JOURNAL OF COMPUTATIONAL PHYSICS 78, 493–508 (1988)

Smoothing, Splines and Smoothing Splines; Their Application in Geomagnetism

C. G. CONSTABLE AND R. L. PARKER

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093

Received August 11, 1987; revised January 4, 1988

We discuss the use of smoothing splines (SS) and least squares splines (LSS) in nonparametric regression on geomagnetic data. The distinction between smoothing splines and least squares splines is outlined, and it is suggested that in most cases the smoothing spline is a preferable function estimate. However, when large data sets are involved, the smoothing spline may require a prohibitive amount of computation; the alternative often put forward when moderate or heavy smoothing is desired is the least squares spline. This may not be capable of modeling the data adequately since the smoothness of the resulting function can be controlled only by the number and position of the knots. The computational efficiency of the least squares spline may be retained and its principal disadvantage overcome, by adding a penalty term in the square of the second derivative to the minimized functional. We call this modified form a penalized least squares spline, (denoted by PS throughout this work), and illustrate its use in the removal of secular trends in long observatory records of geomagnetic field components. We may compare the effects of smoothing splines, least squares splines, and penalized least squares splines by treating them as equivalent variable-kernel smoothers. As *Silverman* has shown, the kernel associated with the smoothing spline is symmetric and is highly localized with small negative sidelobes. The kernel for the least squares spline with the same fit to the data has large oscillatory sidelobes that extend far from the central region; it can be asymmetric even in the middle of the interval. For large numbers of data the penalized least squares spline can achieve essentially identical performance to that of a smoothing spline, but at a greatly reduced computational cost. The penalized spline estimation technique has potential widespread applicability in the analysis of geomagnetic and paleomagnetic data. It may be used for the removal of long term trends in data, when either the trend or the residual is of interest.

1. INTRODUCTION

In this paper we discuss a practical problem that frequently arises in the analysis of geomagnetic data, namely fitting a smooth curve of unknown parametric form to a time series of observations. The goal may be to interpolate the data providing a means of studying long term trends or to remove a trend and look at the remaining part of the signal. The approach that we advocate is that of using cubic smoothing splines to estimate the unknown function, i.e., finding the smoothest twice continuously differentiable function fitting the observations to a specified tolerance. The use of smoothing splines (SS) has particular merit in cases where one is dealing

0021-9991/88 $3.00

with irregularly spaced noisy data; then the standard filtering techniques used in time series analysis become somewhat awkward to implement. SS have been used in the analysis of historical geomagnetic and paleomagnetic data (see, e.g., Clark and Thompson [2]; Parker and Denham [7]; Malin and Bullard [6]) for the purposes of obtaining smooth curves from noisy data records. Here, we discuss the possibility of using them in the study of ionospheric signals in geomagnetic observatory data, for which we found it necessary to remove the long period secular trends associated with variations in the internal part of the geomagnetic field. Many of the records with which we were dealing extended back further than the available secular variation models obtained from global geomagnetic data, necessitating a means of estimating the variation directly from data at a single site. The data used were hourly mean values of the geomagnetic field, so that for records 30 years long the computational problem of finding the SS is considerable. One approach to reducing the computational task is to use least squares splines (denoted here by LSS) to estimate the function (e.g., [12, 13]). Algorithms for this type of estimation (and also for SS) are widely available in the standard mathematical software libraries such as IMSL and SLATEC. This method is an example of collocation; the only control over the smoothness of the resulting function comes from the choice of

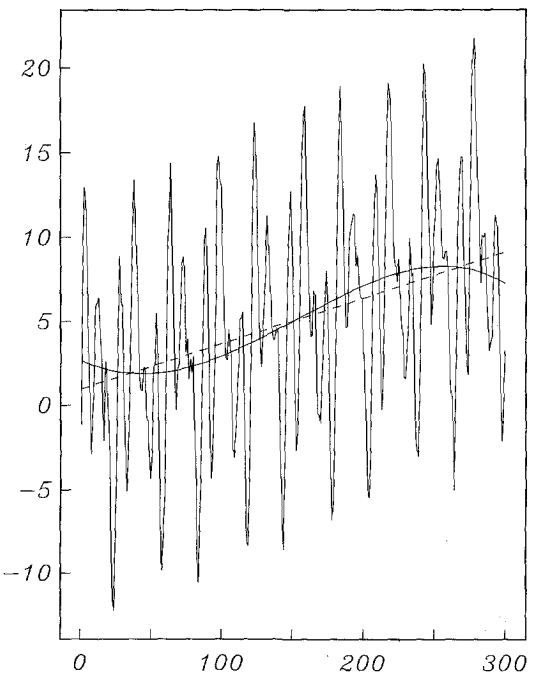

FIG. 1. The synthesized data series shown here is the sum of a linear trend, three high frequency cosinusoids, and Gaussian noise. The solid line is a LSS with a single interior knot. The SS, indicated by the dashed line, does not exhibit the unnecessary (and undesirable) structure put in by the LSS.

number and positions of the knots and the connection points for the spline basis functions (which are piecewise cubic polynomials).

There are situations in which SS may be judged too expensive and LSS cannot model the data adequately. This is perhaps best illustrated by an example where the LSS fails to perform adequately. Figure 1 shows a synthetic data series consisting of the sum of a linear trend, three high frequency cosinusoids, and Gaussian noise. The solid curved line is a LSS with a single interior knot, and does a very poor job of recovering the linear trend. The SS, shown dashed, on the other hand does a satisfactory job, because of the additional smoothing constraint imposed by minimizing the second derivative. This has led us to develop the procedure described here, which we shall refer to as penalized least squares splines (and abbreviate by PS), and provides an excellent approximation to the SS at a greatly reduced computational cost.

In the next section we outline the distinctions between these three methods, LSS, PS, and SS. Then we present an algorithm for the solution to the PS problem. The last section illustrates the properties of the three splines when they are regarded as kernel smoothers or filters and shows why the LSS performs so poorly under many circumstances.

2. Cubic Spline Smoothing

Let us suppose that we have N observations y_i, $i=1, ..., N$ at points x_i, from which we wish to determine a function $f(x)$, whose parametric form is not known, and that

$$y_i = f(x_i) + \varepsilon_i$$

We assume that $x_1 < x_2 < \cdots < x_N$ and that ε_i are random, uncorrelated errors with zero mean and variance σ_i^2. The appoach that we favor is the following: we seek the smoothest possible function in the class $C^2[x_1, x_N]$ (the class of twice continuously differentiable functions on the closed interval x_1, x_N) fitting the observations to a specified tolerance. The misfit to the data is measured by the weighted sum of squared discrepancies:

$$\sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2$$

and the roughness, which is to be minimized, is just the square of the two-norm of the second derivative:

$$\int_{x_1}^{x_N} [\partial_x^2 f(x)]^2 \, dx.$$

It is well known [9, 8] that the solution to this problem is the minimizer, $f_s(x)$, over functions $f \in C^2[x_1, x_N]$ of the functional

$$\sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2 + \lambda \int_{x_1}^{x_N} [\partial_x^2 f(x)]^2 \, dx, \tag{1}$$

where $\lambda > 0$ is a smoothing parameter controlling the trade-off between smoothness of f and goodness of fit to the data; λ is not specified directly, but must be discovered by finding the minimizer f_s with the correct misfit. The functions f_s are called the smoothing spline (here abbreviated to SS) estimators of the data. Upon application of calculus of variations the following property of the SS is immediately apparent: for every choice of $\lambda > 0$, f_s is a separate cubic polynomial within each interval $[x_i, x_{i+1}]$. In the terminology of the spline literature, the transition points x_i between one polynomial and the next are called the "knots" of the spline; for f_s the knots are identical to the points where the data are measured, but in general this need not be the case.

When heavy smoothing is appropriate we often find a very simple-looking minimizer f_s, but that function requires a seemingly excessive number of parameters to specify it: there are almost as many cubic polynomials $(N-1)$ as original data (N). Furthermore, when N is very large the computational cost of finding the optimally smooth curve in this way may be judged too high. It seems obvious that a smooth curve can be quite accurately represented by a much smaller number of parameters than the number required to match every wiggle of the uncorrelated random component of the measurements. This idea corresponds to LSS: here the number of knots is reduced to a small fraction of the number of data and least squares regression is performed in the B-spline basis fuctions associated with these knots [1, Chap. XIV]. The smoothness of the resulting function is controlled by the number and position of the knots instead of by means of a continuous parameter. There is no explicit roughness penalty in the minimized functional, corresponding to the second term in (1). In our experience, LSS have a number of drawbacks. The question of exactly how many knots should be used and their optimum location is a delicate problem, not yet satisfactorily solved. Furthermore, we have discovered practical situations in which it is impossible to enforce a high enough degree of smoothness, even with very small numbers of knots (e.g., the example of Fig. 1). When the action of the LSS is viewed as a kernel operator, the kernel is found to have excessive subsidiary peaks and asymmetry, properties quite undesirable in a good smoothing representation of the original data.

We present here a modification of the LSS technique that overcomes its disadvantages while preserving, in large measure, its computational economy and descriptive parsimony. We return to the minimization of (1), but instead of minimizing over the whole function space $C^2[x_1, x_N]$, we restrict attention to a subspace spanned by a collection of B-splines defined on a set of equally spaced knots. The idea is to use only as many knots as necessary to give a reasonable approximation to the SS f_s; in the case of strong smoothing this will be a tiny

fraction of N. In the next section we present a numerical implementation of this penalized least squares spline (PS) problem, taking advantage of the banded nature of the resulting matrices. We show that after an initial "least squares" phase of computation in which all the data are involved, all further calculations (for example, those to select the proper value of λ) can be performed on matrices with sizes governed by the number of knots. Also, it is possible at all times to keep computer memory requirements to the order of L^2, where L is the number of B-splines used. In Section 3 by means of Silverman's [10] asymptotic theory we are able to check the adequacy of our approximation to the optimally smooth f_s: if the number of knots in a preliminary computation is too small the resulting curve at a specified level of misfit will be too rough and then it is necessary to repeat the calculation with more knots so as to approach the desired smoothness. We compare SS, LSS, and PS regarded as kernel smoothers and illustrate some of the undesirable attributes of the ordinary LSS solution. The results show that, when large numbers of data are involved, the PS is a highly satisfactory alternative to the SS.

3. Solution to the Penalized Least Squares Spline Problem

The cubic splines may be defined for our purposes as the set of functions in $C^2[\xi_1, \xi_L]$ comprised of cubic polynomials on the intervals $[\xi_1, \xi_2]$, $[\xi_2, \xi_3]$,..., $[\xi_{L-1}, \xi_L]$. The points $\xi_1 < \xi_2 < \cdots < \xi_L$ are the knots of the splines. The B-splines constitute a basis for Γ_L, the space of cubic splines (see [1, Chap. IX]). As is well known, each element of the B-spline basis consists of a nonnegative function $b_j(x)$ with support $[\xi_{j-2}, \xi_{j+2}]$ composed of cubic polynomial sections between the knots with continuity of the function and its first two derivatives at the knots, so that $b_j \in C^2[\xi_1, \xi_L]$. A consequence of the continuity requirements is that, provided $j > 2$, $b_j(\xi_{j-2}) = \partial_x b_j(\xi_{j-2}) = \partial_x^2 b_j(\xi_{j-2}) = 0$ and similarly at ξ_{j+2}. If the knots ξ_j are chosen to coincide with the measurement points x_i the optimal SS solution f_s can be expanded in the B-spline basis. But, as we remarked in the Introduction, our intent is to use a much sparser set of knot points upon which to erect a representation of the smooth function:

$$f(x) = \sum_{j=1}^{L} \alpha_j b_j(x),$$

where $L \ll N$. We substitute the B-spline expansion into (1) and seek the minimizer over real α_j of the function F:

$$F = \sum_{i=1}^{N} \left[y_i - \sum_{j=1}^{L} \alpha_j b_j(x_i) \right]^2 + \lambda \int_{x_1}^{x_N} \left[\partial_x^2 \sum_{j=1}^{L} \alpha_j b_j(x) \right]^2 dx. \tag{2}$$

Here we have assumed a simplified random component with identical variances σ_i^2,

it is an easy matter to include the proper weighting if they are different but we have treated the special case to avoid unnecessarily cluttering up the notation. We introduce matrix notation with vectors and arrays having the following meanings: the matrix $B \in M(N \times L)$, the space of N by L matrices, has elements $B_{ij} = b_j(x_i)$; $y \in \mathbb{R}^N$ is the vector of data values; $H \in M(L \times L)$ is a matrix of inner products of second derivatives of B-splines:

$$H_{jk} = \int_{x_1}^{x_N} \partial_x^2 b_j(x)\, \partial_x^2 b_k(x)\, dx.$$

The elements of H are readily obtained in closed form and are particularly simple if the knots are evenly spaced in $[x_1, x_N]$ as we shall assume from now on. In matrix notation (2) may be written

$$F = \| y - B\alpha \|^2 + \lambda \alpha^T H \alpha, \tag{3}$$

where $\| \cdot \|$ is the Euclidean norm. Straightforward differentiation yields the following equation for $\alpha_0 \in \mathbb{R}^L$, the F-minimizing vector of expansion coefficients for fixed λ:

$$(B^T B + \lambda H)\, \alpha_0 = B^T y. \tag{4}$$

Although the matrix on the left of (4) is only $L \times L$ and therefore smaller than the one needed for the full SS solution, it runs the risk of poor conditioning because of its close connection to the normal equations (obtained in the limit as λ tends to zero). We develop a procedure that avoids the unnecessary numerical instability associated with the normal equations by solving a related system through QR factorization.

The first term in (3) can be written as

$$\| y - B\alpha \|^2 = \| y_1 - R\alpha \|^2 + \| y_2 \|^2,$$

where $R \in M(L \times L)$ is an upper triangular matrix, the upper square portion of the right factor in the QR decompositions of B; $y_1 \in \mathbb{R}^L$ is the upper part of the rotated data vector y and $y_2 \in \mathbb{R}^{N-L}$ the lower part of that vector. This is just the result of the application of the standard QR process to the first term of (3). As described in great detail by Lawson and Hanson [5, Chap. 27], the reduction to upper triangular form of B and the rotation of y can be performed very efficiently because B is banded; this is a direct result of the property that the support of $b_j(x)$ is $[\xi_{j-2}, \xi_{j+2}]$. Furthermore, it is unnecessary to hold all of B or y in the computer memory at one time since blocks of data may be brought from the disk and condensed to upper triangular form in a sequence of operations. The Householder triangularization of B in its banded form takes between about $25N$ and $50N$ operations. The actual count depends on the size of the data blocks, the least efficient result occurring when data are added to the system one row at a time. Because the value of λ required to obtain the desired tolerance is not known at the

outset, it is always necessary to solve the system several times with different values, but the bulk of the numerical work is over once this least squares phase has been accomplished.

Since the term $\|y_2\|^2$ is unaffected by variations in α (it is the squared misfit of the best-fitting spline to the data), the function to be minimized is now

$$F_1 = \|y_1 - R\alpha\|^2 + \lambda \alpha^T H \alpha. \tag{5}$$

If we can rewrite (5) as the sum of two terms, each of which is the square of a norm, we can continue to take advantage of the superior stability of QR decomposition for the solution of the minimization problem. In order to proceed we need a kind of Cholesky or square root factorization of the roughness penalty matrix H, but this is not quite straightforward because H is of rank $L-2$; this follows from the fact that a constant function and a linearly varying one are both in the space Γ_L, but the associated roughness $\alpha^T H \alpha$ vanishes in both cases. To solve the problem we transform the vector of unknown coefficients to another vector $a \in \mathbb{R}^L$ as follows. Let

$$\alpha = Va,$$

where

$$V = [I' \vdots v_1 \vdots v_2],$$

where $I' \in M(L \times L-2)$ is an $L-2 \times L-2$ unit matrix above two rows of zeroes, and $v_1, v_2 \in \mathbb{R}^L$ are given by

$$v_1 = (L-1, L-2, ..., 0)[L(L-1)(2L-1)/6]^{-1/2}$$
$$v_2 = (1, 1, ..., 1)\, L^{-1/2}$$

so that $\|v_1\| = \|v_2\| = 1$ and the matrix V is upper right triangular. Notice also that $Hv_1 = Hv_2 = 0$ because v_1 is the coefficient vector associated with a linearly varying function in an evenly distributed B-spline basis, and v_2 is associated with a constant function. In terms of a the roughness functional becomes

$$\alpha^T H \alpha = a^T V^T H V a,$$

where the matrix $V^T H V$ is composed of a positive definite, symmetric part in the top left corner, which we shall call $\tilde{H} \in M(L-2 \times L-2)$, and a lower and right border of zeros. Suppose $\tilde{H}$ has the Cholesky factorization $\tilde{H} = \tilde{J}^T \tilde{J}$, where $\tilde{J} \in M(L-2 \times L-2)$ is lower triangular; then the roughness is

$$\alpha^T H \alpha = \|[\tilde{J} \vdots 0 \vdots 0]a\|^2$$
$$= \|Ja\|^2,$$

where $J \in M(L-2 \times L)$ is the Cholesky factor $\tilde{J}$ with two columns of zeros at the right.

Returning to (5) we may now write the function to be minimized as

$$\|y_1 - RVa\|^2 + \lambda \|Ja\|^2 = \left\| \begin{bmatrix} \lambda^{1/2} J^T \\ RV \end{bmatrix} a - \begin{bmatrix} 0 \\ y_1 \end{bmatrix} \right\|^2 \tag{6}$$

which is just an ordinary, overdetermined least squares problem as we required. Since (6) must be minimized a number of times with different values of λ it pays to make the solution of this part of the problem efficient also. Notice that both R and V are in upper triangular form, so their product RV is also upper triangular; J is in that form too. Consider reordering the rows of coefficient matrix by interleaving rows of RV with those of J so that in the reordered array everything in column j below row $2j$ is zero; the same reordering must be done on the vector, of course. When the QR decomposition of the reordered system is performed, there is no need to treat the already zero, lower left portion of the coefficient matrix; this reduces the usual operations count for the QR minimization of about $5L^3/3$ to only $L^3/3$, a worthwhile improvement of a factor of five.

In practice the value of λ which will yield the required tolerance S^2 is not known *a priori* and must be found using an iterative procedure. The squared misfit for a given value of λ is

$$S^2 = \|B\alpha(\lambda) - y\|^2 = \|RVa(\lambda) - y_1\|^2 + \|y_2\|^2. \tag{7}$$

For computational purposes this may be written as

$$S^2 = \left\| \begin{bmatrix} \lambda^{1/2} J^T \\ RV \end{bmatrix} a - \begin{bmatrix} 0 \\ y_1 \end{bmatrix} \right\|^2 - \|\lambda^{1/2} J^T a\|^2 + \|y_2\|^2.$$

The first and third terms in this expression are obtained during the solution of Eq. (6) and the initial QR decomposition of the B matrix, respectively. The second term is the part of the misfit obtained from the solution of (6) which arises from the roughness misfit and thus contributes nothing to the data misfit and must therefore be subtracted off.

In addition to possessing a lower bound of $S^2_{\min} = \|y_2\|^2$, the tolerance function is bounded above in the limit as $\lambda \to \infty$. Then the roughness of the resulting function is penalized to the maximum possible extent and the best-fitting straight line (in the least squares sense) will be obtained. We make use of this upper bound in the iterative procedure used to find the value of λ corresponding to the desired misfit. The value of $S^2_{\max}$ can readily be found by performing the least squares fit to the straight line part of the spline, i.e., to v_1 and v_2. Once the initial QR decomposition for B has been performed this is a straightforward QR solution to the $L \times 2$ system of equations

$$[Rv_1 \vdots Rv_2]\beta \approx y_1.$$

In order to estimate the value of λ corresponding to the required tolerance we need $\partial S^2/\partial\lambda$. Making use of the fact that the equation actually solved in (6) is

$$[(RV)^T(RV)+\lambda\tilde{H}]=(RV)^T y_1$$

and defining

$$\tilde{R}=RV$$
$$M=(\tilde{R}^T\tilde{R}+\lambda\tilde{H})^{-1}$$

one can show that $S^2(\lambda)$ has the properties

(1) $\lambda\geqslant 0$
(2) $S^2(\lambda)=\lambda^2\,\|\tilde{R}^{-T}\tilde{H}a(\lambda)\|^2+\|y_2\|^2\geqslant 0$
(3) $\partial S^2/\partial\lambda=2\lambda a^T\tilde{H}M\tilde{H}a=2\lambda\,\|P^{-T}\tilde{H}a\|^2\geqslant 0$
(4) $\partial^2 S^2/\partial\lambda^2=-2a^T\tilde{H}M[2\lambda\tilde{H}-\tilde{R}^T\tilde{R}]\,M\tilde{H}a,$

where P is the upper triangular matrix obtained from the QR decomposition of

$$\begin{bmatrix}\lambda^{1/2}J^T\\ RV\end{bmatrix}.$$

A plot of a typical $S^2(\lambda)$ is shown in Fig. 2. We can exploit the generally concave shape of this curve and the fact that $\partial^2 S^2/\partial\lambda^2<0$ everywhere except close to the origin to devise a modified Newton scheme for finding iteratively the λ_* corresponding to the desired misfit, S_*^2. We approximate $S^2(\lambda)$ locally by a hyperbola, i.e., we assume that

$$S^2(\lambda)\approx S^2_{\max}+\frac{c_1}{1+\lambda c_2}.$$

Since for any value of λ, $S^2(\lambda)$, $\lambda S^2/\partial\lambda$, and $S^2_{\max}$ can readily be found from the expressions above, we may solve for c_1 and c_2 to obtain the local behavior of the misfit function; then only a linear equation need be solved to find λ_* by inverse interpolation in the hyperbolic model. It is easily seen that sufficiently near λ_* the iterative scheme gives the same approximants as Newton's method, but far from the solution, the modified method is much better behaved. We start at $\lambda_0=0$, where it may be shown that

$$\left.\frac{\partial S^2}{\partial\lambda}\right|_{\lambda=0}=\|(RV)^{-T}\tilde{H}a\|^2.$$

Using the above modified Newton method, a satisfactory solution can usually be obtained in four or five iterations; the standard Newton's method typically requires 10 or 15 iterations. Since each iteration involves resolving the $2L\times L$ system of

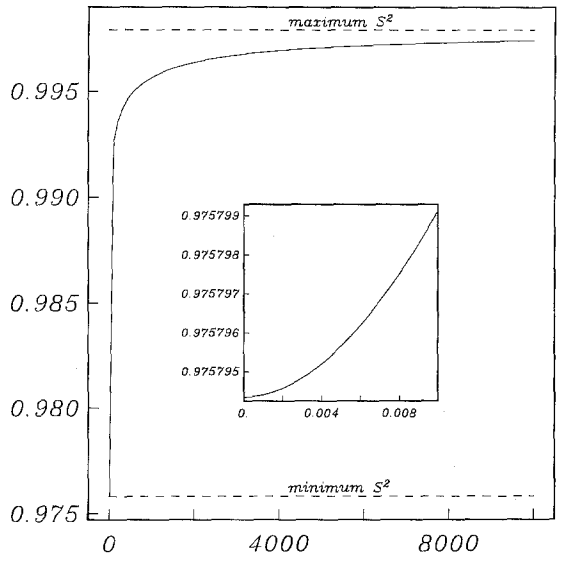

FIG. 2. Plot of a typical $S^2(\lambda)$. The inset shows an enlargement of the behavior close to the origin.

equations specified by (6) the saving in computer time can be substantial. Monotonic convergence from below of the series λ_j, $j=0, 1, ...,$ to λ_* is not guaranteed, and in practical tests we found rare cases of severe overshoot that significantly degraded the performance. To avoid this problem we found it useful to check the sign of $g(\lambda)=S^2(\lambda)-S_*^2$ after each iteration and if an overshoot is detected, we fit a polynomial of the form $g(\lambda)=(\lambda-c)/(d_0+d_1\lambda+d_2\lambda^2)$ using the known values of g and its derivatives at λ_n and λ_{n+1}. The zero at $\lambda=c$ provides the next estimate for λ_*.

The PS algorithm described here presumes that a reasonable method is available for determining S^2, the squared misfit to the data. Craven and Wahba [4] suggest using generalized cross validation to estimate the correct degree of smoothing for SS when the error variance σ^2 is unknown.

4. SPLINES AS VARIABLE-KERNEL SMOOTHERS

To assess the performance of penalized least squares cubic splines we turn to the equivalent variable-kernel representation, an approach applied by *Silverman* [10, 11] in his work on smoothing splines. The kernel function representation is possible for any of the spline smoothers because the smooth function estimate $f_s(x)$ is linear in the observations y_i; thus one may write

$$f_s(x)=N^{-1}\sum_{i=1}^{N} W(x, x_i; \lambda)\, y_i, \tag{8}$$

where x_i, $i=1, ..., N$ are the data sampling positions as before which for the SS coincide with the knot positions, ξ_i. The weight function $W(x, x_i; \lambda)$ shows how the data are averaged together to obtain the smoothed function estimate at any x. We regard the weight function generated by the SS as the optimal one, and the degree to which the PS can reproduce that behavior will be taken as the measure of success. Obviously for the purposes of practical comparison we cannot afford to compute the true W for the SS; fortunately, the important properties of the weight function can be obtained without resorting to any heavy computations at all.

Let the knot points have local density $\phi(\xi)$ so that the proportion of ξ_i in an interval of size $d\xi$ near ξ is approximately $\phi(\xi)\, d\xi$. Then Silverman [11] gives the asymptotic form of the weight function

$$W(x, \xi; \lambda) = \frac{1}{\phi(\xi)\, h(\xi)} K\left(\frac{x-\xi}{h(\xi)}\right), \tag{9}$$

where $K(u)$ is the kernel function

$$K(u) = \frac{1}{2} \exp\left(-\frac{|u|}{\sqrt{2}}\right) \sin\left(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right)$$

and the local averaging length scale or bandwidth $h(\xi)$ satisfies

$$h(\xi) = \lambda^{1/4} \phi(\xi)^{-1/4}. \tag{10}$$

Numerically $h(\xi)$ is a somewhat deceptive measure of the length scale over which the averaging kernel acts because it is too narrow. The distance from the central peak to the first zero of $K(u)$ is $(3\pi\sqrt{2}/4)h \approx 3.332h$. Silverman's approximation for $W(x, \xi; \lambda)$ holds provided x is not too close to the edge of the interval $[x_1, x_N]$, λ is not too large or too small, and N is sufficiently large. Equation (9) is not valid near the boundaries of $[x_1, x_N]$ but Silverman [10] describes a modification to account for this, under the same restrictions on N and λ as before. He shows that these asymptotic forms agree well with the exact weight functions obtained for SS and we shall assume that this is true for all the cases of interest to us.

The above result provides a means of determining whether we have used enough knots and basis functions in the PS. Recall that we regard the SS as a function that we wish to approximate with the PS; if the kernel function can be well represented in our approximation, it follows from (8) that the smooth curve f_s will be closely matched. We substitute λ_*, the smoothing parameter calculated for the PS, into (10) to find the averaging length of the SS with the same parameter. For the PS basis to be capable of a good representation of the ideal averaging kernel the bandwidth of the kernel should be significantly greater than the spacing between the knots. A more quantitative assessment is made possible by calculating the error of a spline approximation to the asymptotic kernel based upon samples with a uniform knot interval of Δ; the maximum error in such an approximation is shown in Fig. 3. Suppose we are content with an error of 10% of the peak amplitude of

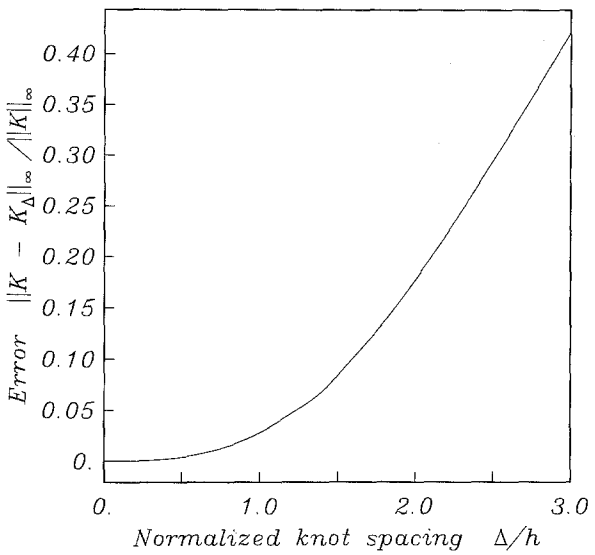

FIG. 3. Maximum error in the approximation of the asymptotic kernel $K(u)$ by a cubic spline K_Δ based upon even knot spacing Δ.

$K(u)$; then Δ/h must be less than 1.65, which corresponds to about four knot intervals under the central positive peak. Thus when $1.65h < \Delta$ we expect the approximation to be poor. Conversely, we assert, when the knot interval is smaller than this, we will obtain a close correspondence between the ideal weight function and the true one. In the event that too few basis functions have been used for the PS, λ_* will be smaller than the λ for the true SS and thus the bandwidth of the resulting smoother will be underestimated.

To illustrate these ideas and to compare LSS with the penalized (PS) kind we carried out some numerical experiments. For simplicity we consider equally spaced data, so that $\phi(\xi) = 1$ and $h(\xi)$ is constant over the whole interval. The smooth curve estimates in Fig. 4 are based upon an impulse data series with $N = 1001$; every $y_i = 0$ with the single exception that $y_{501} = 1$. With evenly spaced data Eq. (9) shows the kernel function is invariant under translation (except near the edges of the interval), and so from (8) it follows the smoothed impulse functions are just the averaging kernels. First, we fit the impulse function with a LSS based upon 10 equally spaced knots (knot spacing $\Delta = 100$); this is the curve drawn with short dashes in Fig. 4. The RMS misfit associated with this smoothing is 0.0314 and the same value is used for the target misfit in the PS and SS fits. The SS with this misfit is shown by the solid curve in the figure; it is indistinguishable from the asymptotic approximation to W with appropriate parameters. Also indistinguishable on the scale of the figure is the PS with $\Delta = 20$. When the knot spacing is increased to $\Delta = 83.33$, however, the PS no longer approximates the SS so well; it is shown by the curve composed of longer dashes. The smoothing parameter $\lambda_* = 1.20 \times 10^6$ and

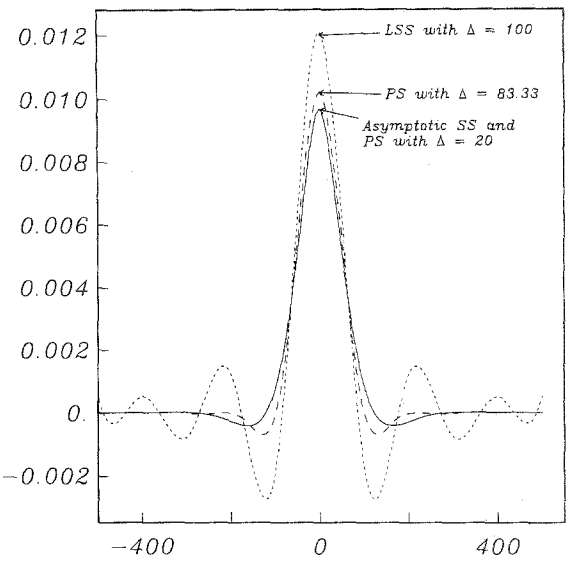

FIG. 4. Comparison of the various smoothings of an impulse function. x_i ranges from -500 to 500, $y_i = 0$ everywhere except $y_{501} = 1$. Short dashed curve: LSS with $\Delta = 100$. Solid curve: SS, PS, and asymptotic kernel with the same misfit as the LSS. Long dashed curve: PS having same misfit but with knot interval set too high.

from (10) this corresponds to $h = 33.1$; thus $1.65h = 55.1 < \Delta$, from which we could predict inadequate agreement between the SS and the PS. On the other hand, when $\Delta = 20$ we find $\lambda_* = 1.79 \times 10^6$ and then $1.65h = 60.4 > \Delta$ which is in accord with the evidence of Fig. 4.

When the PS approximates the true SS and they are both in agreement with the asymptotic theory, we noted earlier that the impulse response resembles the averaging kernel because the kernel function is invariant under translation. This does not apply to the LSS: the impulse response suffers considerable distortion if its knot points are moved only slightly; Fig. 5 illustrates this fact. In Fig. 4 the nonzero datum fell exactly on a knot point of the LSS; in the next figure the impulse lies halfway between the knots, which have the same spacing, $\Delta = 100$. While the RMS misfit is almost unchanged, the shape of the smoothed approximation, given by the short dashed curve, is radically different. The same kind of asymmetry appears in the PS when the knot spacing is too large; again $\Delta = 83.33$ is too wide an interval for the PS with the same misfit. The SS and its faithful approximation by a PS remain symmetric about the impulse—they are indeed accurate translations of the equivalent functions in Fig. 4.

The equivalent variable-kernel representation may be used as a means of determining whether the number of knots used in the fitting procedure was adequate but only after a trial calculation has been done so that the smoothing parameter is known. This number is obviously dependent on both the nature of the data and

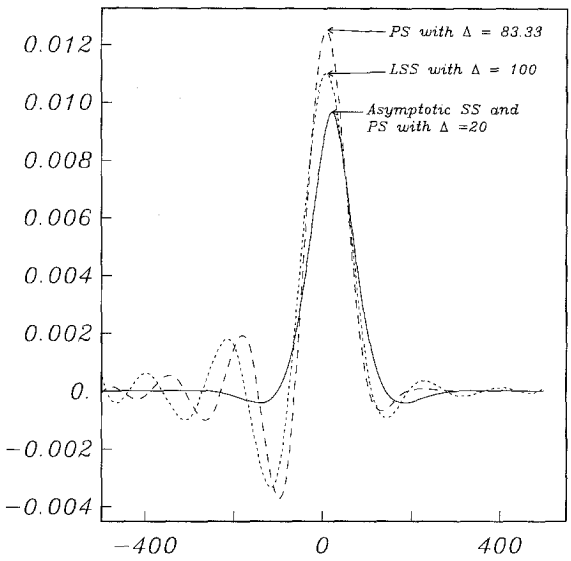

FIG. 5. Same as Fig. 4 except that the impulse is displaced and does not lie on a knot of the LSS or the PS.

how closely one requires the fitted curve to follow the data. Clearly the recovery of a trend from data will require fewer knots than fitting a curve which follows the data closely. We have used the PS successfully to recover slow secular trends from observatory records of the geomagnetic field which may contain up to 300,000 points. The smooth curves generally require less than 50 knot points. An example of the fitting procedure is shown in Fig. 6 for hourly values of the X-component of the geomagnetic field measured at Boulder, Colorado between January 1967 and March 1978. The problem is to remove the very long period secular trend due to the geomagnetic field of internal origin, while trying to preserve intact the ionospheric and magnetospheric variations, some of which have periods as long as a year. The upper portion of the figure shows the PS (solid line) fit to the data with an RMS misfit of 18.5 nT. The dashed line shows the LSS fit for the same number of knots; it has an RMS misfit of 17.45 nT. The lower portion of the figure shows an enlarged version of the section outlined by the box in the upper part. The data are also plotted here for comparison with the models; sheer numbers preclude plotting all the data in the upper part. Some readers may have reservations about using least squares fitting for data that is clearly non-Gaussian and has many outlying points in sections of the record associated with geomagnetic storms. Constable [3] discusses this problem in some detail for an alternative least squares problem associated with this type of data and describes a method for obtaining maximum likelihood parameter estimates. This could easily be used in the PS formalism described here. The sample of data shown here indicates how the PS performs

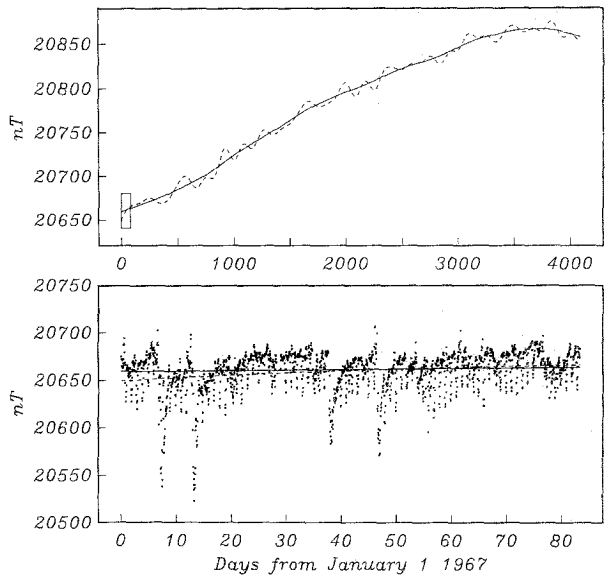

FIG. 6. Fitting splines to geomagnetic observatory data from Boulder, Colorado (Boulder *X* spline). The upper portion of the figure shows the LSS (dashed line) and PS (solid line) each with 50 knots. The lower part shows an enlargement of the area outlined by the box, along with the data from that region.

better in obtaining a smooth curve to represent the secular variation than the LSS, especially near the end of the record. We were unable to obtain satisfactory results with LSS no matter where the knots were placed or how many were used. It was impossible to obtain a smooth enough function with the desired level of misfit. This experience led to the development of the method described here.

ACKNOWLEDGMENTS

This work was funded by National Science Foundation Grant EAR-84-16212. We are grateful to John

REFERENCES

1. C. DE BOOR, *A Practical Guide to Splines* (Springer-Verlag, New York, 1978), Chap. XIV.
2. R. M. CLARK AND R. THOMPSON, *Geophys. J. Roy. Astron. Soc.* **52**, 205 (1978).
3. C. G. CONSTABLE, *Geophys. J. Roy. Astron. Soc.* **95**, (1988).
4. P. CRAVEN AND G. WAHBA, *Numer. Math.* **31**, 377 (1979).
5. C. L. LAWSON AND R. J. HANSON, *Solving Least Squares Problems* (Prentice–Hall, Englewood Cliffs, NJ, 1974).
6. S. R. C. MALIN AND SIR EDWARD BULLARD, *Philos. Trans. Roy. Soc. London A* **299**, 357 (1981).
7. R. L. PARKER AND C. R. DENHAM, *Geophys. J. Roy. Astron. Soc.* **58**, 685 (1979).

8. C. REINSCH, *Numer. Math.* **10**, 177 (1967).
9. I. J. SCHOENBERG, *Proc. Natl. Acad. Sci. USA* **52**, 947 (1964).
10. B. W. SILVERMAN, *Ann. Stat.* **12**, No. 3, 898 (1984).
11. B. W. SILVERMAN, *J. Roy. Stat. Soc. B* **47**, 1 (1985).
12. R. THOMPSON AND R. M. CLARK, *Phys. Earth Planet. Inter.* **27**, 1 (1981).
13. R. THOMPSON AND D. R. BARRACLOUGH, *J. Geomagn. Geolectr.* **34**, 245 (1982).