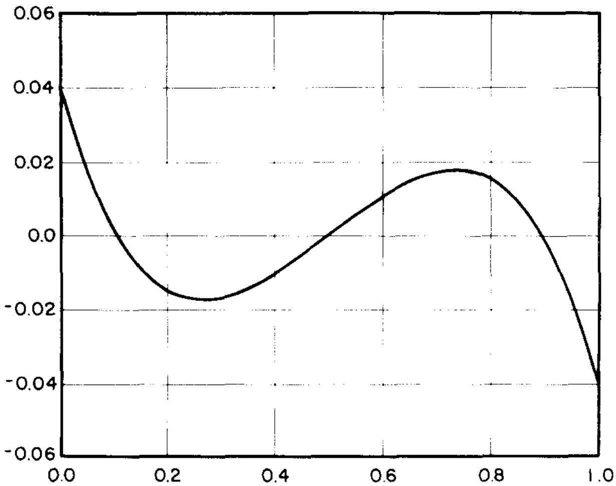
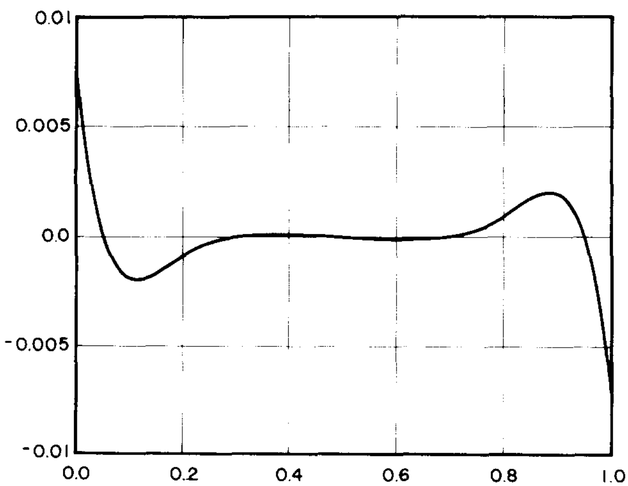


FIG. 8. Bias of modified smoothing spline. $g(t) = (t + 0.1)^{1/2}(1.1 - t)^{1/4}$, $n = 100$. $\lambda/(2\pi)^2 = 10^{-3}$.



Bias of smoothing spline. $g(t) = (t - \frac{1}{4})(t - \frac{1}{2})(t - \frac{3}{4})$, $n = 100$, $\lambda/(2\pi)^2 = 10^{-3}$.



Bias of smoothing spline. $g(t) = (t - \frac{1}{4})(t - \frac{1}{2})(t - \frac{3}{4})$, $n = 100$, $\lambda/(2\pi)^2 = 10^{-5}$.

4. Proofs of Theorems

We first find a Fourier representation of the function f which minimizes (3). We let

$$f(t) = \sum_{j=-\infty}^{\infty} c_j \exp(2\pi ijt),$$

$$\hat{x}_j = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} x_k \exp(-2\pi ijk/n),$$

$$\tilde{c}_j = \sum_{k=-\infty}^{\infty} c_{j+kn}.$$

Since the fitted function is real, $\bar{c}_j = c_{-j}$. Inserting this representation into Eq. (3) we have, after some simplifications using the orthogonality of the exponential functions,

$$(3) = T(\underset{\sim}{c}) = \sum_{j=0}^{n-1} \left| \tilde{c}_j - \frac{\hat{x}_j}{\sqrt{n}} \right|^2 + \lambda \sum_{j=0}^{n-1} \sum_{s=-\infty}^{\infty} |j+sn|^4 \, |c_{j+sn}|^2.$$

Since this is a sum of nonnegative terms, we minimize for each fixed j. For $j=0$ it is clear that

$$c_0 = \hat{x}_0/\sqrt{n}, \qquad c_{sn} = 0, \quad s \neq 0. \tag{24}$$

For each $j \geqslant 1$, the minimizing coefficients c_{j+sn} are

$$c_{j+sn} = \frac{-1}{\lambda(j+sn)^4} (\tilde{c}_j - \hat{x}_j/\sqrt{n}).$$

Summing over all s,

$$\tilde{c}_j = -\frac{(\tilde{c}_j - \hat{x}_j/\sqrt{n})}{\lambda} \sum_{s=-\infty}^{\infty} \frac{1}{(j+sn)^4}$$
$$= -(\tilde{c}_j - \hat{x}_j/\sqrt{n}) \cdot \frac{r_j}{\lambda},$$

say, where

$$r_j = \sum_{s=-\infty}^{\infty} \frac{1}{(j+sn)^4}.$$

Thus

$$\tilde{c}_j = \frac{r_j}{\lambda + r_j} \frac{\hat{x}_j}{\sqrt{n}}$$

and

$$c_{j+sn} = \frac{1}{(j+sn)^4} \frac{1}{\lambda + r_j} \frac{\hat{x}_j}{\sqrt{n}}. \tag{25}$$

It can be verified that the $\underset{\sim}{c}$ thus determined is in fact the minimizer of $T(\underset{\sim}{c})$ by checking that any variation increases $T(\underset{\sim}{c})$. The minimizer $\underset{\sim}{c}$ gives

$$T(\underset{\sim}{c}) = \frac{\lambda}{n} \sum_{j=1}^{n-1} \frac{|\hat{x}_j|^2}{\lambda + r_j}.$$

We now express the IMSE in terms of $\underset{\sim}{c}$. Denoting the Fourier coefficients of g by

$$a_j = \int_0^1 g(t) \exp(-2\pi ijt)\, dt$$

we can express the IMSE as

$$E \int_0^1 [f(t) - g(t)|^2\, dt = E \sum_{j=-\infty}^{\infty} |c_j - a_j|^2,$$

and we have

$$E\,|c_j - a_j|^2 = |Ec_j - a_j|^2 + \sigma^2(c_j).$$

In the case $j = 0$, $\sigma^2(c_0) = \sigma^2/n$ and $Ec_0 = \tilde{a}_0$. Also for $s \neq 0$, $c_{sn} = 0$, so that $\sigma^2(c_{sn}) = 0$.

For $j = 1,\dots, n-1$, noting that $E(\hat{x}_j/\sqrt{n}) = \tilde{a}_j$, we have

$$Ec_{j+sn} = \frac{1}{(j+sn)^4} \frac{1}{\lambda + r_j} \tilde{a}_j$$

and

$$\sigma^2(c_{j+sn}) = \frac{1}{(j+sn)^8} \frac{1}{(\lambda + r_j)^2} \frac{\sigma^2}{n}.$$

To summarize, we have

LEMMA 1. *If* $g \in C^2[0, 1]$ *and* $g(0) = g(1)$, $g'(0) = g'(1)$, *the* ς *minimizing* $T(\varsigma)$ *is given by Eqs.* (24) *and* (25). *The resulting IMSE is*

$$\int_0^1 |Ef(t) - g(t)|^2 \, dt + \int_0^1 \sigma^2 |f(t)| \, dt = B^2 + V,$$

say, where

$$B^2 = |\tilde{a}_0 - a_0|^2 + \sum_{j=1}^{n-1} \sum_{s=-\infty}^{\infty} \left| \frac{\tilde{a}_j}{(j+sn)^4(\lambda + r_j)} - a_{j+sn} \right|^2,$$

$$V = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} \sum_{j=1}^{n-1} \sum_{s=-\infty}^{\infty} \frac{1}{(j+sn)^8} \frac{1}{(\lambda + r_j)^2}.$$

Proof of Theorem 1. We approximate the expression for V^2 above by a simpler sum and then approximate this latter sum by an integral. We first note that we may rewrite

$$V^2 = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} \sum_{|j| \leqslant n/2}' \frac{1}{(\lambda + r_j)^2} \sum_{s=-\infty}^{\infty} \frac{1}{(j+sn)^8} + O(n^{-8}\lambda^{-2}),$$

where $\sum'$ denotes summing over $j \neq 0$. This follows since V^2 is of the form $\sum_1^{n-1} u_j$ and $u_{n-k} = u_{-k}$. The error is incurred in discarding a single term. The sum can be written as

$$\frac{\sigma^2}{n} \sum_{|j| \leqslant n/2}' \frac{1}{(\lambda + r_j)^2} \left(\frac{1}{j^8} + e_j \right) = \frac{\sigma^2}{n} \sum_{|j| \leqslant n/2}' \frac{1}{(\lambda + r_j)^2} \frac{1}{j^8} + R.$$

$e_j = O(n^{-8})$ uniformly in j so that

$$R = O\left(n^{-9} \sum_{|j| \leqslant n/2}' \frac{1}{(\lambda + r_j)^2} \right) = O(n^{-8}\lambda^{-9/4}).$$

We now estimate $(1/(\lambda + r_j)^2)(1/j^8)$ by $1/(1 + \lambda j^4)^2$:

$$\frac{1}{j^8(\lambda + r_j)^2} = \frac{1}{(1 + \lambda j^4 + \sum_{s \neq 0} (j/(j+ns))^4)^2}$$

$$= \frac{1}{(1 + \lambda j^4)^2} + d_j,$$

where

$$|d_j| \leqslant \frac{K}{(1 + \lambda j^4)^3} \left(\frac{j}{j+n} \right)^4,$$

K is a constant. Also

$$\frac{\sigma^2}{n}\sum_{|j|\leqslant n/2}{}' d_j \leqslant \frac{\sigma^2}{n}\sum_{|j|\leqslant n/2}{}' \left(\frac{j/n}{j/n+1}\right)^4 \frac{1}{(1+\lambda j^4)^3}$$
$$\leqslant \frac{\sigma^2}{n^5}\sum_{|j|\leqslant n/2}{}' \frac{j^4}{(1+\lambda j^4)^3}$$
$$= O(n^{-5}\lambda^{-5/4}),$$

where we have estimated the sum by an integral.

Finally, another integral estimate gives

$$\frac{\sigma^2}{n}\sum_{|j|\leqslant n/2} \frac{1}{(1+\lambda j^4)^2} \cong \frac{\sigma^2}{n^{1/4}}\int_{-\infty}^{\infty} \frac{dx}{(1+x^4)^2}.$$

This establishes the theorem.

Proof of Theorem 2. We first approximate

$$|a_0-\tilde{a}_0|^2 + \sum_{j=1}^{n-1}\sum_{s=-\infty}^{\infty}\left|\frac{\tilde{a}_j}{(j+sn)^4(\lambda+r_j)} - a_{j+sn}\right|^2$$

by

$$\sum_{|k|\leqslant n/2}{}' \left|\frac{a_k}{k^4(\lambda+r_k)} - a_k\right|^2.$$

Writing $\tilde{a}_j = a_j + (\tilde{a}_j - a_j)$, the error in this approximation is of the order

$$|a_0-\tilde{a}_0|^2 + \sum_{j=1}^{n-1}\sum_{|j+sn|>n/2} \frac{|\tilde{a}_j|^2}{(j+sn)^8(\lambda+r_j)^2}$$
$$+ \sum_{|k|>n/2}|a_k|^2 + \sum_{j=1}^{n-1}\sum_{s=-\infty}^{\infty} \frac{|\tilde{a}_j-a_j|^2}{(j+sn)^8(\lambda+r_j)^2}.$$

The first term is $O(n^{-5-\epsilon})$, and the third term is $O(n^{-4-\epsilon})$. The second term is

$$\sum_{j=1}^{n-1}\frac{|\tilde{a}_j|^2}{(\lambda+r_j)^2}\sum_{|j+sn|>n/2}\frac{1}{(j+sn)^8} = O\left(n^{-8}\sum_{j=1}^{n-1}\frac{|\tilde{a}_j|^2}{(\lambda+r_j)^2}\right).$$

Using $|a_j|^2 = O(j^{-5-\epsilon})$ and estimating $(\lambda+r_j)^{-2}$ as in the proof of Theorem 1 by $j^8(1+\lambda j^4)^2$ this expression is

$$O\left(n^{-8}\sum_{|j|\leqslant n/2}\frac{j^{3-\epsilon}}{(1+\lambda j^4)^2}\right) = O(n^{-8}\lambda^{-1+\epsilon/4})$$

by an integral estimate. The last term is from similar estimates

$$O\left(n^{-5-\epsilon} \sum_{|j| \leqslant n/2} \frac{1}{j^8(\lambda + r_j)^2}\right) = O(n^{-5-\epsilon}\lambda^{-1/4}).$$

Now

$$\sum_{|k| \leqslant n/2}' \left|\frac{a_k}{k^4(\lambda + r_k)} - a_k\right|^2 = \sum_{|k| \leqslant n/2}' |a_k|^2 \left(\frac{1}{k^4(\lambda + r_k)} - 1\right)^2$$

and

$$\left(\frac{1}{k^4(\lambda + r_k)} - 1\right)^2 = \left(\frac{1}{\lambda k^4 + 1 + \delta_k} - 1\right)^2$$
$$= \left(\frac{\lambda k^4 + \delta_k}{\lambda k^4 + \delta_k + 1}\right)^2,$$

where $\delta_k = O((k/(k+n))^4)$. Now if $f(x) = (x/(x+1))^2$ then $f'(x) = 2x/(x+1)^3$ is bounded for $x > 0$, and $f(x+\delta) = f(x) + O(\delta)$ for $\delta > 0$. Thus

$$\sum_{|k| \leqslant n/2}' |a_k|^2 \left(\frac{1}{k^4(\lambda + r_k)} - 1\right)^2 = \sum_{|k| \leqslant n/2}' |a_k|^2 \frac{\lambda^2 k^8}{(1 + \lambda k^4)^2}$$
$$+ O\left(\sum_{|k| \leqslant n/2}' |a_k|^2 \left(\frac{k/n}{k/n + 1}\right)^4\right)$$
$$= \sum_{|k| \leqslant n/2}' |a_k|^2 \frac{\lambda^2 k^8}{(1 + \lambda k^4)^2} + O(n^{-4}).$$

To estimate the magnitude of the sum we split the range of summation into two parts, summing over $k \leqslant \lambda^{-1/4}$ and $\lambda^{-1/4} \leqslant k \leqslant n/2$. The resulting sums are of the same order of magnitude as

$$\sum_{|k| \geqslant \lambda^{-1/4}}' |a_k|^2 + \lambda^2 \sum_{|k| \leqslant \lambda^{-1/4}}' |a_k|^2 k^8.$$

If $|a_k|^2 \cong k^{-5-\epsilon}$, $0 < \varepsilon < 4$, then each of these terms is of the order $\lambda^{1+\epsilon/4}$. If $\sum |a_k|^2 k^8 < \infty$ then the primary contribution is from the second term which is of order λ^2.

Proof of Corollary. (1) If $|a_k|^2 \cong k^{-5-\epsilon}$, $0 < \varepsilon < 4$ then the IMSE is of the order

$$c_1 \lambda^{1+\epsilon/4} + c_2 n^{-1} \lambda^{-1/4}.$$

If $\lambda = n^{-\alpha}$ the optimal α is seen to be $\alpha = 4/(5+\varepsilon)$ in which case the IMSE is of order $n^{-(4+\epsilon)/(5+\epsilon)}$.

(2) If $\sum |a_k|^2 k^8 < \infty$ the IMSE is of the order

$$c_1 \lambda^2 + c_2 n^{-1} \lambda^{-1/4}.$$

The optimal λ in this case is $\lambda \cong n^{-4/9}$ in which case IMSE $\cong n^{-8/9}$.

5. Final Remarks

One should note that the conditions

$$f^{(k)}(0) = f^{(k)}(1), \qquad k = 0, 1, \ldots, j-1$$

are not usually imposed on smoothing splines minimizing Eq. (16). Since this information is incorporated in our approximation, one would expect the asymptotic behavior of our approximation (in terms of expected mean square error) to be at least as good as that of the standard smoothing spline.

References

1. G. Agarwal and W. Studden, Asymptotic integrated mean square error using least squares and bias minimizing splines, *in* "Department of Statistics Mimeograph Series No. 79-17," Purdue University, 1980
2. P. Craven and G. Wahba, Smoothing noisy data with spline functions, *Numer. Math.* **31** (1979), 377–403.
3. P. Hall, On trigonometric series estimates of densities, *Ann. Stat.*, in press.
4. T. Gasser and H.-G. Müller, Kernel estimation of regression functions, *in* "Smoothing Techniques for Curve Estimation" (Gasser and Rosenblatt, Eds.), Lecture Notes in Mathematics No. 757, pp. 23–68, Springer-Verlag, New York/Berlin, 1979.
5. C. Reinsch, Smoothing by spline functions, *Numer. Math.* **10** (1967), 177–183.
6. M. Rosenblatt, Asymptotics and representation of cubic splines, *J. Approx. Theory* **17** (1976), 332–343.
7. J. Schoenberg, Spline functions and the problem of graduation, *Proc. Nat. Acad. Sci. U.S.A.* **52** (1964), 947–950.
8. G. Wahba, Smoothing noisy data with spline functions, *Numer. Math.* **24** (1975), 383–393.
9. E. Whittaker, On a new method of graduation, *Proc. Edinburgh Math. Soc.* **41** (1923), 63–75.