
DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin 53706

Technical Report #380

August, 1974

SMOOTHING NOISY DATA BY SPLINE FUNCTIONS II

by

Grace Wahba

University of Wisconsin
Madison, Wisconsin

Typist: Mary E. Arthur

This research was supported by the United States Air Force under Grant No. AFOSR 72-2363-B.

Note: This report contains a part of TR 340, by the same title. In TR 340, the true curve $f \in W_2^{(m+p)}$ for $p = 0, 1, \ldots, m$. Here discussion of $p = 0, 1, \ldots, m-1$ is eliminated and more precise results obtained in the $p = m$ case. In TR 340, an upper bound on the expected square error was minimized, here, the expected square error itself is minimized. The optimum smoothing parameter is the same.

ABSTRACT

It is shown how to choose the smoothing parameter when a smoothing spline of degree 2m-1 is used to reconstruct a smooth curve from noisy ordinate data. The noise is assumed "white", and the true curve is assumed to be in the Sobolev space $W_2^{(2m)}$ of functions with absolutely continuous νth derivative, $\nu = 0, 1, \ldots, 2m-1$, square integrable 2mth derivative, and satisfying some boundary conditions. The criteria is minimum expected square error, averaged over the data points. The dependency of the optimum smoothing parameter on the sample size, the noise variance, and the smoothness of the true curve is found explicitly.

1. Introduction

In the Sobolev space $W_2^{(m)}$ of real-valued functions on $[0, 1]$ with absolutely continuous m-1 st derivative and square integrable m-th derivative, define the two functionals

$$J(f) = \int_0^1 (f^{(m)}(u))^2 \, du$$

$$R(f) = \frac{1}{n} \sum_{i=1}^{n} (f(t_i) - y_i)^2 , \qquad t_i \in [0, 1] .$$

It is well known, (see [6] [7]), that the two problems

(A) minimize $R(f) + \lambda J(f)$ $\qquad 0 < \lambda < \infty$ $\qquad$ (given)

(B) minimize $J(f)$ subject to the constraint $R(f) \leq S$, $\quad 0 \leq S < \infty$ $\quad$ (given)

are closely related. If $n \geq m$ (A) always has a unique solution for any λ, and, if $S \leq R(p_{m-1})$ where p_{m-1} is the polynomial of degree m-1 which minimizes $R(f)$, then there exists a unique λ so that the solution to (B) is also the solution to (A). The parameter λ is used in the computation of the solution (see [5]), whereas S, being $\frac{1}{n}$ times the apparent residual sum of squares after the smoothing is performed, is a far more intuitive parameter to work with. In [6], Reinsch has shown how to compute λ given S. Conversely, given λ, $g_{n,\lambda}$ is the solution to (B) with $S = R(g_{n,\lambda})$.

Let $g_{n,\lambda}$ be the solution to (A), for given λ. If the data $\{y_i\}_{i=1}^{n}$ represent some true but unknown smooth curve, say g, measured with error,

$$y_i = g(t_i) + \epsilon_i, \qquad i = 1, 2, \ldots n, \qquad t_i \in [0, 1]$$

then it is desired to choose λ so that $g_{n,\lambda}$ approximates g. It is well known [7] [8], that $g_{n,\lambda}$ is a polynomial spline (of degree $2m-1$), and, if $g_{n,\lambda}$ approximates g, and $g \in W_2^{(m)}$, then $g_{n,\lambda}^{(\nu)}$ will approximate $g^{(\nu)}$ for $\nu = 0, 1, \ldots, m-1$. For this reason, the use of the smoothing spline $g_{n,\lambda}$ is much in favor by researchers who are interested in recovering functional values as well as derivatives from noisy ordinate values, when nothing is known about the solution other than $g \in W_2^{(m)}$. See, for example [9][11][13].

Suppose that the errors $\{\epsilon_i\}_{i=1}^n$ may be considered to be values of random variables $\{\epsilon_i\}_{i=1}^n$ with

$$
\begin{aligned}
E\,\epsilon_i &= 0, && i = 1, 2, \ldots, n \\
E\,\epsilon_i \epsilon_j &= \sigma^2, && i = j \\
&= 0, && i \neq j,
\end{aligned}
$$

where E is mathematical expectation. This is a reasonable model for what might occur in practice. It has been suggested by Reinsch [5] that one should choose S to satisfy $\sigma^2(1 - \sqrt{\frac{2}{n}}) \leq S \leq \sigma^2(1 + \sqrt{\frac{2}{n}})$. On the other hand, experimental results by Wold [12] indicate that S should be chosen less than σ^2, say $.7\sigma^2 \leq S \leq .95\sigma^2$.

It is the purpose of this note to investigate, from a theoretical point of view, the optimum choice of λ and hence S. We adopt the criteria of minimizing the expected squared error, averaged over the data points. This is $E\,\psi(\lambda)$ where

$$\psi(\lambda) = \frac{1}{n} \sum_{i=1}^{n} (g_{n,\lambda}(t_i) - g(t_i))^2 .$$

Other "least squares" criteria are reasonable, for example, to choose λ to minimize

$$E \| g - g_{n,\lambda} \|^2_{W_2^{(m)}} .$$

where $\| \cdot \|_{W_2^{(m)}}$ is some norm in $W_2^{(m)}$. This is useful if one wishes to estimate $g^{(\nu)}(t)$ say, for some $\nu = 1, 2, \ldots, m-1$. The linear functionals which map $g \in W_2^{(m)} \rightarrow g^{(\nu)}(t)$, $t \in [0,1]$ are uniformly bounded in $W_2^{(m)}$, for $\nu \leq m-1$, and so we have

$$\sup_t E \, | g^{(\nu)}(t) - g^{(\nu)}_{n,\lambda}(t) |^2 \leq K E \| g - g_{n,\lambda} \|^2_{W_2^{(m)}}$$

for some K.

The choice of λ to minimize $E \| g - g_{n,\lambda} \|^2$ is discussed in some generality in [10], in the context of the numerical solution of linear operator equations, and we will rely on some of these results. The relationship to the "intuitive" parameter S is not discussed there, however. We remark that a related study has been made in connection with the use of smoothing splines to estimate spectral densities [2].

Note that $J^{\frac{1}{2}}$ is a semi-norm on $W_2^{(m)}$. In order to obtain our results in a reasonably direct manner, we have been forced to modify the problem so that $J(f) = 0 \Rightarrow f = 0$. This may be done by requiring that $g - g_{n,\lambda}$ is in a subspace of $W_2^{(m)}$ whose members satisfy m suitably chosen boundary conditions. For convenience, we take the boundary conditions as,

$$\mathcal{B}_m: \quad f(0) = f'(1) = f''(0) = \ldots = f^{(m-1)}(0) = 0, \ m \text{ odd}$$
$$f(1) = f'(0) = f''(1) = \ldots = f^{(m-1)}(0) = 0, \ m \text{ even} .$$

That is, it is assumed that $g^{(\nu)}$ is known for $\nu = 0, 1, \ldots, m-1$.

The results are apparently true for any m boundary conditions $\widetilde{\mathcal{B}}_m$ such that $f \in \mathcal{B}_m$ and $J(f) = 0 \Rightarrow f = 0$. The solution to the original problem without the boundary conditions is a natural spline. We remark that for interpolation problems, the convergence properties of natural splines tend to be the same as those for splines satisfying boundary conditions, in the interior of $[0,1]$, see [1] [3] [8].

The results are as follows:

Let $g_{n,\lambda}$ be the solution to the problem: Find $f \in W_2^{(m)}$ to minimize $R(f) + \lambda J(f)$, subject to $g - g_{n,\lambda} \in \mathcal{B}_m$, where $t_i = (2i-1)/2n$, $i = 1, 2, \ldots, n$. Suppose further that $g \in W_2^{(2m)}$ and in addition satisfies the boundary conditions

$$\mathcal{B}_m^*: \quad g^{(m)}(1) = g^{(m+1)}(0) = \ldots = g^{(2m-1)}(1), \quad m \text{ odd} \tag{1.1}$$

$$g^{(m)}(1) = g^{(m+1)}(0) = \ldots = g^{(2m-1)}(0), \quad m \text{ even} .$$

Then, as $n \to \infty$, $E\,\psi(\lambda)\,(1 + o(1))$ is minimized by $\lambda = \lambda^*$ given by

$$\lambda^* = \left[\frac{a_m}{\theta}\right]^{2m/(4m+1)} \frac{1}{n^{2m/(4m+1)}} (1+o(1))$$

where a_m is a constant depending on m, given before Theorem 1,

$$\theta = ||g^{(2m)}||_2^2/\sigma^2 .$$

and $||\cdot||_2$ is the $\mathcal{L}_2$ norm. Furthermore, $S^* \equiv R(g_{n,\lambda^*})$ satisfies

$$E\, S^* \leq \sigma^2 \left[1 - c_m \frac{\theta^{1/(4m+1)}}{n^{4m/(4m+1)}}\right] (1+o(1)).$$

where c_m is a constant depending on m, given in Theorem 3. Thus, S should be chosen <u>less</u> than σ^2 by a factor no larger than κ given by

$$\kappa = \left[1 - c_m \frac{\theta^{1/(4m+1)}}{n^{4m/(4m+1)}}\right] (1+o(1)).$$

Note that κ tends to 1 from below as n becomes large and as the ratio of a measure of the fluctuation in the "signal" to σ^2 becomes small. We also find a bound on the expected average square error at the data points when λ^* is used. It is

$$E\,\psi(\lambda_*) \le b_m \quad \sigma^2 \; \frac{\theta^{1/(4m+1)}}{n^{4m/(4m+1)}}\,(1+o(1)) \ ,$$

where b_m is a constant depending on m, given in Theorem 2. Thus, if λ^* (equivalently, S^*) is used, then as more data is gathered the average expected square error tends to zero at the rate $n^{-4m/(4m+1)}$. For practical purposes with moderate sample sizes, if one wishes to estimate derivatives no higher than the first, it is reasonable to take $m = 2$, then the smoothing function is a cubic spline.

2. The Optimal Choice of λ

We first state some standard results, good for any reproducing kernel Hilbert space $\mathcal{H}_Q$ of real-valued functions defined on [0, 1]. (See, for example [4]). Let $\mathcal{H}_Q$ possess the norm $\|\cdot\|_Q$, the inner product $<\cdot,\cdot>$, and the reproducing kernel $Q(s,t)$. Let Q_t be the representer of the evaluation functional at t in $\mathcal{H}_Q$, $Q_t(\cdot) \equiv Q(t,\cdot)$, $\langle Q_t, f\rangle \equiv f(t)$, $f \in \mathcal{H}_Q$, $t \in [0,1]$. Let Q_n be the n×n Grammian matrix with i, jth entry $\langle Q_{t_i}, Q_{t_j}\rangle \equiv Q(t_i, t_j)$, and let I be the n × n identity matrix. Let

$$\bar{g}_n = (g(t_1), g(t_2), \ldots, g(t_n))'$$
$$\bar{\epsilon}_n = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$$
$$\bar{y}_n = (y_1, y_2, \ldots, y_n)', \quad \bar{y}_n = \bar{g}_n + \bar{\epsilon}_n$$
$$\bar{g}_{n,\lambda} = (g_{n,\lambda}(t_1), g_{n,\lambda}(t_2), \ldots, g_{n,\lambda}(t_n))'.$$

The solution $g_{n,\lambda}$ to the problem: Find $f \in \mathcal{H}_Q$ to minimize

$$R(f) + \lambda \|f\|_Q^2$$

is

$$g_{n,\lambda} = (Q_{t_1}, Q_{t_2}, \ldots, Q_{t_n})(Q_n + n\lambda I)^{-1} \bar{y}_n .$$

Then

$$\bar{g}_{n,\lambda} - \bar{g}_n = Q_n(Q_n + n\lambda I)^{-1} \bar{y}_n - \bar{g}_n$$

$$= -n\lambda(Q_n + n\lambda I)^{-1} g_n + Q_n(Q_n + n\lambda I)^{-1} \bar{\epsilon}_n ,$$

and

$$E\frac{1}{n}\sum_{i=1}^{n}(g_{n,\lambda}(t_i) - g(t_i))^2 = \frac{1}{n} E\{(n\lambda)^2 \bar{g}_n'(Q_n + n\lambda I)^{-2} \bar{g}_n - 2n\lambda \bar{\epsilon}_n'(Q_n + n\lambda I)^{-1} Q_n(Q_n + n\lambda I)^{-1} \bar{g}_n + \bar{\epsilon}_n'(Q_n + n\lambda I)^{-1} Q_n^2 (Q_n + n\lambda I)^{-1} \bar{\epsilon}_n\}$$

$$= n\lambda^2 \bar{g}_n'(Q_n + n\lambda I)^{-2} \bar{g}_n + \frac{\sigma^2}{n} \text{Trace}\,(Q_n + n\lambda I)^{-2} Q_n^2 . \tag{2.1}$$

The first term on the right of (2.1) is called the bias term, the second, the variance term.

We now specialize to a particular space $\mathcal{H}_Q$. Let $\mathcal{H}_Q$ be the subspace of $W_2^{(m)}$ satisfying the boundary conditions $\mathcal{B}_m$, with $\|\cdot\|_Q^2 = J(\cdot)$. Without further loss of generality, we may assume that both g and $g_{n,\lambda}$ are in $\mathcal{H}_Q$. If not, since $\{M_\nu g\}_{\nu=0}^{m-1}$ are assumed known, we may replace g by $g - h$, $\epsilon\, \mathcal{H}_Q$ where h the polynomial of degree at most $m-1$ satisfying $h - g \; \epsilon \; \mathcal{B}_m$, and similarly for $g_{n,\lambda}$.

$\mathcal{H}_Q$ has the reproducing kernel $Q(s,t)$ given (see [10] for details) by

$$Q(s,t) = \sum_{\nu=1}^{\infty} \lambda_\nu \phi_\nu(s)\phi_\nu(t), \qquad m \text{ odd} \tag{2.2}$$

$$= \sum_{\nu=1}^{\infty} \lambda_\nu \psi_\nu(s)\psi_\nu(t), \qquad m \text{ even} \tag{2.3}$$

where

$$\lambda_\nu = [(\nu - \tfrac{1}{2})\Pi]^{-2m}$$

$$\phi_\nu(s) = \sqrt{2}\sin(\nu - \tfrac{1}{2})\Pi s$$

$$\psi_\nu(s) = \sqrt{2}\cos(\nu - \tfrac{1}{2})\Pi s .$$

An alternative formula for Q is given recursively by

$$Q(s,t) \equiv Q^m(s,t)$$

$$Q^1(s,t) = \min(s,t),$$

$$Q^m(s,t) = \int_s^1\int_t^1 Q^{m-1}(u,v)\,du\,dv, \qquad m \text{ even},$$

$$Q^m(s,t) = \int_0^s\int_0^t Q^{m-1}(u,v)\,du\,dv, \qquad m \text{ odd}.$$

$Q_t(\cdot)$ is a polynomial spline of degree $2m-1$ with a single knot at t, and $Q_t(\cdot) \in \mathcal{B}_m$.

Next, let

$$t_i = \frac{2i-1}{2n}, \quad i = 1, 2, \ldots, n \tag{2.4}$$

and let Q_n^n be the $n \times n$ matrix with i, jth entry

$$\sum_{\nu=1}^{n} \lambda_\nu \phi_\nu(t_i) \phi_\nu(t_j).$$

It can be shown that the vectors $\{\overline{\phi}_{\nu n}\}_{\nu=1}^{n}$ given by

$$\overline{\phi}_{\nu n} = \frac{1}{\sqrt{n}} (\phi_\nu(t_1), \phi_\nu(t_2), \ldots, \phi_\nu(t_n))$$

are the normalized eigenvectors, and $\{n\lambda_\nu\}_{\nu=1}^{n}$ are the eigenvalues of Q_n^n.

Any g in $\mathcal{H}_Q$ with Q given by (2.2) has a representation

$$g = \sum_{\nu=1}^{\infty} g_\nu \phi_\nu$$

where, letting $(\cdot, \cdot)$ be the $\mathcal{L}_2$ inner product,

$$g_\nu = (g, \phi_\nu), \qquad \nu = 1, 2, \ldots,$$

and

$$\sum_{\nu=1}^{\infty} \frac{g_\nu^2}{\lambda_\nu} \equiv \|g\|_Q^2 = \|g^{(m)}\|_2^2,$$

similarly, for Q given by (2.3).

$$\text{Let } g^n = \sum_{\nu=1}^{n} g_\nu \phi_\nu, \text{ and}$$

$$\bar{g}_n^n = (g^n(t_1), g^n(t_2), \ldots, g^n(t_n))'.$$

Then

$$n\lambda^2 \bar{g}_n^{n'} (Q_n^n + n\lambda I)^{-2} \bar{g}_n^n = \sum_{\nu=1}^{n} \frac{(n\lambda)^2 g_\nu^2}{(n\lambda_\nu + n\lambda)^2} = \sum_{\nu=1}^{n} \frac{\lambda^2 g_\nu^2}{(\lambda_\nu + \lambda)^2}. \tag{2.5}$$

It can be shown that the bias term of (2.1) differs little from the left hand side of (2. 5), as $n \to \infty$, $\lambda \to 0$ in such a way that $n\lambda \to \infty$, that is,

$$n\lambda^2 \bar{g}_n' (Q_n + n\lambda I)^{-2} \bar{g}_n = n\lambda^2 \bar{g}_n^{n'} (Q_n^n + n\lambda I)^{-2} \bar{g}_n^n (1 + o(1)).$$

For examples of this type of calculation in detail, see [10]. We omit a proof here and state only the result, as

<u>Lemma 1</u>

Let Q be given by (2.2) or (2.3) and $\{t_i\}_{i=1}^n$ given by (2.4). Let $g \in W_2^{(m)} \cap \mathcal{B}_m$. Then, as $n \to \infty$, $\lambda \to 0$ in such a way that $n\lambda \to \infty$,

$$n\lambda^2 \bar{g}_n' (Q_n + n\lambda I)^{-2} \bar{g}_n = \sum_{\nu=1}^{n} \frac{\lambda^2 g_\nu^2}{(\lambda_\nu + \lambda)^2} (1 + o(1)). \tag{2.6}$$

Let $\{\lambda_{\nu n}\}_{\nu=1}^{n}$ be the n eigenvalues of Q_n. For the variance term, similarly, one shows that

$$\text{Trace}\,(Q_n+n\lambda I)^{-2}\,Q_n^2 = \sum_{\nu=1}^{n} \frac{\lambda_{\nu n}^2}{(\lambda_{\nu n}+n\lambda)^2} \quad \text{differs little from}$$

$$\text{Trace}\,(Q_n^n+n\lambda I)^{-2}(Q_n^n)^2 = \sum_{\nu=1}^{n} \frac{(n\lambda_\nu)^2}{(n\lambda_\nu+n\lambda)^2} = \sum_{\nu=1}^{n} \frac{1}{(1+\lambda[\Pi(\nu-\frac{1}{2})]^{2m})^2}$$

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

where $o(1)$ is as before. Thus we state

Lemma 2:

Let Q be given by (2.2) or (2.3) and $\{t_i\}_{i=1}^{n}$ by (2.4). Then

$$\text{Trace}\,(Q_n+n\lambda I)^{-2}\,Q_n^2 = \frac{k_m}{\lambda^{1/2m}}(1+o(1)). \tag{2.7}$$

where

$$k_m = \frac{1}{\Pi}\int_0^\infty \frac{dy}{(1+y^{2m})^2},$$

and $o(1)$ is as before.

Thus, combining (2.1) andLemmas 1 and 2 gives

Lemma 3:

Let $g \in W_2^{(m)} \cap \mathcal{B}_m$. Then

$$E\,\frac{1}{n}\sum_{i=1}^{n}(g_{n,\lambda}(t_i)-g(t_i))^2 = \left\{\sum_{\nu=1}^{n}\frac{\lambda^2 g_\nu^2}{(\lambda_\nu+\lambda)^2} + \frac{k_m\sigma^2}{n\lambda^{1/2m}}\right\}(1+o(1)), \tag{2.8}$$

where $o(1)$ is as before.

We now suppose that $g \in W_2^{(2m)} \cap \mathcal{B}_m \cap \mathcal{B}_m^*$. Then

$$\|g^{(2m)}\|_2^2 \equiv \int_0^1 (g^{(2m)}(u))^2 \, du \equiv \sum_{\nu=1}^{\infty} \frac{g_\nu^2}{\lambda_\nu^2} < \infty .$$

Note that since $\sum_{\nu=1}^{n} \frac{g_\nu^2}{(\lambda_\nu+\lambda)^2} \leq \sum_{\nu=1}^{\infty} \frac{g_\nu^2}{\lambda_\nu^2} \equiv \|g^{(2m)}\|_2^2$, the right hand side of (2.8) is less than

$$\{\lambda^2 \|g^{(2m)}\|_2^2 + \frac{k_m \sigma^2}{n\lambda^{1/2m}}\} (1+o(1)) \tag{2.9}$$

and that the quantity in brackets in (2.9) is minimized by $\lambda' = \lambda'(n)$ given by

$$\lambda' = [\frac{a_m}{\theta}]^q \frac{1}{n^q} \tag{2.10a}$$

where

$$\begin{aligned} a_m &= k_m/4m \\ \theta &= \|g^{(2m)}\|_2^2/\sigma^2 \\ q &= 2m/(4m+1) . \end{aligned} \tag{2.10b}$$

and

Theorem 1 below says that λ^* which minimizes (2.8) satisfies

$$\lambda^* = \lambda'(1+o(1)) .$$

Theorem 1.

Let Q be given by (2.2) or (2.3) and $\{t_i\}_{i=1}^n$ given by (2.4). Let $g \in W_2^{(2m)} \cap \mathcal{B}_m \cap \mathcal{B}_m^*$. Then, as $n \to \infty$

$$E \frac{1}{n} \sum_{\nu=1}^{n} (g_{n,\lambda}(t_i) - g(t_i))^2$$

is minimized for $\lambda = \lambda^*$ given by

$$\lambda^* = [\frac{a_m}{\theta}]^q \frac{1}{n^q} (1+o(1)) .$$

Proof

Differentiating the expression in brackets in (2.8) with respect to λ and setting the result equal to 0 gives, after some algebra,

$$H_n(\lambda) \equiv \sum_{\nu=1}^{n} h(\lambda/\lambda_\nu)\, w_\nu = \frac{a_m \sigma^2}{n} \tag{2.11}$$

where

$$h(x) = x^{1/q}/(1+x)^3 ,$$

$$w_\nu = \rho_\nu \lambda_\nu^{1/q}$$

$$\rho_\nu = g_\nu^2/\lambda_\nu^2$$

Now H_n is continuous, $H_n(0) = 0$, $H_n(\lambda) > 0$ for $\lambda > 0$, and, since $h(x)$ is monotone decreasing for $x \geq 1/(3q-1)$, $H_n(\lambda)$ is a decreasing function of λ for $\lambda \geq \lambda_1/(3q-1) = \alpha$, say. Furthermore, $H_n(\lambda)$ is a non-decreasing function of n for each fixed λ. Thus, it follows, that for n sufficiently large, there is always one root of (2.11) less than α and one root larger than α, which must tend to ∞ with n. As $\lambda, n \to \infty$, the right hand side of (2.8) tends to $||g^{(2m)}||_2^2$, which will be seen not to be the minimum. We now look at the root(s) $\lambda^* = \lambda^*(n) \leq \alpha$. It can be checked that, the second derivative of the term in brackets in (2.8) is positive for $\lambda = 0(\frac{1}{n^q})$, so that these roots will be minima if they are $0(\frac{1}{n^q})$.

Rearranging (2.11) gives, for the solution $\lambda = \lambda^*$,

$$\lambda^* = \lambda' \left[\frac{\sum_{\nu=1}^{\infty} \rho_\nu}{\sum_{\nu=1}^{n} \frac{\rho_\nu}{(1+\lambda^*/\lambda_\nu)^3}} \right]^q \tag{2.12}$$

Since the term in brackets in (2.12) is always greater than 1, λ^* must satisfy

$$\lambda^* \geq \lambda' = [\frac{a_m}{\theta}]^q \frac{1}{n^q} . \tag{2.13}$$

Furthermore, for any $k \leq n$, and $\lambda^* \leq \alpha$,

$$\lambda^* \leq \lambda' \left[\frac{\sum_{\nu=1}^{\infty} \rho_\nu}{\sum_{\nu=1}^{k} \frac{\rho_\nu}{(1+\lambda^*/\lambda_\nu)^3}} \right]^q \leq \lambda'(1+\lambda^*/\lambda_k)^3 \left[\frac{\sum_{\nu=1}^{\infty} \rho_\nu}{\sum_{\nu=1}^{k} \rho_\nu} \right]^q \tag{2.14}$$

$$\leq \lambda'(1+\alpha/\lambda_k)^3 \left[\frac{\sum_{\nu=1}^{\infty} \rho_\nu}{\sum_{\nu=1}^{k} \rho_\nu} \right]^q . \tag{2.15}$$

We can now use (2.14) and (2.15) to show

$$\lambda^* \leq \lambda'(1+o(1)).$$

Let $n \geq n_0$ where $\sum_{\nu=1}^{n_0} \rho_\nu > 0$. Then, by (2.15)

$$\lambda^* \leq \lambda' \, c_{n_0}(g) \tag{2.16a}$$

for all $n \geq n_0$, where

$$c_{n_0}(g) = \min_{k \le n_0} (1+\alpha/\lambda_k)^3 \left[\frac{\sum_{\nu=1}^{\infty} \rho_\nu}{\sum_{\nu=1}^{k} \rho_\nu}\right]^q \tag{2.16b}$$

Substituting (2.16) into (2.14) gives, for all $n \ge n_0$ and any $k \le n$,

$$\lambda^* \le \lambda' (1 + [\frac{a_m}{\theta n}]^q \; c_{n_0}(g)/\lambda_k)^3 \left[1 + \frac{\sum_{\nu=k+1}^{\infty} \rho_\nu}{\sum_{\nu=1}^{k} \rho_\nu}\right]^q .$$

Now let

$$k = \frac{1}{\pi} n^{\frac{q}{2m}-\varepsilon} (1+o(1)) \text{ for some } 0 < \varepsilon < \frac{q}{2m}. \quad \text{Then}$$

$$\lambda^* \le \lambda'(1+o(1)),$$

and the Theorem is proved.

Next, substitute λ^* of Theorem 1 into (2.8). After performing some calculations one obtains

Theorem 2.

Under the conditions of Theorem 1

$$E \frac{1}{n} \sum_{i=1}^{n} (g_{n,\lambda^*}(t_i) - g(t_i))^2 = b_m \sigma^2 \frac{\theta^{1/(4m+1)}}{n^{2m/(4m+1)}} (1+o(1)),$$

where

$$b_m = (\frac{k_m}{4m})^{4m/(4m+1)} (4m+1)^2 .$$

Thus we have convergence for the true average mean square error, at the rate $n^{-2m/(4m+1)}$, provided λ^* is used.

We next turn to the apparent average residual sum of squares $R(g_{n,\lambda})$.

$$R(g_{n,\lambda}) = \frac{1}{n}\sum_{i=1}^{n}(g_{n,\lambda}(t_i)-y_i)^2 = \frac{1}{n}\,||Q_n(Q_n+n\lambda I)^{-1}(\bar{g}_n+\bar{\varepsilon}_n)-(\bar{g}_n+\bar{\varepsilon}_n)||^2_{2,n} ,$$

where $\|\cdot\|_{2,n}$ is the norm in Euclidean n-space and $\bar{\epsilon}_n = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$. Thus, under the conditions of Theorem 1,

$$E\,R(g_{n,\lambda}) = n\lambda^2\,\bar{g}'_n\,(Q_n + n\lambda I)^{-2}\bar{g}_n + n\lambda^2\sigma^2\,\mathrm{Trace}\,(Q_n + n\lambda I)^{-2} .$$

$$\leq \lambda^2 \quad \|g^{(2m)}\|_2^2\,(1+o(1)\,) + \frac{\sigma^2}{n}\sum_{\nu=1}^{n}\frac{(n\lambda)^2}{(\lambda_{\nu n}+n\lambda)^2} . \qquad (2.17)$$

Now

$$\frac{1}{n}\sum_{\nu=1}^{n}\frac{n\lambda^2}{(\lambda_{\nu n}+n\lambda)^2} = 1 - 2\lambda\sum_{\nu=1}^{n}\frac{\lambda_{\nu n}}{(\lambda_{\nu n}+n\lambda)^2} - \frac{1}{n}\sum_{\nu=1}^{n}\frac{\lambda^2_{\nu n}}{(\lambda_{\nu n}+n\lambda)^2} . \qquad (2.18)$$

We have

<u>Lemma 4.</u>

Under the conditions of Lemma 2,

$$\sum_{\nu=1}^{n}\frac{\lambda_{\nu n}}{(\lambda_{\nu n}+n\lambda)^2} = \sum_{\nu=1}^{n}\frac{n\lambda_\nu}{(n\lambda_\nu+n\lambda)^2}\,(1+o(1)) = \frac{\ell_m}{2n\lambda^{1+1/2m}}\,(1+o(1)) \qquad (2.19)$$

where

$$\ell_m = \frac{2}{\Pi}\int_0^\infty \frac{x^{2m}\,dx}{(1+x^{2m})^2} .$$

and o(1) is as before.

A complete proof of Lemma 4 is given in [10]. A proof of Lemma 2 may be obtained by following the proof there. Substituting (2.7) and (2.19) into (2.18), then substituting (2.18) and λ^* of Theorem 1 into (2.17) gives

<u>Theorem 3.</u>

Under the conditions of Theorem 1,

$$E\,R(g_{n,\lambda}) \leq \lambda^2 \quad \|g^{(2m)}\|_2^2\,(1+o(1)) + \sigma^2\left(1 - \frac{(k_m+\ell_m)}{n\lambda^{1/2m}}\,(1+o(1)\,)\right) ,$$

and

$$E\,S^* \equiv E\,R(g_{n,\lambda^*}) \leq \sigma^2 \left[1 - c_m \frac{\theta^{1/(4m+1)}}{n^{4m/(4m+1)}}\right] (1+o(1)) ,$$

where

and $\quad c_m = (k_m/4m)^{4m/(4m+1)} \left[4m(1+\ell_m/k_m)-1\right]$.

$o(1) \to 0$ as $n \to \infty$.

REFERENCES

[1] Ahlberg, T. H., Nilson, E. W., and Walsh, J. L. "The Theory of Splines and Their Applications", Academic Press, New York (1967).

[2] Cogburn, R., and Davis, H. T., Periodic Splines and Spectral Estimation, University of New Mexico, Statistics Department, Technical Report # 253 (revised) (1973).

[3] Kershaw, D., A note on the convergence of interpolatory cubic splines, SIAM J. Numer. Anal., 8, 1, 67-74 (1971).

[4] Kimeldorf, G. and Wahba, G. Some results on Tchebycheffian spline functions, J. Math. Anal. Appl. 33, 82-95 (1971).

[5] Reinsch, C. H., Smoothing by spline functions. Numer. Math. 16, 177-183 (1967).

[6] Reinsch, C. H., Smoothing by Spline Functions. II. Numer. Math. 16, 451-454 (1971).

[7] Schoenberg, I. J., Spline functions and the problem of graduation. Proc. Nat. Acad. Sci. (U. S. A.) 52, 947-950 (1964).

[8] Schultz, Martin. Error bounds for polynomial spline interpolation. Math. Comp. 24, 111, 507-515 (1970).

[9] Snowden, Marion. M. S. Thesis, in preparation.

[10] Wahba, Grace. Convergence properties of the method of regularization for noisy linear operator equations, Univ. of Wisconsin Mathematics Research Center TSR #1132 (1973).

[11] Wiencke, Bonnie E., "The roundoff: A Mechanical Analysis of a Skillfully Executed Gymnastic Stunt", unpublished M. S. thesis, Women's Physical Education Department, University of Wisconsin, Madison (1972)

[12] Wold, Svante, Spline functions in data analysis, Technometrics 16, 1, 1-11(1974).

[13] Wold, Svante, Estimation of activation parameters from one kinetic experiment (Varytemp method). Error analysis and revised computer program. Acta. Chem. Scand. 25, 336-339 (1971)