-----------------------
DEPARTMENT OF STATISTICS
-----------------------
University of Wisconsin
1210 W. Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 806

April 1987

# SPLINES

An Entry for the Encyclopedia of Statistical Sciences

by

Grace Wahba

This research sponsered by the National Science Foundation under Grant No. DMS-8404970.

## SPLINES

### UNIVARIATE POLYNOMIAL SPLINES

The name "spine function" was given by I. J. Schoenberg to the piecewise polynomial functions now known as univariate polynomial splines, because of their resemblance to the curves obtained by draftsmen using a mechanical spline - a thin flexible rod with weights or "ducks", used to position the rod at points through which it was desired to draw a smooth interpolating curve. See Schoenberg (1964). A univariate natural polynomial (unp) spline, $f$, is a function on [0,1] (any interval will do, of course), with the following properties: given the positive integer $m$, and $n \geq m$ points $0<t_1<t_2<...<t_n<1$, called "knots"

$$f \in \pi^{m-1},\ t \in [0,t_1],\ \ t \in [t_n,1]$$

$$f \in \pi^{2m-1},\ t \in [t_i,t_{i+1}],\ \ i=1,\ldots,n-1$$

$$f \in C^{2m-2},\ t \in [0,1],$$

where $\pi^k$ is the class of polynomials of at most degree $k$ and $C^k$ is the class of functions with k continuous derivatives. Thus, $f$ is a piecewise polynomial of degree $2m-1$ with the pieces joined at the knots so that $f$ has $2m-2$ continuous derivatives, and satisfying the $m$ boundary conditions $f^{(k)}(t_1)=f^{(k)}(t_n)=0$ for $k=m,m+1,\ldots,2m-1$. "Natural" was the term given by Schoenberg to functions satisfying these (Neumann) boundary conditions which arise "naturally" from the solution to a variational problem, to be described below.

If $f$ is represented by its polynomial coefficients, it is seen that it requires $2m$ coefficients to describe $f$ in $[0,t_1]$ and $f$ in $[t_n,1]$, $(n-1)2m$ coefficients to describe $f$ in the $n-1$ intervals $[t_i,t_{i+1}]$, $i=1,\ldots,n-1$, for a total of $2mn$ unknowns. The continuity conditions provide $(2m-1)n$ conditions, which can be shown to be linearly independent, leaving $n$ conditions to completely specify $f$. These conditions can be provided by specifying the values of $f$ at $t_1,\ldots,t_n$. Other piecewise polynomial functions obtained by modifying the boundary conditions and, or relaxing the continuity conditions at the knots are sometimes called spline functions. See DeBoor (1978), Schumaker (1981). However the unp splines have been the splines of greatest interest to statisticians because of their dual role as solutions to penalized least squares problems (variational problems) and as Bayes estimates.

The variational problem which gives rise to the ump spline takes place in the space of real valued functions on [0,1] known as $W_2^m$, the Sobolev space of functions which have $m-1$ absolutely continuous derivatives and square integrable $m$th derivative. Functions in $W_2^m$ have a Taylor series expansion with remainder to order $m-1$,

$$f(t) = \sum_{\nu=0}^{m-1} f^{(\nu)}(0)\frac{t^\nu}{\nu!} + \int_0^1 \frac{(t-s)_+^{m-1}}{(m-1)!} f^{(m)}(s)ds \tag{1}$$

with

$$\int_0^1 (f^{(m)}(s))^2 ds < \infty,$$

where $(u)_+=u$ for $u \geq 0$ and $(u)_+=0$ for $u<0$. Schoenberg showed that the solution to the following minimization problem: Find $f \in W_2^m$ to minimize $\int_0^1 (f^{(m)})^2 dt$ subject to $f(t_i)=f_i, i=1,\ldots,n$, is the unique unp spline satisfying the interpolating conditions $f(t_i)=f_i$.

The penalized least squares, or variational problem of interest for smoothing arises from the data model:

$$y_i = f(t_i) + \varepsilon_i, \quad i=1,\ldots,n,$$

where $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)' \sim N(0,\sigma^2 I)$. One estimates $f$ as the minimizer in $W_2^m$ of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda J_m(f) \tag{2}$$

where

$$J_m(f) = \int_0^1 (f^{(m)}(s))^2 ds.$$

The minimizer, call it $f_\lambda$, is called a smoothing spline. It is unique and is a unp spline. The visual character of $f_\lambda$ very much depends on the smoothing or bandwidth parameter $\lambda$. It can be shown that as $\lambda \to \infty$, $f_\lambda$ tends to the least squares regression on the polynomials of degree less than $m$, that is, on the null space of the penalty functional $J_m$, and, as $\lambda \to 0$ the solution tends to the unp interpolating spline mentioned previously.

The minimizer of (2) may be viewed as a generalization of the output of a low pass filter known in the Engineering literature as a Butterworth filter. To see the nature of this filter, we assume that $f$ has a representation as

$$f(t) = a_0 + 2\sum_{\nu=1}^{k} a_\nu \cos\pi\nu t + \sum_{\nu=1}^{k} b_\nu \sin 2\pi\nu t \tag{3}$$

where $k<n/2$. Then

$$\int_0^1 (f^{(m)}(s))^2 ds = 2\sum_{\nu=1}^{k} (a_\nu^2 + b_\nu^2)(2\pi\nu)^{2m}. \tag{4}$$

Letting $t_i = \frac{i}{n}$ and substituting (3) into (2), one obtains that

$$f_\lambda(t) = \hat{a}_0(\lambda) + 2\sum_{\nu=1}^{k} \hat{a}_k(\lambda)\cos\pi\nu t + \sum_{\nu=1}^{k} \hat{b}_k \sin 2\pi\nu t \tag{5}$$

where

$$\hat{a}_0(\lambda)=\hat{a}_0,\ \hat{a}_\nu(\lambda)=\frac{\hat{a}_\nu}{1+\lambda(2\pi\nu)^{2m}},\ \hat{b}_\nu(\lambda)=\frac{\hat{b}_0}{1+\lambda(2\pi\nu)^{2m}},$$

with

$$\hat{a}_0=\frac{1}{n}\sum_{j=1}^{n}y_j,\ \hat{a}_\nu=\frac{1}{n}\sum_{j=1}^{n}\cos 2\pi\nu\frac{j}{n}y_j,\ \hat{b}_\nu=\frac{1}{n}\sum_{j=1}^{n}\sin 2\pi\nu\frac{j}{n}y_j.$$

In words, the restriction of the smoothing spline to the periodic, equally spaced data case is obtained by taking the sample Fourier coefficients of the data, downweighting them by the filter weights $\frac{1}{1+\lambda(2\pi\nu)^{2m}}$ and taking the resulting downweighted sample Fourier coefficients as the Fourier coefficients of the estimate. For details of this kind of calculation, see Wahba (1982c).

## SPLINES AS BAYES ESTIMATES

The unp spline is a Bayes estimate under a certain *parsimonious* prior, as follows: Let $F(t), t\in[0,1]$, be the stochastic process

$$F(t)=\sum_{\nu=0}^{m-1}\theta_\nu\frac{t^\nu}{\nu!}+b^{1/2}\int_0^1\frac{(t-s)_+^{m-1}}{(m-1)!}dW(s). \tag{6}$$

Here $dW$ is the differential of the Wiener process(see e.g. Parzen (1962)), $\theta=(\theta_0,\ldots,\theta_{m-1})'$ is a parameter to be discussed, and $F(t)$ is the $m-1$ fold integrated Wiener process of Shepp (1966). Recalling (1), we may, formally, write

$$F^{(\nu)}(0)=\theta_\nu,\quad F^{(m)}(t)=b^{1/2}dW(t) \tag{7}$$

(the $m$th derivative of $F$ is "white noise") although $dW$ itself does not properly exist.

The Bayesian character of $f_\lambda$, the minimizer of (2), appears in both of the following situations:

i) Fixed effects model. Let

$$y_i=F(t_i)+\varepsilon_i,\quad i=1,\ldots,n, \tag{8}$$

where $\varepsilon$ is as before, and $\theta$ is a fixed, unknown parameter. Let $\hat{U}(t)=\sum u_i(t)y_i$ be an estimate of $F(t)$ in the model (8), linear in the data. $\hat{U}(t)$ will be called conditionally unbiassed with respect to $\theta$ if

$$E([\hat{U}(t)-F(t)]/\theta)=0. \tag{9}$$

Let $F_\lambda(t)$ be the minimum variance conditionally unbiassed linear estimate of $F(t)$, given the data $y$, that is, $E(F_\lambda(t)-F(t))^2$ is minimized, subject to $E([F_\lambda(t)-F(t)]/\theta)=0$. Then

$$F_\lambda(t)=f_\lambda(t),\quad t\varepsilon[0,1],$$

for $\lambda=\frac{\sigma^2}{nb}$, where $f_\lambda$ is the minimizer of (2). See Kimeldorf and Wahba (1971).

ii) Random effects model. Same as the fixed effects model except that $\theta_i\sim N(0,\xi)$. Let $\hat{F}_\lambda^\xi(t)$ be the conditional expectation of $F(t)$ given $y$ under this model. Then

$$\lim_{\xi \to \infty} \hat{F}_{\lambda}^{\xi}(t) = f_{\lambda}(t)$$

for $\lambda = \frac{\sigma^2}{nb}$.

MULTIVARIARIATE SPLINES

Univariate polynomial splines admit (at least) three generalizations to functions of several variables. They are, tensor product splines (see DeBoor (1978) and below), multivariate B- splines, and thin plate splines. The multivariate B- splines are generalizations of univariate B-splines (see below), as piecewise polynomials. They have many elegant properties, and have found recent application in industrial applications in surface interpolation. However, they do not appear to be related to solutions of variational problems which will lead to an interpretation as a Bayes estimate and we will not discuss them further. See Hollig (1986) and references cited there.

THIN PLATE SPLINES

The thin plate splines generalize the unp splines as solutions to a variational problem. In two dimensions ($d$=2), with $m$=2, the variational problem leading to the thinplate splines is: Find $f \varepsilon X$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_1(i), x_2(i)))^2 + \lambda J_2^2(f), \tag{10}$$

where

$$J_2^2(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{x_1x_1}^2 + 2f_{x_1x_2}^2 + f_{x_2x_2}^2 dx_1 dx_2. \tag{11}$$

$X$ is an abstract function space of functions of two variables defined in Meinguet (1979). In $d$ dimensions with general $m$, the variational problem becomes: Find $f$ in $X$, (a space of functions of $d$ variables) to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_1(i), \ldots, x_d(i)))^2 + \lambda J_m^d(f), \tag{12}$$

where

$$J_m^d(f) = \sum_{\alpha_1 + \ldots + \alpha_d} \frac{m!}{\alpha_1! \cdots \alpha_d!} \times$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[ \frac{\partial^m f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right]^2 dx_1 \cdots dx_d. \tag{13}$$

The solution to this variational problem has been given in Duchon (1977) and Meinguet (1979). Letting $t = (x_1, \ldots, x_d), t_i = (x_1(i), \ldots, x_d(i))$, and $|t| = |\sum x_i^2|^{1/2}$,

$$f_\lambda(t) = \sum_{\nu=1}^{M} d_\nu \phi_\nu(t) + \sum_{i=1}^{n} c_i E_m(t,t_i) \qquad \textbf{(14a)}$$

where $\phi_1, \ldots, \phi_M$ are the $\begin{bmatrix} d+m-1 \\ d \end{bmatrix}$ monomials in the $d$ variables $x_1, \ldots, x_d$ of total degree less than $m$, and

$$E_m(t,t_i) = const \ |t-t_i|^{2m-d} \ln |t-t_i|, \ 2m-d \ an \ even \ integer$$

$$E_m(t,t_i) = const \ |t-t_i|^{2m-d}, \ otherwise \qquad (14b)$$

It is necessary that $2m-d>0$. The coefficients $c$, $d$ satisfy the $n+M$ equations

$$(K + n\lambda I)c + Td = y$$

$$T'c = 0 \qquad (14c)$$

where $K$ is the $n \times n$ matrix with $ij$th entry $E_m(t_i,t_j)$, $T$ is the $n \times n-M$ matrix with $i\nu$th entry $\phi_\nu(t_i)$, and the solution is known to be unique if $T$ is of rank $M$. (See Wahba and Wendelberger (1980).) As $\lambda \to \infty, f_\lambda$ tends to the least squares regression of the data onto the polynomials of total degree less than $M$ ( that is, onto the null space of $J_m^d$), and as $\lambda \to 0, f_\lambda$ tends to the minimizer of $J_m^d(f)$ subject to the interpolating conditions $f(t_i)=y_i$. The special case $d=1$ can be shown to reduce to the unp spline on [0,1] by using the fact that in that case $f_\lambda^{(m)}$ is 0 outside $[t_1,t_n]$.

We remark that $E_m(\cdot,\cdot)$ is a Green's function for the $m$th iterated Laplacian, $\Delta^m$, where $\Delta = \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_d^2}$, and, if $f$ has compact support, one obtains, by integration by parts that

$$J_m^d(f) = \int \cdots \int f \cdot \Delta^m f$$

Furthermore, $\Delta^m f_\lambda(t)=0$ ,for $t \neq t_1, \ldots, t_n$. (Compare the $d=1$ case.)

The integral over the plane in (11) can be replaced by an integral over a region $\Omega$ containing the data points (see Dyn and Wahba (1982)) but then the minimizer is a solution to a partial diffenential equation with Neumann boundary conditions, on the boundary of $\Omega$, and the corresponding Bayes estimate will have lost its stationary character.

An intuitive idea of the properties of the thin plate spline as the output of a low pass filter may be obtained by condsidering $f(x_1,..,x_d)$ to be a multiply periodic function on the $d$-dimensional torus and replacing the limits $(-\infty,\infty)$ by [0,1], and repeating the arguments of (3)-(5). One finds that $\hat{a}_\nu(\lambda)$ is replaced by

$$\hat{a}_{\nu_1,\ldots,\nu_d}(\lambda) + \frac{\hat{a}_{\nu_1,\ldots,\nu_d}}{\left[1+\lambda \left[\sum_{j=1}^{d}(2\pi\nu_j)^m\right]\right]^2}$$

and so forth.

THE THIN PLATE SPLINE AS A BAYES ESTIMATE

The Bayes estimates which generalize those given in the $d$=1 case can be established with the help of the notion of a generalized divided difference (gdd) of order $m$ in $d$ variables with repect to the points $t_1, \ldots, t_n$. Recall that an ordinary first order divided diference of a function $f$ of one variable, with respect to $t_1, t_2$ is $[(f(t_2)-f(t_1)]/(t_2-t_1)$, second divided differences are defined as divided differences of first divided differences, and so forth. If $f$ is a polynomial of degree $k$, then its divided differences of order greater than $k$ are 0. Let $T$ be the $n \times M$ matrix defined in (14), and suppose that $T$ is of rank $M$. Let $c$ be any $n$ vector satisfying $T'c=0$. Then $c=(c_1, \ldots, c_n)$ is called a generalized divided difference (with respect to $t_1, \ldots, t_n$) because it annihilates all the polynomials of order less than $m$ in $d$ variables, by the definition:

$$\sum_{i=1}^{n} c_i \phi_\nu(t_i) = 0, \quad \nu=1, \ldots, M.$$

Now let $F(t), t \in E^d$ be a Gaussian stochastic process defined as follows:

$$F(t) = \sum_{\nu=1}^{M} \theta_\nu \phi_\nu(t) + b^{1/2} X(t)$$

where the $\theta$'s are as in either the fixed effects or the random effects model. $X(t), t \varepsilon [-\infty, \infty]$ is a zero mean Gaussian stochastic process such that, for any gdd $(c_1, \ldots, c_n)$ with respect to $(t_1, \ldots, t_n)$,

$$E \sum_{i,j} c_i c_j X(t_i) X(t_j) = \sum_{i,j} c_i c_j E_m(t_i, t_j).$$

This can be shown to define a legitimate covariance on the family of all gdd's of $X$ since $E_m(t_i - t_j)$ is known to be positive definite on all gdd's, that is for all $c$ such that $T'c=0$. See Matheron (1973), Duchon (1977). Any two zero mean Gaussian stochastic processes with the property that their gdd's have the same covariance must differ only by a random polynomial of total degree less than m, thus $X$ is being defined here only up to a random polynomial.

The theorems are as follows:

i) Let

$$y_i = F(t_i) + \varepsilon_i$$

where the $\varepsilon_i$ 's are as before and $\theta$ is fixed. Let $F_\lambda(t)$ be the minimum variance conditionally unbiassed estimate of $F(t)$ given the data $y$. Then

$$F_\lambda(t) = f_\lambda(t), \text{for } \lambda = \frac{\sigma^2}{nb}$$

See Kimeldorf and Wahba (1971).

ii) Let $y_i = f(t_i) + \varepsilon_i$ as before except that the $\theta_i$ 's are i. i. d $N(0, \xi)$. Let $\hat{F}_\lambda^\xi(t)$ be the conditional expectation of of $F(t)$ given $y$. Then

$$\lim_{\xi \to \infty} \hat{F}_\lambda^\xi(t) = f_\lambda(t), \text{for } \lambda = \frac{\sigma^2}{nb}$$

See Wahba (1978).

SPLINES ON THE CIRCLE AND THE SPHERE

Splines on the circle can be obtained by supposing that f is in the periodic subspace of $W_2^m$ of functions of the form

$$f(t)=a_0+\sum_{\nu=1}^{\infty}a_\nu\cos 2\pi\nu t+\sum_{\nu=1}^{\infty}b_\nu\sin 2\pi\nu t,\quad \sum_{\nu=1}^{\infty}(a_\nu^2+b_\nu^2)(2\pi\nu)^{2m}<\infty,$$

and finding $f$ to minimize

$$\frac{1}{n}\sum_{\nu=1}^{n}(y_i-f(t_i))^2+\int_o^1(f^{(m)}(u))^2du.$$

The equivalent Bayes model is $y_i=F(t_i)+\varepsilon_i$, where $F(t)=\theta+X(t)$, and

$$X(t)=\sum_{\nu=1}^{\infty}\alpha_\nu\cos 2\pi\nu t+\beta_\nu\sin 2\pi\nu t$$

where the $\alpha_\nu$, $\beta_\nu$ are independent, zero mean normal random variables with $E\alpha_\nu^2=E\beta_\nu^2=(2\pi\nu)^{-2m}$. A closed form expression for $f_\lambda$ as a piecewise polynomial may be obtained by using the fact that

$$EX(s)X(t)=\sum_{\nu=1}^{\infty}(\cos 2\pi\nu s\cos 2\pi\nu t+\sin 2\pi\nu s\sin 2\pi\nu t)(2\pi\nu)^{-2m}$$

$$=\sum_{\nu=1}^{\infty}\cos 2\pi\nu(s-t)(2\pi\nu)^{-2m}$$

and this latter infinite series has a closed form expression in terms of the $2m$ th Bernoulli polynomial, see Craven and Wahba (1979).

The spherical harmonics $Y_{ls}, s=-l,\ldots,l, l=0,1,\ldots$ play the same role on the sphere as sines and cosines on the circle. See Sansone (1959) for more on spherical harmonics. The spherical harmonics are the eigenfunctions of the surface Laplacian $\Delta$ on the sphere, with

$$\Delta Y_{ls}=-l(l+1)Y_{ls}$$

which is analogous to

$$\frac{d^2}{dt^2}\cos 2\pi\nu t=-(2\pi\nu)^2\cos 2\pi\nu t$$

Splines on the sphere are defined as the solution to the variational problem: find $f\varepsilon X$ (an appropriate space) to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-f(P_i))^2+\lambda\int_S(\Delta^{\frac{m}{2}}f)^2dP \tag{16}$$

where $S$ is the sphere, $P$ is a point on the sphere, and

$$\int_S (\Delta^{\frac{m}{2}} f)^2 dP = \int_S f \cdot \Delta^m f dP.$$

For comparison with the circle note that for periodic functions

$$\int_0^1 (f^{(m)}(u))^2 du = \int_0^1 f \cdot f^{(2m)} du.$$

A closed form expression is available for $f_\lambda$ in the cases $m$=2,3 (Wendelberger (1982)). Approximate closed form expressions may be found in Wahba (1981b), Wahba (1982a). The corresponding Bayes model is

$$y_i = \theta + \sum_{l,s} f_{l,s} Y_{l,s}(P_i) + \varepsilon_i$$

where $Ef_{l,s}^2 = l(l+1)^{-m}$. Splines on the sphere have a number of interesting applications in geophysics and meteorology, see for example Shure, Parker, and Backus (1982). Vector smoothing splines can also be defined on the sphere and are useful in estimating horizontal vector fields from discrete, noisy measurements on, for example, the horizontal wind field, the magnetic field, etc., see Wahba (1982b).

CHOOSING THE SMOOTHING PARAMETER, DIAGNOSTICS,CONFIDENCE INTERVALS

Generalized cross validation(GCV) appears to be the most popular mehtod for choosing the smoothing parameter $\lambda$ from the data in the context of smoothing splines, for various theoretical and practical reasons, see Craven and Wahba (1979), Li (1986), Speckman (1982), Utreras (1978), Wahba (1975), Wahba (1985b). GCV is obtained from an ordinary leaving out one or ordinary cross validation procedure by an invariance argument.(Ordinary leaving out one is not invariant under rotations of the observation coordinate system.)

Let $A(\lambda)$ be the influence matrix associated with $f_\lambda$, that is, $A(\lambda)$ satisfies

$$\begin{pmatrix} f_\lambda(t_1) \\ \cdot \\ \cdot \\ \cdot \\ f_\lambda(t_n) \end{pmatrix} = A(\lambda) \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}.$$

The GCV estimate $\hat{\lambda}$ of $\lambda$ is obtained as the minimizer of

$$V(\lambda) = \frac{\frac{1}{n} \| (I - A(\lambda)) y \|^2}{\left[ \frac{1}{n} Tr(I - A(\lambda)) \right]^2}.$$

$A(\lambda)$ has many of the properties of the influence or hat matrix in ordinary least squares regression, and this can be used to build a theory of spline regression diagnostics. See Eubank (1986).

Trace $A(\hat{\lambda})$ can be viewed as the "degrees of freedom for signal" The posterior covariance matrix of $(f_\lambda(t_1), \ldots, f_\lambda(t_n))'$ is $\sigma^2 A(\lambda)$ and this fact has been used to propose posterior "confidence intervals" based on $\hat{\sigma}^2 A(\hat{\lambda})$ where $\hat{\sigma}^2 = RSS(\hat{\lambda})/Tr(I - A(\hat{\lambda}))$. See Wahba (1983). These "confidence intervals" appear to have useful frequentist properties if interpreted "across the function", rather than pointwise, see Hall and Titterington (Nov. 1985), Nychka (1986), Silverman (1985), Wahba (1985a).

PARTIAL SPLINES

Consider the model

$$y_i = f(t_i) + \sum_{j=1}^{p} \theta_j \Phi_j(t_i, z_i) + \varepsilon_i .$$

Here $t_i \varepsilon E^d, z_i=(z_1(i),, \ldots, z_q(i))'$, $f$ is assumed to be "smooth" and the $\Phi_j$'s are known functions of $t \varepsilon E^d$ and the comcommittant variables $z$. An estimate of $\theta$ and $f$ may be obtained by finding $f \varepsilon X$ and $\theta \in E^p$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i) - \sum_{j=1}^{p}\theta_j \Phi_j(t_i, z_i))^2 + J_m^d(f).$$

Let $S_{n \times p}$ be the matrix with $i,j$th entry $\Phi_j(t_i, z_i)$. If the $n \times (M+p)$ matrix $(T:S)$ is of full column rank, then there will be a unique minimizer $(f_\lambda, \hat{\theta})$, and $f_\lambda$ will be a thin plate spline (in the case $d$=1, a unp spline). Such models are extremely flexible and are attractive in a variety of applications, see Engle et al. (1986), Ansley and Wecker (1981), Shiau, Wahba, and Johnson (1986), Wahba (1984) and references cited there. Properties of the estimate of $\theta$ are an area of active research, see, for example, Heckman (1986), Rice (1986).

SPLINES AS PENALIZED LIKELIHOOD ESTIMATES

Let

$$y_i \sim \textit{Binomial}\,(1, p(t_i)),$$

where the logit $f(t) = p(t)/(1-p(t), t \varepsilon E^d$ is assumed to be a smooth function of $t$. The likelihood of $y$ is $p^y(1-p)^{1-y} = e^{y\frac{p}{1-p} - \ln\frac{1}{1-p}} = e^{yf - \ln(1+e^f)}$. The log likelihood of $y_1, \ldots, y_n$ then becomes

$$\log L = Q(y, f) = \sum_i y_i f(t_i) - \ln(1+e^{f(t_i)}).$$

A penalized log likelihood estimate of $f$ is then the minimizer in $X$ of

$$Q(y, f) + \lambda J_m^d(f). \tag{15}$$

The minimizer $f_\lambda$ can be shown to be a thin plate spline. If $t \varepsilon S$, then the spline penalty functional on the sphere can be used, and the solution is a spline on the sphere, and so forth. The binomial distribution above can be replaced by the Poisson distribution $\Lambda(t)$ or, any member of the exponential family. These penalized likelihood estimates generalize the GLIM (generalized linear) models described in McCullagh and Nelder (1983), and reduce to the parametric GLIM model in the null space of the penalty functional in the case $\lambda=\infty$. Equation (15) generally has to be solved by numerical methods, typically in a sufficiently large but finite dimensional space of convenient basis functions. GCV estimates of $\lambda$ are defined for these models as follows: For a trial value of $\lambda$, (15) is minimized by a Gauss-Newton iteration, equivalently, by minimizing an approximating sequence of quadratic optimization problems. At convergence, one evaluates the GCV function $V(\lambda)$ for the final quadratic optimization problem, repeats the calculation for another trial value of $\lambda$, until a minimizer is found, see O'Sullivan, Yandell, and Raynor (1986). A discussion of thin plate basis functions is in Wahba (1980). Silverman (1982) has proposed a density estimate in the context of penalized likelikhood with the spline penalty functional.

REGRESSION SPLINES

Returning to the univariate case, let $h>0$ and consider the density function obtained as the convolution of $2m$ uniform density functions on $[0,h]$. This density is supported on $[0,2m]$, and can in fact be shown to be a piecewise polynomial of degree $2m-1$ with knots at $h, 2h, \ldots, (2m-1)h$ and having $2m-2$ continuous derivatives. For $m$ greater than 2 or 3 this density very much resembles a normal curve over most of its support. Such piecewise polynomials satisfying continuity conditions and having compact support are known as B-splines, and can be defined for arbitrarily spaced knots, see DeBoor (1978), Lyche and Schumaker (1973). The B-splines are chosen so that their supports are overlapping and so that they have common knots wherever possible. Shifted and scaled B-splines are frequently used as basis functions to obtain approximate solutions to variational problems. Consider the model

$$y_i = f(t_i) + \varepsilon_i, \quad i=1, \ldots, n,$$

where $f$ is "smooth". Various authors have considered estimating $f$ by doing ordinary least squares regression on a set of $N$ B-splines, where $N << n$. Here the number $N$ of B-splines acts as a smoothing parameter, if $N$ is small then the estimate of $f$ will be very smooth, while if $N$ is sufficiently large, the data can be interpolated. Such splines are known as regression splines (as opposed to smoothing splines). If $f \in W_2^2$, then as $n \to \infty$, the optimal value of $N$ for integrated mean square error is $O(n^{1/5})$, see Agarwal and Studden (1980), so that for $n$ of the order of 100, typically the optimal value of $N$ may be as small as 3 or 4. In this kind of situation smoothing splines are likely to provide better resolution of multiple peaks in $f$ than regression splines. If $f$ is already near the span of a small number of basis functions, or $n$ is extremely large, the results with the two kinds of splines is likely to be similar, given that $\lambda$ and $N$ are both chosen optimally. Given $N$, it has been suggested that the knots in a B-spline basis for regression splines be chosen by least squares. Although apparently it is possible to choose knots by "eyeball", this is a notoriously difficult numerical problem even for a small number of knots, and multiple minima of the objective function and other problems can be expected.

INTERACTION SPLINES

Let $t=(x_1, \ldots, x_d)$ be in the unit cube in $E^d$. Stone (1985) and others have suggested modeling $f$ as

$$f(x_1, \ldots, x_d) = f_0 + \sum_{\alpha=1}^{d} f_\alpha(x_\alpha)$$

where $\int_0^1 f_\alpha(x_\alpha)dx_\alpha = 0$, $\alpha=1, \ldots, d$. This idea has been extended to model $f$ as

$$f(x_1,\ldots,x_d)=f_0+\sum_{\alpha=1}^{d} f_\alpha(x_\alpha)+\sum_{\alpha<\beta} f_{\alpha\beta}(x_\alpha,x_\beta)+\sum_{\alpha<\beta<\gamma} f_{\alpha\beta\gamma}(x_\alpha,x_\beta,x_\gamma)+\cdots$$

where appropriate marginal integrals are 0 to guarantee identifiability, and the number of terms to be included is to be determined. Such splines are known as interaction splines, by analogy with analysis of variance. The individual component functions $f_{\alpha\beta}, f_{\alpha\beta\gamma}$, etc. are tensor product splines. See Barry (1983), Barry (1986), Wahba (1986), Chen (October 1986).

SPLINES WITH LINEAR INEQUALITY CONSTRAINTS

Splines satisfying a family of linear inequality constraints can be found as the solution to the problem: Find $f \varepsilon X$ to minimize

$$\frac{1}{n}\sum(y_i-f(t_i))^2+\lambda J(f)$$

subject to

$$a_i \le L_i f \le b_i,$$

where $L_i$ is is a bounded linear functional. Included are discretized positivity and monotonicity constraints, see Utreras (1985), Villalobos and Wahba (March 1987).

HISTOSPLINES

Histosplines arise when one observes

$$y_i=\int_{\Omega_i} f\,dt+\varepsilon_i$$

and chooses $f$ as the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\int_{\Omega_i} f(t)dt)^2+\lambda J_m^d(f), \tag{16}$$

See Wahba (1981a), Dyn, Wahba, and Wong (1979) and references cited there.

INDIRECT SENSING PROBLEMS

These occur when one observes

$$y_i=\int K(t_i,s)f(s)dt+\varepsilon_i$$

and are of major practica importance. $f$ can be estimated by solving a variational problem analogous to (16). See O'Sullivan (1986) and references cited there.

ALGORITHMS AND SOFTWARE

This is an area of active research and we only briefly mention a few results. In one dimension the unp spline has special structure which allows fast algorithms for computing both the spline and the GCV estimate of $\lambda$. Transportable code CUBGCV based on the fast algorithm proposed by Hutchinson and deHoog (1985) may be found in Hutchinson (1985), this algorithm with some additions is incorporated in GCVSPL of Woltring (1986). Literature connecting the Markov properties of the unp spline and its relationship to Kalman filtering has suggested fast algorithms, an early reference is Weinert and Kailath (1974). Older code (ICSSCV) for the unp spline with the GCV estimate of $\lambda$ can be found in the IMSL library (1986). In more than one variable, the special structure of the one dimensional case does not appear to exist and more general methods are required. The bidiagonalization approach of Elden (1984 ) and the truncated singular value decomposition, Bates and Wahba (1982) may be used to speed the calculation. Transportable code for thin plate splines using thin plate basis functions is available in Hutchinson (April 1984), and for partial thin plate splines and general problems using the truncated singular value decomposition, in GCVPACK (Bates et al. (October 1986)). GCVSPL, GCVPACK, code for generating B-splines based on DeBoor (1978) and other spline code may be obtained via an electronic mail daemon on the arpanet by writing netlib@anl-mcs.arpa. The message "send index" will cause instructions for the use of the system to be returned to the sender.

## REFERENCES


Agarwal, G., and Studden, W. J. (1980), "Asymptotic Integrated Mean Square Error Using Least Squares and Bias Minimizing Splines," *Annals of Statistics*, 8, 1307-1325.

Ansley, C. F., and Wecker, W. E. (1981) "Extensions and Examples of the Signal Extraction Approach to Regression." *Proceedings of ASA-CENSUS-NBER Conference on Applied Time Series Analysis of Economic Data*, Washington, D. C.181-192.

Barry, D. (1983) "Nonparametric Bayesian Regression." thesis, Yale University.

Barry, D. (1986), "Nonparametric Bayesian Regression," *Ann. Statist.*, 14, 934-953.

Bates, D. M., and Wahba, G. (1982), "Computational Methods for Generalized Cross-Validation with Large Data Sets," in *Treatment of Integral Equations by Numerical Methods*, eds. C. T. Baker, and G. F. Miller, London: Academic Press, 283-296.

Bates, D. M., Lindstrom, M. J., Wahba, G., and Yandell, B. (October 1986), "GCVPACK - Routines for Generalized Cross Validation," *Commun. Statist.*, to appear, University of Wisconsin-Madison Statistics Dept. Technical Report No. 775(rev.).

Chen, Z. (October 1986), "A Stepwise Approach for the Purely Periodic Interaction Spline Model," *Commun. Statist. Theory and Methods*, to appear, University of Wisconsin-Madison Statistics Dept. Technical Report No. 792.

Craven, P., and Wahba, G. (1979), "Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numer. Math.*, 31, 377-403.

DeBoor, C. (1978), *A Practical Guide to Splines*, New York: Springer-Verlag.

Duchon, J. (1977), "Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces," in *Constructive*

*Theory of Functions of Several Variables*, eds. W. Schempp, and K. Zeller Springer-Verlag Berlin, 85-100.

Dyn, N., Wahba, G., and Wong, W. (1979), "Comment on "Smooth Pyhchnophylactic Interpolation for Geographical Regions" by W. Tobler," *J. Am. Statist. Assoc.*, 74, 530-535.

Dyn, N., and Wahba, G. (1982), "On the Estimation of Functions of Several Variables from Aggregated Data," *SIAM J. Math. Anal.*, 13, 134-152.

Elden, L. (1984 ), "A Note on the Computation of the Generalized Cross-Validation Function for Ill-Conditioned Least Squares Problems," *BIT*, 24.

Engle, R., Granger, C., Rice, J., and Weiss, A. (1986), "Semiparametric Estimates of the Relation Between Weather and Electricity Sales.," *J. Am. Stat. Soc.*, 81, 310-320.

Eubank, R. L. (1986), "Diagnostics for Smoothing Splines," *J. Roy. Stat. Soc. Ser. B*, 47. (to appear)

Hall, P., and Titterington, D. M. (Nov. 1985) "On the Structure of Smoothing Techniques in Regression Problems." *submitted to Journal of the Royal Statistical Society, Ser. B*, .

Heckman, N. E. (1986), "Spline Smoothing in a Partly Linear Model," *J. Roy. Stat. Soc., Ser. B*, 48, 244-248.

Hollig, K. (1986), "Multivariate Splines," in *Approximation Theory*, ed. C. deBoor, Providence, RI: American Mathematical Society, 103-127.

Hutchinson, M. F., and deHoog, F. R. (1985), "Smoothing Noisy Data with Spline Functions," *Numer. Math.*, 47, 99-106 .

Hutchinson, M. F. (1985), "Algorithm 642 A Fast Procedure for Calculating Minimum Cross Validation Cubic Smoothing Splines," *ACM Trans. Math. Software*, 12, 150-153.

Hutchinson, M. F. (April 1984), *A Summary of Some Surface Fitting and Contouring Programs for Noisy Data*, Canberra, Autralia: CSIRO Division of Mathematics and Statistics. (Consulting Report No. ACT 84/6)

Kimeldorf, G., and Wahba, G. (1971), "Some Results on Tchebycheffian Spline Functions," *J. Math. Anal. Applic.*, 33, 82-95.

Li, K. (1986), "Asymptotic Optimality of C Sub L and Generalized Cross-Validation in Ridge Regression with Application to Spline Smoothing," *Ann. Statist.*, 14, 1101-1112.

IMSL (1982), International Mathematical and Statistical Library, Houston, TX.

Lyche, T., and Schumaker, L. L. (1973), "Computation of Smoothing and Interpolating Natural Splines Via Local Bases," *SIAM J. Numer. Anal.*, 10, 1027-1038.

Matheron, G. (1973), "The Intrinsic Random Functions and Their Applications," *Adv. Appl. Prob.*, 5, 439-468.

McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman and Hall.

Meinguet, J. (1979), "Multivariate Interpolation at Arbitrary Points Made Simple," *J. App. Math. Phys. (ZAMP)*, 30, 292-304.

Nychka, D. (1986), *A Frequency Interpretation of Bayesian "Confidence" Intervals for Smoothing Splines*, Raleigh, N. C.: The Institute of Statistics, North Carolina State University Mimeo Series 1699.

O'Sullivan, F., Yandell, B., and Raynor, W. (1986), "Automatic Smoothing of Regression Functions in Generalized Linear Models," *J. Am. Stat. Soc.*, 81, 96-103.

O'Sullivan, F. (1986), "A Statistical Perspective on Ill-Posed Inverse Problems," *Statistical Science*, 1, 502-527.

Parzen, E. (1962), *Stochastic Processes*, San Fransisco: Holden-Day.

Rice, J. (1986), "Convergence Rates for Partially Splined Models," *Statistics & Probability Letters*, 4, 203-208.

Sansone, G. (1959), *Orthogonal Functions*, New York: Interscience.

Schoenberg, I. (1964), "Spline Functions and the Problem of Graduation," *Proc. Nat. Acad. Sci. U. S. A.*, 52, 947-950.

Schumaker, L. (1981), *Spline Functions*, Wiley.

Shepp, L. A. (1966), "Radon-Nikodym Derivatives of Gaussian Measures," *Ann. Math. Statist.*, 37, 321-354.

Shiau, J., Wahba, G., and Johnson, D.R. (1986), "Partial Spline Models for the Inclusion of Tropopause and Frontal Boundary Information in Otherwise Smooth Two and Three Dimensional Objective Analysis," *J. Atmos. Ocean Tech.*, 3, 714-725.

Shure, L., Parker, R.L., and Backus, G.E. (1982), "Harmonic Splines for Geomagnetic Modelling," *J. Physics of the Earth and Planteary Interiors*, 28, 215-229.

Silverman, B. (1982), "On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method," *Ann. Statist.*, 10, 795-810.

Silverman, B.W. (1985), "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting," *J. Roy. Stat. Soc. B*, 46, 1-52.

Speckman, P. (1982) "Efficient Nonparametric Regression with Cross- Validated Smoothing Splines." Dept. of Statistics,University of Missouri-Columbia.

Stone, C.J. (1985), "Additive Regression and Other Nomparametric Models," *Ann. Statist.*, 13, 689-705.

Utreras, F. (1978) "Quelques Resultats D'Optimalite Pour La Methode De Validation Croissee." Seminaire d'Analyse Nuumerique, 301, Grenoble:.

Utreras, F. (1985), ''Smoothing Noisy Data Under Monotonicity Constraints, Existence, Characterization and Convergence Rates,'' *Numer. Math*, 47, 611-625.

Villalobos, M., and Wahba, G. (March 1987), ''Inequality Constrained Multivariate Smoothing Splines with Application to the Estimation of Posterior Probabilities,'' *J. Am. Statist. Assoc.*, 82, 239-248.

Wahba, G. (1975), ''Smoothing Noisy Data by Spline Functions,'' *Numer. Math.*, 24, 383-393.

Wahba, G. (1978), ''Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression,'' *J. Roy. Stat. Soc., Ser.B*, 40, 364-372.

Wahba, G. (1980), ''Spline Bases, Regularization, and Generalized Cross Validation for Solving Approximation Problems with Large Quantities of Noisy Data,'' in *Approximation Theory III*, ed. W. Cheney Academic Press, 905-912.

Wahba, G., and Wendelberger, J. (1980), ''Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross-Validation,'' *Monthly Weather Review*, 108, 1122-1145.

Wahba, G. (1981a), ''Numerical Experiments with the Thin Plate Histospline,'' *Communications in Statistics - Theory and Methods*, 10, 2475-2514.

Wahba, G. (1981b), ''Spline Interpolation and Smoothing on the Sphere,'' *SIAM J. Sci. Stat. Comput.*, 2, 5-16.

Wahba, G. (1982a), ''Erratum: Spline Interpolation and Smoothing on the Sphere,'' *SIAM J. Sci. Stat. Comput*, 3, 385-386.

Wahba, G. (1982b), ''Vector Splines on the Sphere, with Application to the Estimation of Vorticity and Divergence from Discrete, Noisy Data,'' in *Multivariate Approximation Theory, Vol. 2*, ed. W. S. K. Zeller Birkhauser Verlag, 407-429.

Wahba, G. (1982c), ''Constrained Regularization for Ill Posed Linear Operator Equations, with Applications in Meteorology and Medicine,'' in *Statistical Decision Theory and Related Topics III, Vol. 2*, eds. S. S. Gupta, and J. O. Berger Academic Press, 383-418.

Wahba, G. (1983), ''Bayesian "Confidence Intervals" for the Cross-Validated Smoothing Spline,'' *J. Roy. Stat. Soc. B.*, 45, 133-150.

Wahba, G. (1984), ''Cross Validated Spline Methods for the Estimation of Multivariate Functions from Data on Functionals,'' in *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory*, eds. H. A. David, and H. T. David Iowa State University Press, 205-235.

Wahba, G. (1985a), ''Comments to "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting by B. W. Silverman,'' *J. Roy. Stat. Soc. B*, 46, 44.

Wahba, G. (1985b), ''A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem,'' *Ann. Statist.*, 13, 1378-1402.

Wahba, G. (1986), ''Partial and Interaction Splines for the Semiparametric Estimation of Functions of Several Variables,'' in *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface*, ed. T. J. Boardman, Washington, D. C.: American Statistical Association, 75-80.

Weinert, H. L., and Kailath, T. (1974), ''Stochastic Interpretations and Recursive Algorithms for Spline Functions,'' *Ann. Statist.*, 2, 787-794.

Wendelberger, J. (1982), *Smoothing Noisy Data with Multidimensional Splines and Generalized Cross-Validation*, University of Wisconsin-Madison Statistics Dept. PhD. Thesis.

Woltring, H. J. (1986), ''A FORTRAN Package for Generalized, Cross-Validatory Spline Smoothing and Differentiation,'' *Advances in Engineering Software*, 8, 104-113.