TECHNICAL REPORT NO. 508

January 1978

IMPROPER PRIORS, SPLINE SMOOTHING AND THE PROBLEM OF GUARDING AGAINST MODEL ERRORS IN REGRESSION

by

Grace Wahba

Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression

by

Grace Wahba

Abstract

Generalized spline smoothing is shown to be equivalent to Bayesian estimation with a partially improper prior. This result is related to a result of Kimeldorf and Wahba, J. Math. Anal. Applic. 33, 1(1971). It supports the idea that spline smoothing is a natural solution to the regression problem when one is given a set of regression functions but one also wants to hedge against the possibility that the true model is not exactly in the span of the given regression functions. A natural measure of the deviation of the true model from the span of the regression functions comes out of the spline theory in a natural way. An appropriate value of this measure can be estimated from the data and used to constrain the estimated model to have the estimated deviation. Some convergence results and computational tricks are also discussed.

Research supported by the U. S. Army under Contract No. DAAG29-77-G-0207

1. Introduction

Consider the model

$$Y(t_i) = g(t_i) + \varepsilon_i, \qquad i = 1,2,\ldots,n \qquad t_i \in T \tag{1.1}$$

where $\underset{\sim}{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_n)' \sim N(0,\sigma^2 I_{n\times n})$ and $g(\cdot)$ is some "smooth" function defined on some index set T. When T is an interval of the real line, cubic polynomial smoothing splines are well known to provide an esthetically satisfying method for estimating $g(\cdot)$, from a realization $\underset{\sim}{y} = (y_1,\ldots,y_n)'$ of $\underset{\sim}{Y} = (Y(t_1),\ldots,Y(t_n))'$. See [18] for a very useful example. Splines are an appealing alternative to fitting a specified set of m regression functions, for example polynomials of degree less than m, when one is uncertain that the true curve $g(\cdot)$ is actually in the span of the specified regression functions. In [11,12,13] certain relationships between Bayesian estimation and spline smoothing are explored. In this note we provide a somewhat different formulation and generalization of the result in [13, Section 7]. Here we prove that polynomial spline (respectively generalized spline) smoothing is equivalent to Bayesian estimation with a prior on g which is "diffuse" on the coefficients of the polynomials of degree $\leq$m (respectively specified set of m regression functions), and "proper" over an appropriate set of random variables not including the coefficients of the regression functions. Since Gauss Markov estimation is equivalent to Bayesian estimation with a prior diffuse over the coefficients of the regression functions, this result leads to the conclusion that spline smoothing is a (the?) natural extension of Gauss-Markov regression with m specified regression functions. We claim that spline smoothing is an appropriate solution to the problem arising when one wants to fit a given set of regression functions to the data but one also wants to "hedge" against model errors, that is, against the possibility that the true model g is not exactly in the span of the given set of regression functions. We show that the spline smoothing approach leads to a natural measure of the deviation of the true g from the span of the regression functions, and furthermore, a good value of this measure can be estimated from the data. The estimated value of the measure is then used to control the deviation of the estimated g.

From another point of view this measure can be viewed as the "bandwidth parameter" which controls the "smoothness" of the estimated g, and so in this approach to non-parametric (or semi-parametric) regression, a good value of the bandwidth parameter can be estimated from the data. See [4,5,14,15,21] for other approaches to the estimation of g in the model (1.1). In the case of polynomial splines integrated mean square error rates of convergence of the estimated g to the true g, as $\max_i |t_{i+1}-t_i| \to 0$, have been recently found and are quoted in Section 5 for comparison with the kernel nonparametric regression results of Benedetti [4] and Priestly and Chao [15].

We also remark upon some approaches to the efficient computation of generalized splines.

2. Polynomial splines as posterior means with a partially improper prior on the polynomials of degree less than m.

Let $T = [0,1]$. Given data $\{y(t_1),\ldots,y(t_n)\}$, $0 < t_1 \ldots < t_n < 1$, the smoothing polynomial spline of degree 2m-1 to the data, call it $g_{n,\lambda}$, is defined as the solution to the minimization problem: Find $g \in W_2^{(m)}$: $\{g:\ g, g', \ldots, g^{(m-1)}$ abs. cont. $g^{(m)} \in L_2[0,1]\}$ to minimize

$$\frac{1}{n}\sum_{j=1}^{n} (g(t_j)-y_j)^2 + \lambda \int_0^1 (g^{(m)}(u))^2 du, \tag{2.1}$$

where $y_j = y(t_j)$, and λ is to be chosen. If $\underset{\sim}{y}$ cannot be interpolated exactly by some polynomial of degree less than m, then the solution is well known to be unique, and to be a polynomial spline of degree 2m-1 (see [19]), that is, it is piecewise a polynomial of degree 2m-1 in each interval $[t_i,t_{i+1}]$, $i = 1,2,\ldots,n-1$, with the pieces joined so that the resulting function has 2m-2 continuous derivatives. An efficient computational algorithm for the cubic polynomial smoothing spline (m=2) is given in [16] and code is available in the IMSL library [10]. We show that the spline solution $g_{n,\lambda}$ to the minimization problem of (2.1) is a Bayesian estimate for g with a "partially diffuse" prior; the quantity $J = \int_0^1 (g_{n,\lambda}^{(m)}(u))^2 du$ is a natural measure of the deviation of $g_{n,\lambda}$ from the span of the polynomials of degree less than m, and furthermore a good value of J can be estimated from the data.

Theorem 1. Let $g(t)$, $t \in [0,1]$ have the prior distribution which is the same as the distribution of the stochastic process $X_\xi(t)$, $t \in [0,1]$,

$$X_\xi(t) = \sum_{j=1}^{m} \theta_j\, \phi_j(t) + b^{1/2} Z(t), \tag{2.2}$$

where $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m)' \sim N(0,\xi I_{m\times m})$, $\phi_j(t) = t^{j-1}/(j-1)!$, $j = 1,2,\ldots,m$, b is fixed, and $Z(t)$ is the m-fold integrated Wiener process [20],

$$Z(t) = \int_0^t \frac{(t-u)_+^{m-1}}{(m-1)!} \quad dW(u) \ . \tag{2.3}$$

Then

(i) The polynomial spline $g_{n,\lambda}(\cdot)$ which is the minimizer of (2.1) has the property

$$g_{n,\lambda}(t) = \lim_{\xi\to\infty} E_\xi\{g(t)|\underset{\sim}{Y} = \underset{\sim}{y}\} \tag{2.4}$$

with $\lambda = \sigma^2/nb$, where E_ξ is expectation over the posterior distribution of $g(t)$ with the prior (2.2). ($\xi=\infty$ corresponds to the "diffuse" prior on $\underset{\sim}{\theta}$).

(ii) Suppose $\underset{\sim}{y}$ cannot be interpolated exactly by some polynomial of degree less than m. Then $\lim_{\lambda\to\infty} g_{n,\lambda}(\cdot)$ is the polynomial of degree m-1 best fitting the data in a least squares sense, $\lim_{\lambda\to 0} g_{n,\lambda}(\cdot)$ is that function in $W_2^{(m)}$ which minimizes $\int_0^1 (g^{(m)}(u))^2 du$ subject to the conditions that it interpolate $\underset{\sim}{y}$, and $J(\lambda) = \int_0^1 (g_{n,\lambda}^{(m)}(u))^2 du$ is a monotone strictly decreasing function of λ.

(iii) Let loss be measured by the mean square prediction error $R(\lambda)$ given by

$$R(\lambda) = \frac{1}{n}\sum_{j=1}^{n} (g(t_j)-g_{n,\lambda}(t_j))^2 \ .$$

Define $\hat{R}(\lambda)$ by

$$\hat{R}(\lambda) = \frac{1}{n}\{||(I-A(\lambda))\underset{\sim}{y}||^2 + \sigma^2 \text{ Trace } A^2(\lambda) - \sigma^2 \text{ Trace } (I-A(\lambda))^2\}$$

where $A(\lambda)$ is the symmetric non-negative definte matrix satisfying

$$\underset{\sim}{g}_{n,\lambda} = A(\lambda)\underset{\sim}{y} \ ,$$

where

$$g_{n,\lambda} = (g_{n,\lambda}(t_1),\ldots,g_{n,\lambda}(t_n))' .$$

If $g = (g(t_1),\ldots,g(t_n))'$ is viewed as fixed, and expectation taken with respect to ε, then

$$E\ \hat{R}(\lambda) = E\ R(\lambda)$$

so that an optimum λ for squared error of prediction loss may be estimated from the data by minimizing $\hat{R}(\lambda)$.

Proof: This theorem is a special case of Theorem 2 which will be proved below. Part (ii) is well known (see [2,13,17,19]), (iii) recently appeared in [6].

We interpret (i) and (ii) as saying that estimation with the polynomial spline $g_{n,\lambda}$ should be viewed as a (the?) natural extension of Gauss-Markov estimation with polynomial regression functions (i.e. estimation with $g_{n,\infty}$). This is because the Gauss-Markov regression estimate can be obtained as the posterior mean of g when g has a prior diffuse on the coefficients of the polynomials; $g_{n,\lambda}$, $\lambda < \infty$ is obtained as the posterior mean of g when g has a diffuse prior on the coefficients of the polynomials modified by the addition of $b^{1/2}Z(\cdot)$ to the prior specification, $b > 0$.

In practice $\lambda = \sigma^2/nb$ is not generally known, so that it is fortunate that λ can be estimated from the data via (iii). If σ^2 is not known an estimate of λ which minimizes $E\ R(\lambda)$ asymptotically for large n for fixed $g \in W_2^{(m)}$ can be obtained by using the method of generalized cross-validation [6].

3. Generalized splines as posterior means with a partially improper prior.

We now consider the general case where polynomials on [0,1] are replaced by some real-valued functions $\{\phi_j(\cdot)\}_{j=1}^m$ defined on some separable index set T. Families of extensions of Gauss-Markov estimates analogous to the smoothing polynomial spline will be found. We require only that the $n \times m$ matrix with jk-th entry $\phi_k(t_j)$ be of rank m.

Let H_K be a reproducing kernel Hilbert space (r.k.h.s.) of real valued functions on T and containing the $\{\phi_j\}$. Recall [3] that an r.k.h.s. is a Hilbert space with the property that, for each fixed $t_* \in T$, the linear functional which maps $g \in H_K$ to $g(t_*)$ is a continuous linear functional. The reproducing kernel (r.k.) and inner product for H_K are denoted respectively by $K(s,t)$, $s, t \in T$, and $<\cdot,\cdot>_K$. It is not hard to show that H_K is the direct sum of span $\{\phi_j\}$ and H_Q, the r.k.h.s. with r.k. $Q(s,t)$, $s, t \in T$ given by (see [24])

$$Q(s,t) = K(s,t) - \sum_{i,j=1}^{m} \phi_i(s)\, k_{ij}\, \phi_j(t)$$

where k_{ij} is the ij-th entry of the inverse of the (necessarily strictly positive definite) matrix with ij-th entry $<\phi_i,\phi_j>_K$. Let P_Q be the orthogonal projection operator in H_K onto H_Q. (That is, $I-P_Q$ is the orthogonal projection in H_K onto span $\{\phi_j\}$.) The analogue of $\int_0^1 (g^{(m)}(u))^2 du$ is $||P_Q g||_K^2$, and this is, of course, a measure of the deviation of g from span $\{\phi_j\}$, being the distance in H_K from g to span $\{\phi_j\}$.

Suppose $\underset{\sim}{y}$ is not in the span of the vectors $\{\underset{\sim}{\phi}_j\}_{j=1}^m$, where $\underset{\sim}{\phi}_j = (\phi_j(t_1),\ldots,\phi_j(t_n))'$. Then [2,13] there is a unique solution, call it $g_{n,\lambda}$ to the minimization problem: Find $g \in H_K$ to minimize

$$\frac{1}{n} \sum_{j=1}^{n} (g(t_j)-y_j)^2 + \lambda \, ||P_Q g||_K^2 \, . \tag{3.1}$$

We shall call any $g_{n,\lambda}$ obtained as a solution of this minimization problem a generalized smoothing spline, or, consistent with the terminology in [2], just a smoothing spline.

Theorem 2. Let $g(t)$, $t \varepsilon T$ have the prior distribution which is the same as the distribution of the stochastic process $X_\xi(t)$,

$$X_\xi(t) = \sum_{j=1}^{m} \theta_j \, \phi_j(t) + b^{1/2} Z(t), \qquad t \varepsilon T \tag{3.2}$$

where $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m) \sim N(0,\xi I_{m\times m})$, b is fixed ≥ 0 and $Z(t)$ is a zero mean Gaussian stochastic process with $E\, Z(s)Z(t) = Q(s,t)$. Then

(i) The generalized spline $g_{n,\lambda}$ which is the minimizer of (3.1) has the property

$$g_{n,\lambda}(t) = \lim_{\xi\to\infty} E_\xi\{g(t)|\underset{\sim}{Y}=\underset{\sim}{y}\}$$

with $\lambda = \sigma^2/nb$, where E_ξ is expectation over the posterior distribution of $g(t)$ with the prior (3.2).

(ii) Suppose $\underset{\sim}{y}$ is not in the span of the $\{\underset{\sim}{\phi}_j\}$. Then $\lim_{\lambda\to\infty} g_{n,\lambda}(\cdot)$ is that element in span $\{\phi_j(\cdot)\}$ best fitting the data in a least squares sense. If Q_n, the $n\times n$ matrix with ij-th entry $Q(t_i,t_j)$ is of full rank, $\lim_{n\to 0} g_{n,\lambda}(\cdot)$ is that function in H_K which minimizes $||P_Q g||_K^2$ subject to the conditions that it interpolate the data, and $J(\lambda) = ||P_Q g_{n,\lambda}||_K^2$ is a monotone decreasing function of λ.

(iii) Let

$$R(\lambda) = \frac{1}{n} \sum_{j=1}^{n} (g(t_j)-g_{n,\lambda}(t_j))^2 \, ,$$

and define $\hat{R}(\lambda)$ by

$$\hat{R}(\lambda) = \frac{1}{n}\{||(I-A(\lambda))\underset{\sim}{y}||^2 + \sigma^2 \text{ Trace } A^2(\lambda) - \sigma^2 \text{ Trace } (I-A(\lambda))^2\}.$$

where $A(\lambda)$ is the symmetric, non-negative definite matrix satisfying

$$\underset{\sim}{g}_{n,\lambda} = A(\lambda)\underset{\sim}{y} .$$

If $\underset{\sim}{g}$ is viewed as fixed, then

$$E\ R(\lambda) = E\ \hat{R}(\lambda)$$

so that an optimum λ for squared error of prediction loss may be estimated from the data by minimizing $\hat{R}(\lambda)$.

Proof: (i) Using Lemma 5.1 of [13], an explicit formula for $g_{n,\lambda}$ is given by

$$g_{n,\lambda}(t) = (\phi_1(t),\dots,\phi_m(t))\ (T'M^{-1}T)^{-1}T'M^{-1}\underset{\sim}{y}$$
$$+ (Q_{t_1}(t),\dots,Q_{t_n}(t))M^{-1}(I-T(T'M^{-1}T)^{-1}T'M^{-1})\underset{\sim}{y} , \tag{3.3}$$

where T is the $n\times m$ matrix of rank m with jk-th entry $\phi_k(t_j)$, $M = n\lambda I_{n\times n}+Q_n$, where Q_n is the $n\times n$ matrix with jk-th entry $Q(t_j,t_k)$, and $Q_{t_i}(t) \equiv Q(t_i,t)$. With the prior of (3.2) it is easily seen that the prior covariances $E\ X_\xi(t)\underset{\sim}{Y}$ and $E\ \underset{\sim}{Y}'\underset{\sim}{Y}$ are

$$E\ X_\xi(t)\underset{\sim}{Y} = \xi\ T(\phi_1(t),\dots,\phi_m(t))' + b(Q_{t_1}(t),\dots,Q_{t_n}(t))'$$

$$E\ \underset{\sim}{Y}'\ \underset{\sim}{Y} = \xi\ T\ T' + bQ_n + \sigma^2\ I .$$

Therefore, from elementary considerations [1]

$$E\{X_\xi(t)|\underset{\sim}{Y}=\underset{\sim}{y}\} = \{\xi T(\phi_1(t),\dots,\phi_m(t))'+b(Q_{t_1}(t),\dots,Q_{t_n}(t))'\}' \cdot (\xi TT'+bQ_n+\sigma^2 I)^{-1}\underset{\sim}{y} .$$

Setting $\lambda = \sigma^2/nb$, $\eta = \xi/b$ and $M = Q_n + n\lambda I$ gives

$$E\{X_\xi(t)|\underset{\sim}{Y}=\underset{\sim}{y}\} = (\phi_1(t),\dots,\phi_m(t))\eta T'(\eta TT'+M)^{-1}\underset{\sim}{y} + (Q_{t_1}(t),\dots,Q_{t_n}(t))(\eta TT'+M)^{-1}\underset{\sim}{y} . \tag{3.4}$$

By comparing (3.3) and (3.4), it remains only to show that

$$\lim_{\eta\to\infty} \eta T'(TT'+M)^{-1} = (T'M^{-1}T)^{-1}T'M^{-1} \tag{3.5}$$

and

$$\lim_{\eta\to\infty} (\eta TT'+M)^{-1} = M^{-1}(I-T(T'M^{-1}T)^{-1}TM^{-1}) . \tag{3.6}$$

Now, it can be verified (see also [9]) that

$$(nTT'+M)^{-1} \equiv M^{-1}-\eta M^{-1}T(I+\eta T'M^{-1}T)^{-1}T'M^{-1}$$

and so

$$\begin{aligned}\lim_{\eta\to\infty} (\eta TT'+M)^{-1} &= \lim_{\eta\to\infty} (M^{-1}-\eta M^{-1}T(I+\eta T'M^{-1}T)^{-1}T'M^{-1}) \\ &= \lim_{\eta\to\infty} (M^{-1}-M^{-1}T(\tfrac{1}{\eta}I+T'M^{-1}T)^{-1}T'M^{-1}) \\ &= M^{-1}-M^{-1}T(T'M^{-1}T)^{-1}T'M^{-1} ,\end{aligned}$$

also

$$\lim_{\eta\to\infty} \eta T'(\eta TT'+M)^{-1} = \lim_{\eta\to\infty} \eta T'\{M^{-1} - \eta M^{-1}T(I+\eta T'M^{-1}T)^{-1}T'M^{-1}\}$$

$$= \lim_{\eta\to\infty} \eta\{I-\eta T'M^{-1}T(I+\eta T'M^{-1}T)^{-1}\}T'M^{-1}$$

$$= \lim_{\eta\to\infty} \eta\{(I+\eta T'M^{-1}T)-\eta T'M^{-1}T\}(I+\eta T'M^{-1}T)^{-1}T'M^{-1}$$

$$= \lim_{\eta\to\infty} \eta(I+\eta T'M^{-1}T)^{-1}T'M^{-1}$$

$$= \lim_{\eta\to\infty} (\tfrac{1}{\eta}I+T'M^{-1}T)^{-1}T'M^{-1}$$

$$= (T'M^{-1}T)^{-1}T'M^{-1} \quad ,$$

giving (3.5) and (3.6).

ii) To show that $\lim_{\lambda\to\infty} g_{n,\lambda}(\cdot)$ is the Gauss-Markov estimate of g, one notes that

$$\lim_{\lambda\to\infty} (T'M^{-1}T)^{-1}T'M^{-1} = (T'T)^{-1}T'$$

$$\lim_{\lambda\to\infty} M^{-1}(I-T(T'M^{-1}T)^{-1}T'M^{-1}) = 0$$

and so $g_{n,\lambda}(t) = \sum_{j=1}^{m} \hat{\theta}_j\phi_j(t)$, where $\hat{\underset{\sim}{\theta}} = (\hat{\theta}_1,\ldots,\hat{\theta}_m)' = (T'T)^{-1}T'\underset{\sim}{y}$

is the Gauss Markov estimate of $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m)$. The conclusion concerning $\lim_{\lambda\to 0} g_{n,\lambda}$ is a consequence of Lemma 3.1 of [13], see also [2]. The decreasing monotonicity of $||P_Q g_{n,\lambda}||_K^2$ is known from [16, Eq.(7)] where it has been derived explicitly from [2]. An easy proof also follows from the representation of $P_Q\, g_{n,\lambda}$ which is given on the left of Eqn. (4.5) below.

iii) Since $R(\lambda) = \frac{1}{n}||\underset{\sim}{g}_{n,\lambda}-\underset{\sim}{g}||^2 = \frac{1}{n}||A(\lambda)\underset{\sim}{y}-\underset{\sim}{g}||^2$,

$$E\ R(\lambda) = \frac{1}{n}\{||(I-A(\lambda))\underset{\sim}{g}||^2 + \sigma^2 \operatorname{Tr} A^2(\lambda)\} .$$

Now

$$E\ ||(I-A(\lambda))\underset{\sim}{y}||^2 = ||(I-A(\lambda))\underset{\sim}{g}||^2 + \sigma^2 \operatorname{Tr} (I-A(\lambda))^2$$

so the result is immediate.

We remark that $A(\lambda)$ is obtained from (3.3) and is $A(\lambda) = T(T'M^{-1}T)^{-1}T'M^{-1}+Q_nM^{-1}(I-T(T'M^{-1}T)^{-1}T'M^{-1})$. Details of computing $\hat{R}(\lambda)$, and the results of some numerical experiments demonstrating the efficacy of $g_{n,\hat{\lambda}}$ as an estimate of g may be found in [6] for the cubic polynomial spline case. ($\hat{\lambda}$ is the minimizer of $\hat{R}(\lambda)$.)

Theorem 1 follows from Theorem 2 by making the identifications $\mathcal{H}_K = W_2^{(m)}$, $Q(s,t) = \int_0^1 \frac{(s-u)_+^{m-1}\ (t-u)_+^{m-1}}{((m-1)!)^2} du$, and $||P_Q g||_K^2 = \int_0^1 (g^{(m)}(u))^2 du$, for complete details concerning this and the remainder of this paragraph see [13]. If $\{\phi_j\}$ is any extended Tchebychev system of functions on [0,1], then T is of rank m whenever $n > m$ and the $\{t_i\}$ are distinct. There will exist an m-th order linear differential operator L_m whose null space is spanned by the $\{\phi_j\}$. If one takes $Z(t)$ formally satisfying $L_m\ Z(t) = dW(t)$, then $Q(s,t) = \int_0^1 G(s,u)G(t,u)du$ where G is a Green's function for L_m, and $||P_Q g||_K^2 = \int_0^1 [(L_m g)(u)]^2 du$.

The results here generalize easily to $\underset{\sim}{\varepsilon} \sim N(0,\sigma^2\Sigma)$, for known σ^2 and Σ via the results in [13].

For related work concerning the use of smoothing splines for smoothing surfaces ($\mathcal{T} = [0,1] \times [0,1] \times \ldots \times [0,1]$) see [25].

4. Representations of $g_{n,\lambda}$ for efficient computing

We believe smoothing splines are appropriate for solving a wide variety of practical problems in practice including smoothing surfaces, once efficient numerical algorithms are developed. If H_K is a space of periodic functions on [0,1] or a tensor product of periodic spaces on $[0,1] \times \ldots \times [0,1]$, and the $\{t_i\}$ are equally spaced or the tensor product of equally spaced points then computing problems are readily solved. (See [26], for a computed example). In general, however, the efficient computation of $g_{n,\lambda}$ presents challenges, if n is very large, as would usually be the case if T is a rectangle in d-space. It will probably be necessary to choose Q with computational ease an important consideration.

Equation (3.3) will generally not be the best representation for computing $g_{n,\lambda}$. We discuss some other representations for $g_{n,\lambda}$ chosen with efficient computing in mind. We assume below that Q_n is of full rank. Since

$$T'M^{-1}(I-T(T'M^{-1}T)^{-1}T'M^{-1} = 0_{m\times n}$$

it is clear that $g_{n,\lambda}$ has a representation

$$g_{n,\lambda} = \sum_{i=1}^{m} \theta_i\phi_i + \sum_{i=1}^{n-m} c_i h_i \tag{4.1}$$

where $\underset{\sim}{\theta}$ and $\underset{\sim}{c} = (c_1,\ldots,c_{n-m})$ are vectors of constants, and

$$h_i(\cdot) = \sum_{j=1}^{n-m} b_{ij}\, Q_{t_j}(\cdot)$$

where the $(n-m)\times n$ dimensional matrix B with ij-th entry b_{ij} satisfies $BT = 0_{(n-m)\times m}$ but is otherwise arbitrary.

We will demonstrate shortly that $\underset{\sim}{c}$, $\underset{\sim}{\theta}$ and $A(\lambda)\underset{\sim}{y} = \underset{\sim}{g}_{n,\lambda}$ satisfy

$$(\textstyle\sum_h + n\lambda BB')\underset{\sim}{c} = B\underset{\sim}{y} \tag{4.2}$$

$$T\underset{\sim}{\theta} = y - MB'\underset{\sim}{c} \tag{4.3}$$

and

$$\underset{\sim}{g}_{n,\lambda} \equiv A(\lambda)\underset{\sim}{y} = y - n\lambda B'\underset{\sim}{c}, \tag{4.4}$$

where $\sum_h$ is the $(n-m)\times(n-m)$ dimensional matrix with jk-th entry $\langle h_i, h_j\rangle_Q$. One attempts to choose B so that $\{h_j\}$, B and $\sum_h$ have convenient properties for computing, and then to obtain $\underset{\sim}{c}$, $\underset{\sim}{\theta}$ $\underset{\sim}{g}_{n,\lambda}$, and $g_{n,\lambda}(\cdot)$ from (4.1)-(4.4) by first solving the linear system (4.2). In the polynomial spline case, by choosing the entries in B corresponding to divided differences, one can obtain $\sum_h$ and B both banded matrices and an efficient code results (see [2,16]). The span of the $\{h_j\}$ can be constructed from B-splines, which are nice hill-like functions (see [7,8]).

Equation (4.2) is equivalent to [2, Eqns. (8.26), (9.1)]. However we provide a direct proof of (4.2) using (3.3) without the elegant but lengthy machinery of [2]. We must show that

$$(h_1,\ldots,h_{n-m})(\textstyle\sum_h + n\lambda BB')^{-1}B\underset{\sim}{y}$$

$$\equiv (Q_{t_1},\ldots,Q_{t_n})(M^{-1} - M^{-1}T(T'M^{-1}T)^{-1}T'M^{-1})\underset{\sim}{y}\ . \tag{4.5}$$

Now since $\langle Q_{t_i}, Q_{t_j}\rangle_K = Q(t_i,t_j)$, we have that $\sum_h = BQ_nB'$ and so the left hand side of (4.5) is given by

$$(Q_{t_1},\ldots,Q_{t_n})B'(BMB')^{-1}B\underset{\sim}{y}\ . \tag{4.6}$$

However

$$B'(BMB')^{-1}B \equiv M^{-1}-M^{-1}T(T'M^{-1}T)^{-1}T'M^{-1} \tag{4.7}$$

as can be seen by observing that the $n\times n$ matrix $X = \begin{bmatrix} T' \\ \cdots \\ BM \end{bmatrix}$ is of full rank and

$$X\{M^{-1}-M^{-1}T\ (T'M^{-1}T)^{-1}TM^{-1}\}X' = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & BMB' \end{array}\right)$$

$$= X\{B'(BMB')^{-1}B\}X' \ . \tag{4.8}$$

Equations (4.3) and (4.4) follow immediately from (4.2) and (3.3).

5. Convergence properties of $g_{n,\lambda}$.

In the case of polynomial splines with $T = [0,1]$ the mean square error convergence properties (of $ER(\lambda)$) are known from [6], and we give them here for comparison purposes. We have, from Theorem 1,

$$E\ R(\lambda) = E\ \frac{1}{n}\sum_{i=1}^{n}(g(t_i)-g_{n,\lambda}(t_i))^2 \equiv \frac{1}{n}\{||(I-A(\lambda))\,\underset{\sim}{g}\,||^2+\sigma^2\ \mathrm{Tr}A^2(\lambda)\}.$$

It is shown in [6] lemmas 4.1 and 4.3 that, if $g \in W_2^{(m)}$, then

$$\frac{1}{n}||(I-A(\lambda))\underset{\sim}{g}||^2 \le \lambda \int_0^1 (g^{(m)}(u))^2 du$$

and

$$\frac{1}{n}\ \mathrm{Tr}\ A^2(\lambda) \le \frac{c\ell_m}{n\lambda^{1/2m}}\ (1+o(1))$$

where

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

and

$$c = \max_i\ [n(t_{i+1}-t_i)]^{1/2m}\ .$$

Thus (ignoring terms of order $o(1)$), an upper bound on $E\,R(\lambda)$ is given by

$$R(\lambda) \leq \lambda \int_0^1 (g^{(m)}(u))^2 du + \frac{c\ell_m \sigma^2}{n\lambda^{1/2m}} .$$

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

$$\lambda^* = \frac{\theta}{n^{2m/(2m+1)}}$$

where

$$\theta = \left[\frac{c\ell_m}{2m} \frac{\sigma^2}{\int_0^1 (g^{(m)}(u))^2 du}\right]^{2m/(2m+1)} .$$

and so

$$\min_\lambda R(\lambda) \leq R(\lambda^*) = O(n^{-2m/(2m+1)}) .$$

We remark on the comparison between this rate and that obtained by Benedetti [4] and Priestly and Chao, [15] for kernel type non-parametric regression estimates. They obtain mean square error at a point convergence rates for their estimate, call it $\hat{g}$, of the form

$$E(g(t)-\hat{g}(t))^2 = O(n^{-2m/(2m+1)})$$

under the assumption that $g^{(m)}(\cdot)$ is well-defined and bounded at t. Their rates and ours are not <u>directly</u> comparable since we assume $g \in W_2^{(m)}$, and compute an estimate of <u>integrated</u> mean square error. However, as in the case of density estimation (see [22,23]) it appears that the same convergence rates under identical assumptions will obtain if the method is matched to "m" <u>and</u> the bandwidth parameter is chosen optimally. Benedetti notes that practical objective estimates for the optimal bandwidth parameter with window estimates are yet to be obtained.

References

[1] Anderson, T. W., An Introduction to Multivariate Analysis, Wiley, New York, 1958.

[2] Anselone, P. M., and Laurent, P. J., A general method for the construction of interpolating or smoothing spline-functions. Numer. Math. 12, 66-82 (1968).

[3] Aronszajn, N., Theory of reproducing kernels, Trans. A.M.S. 68, 337-404 (1950).

[4] Benedetti, Jacqueline K., On the nonparametric estimation of regression functions, J.R.S.S. (B), 39, 2, 248-253, (1977).

[5] Clark, R. M., Non-parametric estimation of a smooth regression function, J.R.S.S. (B), 39, 1, 107-113 (1977).

[6] Craven, P., and Wahba, G., Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. University of Wisconsin, Statistics Department Technical Report #445, October 1977.

[7] Curry, H. B., and Schoenberg, On Polya frequency functions IV; the fundamental spline functions and their limits, J. Analyse Math., (7) 71-107 (1966).

[8] de Boor, C., On calculating with B-splines, J. Approximation Theory, 6, 50-62 (1972).

[9] Householder, A., The Theory of Matrices in Numerical Analysis, Blaisdell, New York (1964).

[10] International Mathematical and Statistical Libraries, Inc., Manual (1977). Subroutine ICSSCU.

[11] Kimeldorf, G., and Wahba, G., A correspondence between Bayesian estimation on stochastic processes and smoothing by splines Ann. Math. Statist., 41, 2, 495-502, (1970).

[12] Kimeldorf, G., and Wahba, G., Spline functions and stochastic processes, Sankhya, Ser. A, Vol. 32, part 2, 173-180, (1970).

[13] Kimeldorf, G., and Wahba, G., Some results on Tchebycheffian spline functions, J. Math. Anal. and Applic., 33, 1, 82-95, (1971).

[14] O'Hagan, A., Curve fitting and optimal design for prediction. J.R.S.S. (B), to appear.

[15] Priestly, M. B., and Chao, M. T., Non-parametric function fitting, J.R.S.S. (B), 34, 3, 385-392 (1972).

[16] Reinsch, C. H., Smoothing by spline functions, Numer. Math. 10, 177-183 (1967).

[17] Reinsch, C. H., Smoothing by spline functions II, Numer. Math. 16, 451-454 (1971).

[18] Rowlands, R. E., Liber, T., Daniel, I. M., Stress analysis of anisotropic laminated plates, J. American Inst. Aeronautics and Astronautics, 12, 7, 903-908 (1974).

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

[20] Shepp, L. A., Radon-Nikodym derivatives of Gaussian measures. Ann. Math. Statist. 37, 321-354 (1966).

[21] Stone, C. J., Consistent nonparametric regression, Ann. Statist. 5, 4, 595-645 (1977).

[22] Wahba, G., Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation, Ann. Statist. 3, 1, 15-29 (1975).

[23] Wahba, G., Histosplines with knots which are order statistics, J.R.S.S. (B), 38, 2, 140-151 (1976).

[24] Wahba, G., A class of approximate solutions to linear operator equations, J. of Approximation Theory, 9, 1, 61-77, (1973).

[25] Wahba, G., A canonical form for the problem of estimating smooth surfaces, Univ. of Wisconsin-Madison, Department of Statistics, Technical Report #420, August 1975.

[26] Wahba, G., Optimal smoothing of density estimates, in Classification and Clustering, J. Van Ryzin, ed., Academic Press, 423-458, (1977).

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 508	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IMPROPER PRIORS, SPLINE SMOOTHING AND THE PROBLEM OF GUARDING AGAINST MODEL ERRORS IN REGRESSION		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Grace Wahba		8. CONTRACT OR GRANT NUMBER(s) DAAG29-77-G-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Wisconsin Madison, Wisconsin		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Research Triangle Park, N.C.		12. REPORT DATE January 1978
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

spline smoothing, improper priors, model errors in regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Generalized spline smoothing is shown to be equivalent to Bayesian estimation with a partially improper prior. This result is related to a result of Kimeldorf and Wahba, J. Math. Anal. Applic. 33, 1(1971). It supports the idea that spline smoothing is a natural solution to the regression problem when one is given a set of regression functions but one also wants to hedge against the possibility that the true model is not exactly in the span of the given regression functions. A natural measure of the deviation of the true model from the span of the regression functions comes out of the spline theory in a natural way. An appropriate value of this measure can be estimated

20 continued:

from the data and used to constrain the estimated model to have the estimated deviation. Some convergence results and computational tricks are also discussed.