DEPARTMENT OF STATISTICS

University of Wisconsin
1210 W. Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 696

December 1982

SURFACE FITTING WITH SCATTERED NOISY
DATA ON EUCLIDEAN d-SPACE AND ON THE SPHERE

Grace Wahba

This research supported by ONR under Contract N00014-77-C-0675 and by NASA under Contract NAG-5-128.
Surface Fitting with Scattered Noisy Data on Euclidean d-Space and on the Sphere

by

Grace Wahba

ABSTRACT

An overview of cross validated spline methods for smoothing noisy data in the plane, in Euclidean d-space, and on the sphere is given. Cross validated thin plate smoothing splines are reviewed and an efficient numerical algorithm for computing them for problems with up to several hundred data points is described. Some numerical results for a two dimensional example are given. A theory of vector splines for smoothing noisy vector data on the sphere is given. The use of generalized cross validation to estimate both the smoothing parameter as well as the relative energy to be assigned to the divergent and nondivergent part of the smoothed vector field is described and tested numerically on simulated upper air horizontal wind fields.


This research supported by ONR under Contract N00014-77-C-0675 and by NASA under Contract NAG-5-128.
1. Introduction: An Overview of Cross Validated Smoothing Splines

It is assumed that data \( y = (y_1, \ldots, y_n)' \) arise according to the model

\[
y_i = f(P_i) + \varepsilon_i \quad i = 1, 2, \ldots, n
\]

where \( P_i \in S \), some index set (e.g. Euclidean d-space, the sphere, torus, etc.). \( f \) is assumed to be a smooth function in some reproducing kernel (r.k.) Hilbert space \( X \) of real-valued or vector-valued functions on \( S \), the \( \{\varepsilon_i\} \) are independent zero mean measurement errors with common unknown variance, and it is desired to recover an estimate of \( f \) given \( y = (y_1, \ldots, y_n)' \). The estimate \( f_\lambda \) of \( f \) will be taken as the minimizer in \( X \) of

\[
\frac{1}{n} \sum_{i=1}^{n} (y_i - f(P_i))^2 + \lambda J(f), \tag{1.1}
\]

where \( J(f) \) is a seminorm on \( X \) with \( M < n \) dimensional null space, spanned by \( \phi_1, \ldots, \phi_M \). Here \( J^{1/2}(f) \) can be taken as the norm of the orthogonal projection of \( f \) onto \( X_1 \) where \( X_1 \) is the orthocomplement of the span of the \( \{\phi_1\} \) in \( X \). If the \( n \times M \) matrix \( T \) with \( i \)th entry \( \phi_i(P_i) \) is of rank \( M \) then the minimizer of (1.1) exists uniquely and always has a representation of the form

\[
f_\lambda(P) = \sum_{i=1}^{n} c_i \xi_i(P) + \sum_{\nu=1}^{M} d_\nu \phi_\nu(P), \quad P \in S, \tag{1.2}
\]

where the \( \xi_i \) can be determined from an r.k. or semikernel for \( X \) with seminorm \( J^{1/2} \), and \( c = (c_1, \ldots, c_n)' \) and \( d = (d_1, \ldots, d_M)' \) satisfy a system of \( n + M \) linear equations. For more details, see Kimeldorf and Wahba (1971), Wahba (1978) and Section 2 below and references cited there.
The most famous special case of $J$ is surely

$$J(f) = \int_0^1 \left( (f(m)(x))^2 \right) dx,$$  \hspace{1cm} (1.3)

which for $m = 2$ leads to the celebrated cubic smoothing splines. In this paper we will primarily be considering generalizations of $J$ of (1.3) to Euclidean $d$-space of the form (for $d = 2$)

$$J(f) = \int \sum_{m=0}^\infty \left( \frac{\partial^m f}{\partial x_1^{m-v} \partial x_2^v} \right)^2 dx_1 dx_2,$$  \hspace{1cm} (1.4)

which lead to the thin plate splines, and to the sphere of the form (for $m$ even)

$$J(f) = \int_\mathbb{S} (\Delta^m f)^2 d\mathbf{p}$$  \hspace{1cm} (1.5)

where $\Delta$ is the Laplacian on the sphere.

The parameter $\lambda$ controls the tradeoff between the "roughness" of the solution as measured by $J(f)$, and the infidelity of the solution to the data, as measured by $\frac{1}{n} \sum_{i=1}^n (y_i - f(P_i))^2$. The visual appearance of $f_\lambda$ can be quite sensitive to the choice of $\lambda$, and estimates of derivatives of $f$ obtained by differentiating $f_\lambda$ even more sensitive. The value $\lambda^*$ of $\lambda$ which minimizes the predictive mean square error defined by $\sum_{i=1}^n (f_\lambda(P_i) - f(P_i))^2$ can be estimated from the data by the method of generalized cross validation (GCV). The GCV estimate $\hat{\lambda}$ of $\lambda$ is the minimizer of the cross validation function $V(\lambda)$ given by

$$V(\lambda) = \frac{\sum_{i=1}^n (f_\lambda(P_i) - f(P_i))^2}{\sum_{i=1}^n (f_\lambda(P_i) - f(P_i))^2}$$

which are generally not computed using (1.2) due to the existence of a local support basis for the span of the $\{\xi_i\}$ in the special case of (1.3).
\[ V(\lambda) = \frac{\frac{1}{n} \| (I - A(\lambda)) \|}{\left( \frac{1}{n} \text{Tr}(I - A(\lambda)) \right)^2}, \] (1.6)

where \( A(\lambda) \) is the so called "influence matrix", which satisfies

\[
\begin{pmatrix}
  f_\lambda(p_1) \\
  \vdots \\
  f_\lambda(p_n)
\end{pmatrix} = A(\lambda)y.
\] (1.7)

For \( n \) moderate to large, \( \hat{\lambda} \) so obtained is known to be a good estimate of \( \lambda^* \). For a discussion of the method and its properties, see Wahba (1977), Utreras (1978), Craven and Wahba (1979), Golub, Heath and Wahba (1979), Speckman (1982).

In general, to find \( f_\lambda \) for given \( \lambda \) requires the solution of a linear system of the order of the number, \( n \), of data points, and to find the minimizer of \( V(\lambda) \) entails the solution of an eigenvector eigenvalue problem of size \( n-M \times n-M \) (see Section 2 below). For \( n \) very large then, it has been proposed (Wahba 1980a, 1980b) that (1.1) be minimized in the span \( X^{(p)} \subset X \) of \( p \) suitably chosen basis functions, where \( p \) is chosen large enough so that \( \lambda \) (and not \( p \)) is the smoothing parameter. More specifically one wants the minimizer of (1.1) in \( X^{(p)} \) to be a good approximation to the minimizer of (1.1) in \( X \) if \( f \) is a "smooth" function. Natural choices of these basis functions are sines and cosines if \( X \) is a space of periodic functions on the circle, and E-splines (deBoor (1978)) of degree \( 2m-1 \) if \( J(f) = \frac{1}{2} \int (f^{(m)}(x))^2 dx \). Basis
functions which are a generalization of B-splines to Euclidean d-space (E^d) have been suggested in Wahba (1980a), see also Dyn and Levin (1981).

The cross validation function V(\lambda) is still defined by (1.6) and (1.7) but now A(\lambda) will depend on the basis functions used. In this case an n - M \times p singular value decomposition can be used to determine \hat{\lambda} instead of an n - M \times n - M eigenvalue-eigenvector decomposition, see Section 3 below. Bates and Wahba (1982) give efficient methods for computing \hat{\lambda} and \hat{f}_X when n is very large and M < p < n basis functions are used, including an efficient truncation procedure for the singular value decomposition.

In Section 2 we describe an efficient algorithm from J. Wendelberger's thesis (1982) for computing cross validated thin plate splines in d dimensions, and show some numerical results. In Section 3 we extend these ideas to vector observations on the sphere and investigate the ability of the GCV method to govern the relative energy assigned to the divergent and nondivergent part of the estimated vector field. The estimates appear to be very good from a mean square error point of view.

A good value of the parameter m governing the number of derivatives in J can also be estimated by GCV as can d-1 relative scale factors in E^d, see Gamber (1979), Wahba and Wendelberger (1980), Wendelberger (1982). It is believed that, if n is large, \lambda as well as several parameters in J can be estimated by GCV provided the several parameters are chosen so that distinct values of \lambda and the several parameters are always associated with r.k.'s which correspond to perpendicular stochastic processes, see Wahba (1981b) for a brief discussion of this point, which we don't pursue further here.
2. Thin plate splines on $\mathbb{E}^d$

In the thin plate spline case $J(f) = J_m^d(f)$ is given by

$$J_m^d(f) = \sum_{i_1=1}^{i_1=d} \ldots \sum_{i_m=1}^{i_m=d} \int \ldots \int \frac{d_m f}{\partial x_{i_1} \ldots \partial x_{i_m}} \, dx_1 \ldots dx_d.$$  

In particular, for $m = 2, d = 2$,

$$J_2^2(f) = \iint \left[ \left( \frac{\partial^2 f}{\partial x_1^2} \right)^2 + 2\left( \frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 + \left( \frac{\partial^2 f}{\partial x_2^2} \right)^2 \right] dx_1 dx_2.$$  

$X$ is the space of all generalized functions for which all partial derivatives of order $\leq m$ exist in $L_2(\mathbb{E}^d)$.

For $X$ to be a reproducing kernel Hilbert space (that is, for the evaluation functionals in $X$ to be bounded with respect to $J$) it is necessary that $2m - d > 0$. The null space of $J_m^d$ is the span of the polynomials in $\mathbb{E}^d$ of total degree less than $m$. Duchon (1976, 1977) has given a semikernel from which the explicit representation of $f$ given in Theorem 1 below has been obtained. (See also Meinguet (1979), Wahba and Wendelberger (1980)).

Before giving a formula for $f_\lambda$ we need some notation. Let $t \in \mathbb{E}^d$, $t = (x_1, x_2, \ldots, x_d)$. If $s = (y_1, \ldots, y_d)$, then $|t \cdot s| = (\sum_{i=1}^{d} (x_i - y_i)^2)^{1/2}$.

Let $E(\tau)$ be defined by

$$E(\tau) = \frac{\tau^{d-2m+1}}{\prod_{i=m}^{d} \log \tau} \quad \text{d even}$$

$$E(\tau) = \frac{\tau^{d-2m}}{\prod_{i=m}^{d}} \quad \text{d odd}$$

where
\[
\theta_m^d = \frac{(-1)^{d/2+1}}{2^{2m-1} \pi^{d/2} (m-1)! (m-d/2)!} \quad d \text{ even}
\]
\[
\theta_m^d = \frac{(-1)^m I(d/2-m)}{2^{2m-d/2} (m-1)!} \quad d \text{ odd}.
\]

Let \( E_{t_i}(t) \) be the function defined by
\[
E_{t_i}(t) = E(|t-t_i|),
\]
and let \( \phi_1, \ldots, \phi_M \) be the M polynomials of total degree less than
\( m \), for example, if \( d = 2, m = 2 \), then \( M = 3 \), and \( \phi_1(x_1, x_2) = 1 \), \( \phi_2(x_1, x_2) = x_1 \), \( \phi_3(x_1, x_2) = x_2 \).

Theorem 1. The minimizer \( f^* \) in \( X \) of
\[
\frac{1}{n} \sum_{i=1}^{n} (y_i - f(t_i))^2 + \lambda J_m^d(f)
\]
is given by
\[
f^*_\lambda = \sum_{i=1}^{n} c_i E_{t_i} + \sum_{i=1}^{M} d_i \phi_i
\]
where \( c = (c_1, \ldots, c_n)^t \) and \( d = (d_1, \ldots, d_M)^t \) are determined by
\[
(K + n\lambda I)c + Td = y \quad (2.3)
\]
\[
T^t c = 0 \quad (2.4)
\]
where \( K \) is the \( n \times n \) matrix with \( ij \)th entry \( E(|t_i - t_j|) \) and \( T \) is the \( n \times M \) matrix with \( ij \)th entry \( \phi_i(t_j) \).

If \( Q \) is any \( n \times n - M \) matrix satisfying \( Q'Q = I_{n-M} \) and \( Q'T = D_{n-M \times M} \),
then it can be shown (Wahba and Wendelberger (1980)) that

\[
\theta_m^d = \frac{(-1)^{d/2+1}}{2^{2m-1} \pi^{d/2} (m-1)! (m-d/2)!} \quad d \text{ even}
\]
\[
\theta_m^d = \frac{(-1)^m I(d/2-m)}{2^{2m-d/2} (m-1)!} \quad d \text{ odd}.
\]
\[ I - A(\lambda) = n\lambda Q(Q'KQ + n\lambda I)^{-1}Q' \]  \hspace{1cm} (2.5)

\[ n\lambda c = (I - A(\lambda))y \]  \hspace{1cm} (2.6)

An efficient numerical algorithm for obtaining \( \lambda, c \) and \( d \) is given in Wendelberger's thesis (1982). It goes as follows:

1. Form the QR decomposition of \( T \) using LINPACK (Dongarra, et al, (1979))

\[ T = (Q_1 : Q_2) \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \]  \hspace{1cm} (2.7)

where the matrix dimensions are \( T: n \times M; Q_1: n \times M; Q_2: n \times n-M; R_1: M \times M \), and \( (Q_1 : Q_2) \) is orthogonal and \( R_1 \) is upper triangular.

\( Q \) in (2.5) will be taken as \( Q_2 \). Let \( B = Q_2'KQ_2 \).

2. Find the eigenvalue eigenvector decomposition of \( B \)

\[ B = Q_2'KQ_2 = UD_U' \],

where \( U \) is orthogonal and \( D \) is diagonal with diagonal entries \( b_\nu \), \( \nu = 1, 2, \ldots, n-M \).

Then

\[ (I - A(\lambda))y = n\lambda Q_2U(D_\nu + n\lambda I)^{-1}U'Q_2'y \]  \hspace{1cm} (2.8)

\[ \text{Tr}(I - A(\lambda)) = \sum_{\nu=1}^{n-M} \frac{n\lambda}{b_\nu + n\lambda}. \]  \hspace{1cm} (2.9)

Letting

\[ w = (w_1, \ldots, w_{n-M}) = U'Q_2'y \]

gives
\[ V(\lambda) = \frac{1}{n} \left\| (I - A(\lambda)) y \right\|^2 = \frac{1}{n^2 \text{Tr}(I - A(\lambda))^2} \sum_{\nu=1}^{n-M} \left( \frac{n\lambda}{b_{\nu} + n\lambda} \right)^2 w_{\nu}^2. \]  

(2.10)

3. Find the minimizer \( \hat{\lambda} \) of \( V(\lambda) \) by global search in \( \log_{10} \lambda \) on the right hand side of (2.10).

4. Obtain \( c \) from

\[ c = Q_2 U (D_B + n\lambda I)^{-1} w. \]  

(2.11)

d may be obtained by noting that

\[ T = Q_1 R_1, \quad Q_1'Q_1 = I_{M \times M} \]

and premultiplying (2.3) by \( Q_1' \) to obtain

\[ Q_1'Kc + R_1d = Q_1' y. \]

5. Since \( R_1 \) is \( M \times M \) upper triangular, \( d \) is easily obtained by solving

\[ R_1d = Q_1' (y - Kc). \]  

(2.12)

A test function \( f(x_1, x_2) \) was generated by Wendelberger and is shown in Figure 2.1. A regular 7x7 square array of 49 points \( t_i \), \( i = 1, 2, \ldots, 49 \) was selected with the middle point \((0, 0)\) and the spacing 1.0. Data \( y_i \), \( i = 1, 2, \ldots, 49 \) was generated as

\[ y_i = f(t_i) + \epsilon_i, \quad t_i = (x_{i1}, x_{i2}), \]

where the \( \epsilon_i \) were pseudorandom normally distributed random variables with 0 mean.
and standard deviation .01. This standard deviation is about $1/8$ of the maximum height of $f$. Figure 2.2 shows $f_{\lambda}$ with $\lambda$ too big, Figure 2.3 shows $f_{\lambda}$ with $\lambda$ too small, and Figure 2.4 shows $f_{\hat{\lambda}}$. In this experiment the effectiveness of $\hat{\lambda}$ can be measured by the inefficiency

$$I = \frac{1}{n} \sum_{i=1}^{n} \frac{(f_{\lambda}(P_i) - f(P_i))^2}{\lambda} \sum_{i=1}^{n} \left( f_{\lambda}(P_i) - f(P_i) \right)^2$$

which was 1.54. Theory as well as other numerical results show that $I+1$ rapidly as $n$ becomes large, if $f$ is smooth. For further numerical results see Utreras (1979).
THE TEST FUNCTION.  
Figure 2.1

THE FITTED SURFACE \( \lambda = 100 \lambda \).  
Figure 2.2

THE FITTED SURFACE \( \lambda = 0.01 \lambda \).  
Figure 2.3

THE FITTED SURFACE \( \lambda = \lambda \).  
Figure 2.4
3. Vector Splines on The Sphere

Wind fields, magnetic fields, etc. are measured on the surface of the earth discretely and with error. From this discrete data it is desired to estimate the field everywhere. In the case of wind fields, it is also desired to estimate the horizontal divergence $D$ and the vorticity $\zeta$ of the fields. We define vector splines on the sphere for this purpose, and show the results of a Monte Carlo study.

We will first describe univariate splines on the sphere and then use the results to define vector splines on the sphere.

Let $P$ be a point on the sphere, $P = (\lambda, \phi)$, where $\lambda =$ longitude ($0 < \lambda < 2\pi$) and $\phi =$ latitude ($-\frac{\pi}{2} < \phi < \frac{\pi}{2}$). (We are using $\lambda$ here for both longitude and the smoothing parameter—which one is meant should be clear from the context). We will use the normalized spherical harmonics $Y_{\ell}^{S}(\lambda, \phi)$, which play the role of sines and cosines on the sphere. The $Y_{\ell}^{S}$ are defined by

$$Y_{\ell}^{S}(\lambda, \phi) = \begin{cases} \cos \lambda P_{\ell}^{S}(\sin \phi) & 0 < s < \ell \\ \sin \lambda P_{\ell}^{S}(\sin \phi) & -\ell < s < 0, \end{cases}$$

$$\theta_{\ell s} = \sqrt{2} \frac{(\ell-|s|)!}{(\ell+|s|)!}$$

$$= \frac{2\ell+1}{4\pi}$$

$$Y_{0}^{0}(\lambda, \phi) = 1$$

where the $P_{\ell}^{S}$ are the Legendre functions (Whittaker and Watson (1958)).
The spherical harmonics are the eigenfunctions of the (spherical) Laplacian

\[ \Delta Y_\ell^S = -\ell(\ell+1)Y_\ell^S, \quad (3.1) \]

where

\[ \Delta f = \frac{1}{a^2} \left[ \frac{1}{\cos^2 \phi} f_{\ell \lambda \lambda} + \frac{1}{\cos \phi} (\cos \phi f_{\ell \phi})_{\phi} \right] \]

and \( a \) is the radius of the sphere. The spherical harmonics form a complete orthonormal system on \( L_2(S) \). If \( f \in L_2(S) \) it has the Fourier Bessel expansion

\[ f \sim \sum_{\ell=0}^{\infty} \sum_{s=-\ell}^{\ell} f_{\ell s} Y_\ell^S \]

where

\[ f_{\ell s} = \int_S f(p) Y_\ell^S(p) dP, \]

see Sansone (1959).

Let

\[ J(f) = \sum_{\ell=1}^{\infty} \sum_{s=-\ell}^{\ell} \frac{f_{\ell s}^2}{\lambda_{\ell s}} \]

where the \( \{\lambda_{\ell s}\} \) are some nonnegative numbers with \( |\lambda_{\ell s}| \to 0 \) as \( \ell \to \infty \).

The set of all functions in \( L_2(S) \) with \( J(f) \) finite can be taken as a Hilbert space \( X \) with \( J^{1/2} \) as a seminorm, with null space the constant functions. \( X \) can be made into a reproducing kernel Hilbert space, provided
\[
\sum_{\lambda=1}^{\infty} (2\lambda+1) \max_{|s| \leq \lambda} \lambda_{s} < \infty.
\]

This follows since

\[
|f(P)| \leq |f_{00}| + \sum_{\lambda=1}^{\infty} \sum_{s=-\lambda}^{\lambda} |f_{\lambda s}| |Y_{\lambda s}(P)|
\]

\[
\leq |f_{00}| + \left( \sum_{\lambda=1}^{\infty} \sum_{s=-\lambda}^{\lambda} \frac{f_{\lambda s}^{2}}{\lambda_{s}} \right)^{1/2} \left( \sum_{\lambda=1}^{\infty} \sum_{s=-\lambda}^{\lambda} \lambda_{s} |Y_{\lambda s}(P)|^{2} \right)^{1/2}
\]

\[
\leq |f_{00}| + J^{1/2}(f) \frac{1}{4\pi} \sum_{\lambda=1}^{\infty} (2\lambda+1) \max_{|s| \leq \lambda} \lambda_{s},
\]

where, in obtaining the last inequality we have applied the addition formula for spherical harmonics (Sansone (1959)).

\[
\sum_{s=-\lambda}^{\lambda} Y_{\lambda s}(P) Y_{\lambda s}(P') = \frac{2\lambda+1}{2\pi} P_{\lambda}(\cos \gamma(P,P')),
\]

where \(\gamma(P,P')\) is the angle between \(P\) and \(P'\), the \(P_{\lambda}\) are the Legendre polynomials, and \(P_{\lambda}(0) = 1\).

It is easy to show that if \(\lambda_{s} = [\lambda(\lambda+1)]^{-m}\), then

\[
J(f) = \int (\Delta^{m/2} f)^{2} dP
\]

\[
= \int \frac{(\Delta^{(m-1)/2} f)^{2}}{\sin^{2}\phi} + (\Delta^{(m-1)/2} f_{\phi})^{2} dP
\]

\[m \text{ even}
\]

\[m \text{ odd}
\]

For \(X\) to be a reproducing kernel space it is necessary that \(m > 1\) (although not necessarily an integer). Later on, we will use the fact that \((\Delta f)(P)\) will be a bounded linear functional in \(X\) for each \(P\) provided
\[ \sum_{\ell=1}^{\infty} \mu^{2(\ell+1)^2} (2\ell+1) \max_{|s| \leq \ell} \lambda_{\ell s} < \infty. \]

We remark that, if \( \lambda_{\ell s} = \lambda_{s} \) independent of \( s \), then it can be shown by the use of (3.2) that \( J(.) \) is isotropic, that is, it is invariant under the group of all rotations of the sphere. The seminorms \( J_{m}^{d} \) of Section 2 are also isotropic, they are invariant under the group of all rotations and translations of \( E^{d} \). The vector splines on the sphere described below also result from isotropic seminorms.

There are, of course, situations when a specifically anisotropic seminorm may be called for (for example, when \( x_{1}, \ldots, x_{d-1} \) are space variables and \( x_{d} \) is a time variable, or, when the earth's Coriolis force is an important factor) but we omit detailed discussion of that case here.

We are now ready to define vector splines on the sphere. Let \( \vec{V} = (U, V) \) be a sufficiently regular horizontal vector field on the sphere, where \( U(P) \) is the eastward component and \( V(P) \) is the northward component at \( P \).

The vorticity \( \zeta \) and the divergence \( D \) of \( \vec{V} \) are given by

\[ \zeta = \frac{1}{\cos \phi} \left[ - \frac{\partial}{\partial \phi} (U \cos \phi) + \frac{\partial V}{\partial \lambda} \right] \]  

(3.4a)

\[ D = \frac{1}{\cos \phi} \left[ \frac{\partial U}{\partial \lambda} + \frac{\partial}{\partial \phi} (V \cos \phi) \right]. \]  

(3.4b)

Then there exists (by Helmholtz' Theorem) two functions \( \psi(P) \) and \( \phi(P) \), \( \text{PeS} \), called the stream function and the velocity potential respectively, with the following properties:
\[ U = \frac{1}{a} \left( -\frac{\partial \psi}{\partial \phi} + \frac{1}{\cos \phi} \frac{\partial \phi}{\partial \lambda} \right) \quad V = \frac{1}{a} \left( -\frac{1}{\cos \phi} \frac{\partial \psi}{\partial \lambda} + \frac{\partial \phi}{\partial \phi} \right) \]

\( \zeta = \Delta \psi \quad D = \Delta \phi. \)  \hspace{1cm} (3.6)

\( \psi \) and \( \phi \) are uniquely determined up to a constant, which we will take to be determined by \( \int_{S} \psi(p) \, dp = \int_{S} \phi(p) \, dp = 0. \)

Let \( \mathcal{X} \) be the collection of all pairs \((\psi, \phi)\) on the sphere which integrate to zero, are square integrable and

\[ J(1)(\psi) = \sum_{s=1}^{\infty} \sum_{s=-s}^{s} \frac{\psi_{s}^{2}}{\lambda_{s}^{(1)}}(1) < \infty \quad \psi_{s} = \int_{S} \psi(p) Y_{s}^{s}(p) \, dp \]

\[ J(2)(\phi) = \sum_{s=1}^{\infty} \sum_{s=-s}^{s} \frac{\phi_{s}^{2}}{\lambda_{s}^{(2)}}(2) < \infty \quad \phi_{s} = \int_{S} \phi(p) Y_{s}^{s}(p) \, dp \]

where \( \{\lambda_{s}^{(1)}\} \) and \( \{\lambda_{s}^{(2)}\} \) are strictly positive sequences satisfying

\[ \sum_{s=1}^{\infty} \lambda_{s}^{2}(s+1)^{2}(2s+1) \max_{s} \lambda_{s}^{(i)}(i) < \infty, \quad i = 1, 2. \]

\( \mathcal{X} \) is clearly a Hilbert space with square norm

\[ ||(\psi, \phi)||^{2} = J(1)(\psi) + \frac{1}{\delta} J(2)(\phi) \]

for any fixed \( \delta > 0 \) and furthermore both members of each pair possess Laplacians everywhere.

Let data \((U_{i}, V_{i})\) be given, where \( U_{i} \) and \( V_{i} \) are supposed to be noisy measurements on the true field \((U(P_{i}), V(P_{i}))\). Let \((\psi, \phi)\) be the streamfunction and velocity potential associated with \((U, V)\) and let \((\psi_{\lambda, \delta}, \phi_{\lambda, \delta})\) be the minimizer of
\[
\frac{1}{n} \sum_{i=1}^{n} \left( -\frac{1}{a} \frac{\partial \psi}{\partial \phi}(P_i) + \frac{1}{\cos \phi_i} \frac{\partial \phi}{\partial \lambda}(P_i) - U_i \right)^2 \\
+ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{\cos \phi_i} \frac{\partial \psi}{\partial \lambda}(P_i) + \frac{1}{a} \frac{\partial \phi}{\partial \lambda}(P_i) - V_i \right)^2 + \lambda [J_1(\psi) + \frac{1}{\delta} J_2(\phi)],
\]

(3.7)

Note that in the residual sum of squares above, \( U(P_i) \) and \( V(P_i) \) are expressed in terms of \( \psi \) and \( \phi \) via (3.5). A unique minimizer \( (\psi_{\lambda, \delta}^*, \phi_{\lambda, \delta}^*) \) exists for each \( \lambda > 0, \delta > 0 \) and the resulting wind field \( (U_{\lambda, \delta}^*, V_{\lambda, \delta}^*) \) constructed from \( (\psi_{\lambda, \delta}^*, \phi_{\lambda, \delta}^*) \) may be termed a vector smoothing spline field. Its vorticity and divergence will be given by \( \zeta_{\lambda, \delta} = \Delta \psi_{\lambda, \delta}^*, \Delta \phi_{\lambda, \delta} = \Delta \phi_{\lambda, \delta}^* \).

The parameter \( \delta \) influences the relative amount of energy that will be assigned to the divergent and nondivergent part of the estimated vector field. It is tempting to set \( \delta = 1 \), however, in some applications this may be a priori not the correct value. (The application we have in mind is to upper altitude horizontal wind fields in mid latitudes, where the divergent part is generally smaller than the nondivergent part.

In the numerical study described below we have investigated the ability of the GCV method to estimate a good value of \( \delta \) as well as a good value of \( \lambda \).

Explicit expressions for the minimizer of (1.1) with \( J \) given by (3.3) and related expressions may be obtained in terms of infinite series, and various methods of approximating the solution by a finite form may be based on approximations to the relevant r.k. See Wahba (1981a), Freeden (1981a,b), Shure, Parker and Backus (1982). Wendelberger (1982) has found a closed form expression for the cases of (3.3) with \( m = 2 \) and 3 in terms of di- and tri-logarithms. These results could have been extended to aid in the minimization of (3.7) for special choices of the \( \lambda_{\delta}(i) \). However, in
this work we have chosen to obtain the minimizer of (3.7) in the span
of two sets of the \( \tilde{N} = N(N+2) \) spherical harmonics. This proved to be
quite feasible for \( N \) up to around 16 on the Amdahl at Goddard Space
Flight Center.

Let

\[
\psi = \sum_{\lambda=1}^{N} \sum_{s=-\lambda}^{\lambda} \alpha_{\lambda s} Y_{\lambda}^{s}
\]

\[
\phi = \sum_{\lambda=1}^{N} \sum_{s=-\lambda}^{\lambda} \beta_{\lambda s} Y_{\lambda}^{s}.
\]

(3.8)

The coefficients \( \{\alpha_{\lambda s}\} \) and \( \{\beta_{\lambda s}\} \) for which (3.7) is minimized, and the
cross validation estimates of \( \lambda \) and \( \delta \) are found as follows.

First, the indices \( (\lambda, s), s = -\lambda, \ldots, \lambda \), are renumbered from 1 to \( \tilde{N} \).

Let \( X_{\phi} \) be the \( n \times \tilde{N} \) matrix with \((i, \lambda s)\) entry

\[
\frac{1}{a} \frac{\partial}{\partial \phi} Y_{\lambda}^{s}(p_i)
\]

and \( X_{\lambda} \) be the \( n \times \tilde{N} \) matrix with \((i, \lambda s)\)th entry

\[
\frac{1}{a} \frac{1}{\cos \phi_i} \frac{\partial}{\partial \lambda} Y_{\lambda}^{s}(p_i)
\]

and let \( X \) be the \( 2n \times 2\tilde{N} \) matrix

\[
X = \begin{pmatrix}
-X_{\phi} & X_{\lambda} \\
X_{\lambda} & X_{\phi}
\end{pmatrix}.
\]

(3.9)

Let \( D_{\delta} \) be the \( 2\tilde{N} \times 2\tilde{N} \) matrix
\[
D_\delta = \begin{pmatrix}
D_1 & 0 \\
0 & \delta D_2
\end{pmatrix}
\]

(3.10)

where \( D_i \) is the \( N \times N \) diagonal matrix with \( \lambda \)'s \( i \)th entry \( \lambda_{ii}(i), i = 1, 2 \).

Letting \( z = (U_1, \ldots, U_n, V_1, \ldots, V_n), \gamma = (\alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_N) \), it is seen by substituting (3.8) into (3.7) that we have to find \( \gamma \) which minimizes

\[
\frac{1}{n} \| z - X\gamma \|^2 + \lambda \gamma^T D_\delta^{-1} \gamma.
\]

The minimizer is

\[
\gamma = (X'X + n\lambda D_\delta^{-1})^{-1}X'z.
\]

(3.11)

By the use of (3.5) and (3.11), it follows that the estimated wind field \((U_\lambda, \delta; V_\lambda, \delta)\) at the data points satisfies

\[
\begin{pmatrix}
U_\lambda, \delta(P_1) \\
\vdots \\
U_\lambda, \delta(P_n) \\
V_\lambda, \delta(P_1) \\
\vdots \\
V_\lambda, \delta(P_n)
\end{pmatrix}
= A(\lambda, \delta)z
\]

(3.12)

where \( A(\lambda, \delta) \) is the \( 2n \times 2n \) "influence" matrix

\[
A(\lambda, \delta) = X(X'X + n\lambda D_\delta^{-1})^{-1}X'.
\]

The cross validation function \( V(\lambda, \delta) \) to be minimized in \( \lambda \) and \( \delta \) is

\[
V(\lambda, \delta) = \frac{\frac{1}{n} \| (I - A(\lambda, \delta))z \|^2}{\frac{1}{n} \text{Trace}(I - A(\lambda, \delta))^2}.
\]

(3.13)
The minimizer of $V(\lambda, \delta)$ was found, in the study described below, as follows. For fixed $\delta$, let $W_\delta = XD_\delta^{1/2}$, and let the singular value decomposition (SVD) of $W_\delta$ be

$$W_\delta = UD_\delta V'$$

(3.14)

where $UU' = U'U = I_{2n \times 2n} = V'V$ and $D_\delta$ is a diagonal matrix with entries $b_1, \ldots, b_{2n}$.

(Here, $\frac{n}{2} \leq n \leq \frac{N}{2}$.) $U, \{b_i\}$ and $V$ are computed using LINPACK. Letting

$$w = (w_1, \ldots, w_{2n})' = Uz,'$$

then

$$V(\lambda, \delta) = \frac{1}{2n} \sum_{i=1}^{2n} \left( \frac{n\lambda}{b_i^2 + n\lambda} \right)^2 w_i^2$$

$$= \left( \frac{1}{2n} \sum_{\nu=1}^{2n} \frac{n\lambda}{b_\nu^2 + n\lambda} \right)^2$$

(3.15)

$$\gamma = D_\delta^{1/2} v \begin{pmatrix} b_1 \\ b_1^2 + n\lambda \\ 0 \\ \vdots \\ 0 \\ b_{2n} \\ b_{2n}^2 + n\lambda \end{pmatrix} w.$$  

(3.16)

For fixed $\delta, \hat{\lambda}(\delta)$, the minimizer of (3.15), was easily found by a global search in increments of $\log \lambda$. Then $V(\hat{\lambda}(\delta), \delta)$ was plotted for 8 values of $\delta$ chosen in powers of $1/6$, and the minimum was readily evident. No doubt more efficient and automatic search procedures can be found. However, $Y$ is not a convex function of $\lambda$, and it is possible to encounter more than one local minimum. We have on occasion seen this when analyzing experimental data with small $n$.

For large $n$, $N$, and $W_\delta$ poorly conditioned, computing the SVD can be expensive, or it can fail to converge in a reasonable time. Some shortcut methods which alleviate this problem somewhat and use less storage have been developed. See Bates and Wahba, (1982).
We designed a Monte Carlo experiment to estimate the accuracy with which the method is able to estimate vorticity and horizontal divergence of the 500 mb. horizontal wind field over North America from observations on the East ($U_i$) and North ($V_i$) components of the wind taken by the North American radiosonde network. We will show the results of one such experiment. This information is useful in meteorological studies.

Since the accuracy will depend on the distribution of data points, and the relations between the energy in the true divergent and nondivergent parts of the wind field as well as the measurement error, pains were taken to make the simulated "truth" and simulated measurement errors as realistic as possible. Nevertheless, the results obtained below must be treated as a lower bound to accuracy expected in practice. In particular, it will be seen below that a particular sequence of $\lambda_{\lambda_S}(i)$ has been chosen to define $J(1)$ and $J(2)$, based on some (partly ad hoc) meteorological considerations, and the simulated "truth" below was generated via a random model based on the same $\lambda_{\lambda_S}(i)$. In practice the matching of the $\lambda_{\lambda_S}(i)$ used in $J_1$ and $J_2$ cannot be expected to match nature as well. (Also energy exists in nature at higher wavenumbers than are being simulated.)

We digress briefly to discuss the choice of the $\lambda_{\lambda_S}(i), i = 1, 2$, in this study. Suppose a (univariate) function on the sphere can be considered to be a random linear combination of spherical harmonics

$$f(p) = \sum \sum f_{\lambda_S} Y_{\lambda S}(p)$$

where the $f_{\lambda_S}$ are supposed to be realizations of zero mean independent (Gaussian) random variables with $Ef_{\lambda_S}^2 = b\lambda_{\lambda_S}$ for some $b$. 
Letting

\[ y_i = f(P_i) + \varepsilon_i \]

where the \( \{\varepsilon_i\} \) are independent and normally distributed with mean 0 and variance \( \sigma^2 \), then \( f_\lambda \) is the Bayes estimate of \( f \), that is

\[ f_\lambda(p) = E\{f(P) \mid y_1, \ldots, y_n\} \]

where \( f_\lambda \) is the minimizer of

\[ \frac{1}{n} \sum_{i=1}^{n} (u_i - f(P_i))^2 + \lambda J(f), \]

where \( \lambda = \sigma^2/nb \) and

\[ J(f) = \frac{f_{ss}^2}{\lambda_{ss}}. \]

For further details, see Kimeldorf and Wahba (1971), Wahba (1978), Wahba (1982b).

With many (possibly oversimplifying) assumptions information concerning \( \{\lambda_{ss}\} \) may be obtained, from historical records. For example, suppose \( \lambda_{ss} = \lambda_s \), independent of \( s \), (that is, the resulting \( J \) will be isotropic). Then

\[
Ef(P)f(P') = E \sum_{\lambda_s} \sum_{\lambda'_{s'}} f_{\lambda_s} f_{\lambda'_{s'}} Y_{\lambda} S(P) Y_{\lambda'} S'(P')
\]

\[ = b \sum_{\lambda_s} \lambda_s Y_{\lambda} S(P) Y_{\lambda} S'(P') \]

\[ = b \sum_{\lambda} \frac{2\lambda+1}{4\pi} p_{\lambda}(\cos \gamma(P,P')) \]

\[ = r(\gamma(P,P')), \text{ say.} \]
Figure 3.1 gives a sequence of idealized \( \{ \lambda_\ell \} \) of which the first 14 were used in this study and Figure 3.2 gives a plot of the correlation function \( \rho(\gamma) = r(\gamma)/r(0) \) associated with the \( \{ \lambda_\ell \} \) of Figure 3.1. Figure 3.3 gives a fitted isotropic correlation function estimated from historical 500 mb. geopotential fields by Julian and Thieboux (1975) which roughly matches that of Figure 3.2. Although in meteorological practice the isotropy and other assumptions made below may be suspect, in principle appropriate sequences \( \{ \lambda_{ks} \} \), or at least appropriate rates of decay of the \( \{ \lambda_{ks} \} \) may be obtained from historical records. See, for example, Kassahara (1976).

We obtained a model streamfunction and velocity potential of the form

\[
\psi = C_1 \sum_{\ell=1}^{N} \sum_{s=-\ell}^{\ell} a_{ks} Y^s_\ell \\
\phi = C_2 \sum_{\ell=1}^{N} \sum_{s=-\ell}^{\ell} b_{ks} Y^s_\ell
\]

by choosing \( a_{ks} \) and \( b_{ks} \) as normally distributed pseudo-random numbers with mean 0 and variances \( \lambda_{ks}(1) = \lambda_{ks}(2) = \lambda_\ell \) given in Figure 3.1. \( C_1 \) and \( C_2 \) were scale factors chosen so that the simulated \( \delta = \Delta \psi \) and \( D = \Delta \phi \) had magnitudes typical of real 500 mb. wind fields.

\[
\left( \frac{1}{4\pi} \int z^2 dp \right)^{1/2} = 6 \times 10^{-5} / \text{sec}, \left( \frac{1}{4\pi} \int D^2 dp \right)^{1/2} = 1 \times 10^{-5} / \text{sec}.
\]

If \( C_1 = C_2 \) then the optimal \( \delta \) would be 1 (or near 1). However, here, the divergent part of the wind field is smaller than the nondivergent part, and the optimal \( \delta \) will be less than 1. The experiment below tests the ability of GCV to estimate a good value of \( \delta \) (as well as \( \lambda_\ell \)).
Figure 3.1 gives a sequence of idealized \( \{ \lambda \} \) of which the first 14 were used in this study and Figure 3.2 gives a plot of the correlation function \( \rho(\gamma) = r(\gamma)/r(0) \) associated with the \( \{ \lambda \} \) of Figure 3.1.

Figure 3.3 gives a fitted isotropic correlation function estimated from historical 500 mb. geopotential fields by Julian and Thiebaux (1975) which roughly matches that of Figure 3.2. Although in meteorological practice the isotropy and other assumptions made below may be suspect, in principle appropriate sequences \( \{ \lambda \} \), or at least appropriate rates of decay of the \( \{ \lambda \} \) may be obtained from historical records. See, for example, Kassahara (1976).

We obtained a model streamfunction and velocity potential of the form

\[
\psi = C_1 \sum_{\ell=1}^{N} \sum_{s=-\ell}^{\ell} a_{\ell s} Y_{\ell s}^s \\
\phi = C_2 \sum_{\ell=1}^{N} \sum_{s=-\ell}^{\ell} b_{\ell s} Y_{\ell s}^s
\]

by choosing \( a_{\ell s} \) and \( b_{\ell s} \) as normally distributed pseudo-random numbers with mean 0 and variances \( \lambda_{s}(1) = \lambda_{s}(2) = \lambda \) given in Figure 3.1. \( C_1 \) and \( C_2 \) were scale factors chosen so that the simulated \( \zeta = \Delta \psi \) and \( D = \Delta \phi \) had magnitudes typical of real 500 mb. wind fields.

\[
\frac{1}{4\pi} \langle \xi^2 \rangle^{1/2} = 6 \times 10^{-5}/\text{sec,} \quad \frac{1}{4\pi} \langle D^2 \rangle^{1/2} = 1 \times 10^{-5}/\text{sec.}
\]

If \( C_1 = C_2 \) then the optimal \( \delta \) would be 1 (or near 1). However, here, the divergent part of the wind field is smaller than the nondivergent part, and the optimal \( \delta \) will be less than 1. The experiment below tests the ability of GCV to estimate a good value of \( \delta \) (as well as \( \lambda \)).
Figure 3.1. Idealized $\lambda_\alpha$.

Figure 3.2. Correlation function for the $\lambda_\alpha$ of Figure 3.1.

Figure 3.3. Sample Correlation Function.
Model wind vectors \( (U(P_i), V(P_i)) \) were computed from the model \((\psi, \phi)\) of (3.17) for \( \{P_i\} \) corresponding to \( n = 114 \) North American radiosonde stations. The data \( z = (U_1, \ldots, U_n, V_1, \ldots, V_n) \), where \( U_i = U(P_i) + \varepsilon_i^U, V_i = V(P_i) + \varepsilon_i^V \), were constructed by adding the measurement errors \( \varepsilon_i^U, \varepsilon_i^V \) as normally distributed pseudo random numbers with mean 0 and standard deviation \( \sigma = 2.5 \) meters/sec., a realistic value for the measurement error standard deviation.

Figure 3.4 shows the simulated wind vectors. Figure 3.5 shows the estimate of the true wind field, plotted on a \( 5^\circ \times 5^\circ \) grid in latitude and longitude. Figures 3.6 and 3.7 show the model and estimated vorticity and divergence, respectively. In Figures 3.5-3.7 \( \delta = 1/36 \) was used, which was the minimizer of \( V(\hat{\lambda}(\delta), \delta) \) found by the search described above. Figure 3.8 gives \( \text{MSE}(\lambda(\delta), \delta) \) and \( \text{MSE}(D(\lambda(\delta), \delta)) \) and their sum, where

\[
\text{MSE}(\lambda, \delta) = \frac{1}{K} \sum_{k=1}^{K} (\lambda(P_k) - \lambda(P_k))^2
\]

\[
\text{MSE}(D, \delta) = \frac{1}{K} \sum_{k=1}^{K} (D(P_k) - D(P_k))^2.
\]

The \( \{P_k\} \) constituted a regular grid inside the U.S. It can be seen from Figure 3.8 that if \( \delta \) is taken as too small (i.e. divergence is suppressed), then the mean square error in the estimated vorticity becomes large, and similarly if \( \delta \) is too large the mean square error in the estimated divergence becomes large. It appears that the GCV estimate of \( \delta \) here is quite close to the \( \delta \) which minimizes \( \text{MSE}(\lambda) + \text{MSE}(D) \).
Figure 3.8

Mean square error in the estimated vorticity and divergence,
REFERENCES


Surface Fitting with Scattered Noisy Data on Euclidean d-space and on the Sphere

Grace Wahba

Department of Statistics
University of Wisconsin
1210 W. Dayton St., Madison, WI 53706

Office of Naval Research
Arlington, VA 22217

Approved for public release; distribution unlimited.

An overview of cross validated spline methods for smoothing noisy data in the plane, in Euclidean d-space, and on the sphere is given. Cross validated thin plate smoothing splines are reviewed and an efficient numerical algorithm for computing them for problems with up to several hundred data points is described. Some numerical results for a two dimensional example are given. A theory of vector splines for smoothing noisy vector data on the sphere is given. The use of generalized cross validation to estimate both the smoothing parameter as well as the relative energy to be assigned to the divergent and nondivergent part of the smoothed vector field is described and tested numerically on simulated upper air horizontal wind fields.