Spline Function: Overview

By *Grace Wahba and Yuedong Wang*

1 Univariate Polynomial Splines

The name "spline 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 [58]. 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 $0 < t_1 < t_2 < \cdots < t_n < 1$, called "*knots*"

$$\begin{aligned} &f \in \pi^{m-1}, t \in [0, t_1], t \in [t_n, 1], \\ &f \in \pi^{2m-1}, t \in [t_i, t_{i+1}], i = 1, \ldots, n-1, \\ &f \in C^{2n-2}, t \in [0, 1], \end{aligned}$$

where π^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, satisfying the $2m$ boundary conditions $f^{(k)}(0) = f^{(k)}(1) = 0$ 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 specify f completely. These conditions can be provided by specifying the values of f at $t_1, \ldots, t_n$.

Interpolating (unp) splines have been of interest to numerical analysts at least since Schoenberg's 1964 work. Suppose that f is some function which possesses a **Taylor expansion** with remainder to order $m - 1$, and let f_n be the unp spline of interpolation to f at

This article was originally published online in 2006 in Encyclopedia of Statistical Sciences, © John Wiley & Sons, Inc. and republished in Wiley StatsRef: Statistics Reference Online, 2014.

the points $t_1, \ldots, t_n$. Then f_n and its first $m-1$ derivatives tend point-wise to f and its first $m-1$ derivatives as n gets large, provided the t_i's are distributed "nicely," and if f possesses $2m$ continuous derivatives and satisfies the Neumann boundary conditions, then all $2m$ of the derivatives of f_n will converge to those of f. Integrals of f_n also converge to integrals of f, and this fact can be used to generate **quadrature** formulas; see Schoenberg [59]. These favorable approximation theoretic properties, as well as the fact that splines are easy to compute, have led to their popularity among numerical analysts. Interpolating splines are frequently the functions of choice when it is desired to represent everywhere a function whose values are given exactly only on a finite set of points. Piecewise polynomial functions satisfying other boundary and continuity conditions are also called splines. The scholarly work of Schumaker [60] provides a history of interpolating splines from a numerical analyst's point of view. Prenter [54] describes their role in the numerical solution of differential equations and DeBoor [11] is the standard reference on algorithms for generating (univariate) splines.

Splines are of interest to statisticians for the same reasons that they are of interest to numerical analysts, as well as because of their favorable properties in smoothing noisy data. The two major types of splines of interest for smoothing noisy data as a function of one variable are regression splines and smoothing splines. We will discuss these in turn, and then go on to some other spline models.

2 Regression Splines

To discuss regression splines, we first want to describe the B-splines (B stands for "basis"). B-splines can be conveniently defined in terms of truncated power functions and divided difference operators. Given a function $f(\cdot)$ and "knots" $t_i, \ldots, t_{i+k}$ define the divided difference operator $[t_i, \ldots, t_{i+k}]f(\cdot)$ as

$$\begin{aligned}
[t_i, t_{i+1}]f(\cdot) &= \frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}, \\
[t_i, t_{i+1}, t_{i+2}]f(\cdot) &= \left[\frac{f(t_{i+2}) - f(t_{i+1})}{t_{i+2} - t_{i+1}} - \frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}\right] / (t_{i+2} - t_i),
\end{aligned}$$

and so forth. For fixed x, we will let $f(\cdot) = (\cdot - x)_+^{k-1}$ be the truncated power function, where $(u)_+ = u$ if $u \geq 0$ and 0 otherwise. Then the B-spline of degree $k-1$ for the knots $t_i, \ldots, t_{i+k}$ is defined as

$$B_i(x) = (t_{i+k} - t_i)[t_i, \ldots, t_{i+k}](\cdot - x)_+^{k-1}.$$

For example, for $k = 2$,

$$\begin{aligned}
B_2(x) &= \frac{(t_{i+2} - x)_+ - (t_{i+1} - x)_+}{t_{i+2} - t_{i+1}} \\
&\quad - \frac{(t_{i+1} - x)_+ - (t_i - x)_+}{t_{i+1} - t_i},
\end{aligned}$$

which is a tent function on $[t_i, t_{i+2}]$. $B_i(x)$ is a piecewise polynomial of degree $k-1$ and can be shown to possess $k-2$ continuous derivatives, be zero outside $[t_i, \ldots, t_{i+k}]$, and positive in the interior. The B-splines of degree $k-1$ can also be obtained as projections on the real line of the volumes of convex polyhedra in k dimensions. Figure 1 illustrates this for $k = 2$. B-splines with coalescing knots are allowed; the effect is to reduce the continuity conditions at the multiple knots. If the t_i's are equally spaced, then one can also get the B-splines by shifting and rescaling the convolution of k uniform densities, and then the B-splines for $k \geq 1$ will be hill functions. Given k and a set of knots $t_1, \ldots, t_{N+k}$, one can define a set of N B-splines $B_1, \ldots, B_N$ where B_j is the B-spline of degree $k-1$ with knots $t_j, \ldots, t_{j+k}$. These N functions provide a basis for a space of smooth functions with $k-2$ continuous derivatives and may be used as regression functions when one wants to fit a smooth function without otherwise specifying its form. Specifically, suppose we are given data $\mathbf{y} = (y_1, \ldots, y_n)$, from the model

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where f is known to be "smooth" and $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \sim \mathrm{N}(\mathbf{0}, \sigma^2 I)$. Given knots $t_1, \ldots, t_{N+k}$ and the corresponding B-splines of degree $k-1$, one may estimate f as f_N, where

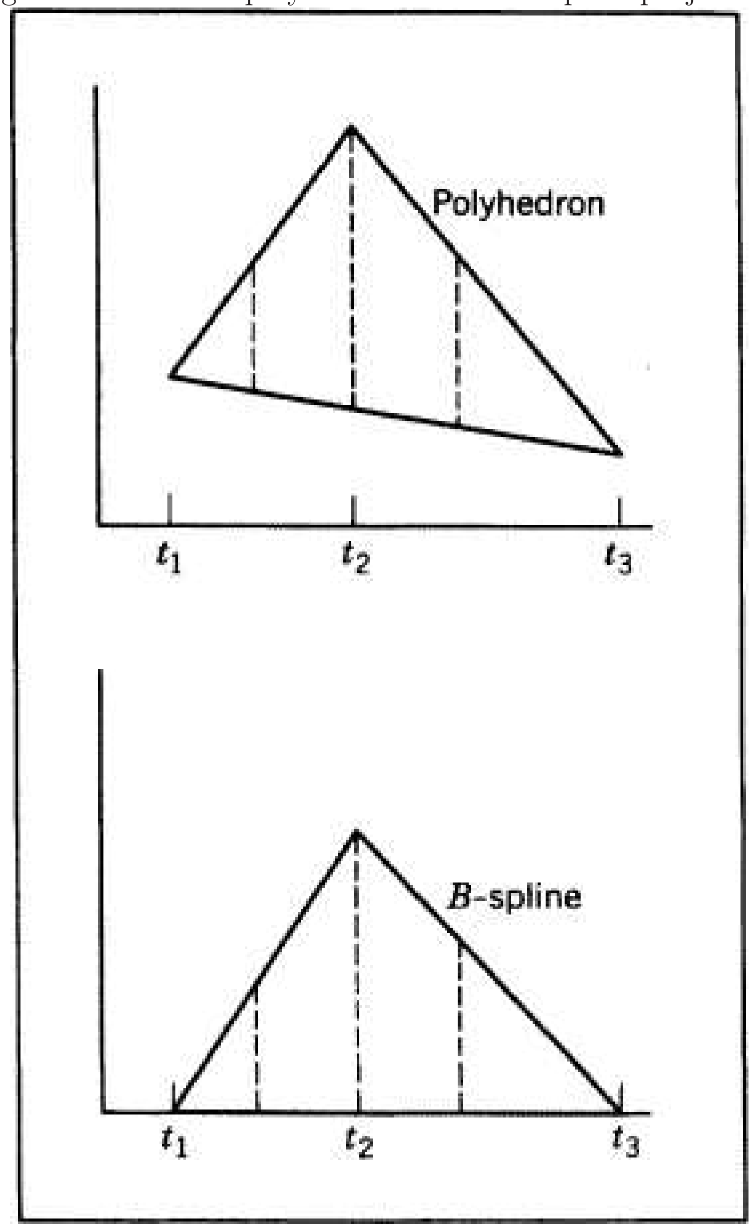
$$f_N \sim \sum_{l=1}^{N} c_l B_l$$

by doing ordinary **least-squares regression**, that is, choosing $\mathbf{c} = (c_1, \ldots, c_N)$ to minimize

$$\sum_{i=1}^{n} \left(y_i - \sum_{l=1}^{N} c_l B_l(x_i) \right)^2 .$$

If $N = n$, then f_N will interpolate the data, and as N becomes much smaller than n, f_N will have an increasingly smooth appearance, and the residuals will tend to increase. In principle, the knots $t_1, \ldots, t_N$ can be left as unknowns and chosen along with the coefficients to minimize the above sum of squares, but in practice, with noisy data, the determination of more than just a few knots this way is difficult and complicated by multiple local minima. Choosing the knots when interpolating a smooth function which is given exactly seems to be easier. See DeBoor [11]. The "eyeball" or trial and error method is also frequently used to choose the knots. Agarwal and Studden [1] give theoretical asymptotic results on the optimal number and location of knots for approximation to f in the model (1). Loosely speaking, if f has two derivatives then the optimal number of knots is of the order of $n^{1/5}$, so that there will be many fewer B-splines than data points. Regression splines are easy to compute using standard regression programs and the B-spline programs given in DeBoor [11], and if the true f is in the span of the B-splines chosen, then the estimate of f shares all of the usual properties of least-squares regression estimates. In general, however, the estimates of f may be biased, with the order of the bias similar to the order of the variance if N is chosen to minimize mean square error. See Buse and Lim [6], Poirier [53], and Winsberg and Ramsay [96] for applications of regression splines.

Figure 1: A convex polyhedron and its $\boldsymbol{B}$-spline projection.



3 Smoothing Splines

The other popular spline method for fitting the model (1) is to find f in an appropriate space of functions to minimize the penalized least-squares (PLS)

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

where

$$J_m = \int_0^1 f^{(m)}(t)dt,$$

for some integer m. The minimizer f_λ to this problem was obtained by Schoenberg [58]. If there are at least m distinct x_i's, the solution, which is known as a *smoothing spline*, is unique and is a unp spline of degree $2m-1$, with knots at distinct data points of $x_1, \ldots, x_n$. The smoothing parameter λ controls the trade-off between the fit to the data as measured by the residual sum of squares and the smoothness, as measured by J_m. For $m = 2$, then $f^{(2)}$ is curvature, and a small J_2 corresponds to visual, or psychological, smoothness. As $\lambda \to \infty$, the solution tends to the polynomial of degree m best fitting the data in a least-squares sense, and as $\lambda \to 0$, f_λ tends to the unp spline which interpolates to the data.

Figure 2 from Wahba and Wold [90] shows a model function f (dotted line), data generated according to the model (1), and a smoothing spline fit to the data (solid line) with a value of λ which is too large. Figure 3 shows the same model f and data, and a smoothing spline with λ too small. Figure 4 shows the same f and data, and the fitted smoothing spline with λ chosen by ordinary **cross-validation** (OCV). OCV involves deleting a data point and solving the optimization problem with a trial value of λ, computing the difference between the predicted value and the deleted observation with this trial value of λ, accumulating the sums of squares of these differences as one runs through each of the data points in turn, and finally choosing the λ for which the accumulated sum is smallest. Generalized cross-validation (GCV) was developed later by Craven and Wahba [9] and Golub et al. [23], and is an improvement over OCV both on asymptotic theoretical grounds and computational ease, although numerical results in an example like the one given can be expected to be quite similar. Although originally these cross-validation methods for choosing λ were computationally expensive, fast $O(n)$ transportable code is now readily available for the smoothing spline with GCV, see the "Algorithms" section below.

4 Choosing between Regression and Smoothing Splines

We make a few remarks on the choice of regression vs. smoothing splines for smoothing data from the model (1). Asymptotic theoretical convergence rates for the two methods are the same under the same assumptions (compare Agarwal and Studden [1] and Wahba [72]), provided the smoothing parameters N and λ are both chosen optimally. For very large data sets (say $n > 1000$), the results for data from the model (1) are likely to be practically

Figure 2: Model function (dashed curve), simulated data (open squares), and smoothing spline with too large value of λ (solid curve).

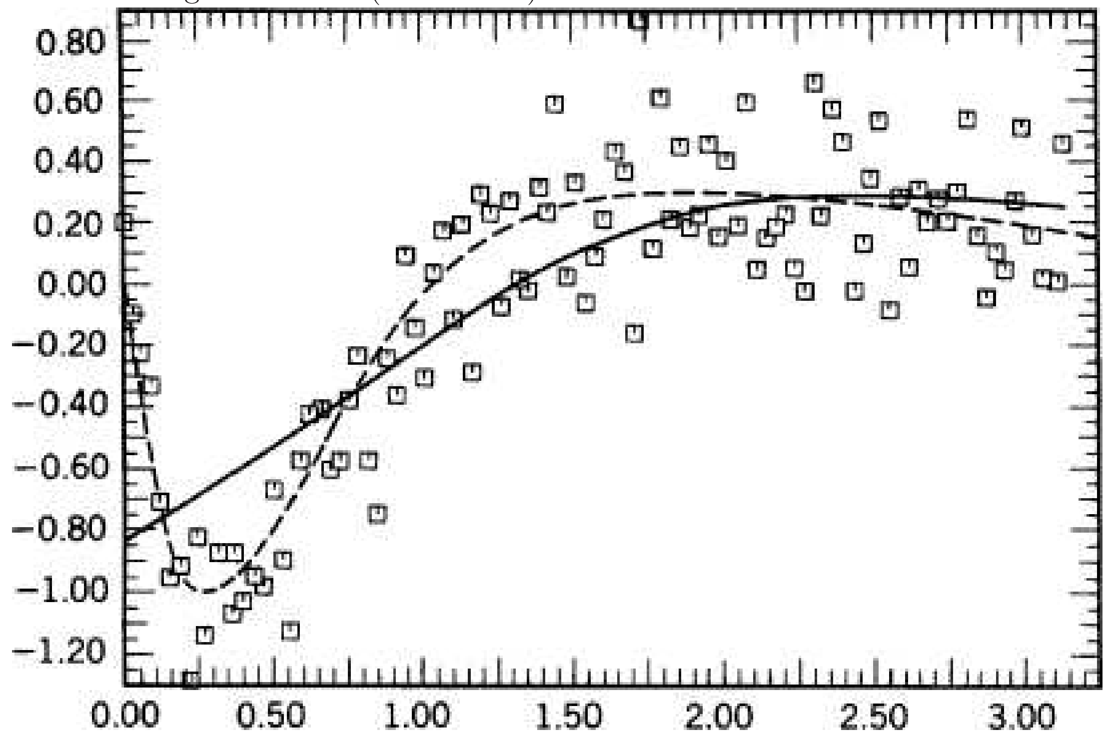


Figure 3: Same as Figure 2, except λ is too small.

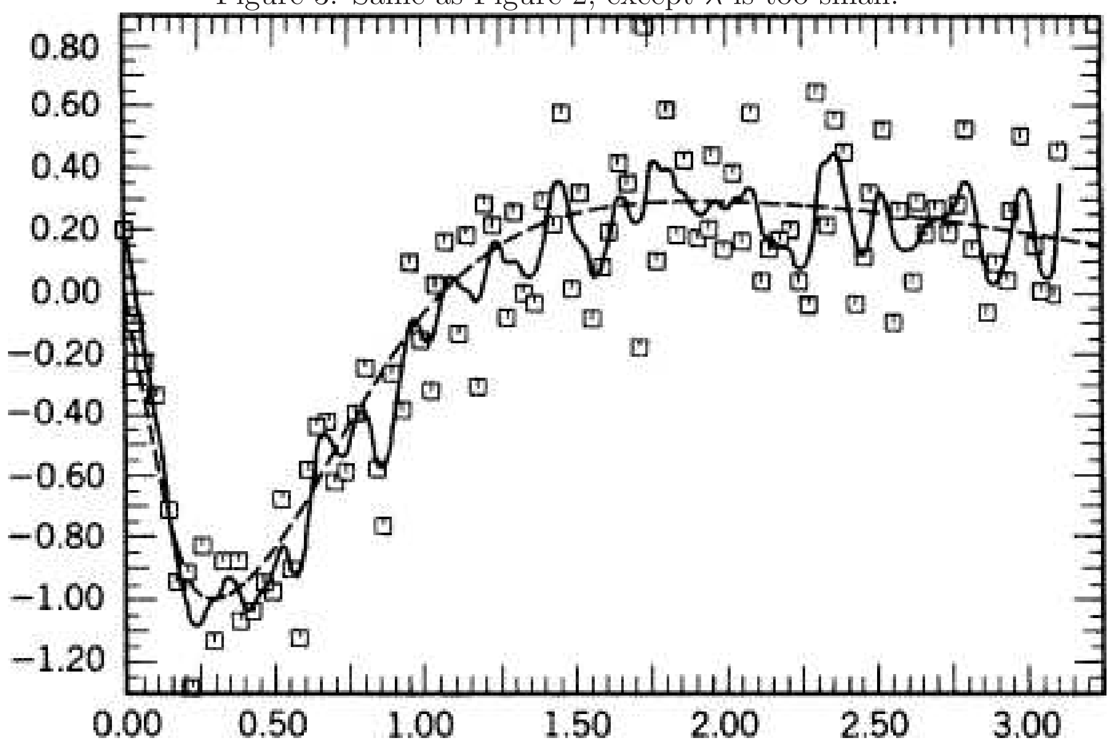
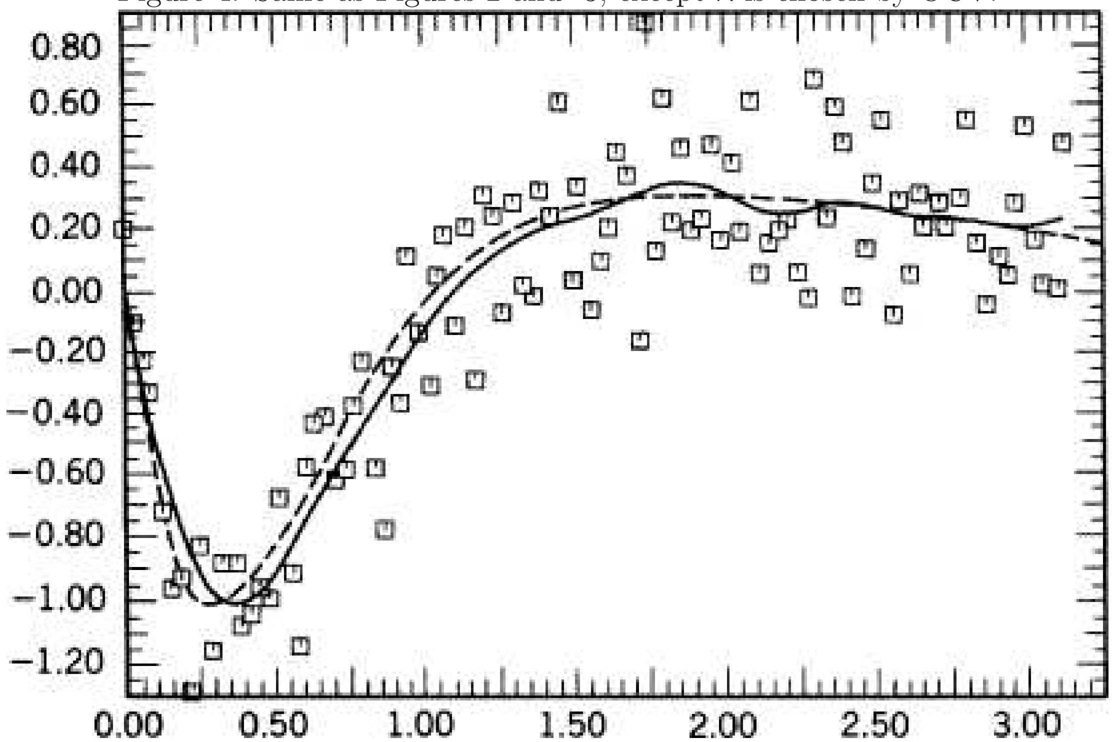


Figure 4: Same as Figures 2 and 3, except λ is chosen by OCV.



nearly the same, that is, indistinguishable on an $8\frac{1}{2}'' \times 11''$ plot, and the regression spline will require less storage to manipulate the results. Similar remarks concerning comparison of the results hold for medium-to-small sample sizes if the underlying f is close to being in the span of a small number of B-splines. (Recall that the optimal number of B-splines is $n^{1/5}$ under typical circumstances.) However it is likely that for examples like the one shown, and for example, the multimodal cases in Craven and Wahba [9], the regression spline with the optimal number of B-splines does not have the resolution to follow local features that can, in fact, be followed by a smoothing spline. A hybrid approach for very large data sets has been suggested by Nychka et al. [48], where the variational problem of (2) is solved in a space spanned by enough B-splines to avoid losing resolution. See also Luo and Wahba [42]. For estimation of derivatives or other local features such as maxima, the cross-validated smoothing spline is probably the method of choice, and in fact good results have been obtained with biomechanical data, which frequently satisfy the assumptions of model (1) well. See Woltring [97]. Some authors have found the use of regression splines to be useful as a smoother in such applications as **projection pursuit ¡stat05862¿** where a smoothing operation needs to be repeatedly carried out and sharp detail in the function is not expected.

5 Splines As Bayes Estimates

Smoothing splines can be interpreted as Bayes estimates if one views f as a sample function from a zero mean **Gaussian stochastic process**. Loosely speaking, the **stochastic process** can be described as $f^{(m)} = b \times$ **white noise**. Then it can be shown that for each x the conditional expectation of $f(x)$ given the data $y_1, \ldots, y_n$ is $f_\lambda(x)$ with $\lambda = \sigma^2/nb$. See Kimeldorf and Wahba [38] and Wahba [74]. It can be seen from these references (and originally from the work of Parzen [52]) that there is a duality between Bayes estimates given discrete data on continuous-time stochastic processes and the solution to variational problems like (2), which extends to a very general class of penalty functionals. In particular, the splines in the remainder of this article which can be obtained as the solution to a variational problem also have interpretations as Bayes estimates.

6 Multivariate Splines

There are several generalizations of univariate splines to several variables. The multivariate B-splines are generalizations of the univariate B-splines, which are piecewise polynomials, satisfy certain continuity conditions, and have compact support. The thin-plate smoothing splines generalize the univariate polynomial splines as the solution to a variational problem, and are popular in **meteorology**, computational vision, and other applications for smoothing two- and three-dimensional noisy data. They are not piecewise polynomials, however. The **tensor** product smoothing splines also generalize the univariate splines as the solution to a variational problem, and are the foundation for the smoothing spline ANOVA. We will discuss each of these separately.

7 Multivariate *B*-Splines

The multivariate B-splines in d variables are piecewise polynomials of degree $k-1$, which are 0 outside a convex polyhedron and positive inside. They can be obtained as projections of the volume of convex polyhedra in $k+d$ dimensions onto Euclidean d-space. See DeBoor [10] and Hollig [29]. Tensor products of B-splines are special cases of multivariate B-splines. In two dimensions, by a tensor product B-spline we mean a function of two variables, say x_1 and x_2 of the form $f(x_1,\, x_2) = B_i(x_1)B_j(x_2)$. The multivariate B-splines have found applications in computer-aided design and other fields where it is desired to model a smooth surface in two or three dimensions given exact values of it at a finite number of points. Tensor products of B-splines have also been used as a basis for bivariate regression where the data are given on a regular grid.

8 Thin-Plate Splines

The *thin-plate* splines are a popular generalization of the unp splines as solutions to a variational problem. In two dimensions ($d = 2$), with $m = 2$, the variational problem leading to the thin-plate spline is: Find $f \in \mathcal{H}$ to minimize the PLS

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

where

$$\begin{aligned} J_2^2(f) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(f_{x_1\,x_1}^2 + 2f_{x_1\,x_2}^2 \\ &\quad + f_{x_2\,x_2}^2)dx_1 dx_2. \end{aligned} \tag{3}$$

$\mathcal{H}$ is an abstract function space of functions of two variables for which (3) is finite, defined in Meinguet [46]. In d dimensions with general m, the variational problem becomes: Find f in $\mathcal{H}$ (a space of functions of d variables) to minimize the PLS

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_1(i), \ldots, x_d(i)))^2 \\ &+\lambda J_m^d(f), \end{aligned} \tag{4}$$

where

$$\begin{aligned} J_m^d(f) &= \sum_{\alpha_1+\cdots+\alpha_d=m}\frac{m!}{\alpha_1!\ldots\alpha_d!} \\ &\quad \times \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{\partial^m f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}\right)^2 \\ &\quad dx_1\ldots dx_d. \end{aligned} \tag{5}$$

It is necessary that $2m - d > 0$. An explicit representation for f_λ, the solution to this variational problem, was given by Duchon [12] and Meinguet [46], and is discussed further in a smoothing context by Wahba and Wendelberger [89]. The solution is known to lie in the span of a certain set of $n + m$ easily generated functions; see the "Algorithms" section for transportable software. Let $\boldsymbol{x}_i = (x_1(i), \ldots, x_d(i))$, and let Δ be the Laplacian operator, that is, $\Delta = \partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_d^2$. Then f_λ has the property that $\Delta^m f_\lambda(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \neq \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. This is a generalization of the analogous property of the unp spline in one variable, namely, $\Delta^m f_\lambda \equiv f_\lambda^{(2m)} = 0$ for $\boldsymbol{x} \neq \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, since f_λ is a polynomial of degree $2m - 1$ in the intervals between the knots. In several dimensions f_λ is a linear combination of $\binom{d+m-1}{d}$ monomials in the d variables $x_1, \ldots, x_d$ of total degree less than m, and n other functions each of which is a Green's function for the mth iterated Laplacian. The thin-plate spline is also a Bayes estimate, and loosely speaking, one can think of the prior as f satisfying $\Delta^{m/2} f = b\times$ white noise. The thin-plate spline appears to be a particularly useful tool for smoothing data from **diffusion processes** and other phenomena that may be thought of as representing the solution to an elliptic partial differential equation driven by white noise. The thin-plate spline is also a special case of an estimate of one of the "intrinsic random functions" of Matheron [44]; see Duchon [12]. The integral over d-space in (5) can be replaced by an integral over a bounded region Ω containing the data points (see Dyn and Wahba [13]) and then the minimizer will satisfy Neumann boundary conditions on the boundary of Ω. For $d = 1$, the solution inside Ω coincides with the unp spline previously described, but for $d > 1$, the minimizer of (4) satisfies the Neumann boundary conditions only at ∞, and the two multivariate thin-plate splines will be different.

9 Tensor Product Smoothing Splines

The space of functions of one variable referred to in connection with the univariate smoothing spline is known in the approximation theory literature as W_2^m and is a reproducing kernel Hilbert space of functions with square integrable mth derivative. To smooth functions of two variables, one can define a tensor product space $\mathcal{H} = W_2^m \otimes W_2^m$ which consists of sums and limits of sums of functions of the form $f(x_1, x_2) = f_1(x_1)f_2(x_2)$ with f_1 and f_2 in W_2^m and find $f \in \mathcal{H}$ to minimize the PLS

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

where now

$$\begin{aligned} J(f) &= \int_0^1 \int_0^1 \left(\frac{\partial^{2m} f}{\partial x_1^m \partial x_2^m}\right)^2 dx_1 dx_2 \\ &\quad + \text{other terms.} \end{aligned}$$

The other terms involve lower-order derivatives and guarantee that the solution will be unique under general conditions. Generalizations to d variables can be made. See Mansfield [43], Wahba [73], and Wahba [85]. These splines are piecewise polynomials in d variables, where the boundaries of the pieces are horizontal and vertical lines drawn through the data points. These splines have interested statisticians because of their role in the development of the smoothing spline ANOVA which is described in Section 14.

10 Splines on the Circle and the Sphere

Splines on the circle can be obtained by supposing that f is periodic with a representation of the form:

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

$$\sum_{\nu=1}^{\infty}(a_\nu^2 + b_\nu^2)(2\pi\nu)^{2m} < \infty,$$

and finding f to minimize the PLS

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \int_0^1 (f^{(m)}(u))^2 du.$$

A closed-form expression for f_λ as a piece-wise polynomial which satisfies m periodic boundary conditions may be obtained by using the fact that

$$\begin{aligned}\sum_{\nu=1}^{\infty}&(\cos 2\pi\nu z \cos 2\pi\nu x \\ &\quad + \sin 2\pi\nu z \sin 2\pi\nu x)(2\pi\nu)^{-2m} \\ &= \sum_{\nu=1}^{\infty} \cos 2\pi\nu(z-x)(2\pi\nu)^{-2m}\end{aligned}$$

and this latter infinite series has a closed-form expression in terms of the $2m$th **Bernoulli polynomial**; see Craven and Wahba [9].

The Bayes model here is

$$f(x) = a_0 + \sum_{\nu=1}^{\infty}(\alpha_\nu \cos 2\pi\nu x + \beta_\nu \sin 2\pi\nu x),$$

where the α_ν, β_ν are independent, zero mean normal random variables with $E\alpha_\nu^2 = E\beta_\nu^2 = b(2\pi\nu)^{-2m}$. This Bayes model also can be thought of as satisfying $f^{(m)} = b\times$ white noise, along with the periodic boundary conditions. A periodic smoothing spline with equally spaced data points can also be shown to be a kernel estimate. With unequally spaced data points in the general case there is an approximately equivalent variable **kernel estimate**; see Silverman [64].

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 [57] for more on spherical harmonics. The spherical harmonics are the eigenfunctions of the surface Laplacian Δ on the sphere, with

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

which is analogous to

$$\frac{d^2}{dx^2}\cos 2\pi\nu x = -(2\pi\nu)^2 \cos 2\pi\nu x.$$

Splines on the sphere are defined as the solution to the variational problem: find $f \in \mathcal{H}$ (an appropriate space) to minimize the PLS

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

where S is the sphere, and P is a point on the sphere. A closed-form expression is available for f_λ in the cases $m = 2, 3$ (Wendelberger [95]). Approximate closed-form expressions may be found in Wahba [76, 77]. The corresponding Bayes model is

$$f(P_i) = \theta + \sum_{l=0}^{\infty} \sum_{s=-l}^{l} f_{ls} Y_{ls}(P_i),$$

where $E[f_{ls}^2] = b[l(l+1)]^{-m}$. This can also be viewed as $\Delta^{m/2} f = b\times$ white noise. Splines on the sphere have a number of interesting applications in geophysics and meteorology; see for example Shure et al. [62]. 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 [79].

11 Reproducing Kernel Hilbert Space and General Smoothing Spline Models

The polynomial smoothing spline, thin-plate spline, spline on the circle and spline on the sphere share a common feature: they are all solutions to the PLS based on noisy data from model (1) in some well-defined model spaces. The domains of these spline functions are different and model spaces are chosen accordingly with suitable smoothness properties. Specifically these model spaces are **reproducing kernel Hilbert spaces ¡stat00479¿** (RKHS) (see Wahba [86], Gu [19] and Wang [92]). For example, the domain and model space for polynomial smoothing splines are $[0, 1]$ and W_2^m respectively. This section presents a general smoothing spline model with a RKHS on an arbitrary domain as the model space. The general smoothing spline model provides a unified framework for the developments of theory, inference and software.

A general smoothing spline model assumes that data are generated from model (1) where the function f belongs to a RKHS $\mathcal{H}$ defined on an arbitrary domain $\mathcal{X}$. With a well-defined quadratic functional penalty $J(f)$, the model space is decomposed into $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ where $\mathcal{H}_0$ consists of functions with $J(f) = 0$ (i.e. not penalized, see Gu [19]). Furthermore $J(f) = ||P_1 f||^2$ where P_1 is the orthogonal projection operator onto $\mathcal{H}_1$. Usually $\mathcal{H}_0$ is a finite dimensional space with basis functions denoted as $\phi_1, \dots, \phi_p$. Both $\mathcal{H}_0$ and $\mathcal{H}_1$ are RKHS's with their reproducing kernels (RK) denoted as R_0 and R_1 respectively.

The smoothing spline estimate f_λ of f is the minimizer in $\mathcal{H}$ of the PLS

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + J(f). \tag{6}$$

The representer theorem (Kimeldorf and Wahba [38] and Wahba [86]) states that, when $T = \{\phi_\nu(x_i)\}_{i=1}^{n}{}_{\nu=1}^{p}$ is of full column rank, the unique minimizer of (6) is given by

$$f_\lambda(x) = \sum_{\nu=1}^{p} d_\nu \phi_\nu(x) + \sum_{i=1}^{n} c_i \xi_i(x), \tag{7}$$

where $\xi_i(x) = R_1(x_i, x)$ are called representers, and c_i's and d_i's are coefficients. Computation of the smoothing spline estimate will be discussed in the "Algorithms" section below.

The prior of the corresponding Bayes model for the general spline model is

$$F(x) = \sum_{\nu=1}^{p} \alpha_\nu \phi_\nu(x) + b^{\frac{1}{2}} U(x),$$

where $\alpha_1, \ldots, \alpha_p \overset{iid}{\sim} \mathrm{N}(0, \kappa)$, $U(x)$ is a zero-mean Gaussian stochastic process with covariance function $R_1(x, z)$, and α_ν's and $U(x)$ are mutually independent. With $\lambda = \sigma^2/nb$, the posterior mean $\lim_{\kappa \to \infty} \mathrm{E}(F(x)|\boldsymbol{y}) = f_\lambda(x)$ (Wahba [86]). Therefore the smoothing spline estimate is a Bayes estimate with an improper prior on elements in the null space $\mathcal{H}_0$.

12 Choosing The Smoothing Parameter, Confidence Intervals, Diagnostics and Hypothesis Tests

GCV appears to be the most popular method for choosing the smoothing parameter λ from the data in the context of smoothing splines, for various theoretical and practical reasons; see Craven and Wahba [9], Li [39], Speckman [65, 66], Utreras [68], Wahba [72, 83], and Wahba and Wang [87]. GCV can be obtained from OCV by an invariance argument by rotating the system to a standard coordinate system, doing OCV, and rotating back. (Ordinary leaving out one is not invariant under rotations of the observation coordinate system.)

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

$$\begin{pmatrix} f_\lambda(x_1) \\ \vdots \\ f_\lambda(x_n) \end{pmatrix} = A(\lambda) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

An explicit representation of A in the unp case can be found in Craven and Wahba [9], and in general in Wahba [86]. The GCV estimate $\hat{\lambda}$ of λ is obtained as the minimizer of

$$V(\lambda) = \frac{(1/n)||(I - A(\lambda))\mathbf{y}||^2}{((1/n)\mathrm{Tr}(I - A(\lambda)))^2}.$$

An alternative approach for choosing the smoothing parameter is the generalized maximum likelihood (GML) method that estimates λ as the minimizer of

$$M(\lambda) = \frac{\mathbf{y}^T(I - A(\lambda))\mathbf{y}}{[\det^+\{(I - A(\lambda))\}]^{\frac{1}{n-p}}},$$

where $\det^+$ represents the product of the nonzero eigenvalues. The GML criterion may be derived using the connection between smoothing spline models and Bayes models, or the connection between smoothing spline models and linear **mixed effects ¡stat05862¿** models (see Wahba [86] and Wang [91]).

$A(\lambda)$ 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 **confidence intervals** and spline **regression diagnostics**.

Trace $A(\lambda)$ can be viewed as the "degrees of freedom for signal." The posterior covariance matrix of $(f_\lambda(x_1), \ldots, f_\lambda(x_n))^T$ is $\sigma^2 A(\lambda)$ and this fact has been used to construct Bayesian confidence intervals based on $\hat{\sigma}^2 A(\hat{\lambda})$, where $\hat{\sigma}^2 = RSS(\hat{\lambda})/\mathrm{Tr}(I - A(\hat{\lambda}))$. See Wahba [80]. These Bayesian confidence intervals appear to have useful frequentist properties if interpreted "across the function," rather than pointwise; see Hall and Titterington [24], Nychka [47], Silverman [64], and Wahba [78]. Note that Bayesian confidence intervals can be constructed for the function f as well as its projections onto $\mathcal{H}_0$ and $\mathcal{H}_1$ at any point in the domain. Details can be found in Wang [92].

Eubank [17] has proposed methods for detecting **outliers** and **influential observations**, by exploiting the influence matrix analogy. See also Silverman [64]. For example, letting $\epsilon_j(\hat{\lambda})$ be the jth residual, it is suggested that the quantities

$$T_j = \epsilon_j(\hat{\lambda})/\{\hat{\sigma}(1 - a_{jj}(\hat{\lambda}))\}^{1/2}$$

be called "studentized residuals" and an observation be considered an outlier if $|T_j|$ exceeds an appropriate critical value from a **Student's t-distribution** with approximate **degrees of freedom** $\mathrm{Tr}(I - A(\hat{\lambda}))$.

Nonparametric models may be used to check whether there is a significant departure from a parametric model. Wahba [86] considered the following hypothesis for the general spline model

$$\mathrm{H}_0: \ f \in \mathcal{H}_0, \qquad \mathrm{H}_1: \ f \in \mathcal{H} \text{ and } f \notin \mathcal{H}_0.$$

For the polynomial smoothing spline with $\mathcal{H} = W_2^m$, H_0 corresponds to $f \in \pi^{m-1}$. Cox, Koh, Wahba and Yandell [8] proposed the **locally most powerful test ¡stat05862¿**, and Wahba [86] proposed GML and GCV tests. See Liu and Wang [40] for a comparison of these tests, and Liu, Meiring and Wang [41] for an extension to non-Gaussian data.

13 Partial Splines

Consider the model

$$y_i = f(x_i) + \sum_{j=1}^{q} \beta_j \Phi_j(x_i, \mathbf{z}_i) + \epsilon_i,$$

where $x_i \in \mathcal{X}$, $\mathbf{z}_i = (z_1(i), \ldots, z_d(i))^T$, f belongs to a RKHS $\mathcal{H}$, and Φ_j's are known functions of x and the **concomitant variables z**. An estimate of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T$ and f may be obtained by finding $f \in \mathcal{H}$ and $\boldsymbol{\beta} \in E^q$ to minimize the PLS

$$\frac{1}{n} \sum_{i=1}^{n} \left(y_i - f(x_i) - \sum_{i=1}^{q} \beta_j \Phi_j(x_i, \mathbf{z}_i) \right)^2 + J(f).$$

Let $S_{n\times q}$ be the matrix with i, jth entry $\Phi_j(x_i, \mathbf{z}_i)$. If the $n \times (p+q)$ matrix $(T : S)$ is of full column rank, then there will be a unique minimizer $(f_\lambda, \hat{\boldsymbol{\beta}})$. Such models are extremely flexible and are attractive in a variety of applications; see Engle et al. [16], Ansley and Wecker [2], Shiau et al. [61], and Wahba [81] and references cited therein. Properties of the estimate of $\boldsymbol{\beta}$ have been studied by many authors; see, for example, Heckman [27] and Rice [56].

14 Smoothing Spline ANOVA

One way to generalize univariate splines is to consider tensor product model spaces as illustrated in the section on tensor product of polynomial smoothing splines. Analogous to the classical multi-factor ANOVA, a smoothing spline ANOVA (SS ANOVA) decomposes a multivariate function (or equivalently a tensor product space) into main effects and interactions which facilitates model selection and interpretation.

Consider model (1) where f is a function of d variables $x_1 \in \mathcal{X}_1, \ldots, x_d \in \mathcal{X}_d$, and $\mathcal{X}_1, \ldots, \mathcal{X}_d$ are arbitrary sets. Let $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$. The construction of an SS ANOVA model starts with a tensor product space $\mathcal{H}^{(1)} \otimes \cdots \otimes \mathcal{H}^{(d)}$ on $\mathcal{X}$ where $\mathcal{H}^{(k)}$ is a RKHS for f as a function of x_k. If each space $\mathcal{H}^{(k)}$ contains constant functions, then f can be decomposed into

$$f = \mu + \sum_{k=1}^{d} f_k(x_k) + \sum_{k<l} f_{kl}(x_k, x_l) + \ldots + f_{1\ldots d}(x_1, \ldots, x_d), \tag{8}$$

where μ represents the grand mean, $f_k(x_k)$ represents the main effect of x_k, $f_{kl}(x_k, x_l)$ represents the two-way interaction between x_k and x_l, and the remaining terms represent higher-order interactions. When all $\mathcal{X}_k$'s are finite discrete sets, the above decomposition corresponds to the classical multi-factor ANOVA. SS ANOVA is a general modeling technique rather than some specific decompositions. Different forms of SS ANOVA decompositions may be constructed for different purposes. The decomposition (8) is perhaps the simplest and most common form of the SS ANOVA decomposition. Other forms of decompositions can be found in Wang [92].

SS ANOVA decomposes the tensor product space $\mathcal{H}^{(1)} \otimes \cdots \otimes \mathcal{H}^{(d)}$ into some orthogonal subspaces (Wang [92]). The number of subspaces increases exponentially as the dimension d increases. To tackle this **curse of dimensionality problem ¡stat05862¿**, high-order interactions are often removed from the model space for ease of interpretation. A model containing any subset of subspaces of the SS ANOVA decomposition is an SS ANOVA model. An SS ANOVA model with main effects only corresponds to the well-known **additive model ¡stat05862¿** (Hastie and Tibshirani [25]). For a given SS ANOVA model, after reorganization the model space can be written as

$$\mathcal{M} = \mathcal{H}^0 \oplus \mathcal{H}^1 \oplus \cdots \oplus \mathcal{H}^q, \tag{9}$$

where $\mathcal{H}^0$ is a finite dimensional space consisting of functions which are not going to be penalized, and $\mathcal{H}^1, \ldots, \mathcal{H}^q$ are orthogonal RKHS's with RKs R^j for $j = 1, \ldots, q$. The estimate f_λ of f is the minimizer in $\mathcal{M}$ of the PLS

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(\boldsymbol{x}_i))^2 + \sum_{j=1}^{q}\lambda_j \|P_j f\|^2,$$

where P_j is the projector operator onto $\mathcal{H}^j$ and λ_j's are smoothing parameters. The representer theorem still holds. The estimate f_λ is again a Bayes estimate with an improper prior on elements in the null space $\mathcal{H}^0$. Bayesian confidence intervals may be constructed for the overall function f as well as its components at any point in the domain. The smoothing parameters can be selected using the GCV or GML method. See Gu [19] and Wang [92] for details.

One major issue when applying the SS ANOVA methodology is the selection of the model space $\mathcal{M}$ since the number of subspaces increases quickly. One may select $\mathcal{M}$ for ease of interpretation and/or using prior knowledge. Bayesian confidence intervals for components may be used to remove "insignificant" ones. Gu [19] developed geometric diagnostic tools for the selection of components based on their estimates. Replacing the squared norm $\|P_j f\|^2$ by the norm $\|P_j f\|$ in the PLS, Zhang and Lin [100] developed the component selection and smoothing operator (COSSO) method for simultaneous selection and estimation of an SS ANOVA model. See also Chen [7].

15 Splines As Penalized Likelihood Estimates

All smoothing spline and SS ANOVA models in the previous sections assume data are generated from the nonparametric regression model (1) where the main interest is to estimate the mean function f. There are many other statistical models where the likelihood depends on a function f and the main interest is to estimate f nonparametrically. Suppose that $f \in \mathcal{M}$ where $\mathcal{M}$ is a RKHS. Then a **penalized likelihood** (PL) estimate of f is the minimizer in $\mathcal{M}$ of

$$-l(f) + \lambda J(f), \tag{10}$$

where $l(f)$ is the log-likelihood. Many problems fall under the framework of penalized likelihood estimation. We discuss a few of them below.

Suppose independent observations $y_1, \ldots, y_n$ are generated from a distribution in the **exponential family ¡stat05862¿** with density

$$g(y_i|\boldsymbol{x}_i) = \exp\left\{\frac{y_i f(\boldsymbol{x}_i) - b(f(\boldsymbol{x}_i))}{a_i(\phi)} + c(y_i, \phi)\right\}, \quad i = 1, \ldots, n,$$

where $\boldsymbol{x}_i$'s are observations of d independent variables $x_1, \ldots, x_d$, and ϕ is a dispersion parameter. A canonical link is assumed for simplicity. A **generalized linear model ¡stat05752¿** assumes that $f(\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_d x_d$ (McCullagh and Nelder [45]).

For flexibility, the multivariate function f may be modeled nonparametrically using a model space based on an SS ANOVA decomposition. Then the model space $\mathcal{M}$ is given by (9). The function f can be estimated using the PL (10) where the log-likelihood $l(f) = \sum_{i=1}^{n}\{[y_i f(\boldsymbol{x}_i) - b(f(\boldsymbol{x}_i))]/a_i(\phi) + c(y_i, \phi)\}$. The representer theorem still holds. The smoothing parameters can be estimated using the GCV, GML or unbiased risk method. See Wahba et al [88].

For density estimation, suppose $y_1, \ldots, y_n$ are iid random samples from a density function $g(y)$ and the goal is to estimate $g(y)$ nonparametrically. Denote the domain of y as $\mathcal{Y}$ and assume that $g(y) > 0$ on $\mathcal{Y}$. To enforce the positivity and unity constraints of a density function, Gu and Qiu [22] proposed to model the logistic transformation f that satisfies $g = e^f / \int_{\mathcal{Y}} e^f dy$. Free of constraints, Gu and Qiu [22] modeled f using a RKHS with constant functions being removed for identifiability. The function f was estimated using the PL (10) where the log-likelihood $l(f) = \sum_{i=1}^{n} f(y_i) - n \log \int_{\mathcal{Y}} e^{f(y)} dy$. Since the log-likelihood involves an integral with respect to y, the minimizer of the PL does not fall in a finite dimensional space. Nevertheless, an approximate solution by minimizing the PL in a finite dimensional space spanned by representers maintains the same convergence rate. See also Silverman [63].

For conditional density estimation, suppose (x_i, y_i) for $i = 1, \ldots, n$ are independent observations from a conditional density $g(y|x)$ and the goal is to estimate $g(y|x)$ nonparametrically. Let $f(x, y)$ be the logistic transformation of $g(y|x)$. Gu [20] modeled f using an SS ANOVA model $f(x, y) = f_2(y) + f_{12}(x, y)$ where the constant and the main effect of x are removed for identifiability. Again, the function f was estimated using the PL (10) where the log-likelihood $l(f) = \sum_{i=1}^{n} f(x_i, y_i) - n \log \int_{\mathcal{Y}} e^{f(x_i, y)} dy$.

Other applications of penalized likelihood estimation include **hazard rate ¡stat05862¿** regression (Gu [21]), **accelerated failure time models stat05862** (Gu [19]), semi-parametric **nonlinear mixed effects models ¡stat05862¿** (Ke and Wang [35]), and generalized nonparametric mixed effects models (Karcher and Wang [34]).

16 Splines with Linear Inequality Constraints

Splines satisfying a family of linear inequality constraints can be found as the solution to the problem: Find $f \in \mathcal{H}$ to minimize

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

subject to

$$a_i \leq \mathcal{L}_i f \leq b_i,$$

where $\mathcal{L}_i$ is a bounded linear functional. Included are discretized positivity and monotonicity constraints; see Utreras [69] and Villalobos and Wahba [71].

17 Histosplines

Histosplines is the name given to splines which are constructed to have a volume matching or volume smoothing property. They were introduced into the statistical literature in the context of **density estimation** by Boneva et al. [5]. They constructed a univariate spline which had the volume matching property

$$\int_{x_i}^{x_{i+1}} f(x)dx = n_i/n,$$

where n_i is the number of observations from a random sample of size n which fell in the bin with boundaries x_i and x_{i+1}. Volume smoothing histosplines arise when one observes

$$y_i = \int_{\Omega_i} f dt + \epsilon_i$$

and chooses f as the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \int_{\Omega_i} f(t)dt\right)^2 + \lambda J_m^d(f). \tag{11}$$

See Wahba [77] and Dyn et al. [14] and references cited therein.

18 Linear and Nonlinear Functionals

Wahba [86] considered the following model

$$y_i = \mathcal{L}_i f + \epsilon_i, \quad i = 1, \ldots, n, \tag{12}$$

where $\mathcal{L}_i$'s are known bounded linear functionals. This model allows indirect observations of f through linear functionals. One example is Fredholm's integral equation of the first kind

$$y_i = \int K(x_i, s)f(s)ds + \epsilon_i, \quad i = 1, \ldots, n, \tag{13}$$

where K is a known function. In this case $\mathcal{L}_i f = \int K(x_i, s)f(s)ds$. Another example is $\mathcal{L}_i f = f'(x_i)$ when observations are made on the first derivative. See O'Sullivan [49] for more examples. The standard nonparametric regression model (1) is a special case with $\mathcal{L}_i f = f(x_i)$ and $\mathcal{L}_i$ in this case is called the evaluational functional.

The goal is to estimate the function f nonparametrically from noisy data. Suppose that f belongs to a RKHS $\mathcal{H}$. Then the smoothing spline estimate f_λ of f is the minimizer in $\mathcal{H}$ of the PLS

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \mathcal{L}_i f)^2 + \lambda J(f).$$

The representer theorem still holds (Wahba [86]).

Sometimes observations of f are made indirectly through nonlinear functionals. Examples in remote sensing and reservoir modeling can be found in O'Sullivan and Wahba [51], Wahba [84] and O'Sullivan [49]. There are often constraints such as positivity and monotonicity to the function f in a nonparametric regression model and these constraints may be enforced by nonlinear transformations. For example, in many situations the function f in model (13) is known to be positive (Vardi and Lee [70]). With the exponential transformation $f = \exp(g)$ the problem reduces to the estimation of g which is constraint free and observed through nonlinear functionals. When a function f is strictly increasing, one may use the transformation $f(x) = f(0) + \int_0^x \exp\{g(s)\}ds$ to enforce the monotonicity constraint (Wang [92]).

Ke and Wang [36] considered the following nonparametric nonlinear regression model

$$y_i = \mathcal{N}_i(f_1, \ldots, f_r) + \epsilon_i, \quad i = 1, \ldots, n, \tag{14}$$

where f_k belongs to an RKHS $\mathcal{H}_k$ on an arbitrary domain $\mathcal{X}_k$ for $k = 1, \ldots, r$, and $\mathcal{N}_i$'s are known nonlinear functionals on $\mathcal{H}_1 \times \cdots \times \mathcal{H}_r$. Model (14) contains both model (12) and the SS ANOVA model as special cases. Assume that $\mathcal{H}_k = \mathcal{H}_{k0} \oplus \mathcal{H}_{k1}$ where $\mathcal{H}_{k0} = \text{span}\{\phi_{k1}, \ldots, \phi_{kp_k}\}$ consists of functions which are not penalized. The smoothing spline estimates of $f_1, \ldots, f_r$ are minimizers of the PLS

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \mathcal{N}_i(f_1, \ldots, f_r))^2 + \sum_{k=1}^{r}\lambda_k \|P_{1k} f_k\|^2,$$

where P_{1k} is a projection operator from $\mathcal{H}_k$ onto $\mathcal{H}_{1k}$, and λ_k's are smoothing parameters.

For a special case when $\mathcal{N}_i$ depends on f_k's through a nonlinear function and some bounded linear functionals, Ke and Wang [36] showed that the representer theorem still holds. In general the solution to the PLS does not fall in a finite dimensional space. Ke and Wang [36] extended the Gauss-Newton method to infinite dimensional spaces and used the Gauss-Seidel algorithm to estimate each function iteratively.

19 Semiparametric Models

A semiparametric model contains both parametric and nonparametric components. Often in practice there is enough knowledge to model some components parametrically while leaving some uncertain and/or nuisance component unspecified. The partial spline is a semiparametric model.

Wang [92] considered the following semiparametric linear regression model

$$y_i = \boldsymbol{s}_i^T \boldsymbol{\beta} + \sum_{k=1}^{r} \mathcal{L}_{ki} f_k + \epsilon_i, \quad i = 1, \ldots, n, \tag{15}$$

where $\boldsymbol{s}$ is a q-dimensional vector of independent variables, $\boldsymbol{\beta}$ is a vector of parameters, $\mathcal{L}_{ki}$ are bounded linear functionals, and f_k's are unknown functions. In addition to the partial spline, model (15) contains the varying-coefficient model (Hastie and Tibshirani [26]) and the functional linear model (Ramsay and Silverman [55]) as special cases. A vector spline model may also be represented as a special case of model (15) (Wang [92]).

Assume that f_k belongs to a RKHS $\mathcal{H}_k$ on an arbitrary set $\mathcal{X}_k$, and $\mathcal{H}_k = \mathcal{H}_{k0} \oplus \mathcal{H}_{k1}$ where $\mathcal{H}_{k0} = \text{span}\{\phi_{k1}, \ldots, \phi_{kp_k}\}$ consists of functions which are not penalized. The PLS estimates of $\boldsymbol{\beta}$ and $f_1, \ldots, f_r$ are minimizers of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{s}_i^T\boldsymbol{\beta} - \sum_{k=1}^{r}\mathcal{L}_{ki}f_k)^2 + \sum_{k=1}^{r}\lambda_k\|P_{1k}f_k\|^2,$$

where P_{1k} is a projection operator from $\mathcal{H}_k$ onto $\mathcal{H}_{1k}$, and λ_k's are smoothing parameters. Again the representer theorem holds.

To allow for nonlinear dependence on both parameters and nonparametric functions, Wang and Ke [93] considered the following semiparametric nonlinear regression model

$$y_i = \mathcal{N}_i(\boldsymbol{\beta}, \boldsymbol{f}) + \epsilon_i, \quad i = 1, \ldots, n, \tag{16}$$

where $\boldsymbol{\beta}$ is a vector of parameters and $\boldsymbol{f} = (f_1, \ldots, f_r)$ are unknown functions. Model (16) contains many existing statistical models such as nonlinear regression, nonlinear nonparametric regression, projection pursuit, and single and multiple index models as special cases. Its extension to clustered data also contains shape invariant models as a special case.

With the same assumptions for model spaces of f_k's as above, the PLS estimates of $\boldsymbol{\beta}$ and $\boldsymbol{f}$ are minimizers of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \mathcal{N}_i(\boldsymbol{\beta}, \boldsymbol{f}))^2 + \sum_{k=1}^{r}\lambda_k\|P_{1k}f_k\|^2.$$

Wang and Ke [93] developed an algorithm using Gauss-Newton and Gauss-Seidel methods.

20 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 λ. Transportable code CUBGCV based on the fast algorithm proposed by Hutchinson and DeHoog [32] may be found in Hutchinson [31]; this algorithm with some additions is incorporated in GCVSPL of Woltring [98]. 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 [56]. Older code (ICSSCV) for the unp spline with the GCV estimate of λ can be found in the IMSL library [4]. 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 [15] and the truncated **singular value decomposition ¡stat02351¿**, Bates and Wahba [3] may be used to speed the calculation. Transportable code for thin-plate splines using thin-plate basis functions is available in Hutchinson [30], and for partial thin-plate splines and general problems using the truncated singular value decomposition, in GCVPACK (Bates et al. [33]). GCVSPL, GCVPACK, code for generating B-splines based on DeBoor [11] 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.

For models under the framework of RKHS, either based on the representer theorem or approximation, the spline estimates are usually represented as linear combinations of basis functions in a finite dimensional space not subject to penalty and representers. Consequently the estimation involves computation of coefficients of the linear combinations and smoothing parameters using the GCV or GML criterion. For example, the estimate of the general smoothing spline model is given in (7) where the coefficients $\boldsymbol{c} = (c_1, \ldots, c_n)^T$ and $\boldsymbol{d} = (d_1, \ldots, d_p)^T$ are solutions to

$$\begin{aligned}(\Sigma + n\lambda I)\boldsymbol{c} + T\boldsymbol{d} &= \boldsymbol{y}, \\ T^T\boldsymbol{c} &= \boldsymbol{0}.\end{aligned}$$

Algorithms can be found in Gu [19] and Wang [92]. Fitting with all representers usually requires $O(n^3)$ operations. Kim and Gu [37] showed that a random subset of n representers of size $q = o(n)$ can achieve the same convergence rate. The resulting algorithms require $O(qn^2)$ operations. Fitting spline models with massive data requires more research. See Helwig and Ma [28].

Several R packages have been developed for fitting various spline models. The ASSIST package (Wang [92]) fits (i) general spline models for independent and correlated Gaussian data, and for independent binomial, Poisson and Gamma data; (ii) nonparametric nonlinear regression models; (iii) semiparametric linear and nonlinear regression models; and (iv) semiparametric linear and nonlinear mixed effects models. Some well known models that may be fitted by functions in the ASSIST package are polynomial splines, periodic splines, spherical splines, thin-plate splines, l-splines, generalized additive models, SS ANOVA models, SS ANOVA mixed effects models, projection pursuit models, multiple index models, varying coefficient models, functional linear models, and self-modeling nonlinear regression models. The gss package (Gu [19]) fits density and conditional density functions, hazard rate regression and accelerated failure time models, in addition to smoothing spline and SS ANOVA regression models. The fields package (Furrer, Nychka and Sain [18]) fits cubic spline, thin-plate spline and other spatial models. The mgcv package (Wood, [99]) may also be used to fit many spline models.

21 Related Articles

See also **Curve Fitting**; **Free-Knot Splines**; **Graduation, Whittaker-Henderson**; **Interpolation**; **Maximum Penalized Likelihood Estimation**; **Moving Averages**; **Nonparametric Regression**; **Osculatory Interpolation**; **Regression**; **Regularization Methods; Scatterplot Smoothers; Semiparametric Regression; Smoothing**;

Smoothing in Environmental Epidemiology; **Smoothness Priors**; **Splines in Nonparametric Regression**; **Spline Smoothing**.

Acknowledgments

Grace Wahba's research is supported in part under National Science Foundation Grant DMS-1308847 and a Consortium of NIH Institutes under Award U54AI117924. Yuedong Wang's research is supported in part under National Science Foundation Grant DMS-1507620.

References

[1] Agarwal, G. and Studden, W. J. (1980). *Ann. Statist.*, **8**, 1307–1325.

[2] 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., pp. 181–192.

[3] 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*, C. T. Baker and G. F. Miller, eds. Academic, London, pp. 283–296.

[4] Bates, D. M., Lindstrom, M. J., Wahba, G. and Yandell, B. (1987). *Commun. Statist. Simul. Comp.*, **16**, 263–297.

[5] Boneva, K., Kendall, D. and Stefanov, I. (1971). *J. R. Statist. Soc. B*, **33**, 1–70.

[6] Buse, A. and Lim, L. (1977). *J. Amer. Statist. Ass.*, **72**, 64–68.

[7] Chen, Z. (1993). Fitting multivariate regression functions by interaction spline models. *J. R. Statist. Soc. B*, **55**, 473–491.

[8] Cox, D. D., Koh, E., Wahba, G. and Yandell, B. (1988), Testing the (Parametric) Null Model Hypothesis in (Semiparametric) Partial and Generalized Spline Model, *Ann. Statist.*, **16**, 113–119.

[9] Craven, P. and Wahba, G. (1979). *Numer. Math.*, **31**, 377–403.

[10] DeBoor, C. (1976). Splines as linear combinations of *B*-splines: A survey. In *Approximation Theory II*, G. G. Lorenz, C. K. Chui and L. L. Schumaker, eds. Academic, New York, pp. 1–48.

[11] DeBoor, C. (1978). *A Practical Guide to Splines*. Springer, New York.

[12] Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller, eds. Springer, Berlin, pp. 85–100.

[13] Dyn, N. and Wahba, G. (1982). *SIAM J. Math. Anal.*, **13**, 134–152.

[14] Dyn, N., Wahba, G. and Wong, W. (1979). *J. Amer. Statist. Ass.*, **74**, 530–535.

[15] Elden, L. (1984). *BIT*, **24**, 467–472.

[16] Engle, R., Granger, C., Rice, J. and Weiss, A. (1986). *J. Amer. Statist. Ass.*, **81**, 310–320.

[17] Eubank, R. L. (1985). *J. R. Statist. Soc. B*, **47**, 332–341.

[18] Furrer, R., Nychka, D., Sain, S. (2009). fields: Tools for spatial data. R package version 5.02. URL: /http://www.image.ucar.edu/Software/Fields.

[19] Gu, C. (2013). *Smoothing Spline ANOVA Models*. Springer-Verlag, New York.

[20] Gu, C. (1995), Smoothing Spline Density Estimation: Conditional Distribution, *Statistica Sinica*, **5**, 709–726.

[21] Gu, C. (1996), Penalized likelihood hazard estimation: a general procedure. *Statistica Sinica*, **6**, 861–876.

[22] Gu, C. and Qiu, C. (1993). Smoothing Spline Density Estimation: Theory. *Ann. Statist.*, **21**, 217-234.

[23] Golub, G. H., Heath, M. and Wahba, G. (1979). *Technometrics*, **21**, 215–224.

[24] Hall, P. and Titterington, D. M. (1987). *J. R. Statist. Soc. B*, **49**, 184–198.

[25] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.

[26] Hastie, T. and Tibshirani, R. (1993). Varying Coefficient Model. *J. R. Statist. Soc. B*, **55**, 757–796.

[27] Heckman, N. E. (1986). *J. R. Statist. Soc. B*, **48**, 244–248.

[28] Helwig, N.E. and Ma, P. (2015) Fast and stable multiple smoothing parameter selection in smoothing spline analysis of variance models with large samples. *J. Comp. Graph. Stat.*, **24**, 715–732.

[29] Hollig, K. (1986). Multivariate splines. In *Approximation Theory*, C. deBoor, ed. American Mathematical Society, Providence, RI, pp. 103–127.

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

[31] Hutchinson, M. F. (1985). *ACM Trans. Math. Software*, **12**, 150–153.

[32] Hutchinson, M. F. and DeHoog, F. R. (1985). *Numer. Math.*, **47**, 99–106.

[33] IMSL Library (1986).

[34] Karcher, P. and Wang, Y. (2001), Generalized Nonparametric Mixed Effects Models,*J. Comp. Graph. Stat.*, **10**, 641–655.

[35] Ke, C. and Wang, Y. (2001), Semi-parametric Nonlinear Mixed Effects Models and Their applications, *J. Amer. Statist. Ass.*, **96**, 1272–1298.

[36] Ke, C. and Wang, Y. (2004), Nonparametric Nonlinear Regression Models, *J. Amer. Statist. Ass.*, **99**, 1166–1175.

[37] Kim, Y. J. and Gu, C. (2004). Smoothing Spline Gaussian Regression: More Scalable Computation via Efficient Approximation, *J. R. Statist. Soc. B*, **66**, 337–356.

[38] Kimeldorf, G. and Wahba, G. (1971). *J. Math. Anal. Appl.*, **33**, 82–95.

[39] Li, K. (1986). *Ann. Statist.*, **14**, 1101–1112.

[40] Liu, A. and Wang, Y. (2004), Hypothesis Testing in Smoothing Spline Models, *J Stat Comput Simul.*, **74**, 581–597.

[41] Liu, A., Meiring, W. and Wang, Y. (2005), Testing Generalized Linear Models Using Smoothing Spline Methods, *Statistica Sinica*, **15**, 235–256.

[42] Luo, Z. and Wahba, G. (1997). *J. Amer. Statist. Ass.*, **92**, 107–116.

[43] Mansfield, L. (1972). *Numer. Math.*, **20**, 99–114.

[44] Matheron, G. (1973). *Adv. Appl. Prob.*, **5**, 439–468.

[45] McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. Chapman and Hall, London, England.

[46] Meinguet, J. (1979). *J. Appl. Math. Phys. (ZAMP)*, **30**, 292–304.

[47] Nychka, D. (1986). Frequency Interpretation of Bayesian “Confidence” Intervals for Smoothing Splines. *Mimeo Series 1699*, Institute of Statistics, North Carolina State University, Raleigh, NC.

[48] Nychka, D., Wahba, G., Goldfarb, S. and Pugh, T. (1984). *J. Amer. Statist. Ass.*, **79**, 832–846.

[49] O’Sullivan, F. (1986). *Statist. Sci.*, **1**, 502–527.

[50] O’Sullivan, F., Yandell, B. and Raynor, W. (1986). *J. Amer. Statist. Ass.*, **81**, 96–103.

[51] O'Sullivan, F. and Wahba, G. (1985). A Cross Validated Bayesian Retrieval Algorithm for Nonlinear Remote Sensing Experiments. *Journal of Computational Physics*, **59**, 441-455.

J. Amer. Statist. Ass., **81**, 96–103.

[52] Parzen, E. (1970). Statistical inference on time series by rkhs methods. In *12th Biennial Seminar Canadian Mathematical Congress Proc*. R. Pyke, ed. Canadian Mathematical Congress, Montreal, pp. 1–37.

[53] Poirier, D. (1973). *J. Amer. Statist. Ass.*, **68**, 515–524.

[54] Prenter, P. (1975). In *Splines and Variational Methods*. Wiley, New York.

[55] Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.

[56] Rice, J. (1986). *Statist. Prob. Lett.*, **4**, 203–208.

[57] Sansone, G. (1959). *Orthogonal Functions*. Wiley, New York.

[58] Schoenberg, I. (1964). *Proc. Natl. Acad. Sci. U. S. A.*, **52**, 947–950.

[59] Schoenberg, I. (1968). Monosplines and quadrature formulae. In *Theory and Application of Spline Functions*, T. N. E. Grenville, ed. University of Wisconsin Press, Madison, WI.

[60] Schumaker, L. (1981). *Spline Functions*. Wiley, New York.

[61] Shiau, J., Wahba, G. and Johnson, D. R. (1986). *Atmos. Ocean Tech.*, **3**, 714–725.

[62] Shure, L., Parker, R. L. and Backus, G. E. (1982). *J. Phys. Earth Planet Interiors*, **28**, 215–229.

[63] Silverman, B. W. (1982). *Ann. Statist.*, **10**, 795–810.

[64] Silverman, B. W. (1985). *J. R. Statist. Soc. B*, **46**, 1–52.

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

[66] Speckman, P. (1985). *Ann. Statist.*, **13**, 970–983.

[67] Stone, C. J. (1985). *Ann. Statist.*, **13**, 689–705.

[68] Utreras, F. (1978). Quelques Resultats D'Optimalité Pour La Methode De Validation Croissée. *Seminaire d'Analyse Nuumerique, 301*, Grenoble, France.

[69] Utteras, F. (1985). *Numer. Math.*, **47**, 611–625.

[70] Vardi, Y. and Lee, D. (1993). From Image Deblurring to Optimal Investments: Maximum Likelihood Solutions for Positive Linear Inverse Problems. *J. R. Statist. Soc. B*, **55**, 569–612.

Utteras, F. (1985). *Numer. Math.*, **47**, 611–625.

[71] Villalobos, M. and Wahba, G. (1987). *J. Amer. Statist. Ass.*, **82**, 239–248.

[72] Wahba, G. (1975). *Numer. Math.*, **24**, 383–393.

[73] Wahba, G. (1978a). Interpolating Surfaces: High Order Convergence Rates and Their Associated Designs, with Applications to X-Ray Image Reconstruction. *Technical Report No. 523*, Statistics Dept., University of Wisconsin, Madison, WI.

[74] Wahba, G. (1978b). *J. R. Statist. Soc. B*, **40**, 364–372.

[75] Wahba, G. (1980). Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data. In *Approximation Theory III*, W. Cheney, ed. Academic, New York, pp. 905–912.

[76] Wahba, G. (1981a). *SIAM J. Sci. Statist. Comp.*, **2**, 5–16.

[77] Wahba, G. (1981b). *Commun. Statist. Theor. Meth.*, **10**, 2475–2514.

[78] Wahba, G. (1982a). *SIAM J. Sci. Statist. Comp.*, **3**, 385–386.

[79] 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**, W. S. K. Zeller, ed. Verlag, Birkhäuser, pp. 407–429.

[80] Wahba, G. (1983). *J. R. Statist. Soc. B*, **45**, 133–150.

[81] 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*, H. A. David and H. T. David, eds. Iowa State University Press, Ames, IA, pp. 205–235.

[82] Wahba, G. (1985a). *J. R. Statist. Soc. B*, **46**, 44.

[83] Wahba, G. (1985b). *Ann. Statist.*, **13**, 1378–1402.

[84] Wahba, G. (1987). Three Topics in Ill Posed Inverse Problems. In *Inverse and Ill-Posed Problems*, M. Engl and G. Groetsch, ed., Academic Press, 37-51.

[85] 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*, T. J. Boardman, ed. American Statistical Association, Washington, D.C., pp. 75–80.

[86] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

[87] Wahba, G. and Wang, Y. (1995) Behavior Near Zero of the Distribution of GCV Smoothing Parameter Estimates for Splines. *Statistics and Probability Letters*, **25**, 105–111.

[88] Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B., Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy, *Ann. Statist*, **23**, 1865-1895.

[89] Wahba, G. and Wendelberger, J. (1980). *Monthly Weather Rev.*, **108**, 1122–1145.

[90] Wahba, G. and Wold, S. (1975). *Commun. Statist. A.*, **4**, 1–17.

[91] Wang, Y. (1998), Smoothing Spline Models with Correlated Random Errors. *J. Amer. Statist. Ass.*, **93**, 341–348.

[92] Wang, Y. (2011). *Smoothing Splines: Methods and Applications*. Chapman and Hall, New York.

[93] Wang, Y. and Ke, C. (2009). Smoothing Spline Semi-parametric Nonlinear Regression Models. *J. Comp. Graph. Stat.*, **18**, 165–183.

[94] Weinert, H. L. and Kailath, T. (1974). *Ann. Statist.*, **2**, 787–794.

[95] Wendelberger, J. (1982). Smoothing Noisy Data with Multidimensional Splines and Generalized Cross-Validation. Ph.D. thesis, Statistics Dept., University of Wisconsin, Madison, WI.

[96] Winsberg, S. and Ramsay, J. O. (1983). University of Wisconsin, WI. *Psychometrika*, **48**, 575–595.

[97] Woltring, H. J. (1985). *Human Movement Sci.*, **4**, 229–245.

[98] Woltring, H. J. (1986). *Adv. Eng. Software*, **8**, 104–113.

[99] Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall, New York.

[100] Zhang, H. H. and Lin, Y. (2006), Component Selection and Smoothing in Smoothing Spline Analysis of Variance Models - COSSO, *Annals of Statistics*, **34**, 2272-2297.