yandell

DEPARTMENT OF STATISTICS

The University of Wisconsin
1210 W. Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 828

May 1988

# LOCALLY ADAPTIVE SMOOTHING SPLINES

by

Joan G. Staniswalis
and
Brian S. Yandell

**Abstract**

Locally adaptive smoothing splines combine features of variable kernels and smoothing splines to allow for local adaptive fitting and for a minimization of integrated mean squared error. Basically, one first adaptively fits a function with a local bandwidth kernel estimator, followed by a global fit to the presmoothed data using a penalized likelihood. One has to be careful to allow the variable kernel bandwidth to converge slightly faster than the spline bandwidth. We present some properties of the estimator and demonstrate its practical use through simulations and data analysis.

# Locally Adaptive Smoothing Splines

Joan G. Staniswalis* Brian S. Yandell†

Virginia Commonwealth U. U. of Wisconsin-Madison

May 23, 1988

## 1 Introduction

Suppose one obtains observations $Y_0, \cdots, Y_{n-1}$ of the form

$$Y_i = f(x_i) + \epsilon_i \ ,$$

with $x_i = i/n$, $i = 0, \cdots, n-1$. It is assumed that the errors $\epsilon_i$ contaminating the observations of $f(x)$ are independent random variates with mean 0 and variance $\sigma^2$. Of interest is the nonparametric estimation of the function $f \in C^k[0,1], k \geq 4$, such that $f^{(k)} \in Lip_\gamma[0,1]$, by a function from the Sobolev space of order two,

$$W_2^2 = \{g | g, g' \in C[0,1] \text{ and } g^{(2)} \in L_2[0,1]\}.$$

*Supported by NSF grant DMS 8717560 while visiting the University of Wisconsin-Madison

†Supported by NSF grant DMS 8704341 and USDA-CSRS grant 511-100

We propose a nonparametric estimate of a curve which is a hybrid of kernel smoothing (Priestly and Chao [**1972**]) and spline smoothing (Reinsch [**1967**]; Wahba and Wold [**1975**]).

We combine the ideas of the computationally simple kernel estimator,

$$f_n(x,b) = \sum_{i=0}^{n-1} w\left(\tfrac{x-x_i}{b(x)}\right) Y_i \Big/ \sum_{i=0}^{n-1} w\left(\tfrac{x-x_i}{b(x)}\right), \ 0 < b(x) < 1/2 \ , \tag{1}$$

with the theoretically and aesthetically appealing cubic smoothing spline (Wahba [**1975**]). The kernel estimator minimizes the following weighted least squares criterion (Cleveland [**1979**]; Staniswalis [**1987**]) at fixed $x$,

$$(nb)^{-1} \sum_{i=0}^{n-1} w\left(\tfrac{x-x_i}{b}\right)[Y_i - f(x)]^2 \tag{2}$$

while the cubic smoothing spline minimizes the penalized least squares (Good and Gaskins [**1971**])

$$n^{-1} \sum_{i=0}^{n-1} (Y_i - g(x_i))^2 + \lambda \int_0^1 [g^{(2)}(x)]^2 dx \tag{3}$$

among all $g \in W_2^2$, with $\lambda \geq 0$ the smoothness constant. Our purpose is to show that the minimizer of

$$U(f) = n^{-1} \sum_{j=0}^{n-1} \left[ [W_j^{-1} \sum_{i=0}^{n-1} w\left(\frac{x_j - x_i}{b(x_j)}\right) (Y_i - f(x_j))^2 \right] + \lambda \int_0^1 [f^{(2)}(x)]^2 dx \ , \tag{4}$$

where $W_j = \sum_{i=0}^{n-1} w\left(\frac{x_j - x_i}{b(x_j)}\right)$, has smaller IMSE than that of the spline estimator of (3). Thus a locally adaptive smoothing spline is proposed, which improves on the IMSE of the global smoothing spline, using ideas and methods from adaptive kernel estimators.

Computer simulations and real data examples are used to show that the asymptotic analysis is actually providing us with useful guidance in the selection of $b(x)$ in the case of finite sample sizes.

## 2 Kernels and Splines

We consider kernel estimators of the form (1) where $w(v)$ with $k$ continuous derivatives is a symmetric kernel of order $k$ with compact support on [-1,1], satisfying the moment conditions

$$\int_{-1}^{1} v^j w(v)dv = \begin{cases} 1 & j=0 \\ 0 & j=1,\cdots,k-1 \\ W_k \neq 0 & j=k \end{cases} .$$

The bandwidth $b \in (0, 1/2]$ is a function of $n$ and $x$ although the notation does not reflect this. If $b \to 0$ and $nb \to \infty$ as $n \to \infty$ then $f_n(x, b)$ is a consistent estimator of $f(x)$ (Gasser and Muller [**1979**]). The bias of the kernel estimator is proportional to $b^k f^{(k)}(x)$ and the variance is inversely proportional to $nb$. Therefore, the bias of $f_n(x, b)$ can only be reduced at the cost of increasing the variance.

Rice [**1984b**] proposed a method for estimating the optimal finite sample global bandwidth $b^g$ which minimizes the integrated squared error $IMSE(b)$ of $f_n(x, b)$. The choice of the global bandwidth $b^g$ is governed largely by the peaks and troughs of $f$. This global bandwidth results in a kernel estimate which tracks the observations $Y_i$ in the flat regions of $f$, rather than averaging out the contaminating noise. Staniswalis [**1985**] proposed a method for estimating the optimal finite sample local bandwidth $b^\ell(x)$ which minimizes the mean squared error $MSE(x; b)$ of $f_n(x, b)$. The local bandwidth $b^\ell(x)$ results in a kernel estimate with a small bandwidth near peaks of $f$ (reducing bias) and a larger bandwidth in the flat regions of $f$ (reducing variance). Kernel estimators using data adaptive global and local bandwidth selection procedures have been shown to exhibit these properties as well (Staniswalis [**1985**]). However, the local bandwidth can be difficult to estimate in places where $f^{(k)}$ has high curvature. Consequently, the greatest incentive to using a kernel estimator with a local bandwidth over a global bandwidth selection procedure is in the reduction in variance realized in places where $f^{(k)}$ is very smooth.

The smoothing spline $\mu_{n,\lambda}$ which minimizes (3) is a piecewise cubic polynomial with knots at $x_1, \cdots, x_n$ and two continuous derivatives satisfying the boundary conditions

$$\mu_{n,\lambda}{}^{(i)}(0) = \mu_{n,\lambda}{}^{(i)}(1) = 0 \; for \; i = 2, 3.$$

The smoothing parameter ($\lambda$) can be chosen from the data by either cross validation or maximum likelihood methods (Wahba [**1985**]). Cogburn and Davies [**1974**] and Silverman [**1984**] showed that in the interior of (0,1), the cubic smoothing spline is asymptotically equivalent to a kernel estimator with a global bandwidth of $h(\lambda) = \lambda^{1/4}$ and kernel of order 4 given by

$$S(u) = 2^{-1} \exp(-|u|/\sqrt{2}) \sin(|u|/\sqrt{2} + \pi/4).$$

Tanner and Wong [**1985**] point out that it is desireable for $h(\lambda)$ to depend on the local curvature of $f$.

The optimal rate of convergence for IMSE of a kernel estimator with a kernel of order $k = 4$ is $O(n^{-8/9})$. For general $f \in C^k[0,1]$, i.e., $f$ that does not necessarily satisfy the boundary conditions

$$f^{(i)}(0) = f^{(i)}(1) = 0 \; for \; i = 2, 3 \; , \tag{5}$$

the optimal rate of convergence for the IMSE of $\mu_{n,\lambda}(x)$ is slower than $n^{-8/9}$ (Utreras [**1987**]). If the above boundary conditions are satisfied, then the IMSE of $\mu_{n,\lambda}(x)$ can attain the optimal rate $n^{-8/9}$.

The penalized weighted least squares criterion (4) combines features of the weighted least squares criterion (2) and the penalized least squares criterion (3). When $b(x_j) \leq n^{-1}$, $j = 0, \cdots, n-1$, $U(f)$ reduces to (3) whose minimizer is the cubic smoothing spline. On the other hand, if $\lambda = 0$ then $U(f)$ reduces to (2) whose minimizer is the kernel estimator. The following lemma shows that the minimizer of $U(f)$ is simply the minimizer of the penalized least squares criterion applied to the kernel-smoothed data. Its proof is relegated to the Appendix.

**Lemma 1** *The unique minimizer of $U(f)$ among $f \in W_2^2$ is also the unique minimizer of*

$$n^{-1} \sum_{i=0}^{n-1} [f(x_i) - \tilde{Y}_i]^2 + \lambda \int_0^1 [f^{(2)}(x)]^2 dx$$

*among all $f \in W_2^2$, where $\tilde{Y}_i = W_i^{-1} \sum_{j=0}^{n-1} w\left(\frac{x_j - x_i}{b(x_j)}\right) Y_j$.*

Let $\hat{f}_n(x; \lambda, b)$ denote the minimizer of $U(f)$. The lemma says that $\hat{f}_n$ is a smoothing spline fit to the presmoothed data $(x_i, \tilde{Y}_i)$, $i = 0, \cdots, n-1$. Therefore, $\hat{f}_n(x; \lambda, b)$ can exploit the local smoothness of $f(x)$ through $b(x)$ and the global smoothness of $f$ through $\lambda$.

# 3 Asymptotic Bias and Variance

Of interest are conditions under which the asymptotic IMSE is smaller for $\hat{f}_n(x; \lambda, b)$ than for $\mu_{n,\lambda}(x)$. In order to get some insight into how to select the bandwidth $b(x)$ to achieve this, asymptotic expressions are derived for

$$B^2(\lambda) = \int [E\hat{f}_n(x; \lambda, b) - f(x)]^2 dx$$

and

$$V(\lambda) = \int E[\hat{f}_n(x;\lambda,b) - E\hat{f}_n(x;\lambda,b)]^2 dx \ ,$$

where $b = b(x)$. Set

$$B_0^2(\lambda) = \int [E\mu_{n,\lambda}(x) - f(x)]^2 dx$$

and

$$V_0(\lambda) = \int E[\mu_{n,\lambda}(x) - E\mu_{n,\lambda}(x)]^2 dx \ .$$

It is of interest to select $b(x)$ in such a way that

$$V(\lambda) \leq V_0(\lambda) \ for \ all \ \lambda > 0$$

without suffering a large increase in $B^2(\lambda)$ relative to $B_0^2(\lambda)$.

Lemma 1 of Rice and Rosenblatt [**1981**] showed that if $f(0) = f(1)$ and $f^{(1)}(0) = f^{(1)}(1)$, then

$$V_0(\lambda) \approx n^{-1}\sigma^2 \sum_{j=0}^{n-1} \lambda_j^2,$$

where

$$\lambda_j^2 = \begin{cases} \sum_{s=-\infty}^{\infty}(j+sn)^{-8}(\lambda' + r_j)^{-2} & ; j = 1,\cdots,n-1 \\ 1 & ; j = 0 \end{cases}$$

with $\lambda' = \lambda(2\pi)^4$ and $r_j = \sum_{s=-\infty}^{\infty}(j+sn)^{-4}$. The following lemma provides an asymptotic expression for $V(\lambda)$.

**Lemma 2** . *If* $f(0) = f(1)$ *and* $f^{(1)}(0) = f^{(1)}(1)$, *then* $V(\lambda) = n^{-1}\sigma^2 \sum_{j=0}^{n-1} q_j \lambda_j^2$ , *where*

$$q_j = n^{-1} \sum_{i=0}^{n-1} \sum_{m=0}^{n-1} \sum_{l=0}^{n-1} W_m^{-1} W_i^{-1} w\left(\frac{x_m - x_l}{b(x_m)}\right) w\left(\frac{x_l - x_i}{b(x_i)}\right) \cos[2\pi j(m-i)/n] \ ,$$

*for* $j = 0,\cdots,n-1$.

The $q_j$ are converging to $\cos(0) = 1$ as $n \to \infty$. Furthermore, from the bias properties of kernel estimators (Muller [**1985**]), for large $n$, $0 < q_{n-1} \leq \cdots \leq q_0 \leq 1$. Therefore, for large $n$ the expected result that $V(\lambda) \leq V_0(\lambda)$ follows.

The following lemma allows us to compare $B^2(\lambda)$ with $B_0^2(\lambda)$. the proof is in the Appendix.

**Lemma 3**

$$B^2(\lambda) = B_0^2(\lambda) + \int [\nu_{n,\lambda}(x)]^2 dx + 2\int E[\mu_{n,\lambda}(x) - f(x)]\nu_{n,\lambda}(x)dx$$

*where* $\nu_{n,\lambda}(\lambda)$ *is the smoothing spline with smoothing parameter* $\lambda$ *which is fit to* $(x_i,\ Bias(x_i))$, $i = 0, \cdots, n-1$. *Here* $Bias(x) = E[f_n(x,b)] - f(x)$.

If $b(x)$ is $O[n^{-1/(2k+1)}]$, the optimal rate for minimizing MSE of the kernel estimator, then $\nu_{n,\lambda}(x) = O[n^{-k/(2k+1)}]$. If $f$ satisfies the boundary conditions (5), then $k > 4$ ensures that $B^2(\lambda) \approx B_0^2(\lambda)$. If $f$ does not satisfy (5), then $k \geq 4$ is sufficient for $B^2(\lambda) \approx B_0^2(\lambda)$.

# 4 Simulations and Data Analysis

## 4.1 Simulations

The simulations were performed on the Statistics Research VAX at the University of Wisconsin-Madison. The purpose was to convincingly demonstrate that the locally adaptive smoothing spline has smaller IMSE that the global smoothing spline. A rescaled version of the function used by Wahba and Wold [**1975**] was selected for the simulations

$$f(x) = 4.26[e^{-3.25x} - 4e^{-6.5x} + 3e^{-9.75x}].$$

Independent identically distributed $N(0, \sigma^2)$ contaminating errors for $n = 50$ were generated with the public domain random number generator RNOR. Noisy observations of $f$ on [-1,2] were used by the kernel smoother in order to avoid boundary modifications to the kernel (Rice [**1984a**]). The spline fit to the presmoothed data (the LASS) and the global spline fit used only the region [0,1].

One hundred independent realizations of size 50 of the locally adaptive smoothing spline (LASS) and the global spline smoother were generated. The LASS were created by generating raw data, presmoothing with the local bandwidth kernel smoother of Staniswalis [**1985**], and then applying a cubic spline smoother. The LASS was applied with the kernels of Muller [**1984**].

As described earlier, the advantage to using a locally adaptive spline smoother over a global spline smoother is the reduction in variance where the underlying curve $f$ is very smooth. The function used in the simulation is a paradigm of the undesireable 'wiggliness' which can result from locally undersmoothing the noisy data. Figure 1 is a realization of the two locally adaptive spline smoothers and the global spline smoother for this mixture of exponentials.

The mean squared error of the locally adaptive spline smoother and the global spline smoother were estimated from these realizations. Figure 2 presents the ratio of local to global MSE. Note the reduction in variance achieved by the local spline smoother over the global spline smoother without an increase in bias, particularly for the kernel of order 6. The average (over x) estimated MSE for the locally adaptive spline smoother are .0063 and .0050 for $k = 4$ and 6, respectively. The average estimated MSE for the global spline smoother is .0097. Smoothing the data with higher order kernels allows the locally adaptive smoothing spline to enjoy a large decrease in variance without a subsequent increase in bias.

### 4.2 Data Analysis

The voltage drop data in Ch. 3 (ex. 14) of Eubank **[1988]** was analysed. Figure 3 is a plot of the locally adaptive spline smoother for $k$ = 4, 6 and 8 and the global spline smoother superimposed on the data. The curve $f$ was assumed to be periodic on [0,1] in order to avoid boundary modifications to the kernel (Rice **[1984a]**). Again, it is evident that the higher order kernels relieve the bias problem of the locally adaptive spline smoother while allowing for a decrease in variance over the global spline smoother.

## 5 Appendix

### 5.1 Proof of Lemma 1

$U(f)$ may be minimized by first solving for $\hat{f}_n(x; \lambda, b)$ such that the Gateaux derivative of $U$ at $\hat{f}_n$

$$\phi(\hat{f}_n; g) = \frac{d}{d\delta} U(\hat{f}_n + \delta g)|_{\delta=0} \ ,$$

is 0 for all $g \in W_2^2$. We proceed in this calculation as in Eubank **[1988]**, treating $\lambda$ and $b(x)$ as fixed smoothing parameters independent of $f$.

Then the Gateaux derivative of $U$ at $f$ in the direction $g$ can be written as

$$\phi(f; g) = -2n^{-1} \sum_{j=0}^{n-1} \left[ [W_j^{-1} \sum_{i=0}^{n-1} w \left( \frac{x_j - x_i}{b(x_j)} \right) (Y_i - f(x_j)) \, g(x_j) \right]$$

$$+ 2\lambda \int_0^1 f^{(2)}(x) g^{(2)}(x) dx \ . \tag{6}$$

We need to solve for $\hat{f}_n$ such that

$$\phi(\hat{f}_n; g) = 0 \text{ for all } g \in W_2^2 \ . \tag{7}$$

If $\hat{f}_n$ were a natural spline of order four, then it could be written as

$$\hat{f}_n(x; \lambda, b) = \sum_{t=0}^{n-1} \beta_t e_t(x) \tag{8}$$

for some $\beta_0, \cdots, \beta_{n-1}$ which would depend on $b$ and $\lambda$, where $e_0, \cdots, e_{n-1}$ form a basis for the natural splines of order four on the interval [0,1] with knots at $x_0, \cdots, x_{n-1}$. In this case equations (6) through (8) imply that we need to solve for $\beta_0, \cdots, \beta_{n-1}$ such that

$$(n\lambda)^{-1} \sum_{j=0}^{n-1} \left[ [W_j^{-1} \sum_{i=0}^{n-1} w\left(\frac{x_j - x_i}{b(x_j)}\right) \left(Y_i - \sum_{t=0}^{n-1} \beta_t e_t(x_j)\right) g(x_j) \right]$$

$$= \int_0^1 \left( \sum_{t=0}^{n-1} \beta_t e_t^{(2)}(x) \right) g^{(2)}(x) dx \tag{9}$$

for all $g \in W_2^2$. Using Lemma 5.1 of Eubank [1988], which is attributed to Lyche and Schumaker [1973], the right hand side of (9) is

$$6 \sum_{j=0}^{n-1} g(x_j) \left[ \sum_{t=0}^{n-1} \beta_t \delta_{jt} \right] ,$$

where

$$e_t(x) = \sum_{j=0}^{1} \theta_{jt} x^j + \sum_{j=0}^{n-1} \delta_{jt} (x - x_j)_+^3 \ .$$

Therefore, $\beta_0, \cdots, \beta_{n-1}$ must satisfy

$$(n\lambda)^{-1} \sum_{j=0}^{n-1} \left[ [W_j^{-1} \sum_{i=0}^{n-1} w\left(\frac{x_j - x_i}{b(x_j)}\right) \left(Y_i - \sum_{t=0}^{n-1} \beta_t e_t(x_j)\right) g(x_j) \right]$$

$$= 6 \sum_{j=0}^{n-1} g(x_j) \left[ \sum_{t=0}^{n-1} \beta_t \delta_{jt} \right] , \tag{10}$$

for all $g \in W_2^2$.

Equation (10) must hold for all $g \in W_2^2$; it must be that

$$(n\lambda W_j)^{-1} \sum_{i=0}^{n-1} w\left(\frac{x_j - x_i}{b(x_j)}\right) \left(Y_i - \sum_{t=0}^{n-1} \beta_t e_t(x_j)\right) = 6 \sum_{t=0}^{n-1} \beta_t \delta_{jt} , \tag{11}$$

for all $j = 0, \cdots, n-1$. Rearranging (11), and using the fact that $W_j^{-1} \sum_{i=0}^{n-1} w\left(\frac{x_j - x_i}{b(x_j)}\right) = 1$, we are reduced to solving for $\beta_0, \cdots, \beta_{n-1}$ such that

$$\sum_{t=0}^{n-1} \beta_t \{e_t(x_j) + 6n\lambda\delta_{jt}\} = \tilde{Y}_j \ , j = 0, \cdots, n-1. \tag{12}$$

Observe that the unique minimizer of

$$n^{-1} \sum_{i=0}^{n-1} (Y_i - f(x_i))^2 + \lambda \int (f^{(2)}(x))^2 dx$$

is $\mu_{n,\lambda}(x) = \sum_{t=0}^{n-1} \gamma_t e_t$, where the $\gamma_0, \cdots, \gamma_{n-1}$ satisfy the equations

$$\sum_{t=0}^{n-1} \gamma_t \{e_t(x_j) + 6n\lambda\delta_{jt}\} = Y_j \ , j = 0, \cdots, n-1, \tag{13}$$

(see ch. 5 of Eubank [**1988**]).

Comparing equations (12) and (13) it is clear that $\hat{f}_n(x; \lambda, b)$ is the natural smoothing spline of order four fit to the presmoothed data $(x_i, \tilde{Y}_i)$, $i = 0, \cdots, n-1$. That is, $\hat{f}_n(x; \lambda, b)$ is the minimizer of

$$n^{-1} \sum_{i=0}^{n-1} (\tilde{Y}_i - f(x_i))^2 + \lambda \int (f^{(2)}(x))^2 dx \ .$$

The proofs in Chapter 5 of Eubank [**1988**] may be adapted to show that $\sum_{t=0}^{n-1} \beta_t e_t(x)$ is in fact the unique minimizer of $U(f)$, the penalized weighted likelihood.

## 5.2 Proof of Lemma 2

Represent both the spline $\hat{f}_n$ fit to the presmoothed data and the smoothe function $f$ in terms of a Fourier series expansion:

$$\hat{f}_n(x;\lambda,b) = \sum_{s=-\infty}^{\infty} c_s e^{2\pi isx}$$

and

$$f(x) = \sum_{s=-\infty}^{\infty} a_s e^{2\pi isx} .$$

For computational convenience, let $\hat{f}_n(x;\lambda,b)$ be the minimizer of

$$n^{-1}\sum_{i=0}^{n-1}(\tilde{Y}_i - f(x_i))^2 + \lambda'(2\pi)^{-4}\int_0^1 (f^{(2)}(x))^2 dx \ , \ \lambda' = \lambda(2\pi)^4.$$

Referring back to the results and methods in Rice and Rosenblatt [**1981**], it can be shown that

$$c_{sn} = \begin{cases} 0 & \text{when } s \neq 0 \\ n^{-1/2}\hat{\tilde{Y}}_0 & \text{when } s = 0 \end{cases}$$

and

$$c_{j+sn} = (j+sn)^{-4}(\lambda' + r_j)^{-1} n^{-1/2}\hat{\tilde{Y}}_j \ , j \geq 1 \ .$$

Here $\hat{\tilde{Y}}_0, \cdots, \hat{\tilde{Y}}_{n-1}$ are the discrete Fourier coefficients of $\tilde{Y}_0, \cdots, \tilde{Y}_{n-1}$; i.e.,

$$\hat{\tilde{Y}}_j = n^{-1/2}\sum_{t=0}^{n-1} \tilde{Y}_t \exp(-2\pi ijt/n)$$

$$= U_j^*(WY) \ ,$$

where $U_j = \{\exp(-2\pi ijt/n)\}_{t=0}^{n-1}$, $W_{ij} = W_i^{-1} w\left(\frac{x_i - x_j}{b(x_i)}\right)$, $i,j = 0,\cdots,n-1$, and $Y = (Y_0,\cdots,Y_{n-1})^T$. It follows that $var(\hat{\tilde{Y}}_j) = \sigma^2 U_j^* W W^T U_j$ and $V = \sum_{s=-\infty}^{\infty} var(c_s)$ by Parseval's Theorem. Therefore

$$V = \sigma^2 n^{-1} \mathrm{trace}(\Lambda U^* W W^T U \Lambda)$$

$$= \sigma^2 n^{-1} \mathrm{trace}(W W^T U \Lambda^2 U^*)$$

where $\Lambda = diag(\lambda_0, \cdots, \lambda_{n-1})$ and $U = [U_0 \cdots U_{n-1}]$.

The elements of $A = U\Lambda^2 U^*$ are of the form

$$A_{jl} = n^{-1} \sum_{t=0}^{n-1} \lambda_t^2 \exp(2\pi i t(j-l)/n) \ .$$

Thus

$$n\sigma^{-2} V = \sum_{l=0}^{n-1} (W W^T A)_{ll} = \sum_{l=0}^{n-1} \sum_{j=0}^{n-1} (W W^T)_{lj} A_{jl}$$

$$= \sum_{l=0}^{n-1} \sum_{j=0}^{n-1} \left[ \sum_{q=0}^{n-1} (W_l W_j)^{-1} w\left(\frac{x_l - x_q}{b_l}\right) w\left(\frac{x_q - x_j}{b_j}\right)\right] \left(n^{-1} \sum_{t=0}^{n-1} \lambda_t^2 \exp(2\pi i t(j-l)/n)\right)$$

$$\equiv \sum_{t=0}^{n-1} \lambda_t^2 q_t \ ,$$

with $q_t$ defined accordingly. Note that $q_t$ is real valued since $w$ is symmetric and $q_t$ is symmetric in $x_j$, $x_l$.

## 5.3 Proof of Lemma 3

Using Parseval's theorem, we can express

$$B^2(\lambda) = \sum_{s=-\infty}^{\infty} Bias^2(c_s) \ .$$

As in Lemma 1 of Rice and Rosenblatt [**1981**],

$$B^2(\lambda) = \sum_{j=0}^{n-1} \sum_{s=-\infty}^{\infty} |\lambda_{js}(\tilde{a}_j + \tilde{d}_j) - a_{j+sn}|^2 \ ,$$

where

$$\tilde{d}_j = n^{-1/2} \sum_{l=0}^{n-1} Bias(x_l) \exp(-2\pi i j l/n) ,$$

$$\tilde{a}_j = \sum_{s=-\infty}^{\infty} a_{j+sn} , \qquad j = 0, \cdots, n-1 ,$$

$\lambda_{js} = (j+sn)^{-4}(\lambda' + r_j)^{-1}$, $j = 1, \cdots, n-1$, $\lambda_{00} = 1$ and $\lambda_{0s} = 0$ for $s \neq 0$. Here $Bias(x)$ is the bias of the kernel estimator of $f(x)$ which uses the kernel $w$ and the bandwidth $b(x)$. From Parseval's theorem, we recognize that

$$B^2(\lambda) = \int \{[E\mu_{n,\lambda}(x) - f(x)] + \nu_{n,\lambda}(x)\}^2 \, dx$$

where $\nu_{n,\lambda}$ is the smoothing spline fit to $(x_i, Bias(x_i))$ and where $\mu_{n,\lambda}$ is the smoothing spline fit to $(x_i, Y_i))$, $i = 0, \cdots, n-1$. Therefore,

$$B^2(\lambda) = B_0^2(\lambda) + \int [\nu_{n,\lambda}(x)]^2 dx + 2 \int [E\mu_{n,\lambda}(x) - f(x)]\nu_{n,\lambda}(x) dx .$$

# 6 References


W. S. Cleveland [**1979**], *Robust locally weighted regression and smoothing scatterplots*, J. Amer. Statist. Assoc. **74**, 829–836.

R. Cogburn and H. T. Davies [**1974**], *Periodic splines and spectral density estimation*, Ann. Statist. **2**, 1108–1126.

R. L. Eubank [**1988**], *Spline Smoothing and Nonparametric Regression*, Marcell Dekker, Inc., New York.

T. Gasser and H. G. Muller [**1979**], *Kernel estimation of regression functions*, in Lecture Notes in Mathematics: Smoothing Techniques for Curve Estimation, T. Gasser and M. Rosenblatt, eds., vol. **757**, 23–68.

I. J. Good and R. A. Gaskins [**1971**], *Non-parametric roughness penalties for probability densities*, Biometrika **58**, 255–277.

T. Lyche and L. L. Schumaker [**1973**], *Computation of smoothing and interpolating natural splines via local bases*, SIAM J. Numer. Anal. **10**, 1027–1038.

H. G. Muller [**1984**], *Smooth optimum kernel estimators of densities, regression curves and modes*, Ann. Statist. **12**, 766–774.

——— [**1985**], *Kernel estimators of zeros and of location and size of extrema of regression functions*, Scand. J. Statist. **12**, 221–232.

M. B. Priestly and M. T. Chao [**1972**], *Nonparametric function fitting*, J. Roy. Statist. Soc. Ser. B **34**, 385–392.

C. H. Reinsch [**1967**], *Smoothing by spline functions*, Numer. Math. **10**, 177–183.

J. Rice and M. Rosenblatt [**1981**], *Integrated mean squared error of a smoothing spline*, J. Approx. Theor. **33**, 353–369.

J. A. Rice [**1984a**], *Boundary modification for kernel regression*, Comm. Statist. Ser. A **13**, 893–900.

J. R. Rice [**1984b**], *Bandwidth choice for nonparametric regression*, Ann. Statist. **12**, 1215–1230.

B. W. Silverman [**1984**], *Spline smoothing: the equivalent variable kernel method*, Ann. Statist. **12**, 898–916.

J. G. Staniswalis [**1985**], *Local bandwidth selection for kernel estimates*, Ph.D. dissertation, Dept. of Math., U. of California, San Diego.

______ [**1987**], *A weighted likelihood formulation for kernel estimators of a regression function with biomedical applications*, TR 5, Dept. of Biostatist., Virginia Commonwealth U.

M. Tanner and W-H. Wong [**1985**], *Contribution to the discussion of the paper by Silverman*, J. Roy. Statist. Soc. Ser. B **47**, 44–45.

Utreras [**1987**], *Convergence rates for multivariate smoothing spline functions* , J. Approx. Theor. **39**.

G. Wahba [**1975**], *Smoothing noisy data by spline functions*, Numer. Math. **24**, 383–393.

______ [**1985**], *A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem*, Ann. Statist. **13**, 1378–1402.

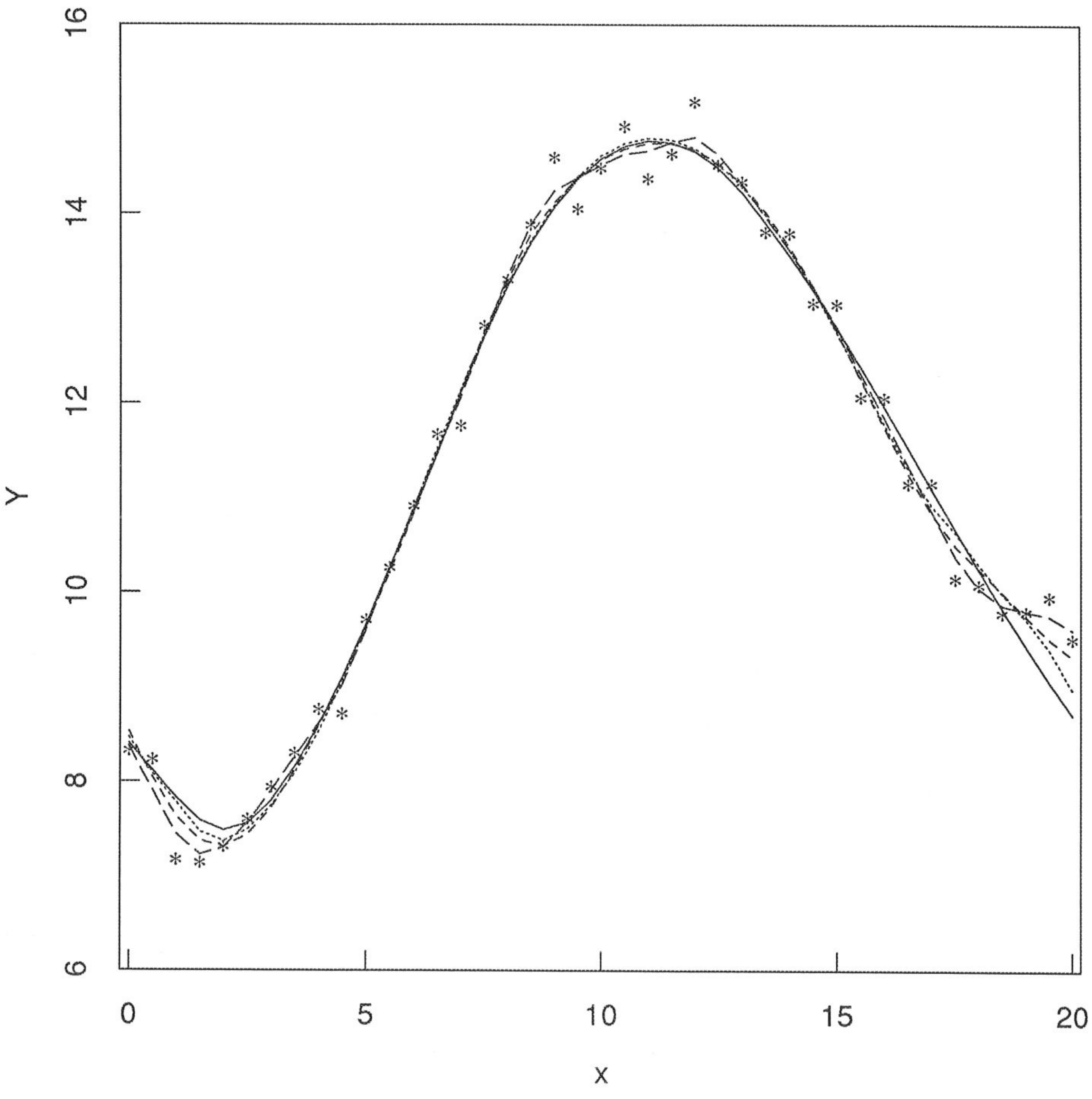
G. Wahba and S. Wold [**1975**], *A completely automatic French curve: Fitting spline functions by cross-validation*, Comm. Statist. **4**, 1–17.

# 7 Figure Captions

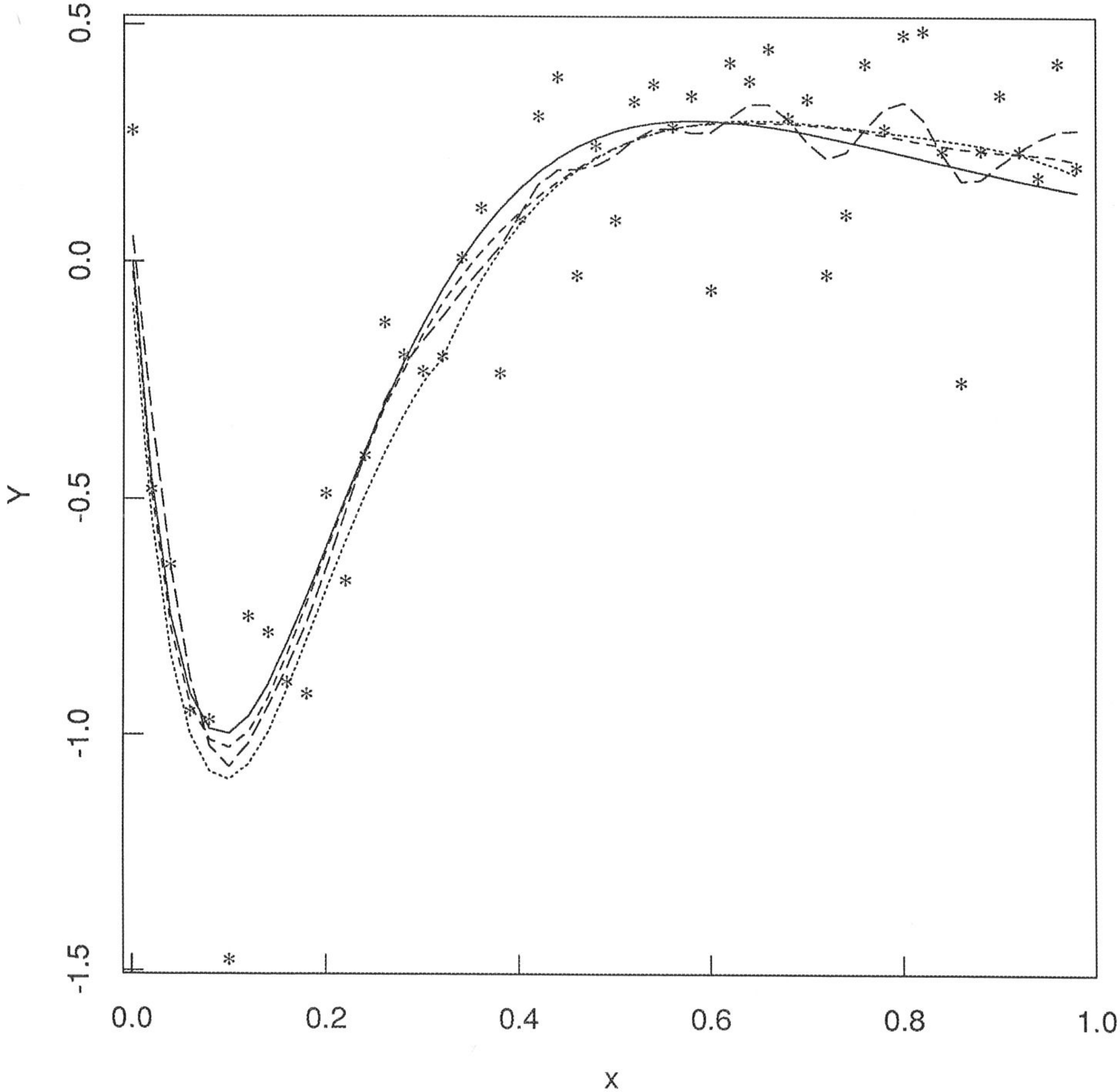
1. Realizations of the LASS and Global Spline Smoothers for Data Simulated from a Mixture of Exponentials. Solid line is true f; dotted line is LASS with $k$=4; short dashed line is LASS with $k$=6; long dashed line is global spline smoother.

2. Estimated MSE of LASS Relative to Estimated MSE of Global Spline Smoother. Solid line is $k$=4; dotted line is $k$=6.

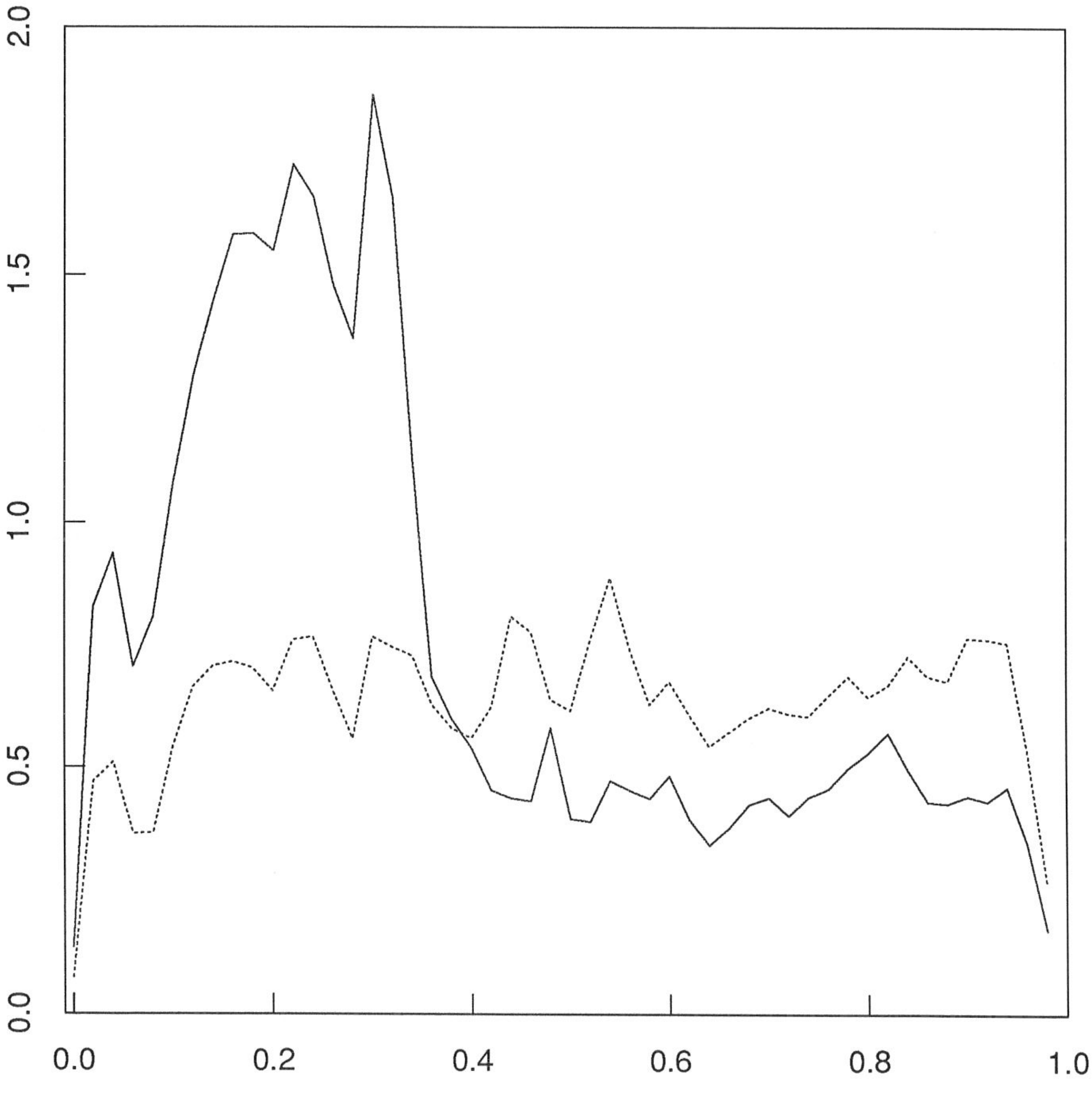
3. Raw Data, LASS, and the Global Spline Smoother for Voltage Data. Solid line is LASS with $k$=4; dotted line is LASS with $k$=6; short dashed line is LASS with $k$=8; long dashed line is global spline smoother.



16
14
12
10
8
6
Y
0
5
10
15
20
x



0.5
0.0
-0.5
-1.0
-1.5
Y
0.0
0.2
0.4
0.6
0.8
1.0
x



0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0