See back page for 1473

TECHNICAL REPORT NO. 555

March 1979

HOW TO SMOOTH CURVES AND SURFACES WITH SPLINES AND CROSS-VALIDATION

Grace Wahba

How to Smooth Curves and Surfaces With Splines and Cross-Validation

by

Grace Wahba

ABSTRACT

We briefly review the use of smoothing splines and the method of generalized cross validation (GCV) for smoothing discrete noisy data from an unknown but smooth curve. Then we describe the use of "plaque mince" or Laplacian smoothing splines with GCV for smoothing discrete noisy data from an unknown but smooth surface. A numerical algorithm for this (non-trivial!) computational problem is described, and an example from a Monte Carlo study is presented to show how the method works on simulated data. The results are extremely promising. Some design problems are briefly mentioned. Some conjectures are made concerning optimality properties of Laplacian smoothing splines and Laplacian histosplines.

Note: This report was written for the Proceedings of the 24th Design of Experiments Conference, held at Madison, Wisconsin, May 3-5, 1978, and sponsored by the U.S. Army Research Office.

TYPIST: Mary E. Arthur

This work was sponsored in part by the U.S. Army under Contract No. ~~DAAG29-77-G-0207~~ and in part by the Office of Naval Research under Grant No. N00014-77-C-0675.

How to Smooth Curves and Surfaces With Splines and Cross-Validation

1. Introduction

In the conference talk we considered four problems. The first two had to do with estimating curves when they are observed discreetly and with error. The model is

$$y_i = f(t_i) + \epsilon_i, \qquad i = 1,2,\ldots,n$$

where $f(t)$, $t \in [0,1]$ is an unknown curve, only known to be "smooth", $0 \le t_1 \le \ldots t_n \le 1$, and ϵ_i are independent zero mean random variables with a common unknown variance σ^2. The $\{y_i\}$ are observed. The first problem is: How should f be estimated nonparametrically from $\underset{\sim}{y} = (y_1,\ldots,y_n)'$? The second (or design) problem is: How should the points $\{t_i\}$ be chosen so that the estimate of f is as good as possible? The third and fourth problems have to do with estimating surfaces. The model is

$$z_i = u(x_i,y_i) + \epsilon_i, \qquad i = 1,2,\ldots,n,$$

where $u(x,y)$, $(x,y) \in$ some region in the plane, is only known to be "smooth". (x_i,y_i), $i = 1,2,\ldots,n$ are n points in this region, the ϵ_i are zero mean independent random variables with common unknown variance σ^2, and $\underset{\sim}{z} = (z_1,\ldots,z_n)'$ is observed. The third problem is: How should u be estimated nonparametrically from $\underset{\sim}{z}$. The fourth (or design) problem is: How should the points (x_i,y_i), $i = 1,2,\ldots,n$ be chosen so that the estimate of u is as good as possible. We will not discuss the design problems here. The design work mentioned in the talk has appeared in Athavale and Wahba (1978) and Wahba (1971, 1974, 1976, 1978c). That work (and the work of others, mentioned there) represents only some first steps in design problems for nonparametric curve and surface fitting. There are many open problems.

Very good and relatively complete results for the first (curve estimation) problem are available (Craven and Wahba (1979), (CW) and Golub, Heath and Wahba (1977) (GHW), and transportable code is available from at least three sources Fleisher, (1979), Merz (1978a), and Paihua (1978). We will briefly summarize those results, because they will aid in understanding our discussion of the third problem, that is, surface smoothing. The remainder of this paper will then be devoted to the problem of smoothing of surface data non-parametrically. Some very nice theoretical results are available, and we have turned them into a computer program which delivers very pleasing pictures. The development of the program is the work of Mr. James Wendelberger, and it and other results will appear in his Ph.D. thesis.

2. Curve Smoothing

For curve smoothing, we recommend that f be estimated by the solution $f_{n,m,\lambda}$ of the minimization problem: Find $f \in H^m = \{f: f, f', \ldots, f^{(m-1)}$ abs. cont., $f^{(m)} \in L_2[0,1]\}$ to minimize

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

The first term represents infidelity of f to the data, and the second term represents "roughness" of the solution. The parameter λ represents the tradeoff between the two. m=2 represents "psychological" smoothness (we think!) and is frequently used, and gives good results. We briefly discuss the determination of m from the data later. The solution $f_{n,m,\lambda}$ is known to be a polynomial spline of degree 2m-1. The parameter λ is chosen from the data by the method of generalized cross-validation (GCV). GCV is derived from CV ("ordinary" cross validation). CV goes as follows: Let $f^{(k)}_{n,m,\lambda}$

be the solution of the minimization problem of (1) with the kth data point omitted. The value λ will be a good choice if $f^{(k)}_{n,m,\lambda}(t_k)$ comes close, on the average, to y_k. We measure this by the "ordinary" cross validation function $V_0(\lambda) = V_0^m(\lambda)$.

$$V_0(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(f^{(k)}_{n,m,\lambda}(t_k)-y_k)^2 \; .$$

For fixed m the parameter λ is chosen by minimizing $V_0^m(\lambda)$. For technical reasons involving convergence proofs, we replace $V_0(\lambda)$ by the generalized cross validation function

$$V(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(f^{(k)}_{n,m,\lambda}(t_k)-y_k)^2 w_k(\lambda)$$

where the $\{w_k(\lambda)\}$ are certain weights to reflect unequally spaced data, end effects, etc. Details are given in CW and GHW. It turns out that $V(\lambda)$ is much easier to compute than $V_0(\lambda)$, and $V(\lambda)$ has the representation

$$V(\lambda) = \frac{\frac{1}{n}\|(I-A(\lambda))\underline{y}\|^2}{(\frac{1}{n}\,\mathrm{Trace}(I-A(\lambda)))^2}$$

where $A(\lambda)$ is the n×n matrix which is uniquely determined by

$$\begin{pmatrix} f_{n,m,\lambda}(t_1) \\ \vdots \\ f_{n,m,\lambda}(t_n) \end{pmatrix} = A(\lambda)\underline{y} \; .$$

Pleasing results have been obtained using smoothing splines with GCV in both Monte Carlo studies and various applications, Benedetti (1977), CW, GHW, Merz (1978a, 1978b), Stutzle (1977), Utreras (1978a), Welch (1979). These results are not surprising in the light of the following theoretical result (CW, GHW). Let $R(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(f_{n,m,\lambda}(t_i)-f(t_i))^2$. $R(\lambda)$ is the "true mean square error" averaged over the data points. Before data are observed both $R(\lambda)$ and $V(\lambda)$

can be considered functions of the unknown f and random functions of the $\{\epsilon_i\}$. Let λ^* be the minimizer of $ER(\lambda)$ and $\hat{\lambda}$ be the minimizer of $EV(\lambda)$. We have under rather general circumstances (see CW, BMW)

$$\lim_{n\to\infty} \frac{ER(\hat{\lambda})}{ER(\lambda^*)} \downarrow 1 \ . \tag{2}$$

Thus (very loosely), the mean square error with the estimated λ tends to the minimum mean square error achieveable with any λ. Let $\hat{\lambda}$ be the minimizer of $V(\lambda)$. Numerical results based on Monte Carlo studies with $m=2$ reported in CW, with $n=50$ and equally spaced data points, show the achieved inefficiency $R(\hat{\lambda})/\min_\lambda R(\lambda)$ in the range 1.01 to 1.42.

Some numerical experiments to assess the effectiveness of choosing m by GCV have been done. (Lucas, 1978). One obtains $V^m(\hat{\lambda})$ for each m and minimizes $V^m(\hat{\lambda})$ over m. The results indicate that this procedure does a good job of picking out the m and $\hat{\lambda}$ which minimize $R(\lambda) = R^m(\lambda)$, and that there are classes of f's for which it is worthwhile to do this, that is, $\min_\lambda R^m(\lambda)$ is usefully less than $\min R^2(\lambda)$ for some $m \neq 2$. Efficient transportable code is not presently available, however. Depending on f, reduction in inefficiency of several percent can be obtained.

3. Surface Smoothing

We now turn to the third problem, that of recovering smooth surfaces. We recommend that u be estimated by the solution $u_{n,m,\lambda}$ of the minimization problem: Find $u \in \mathcal{H}$ (an appropriate space, to be described) to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(u(x_i,y_i)-z_i)^2 + \lambda \sum_{j=1}^{m} \iint \binom{m}{j}\left(\frac{\partial^m u}{\partial x^j \partial y^{m-j}}\right)^2 dxdy \ . \tag{3}$$

and that λ (and possibly m) be estimated by GCV. We now describe how to do this. For mathematical convenience the limits on the double integral in

(3) are taken to be $-\infty$ and ∞. $\mathcal{H}$ is taken as $\mathcal{H} = \mathcal{H}^m(R^2) = \{u:\ u \in \mathcal{D}'$, $\frac{\partial^m u}{\partial x^j \partial y^{m-j}} \in L_2(R_2)$, $j = 0,1,\ldots,m$. ($\mathcal{D}'$ is the dual of the Schwartz space $\mathcal{D}$ of infinitely differentiable functions with compact support, this need not concern us here, see Meinguet (1978, 1979), Schwartz (1966).)

Theorem: Let $t_i = (x_i,y_i)$, $t = (x,y)$ and $|t-t_i| = ((x-x_i)^2+(y-y_i)^2)^{1/2}$. Let $m \geq 2$ and $n \geq M = \binom{m+1}{2}$. The solution $u_{n,m,\lambda}$ to the problem: Find $u \in \mathcal{H}$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(u(t_i)-z_i)^2 + \lambda \iint \sum_{j=0}^{m} \binom{m}{j} \left(\frac{\partial^m u}{\partial x^j \partial y^{m-j}}\right)^2 dxdy \ . \tag{4}$$

is given by

$$u_{n,m,\lambda}(t) = \sum_{j=1}^{n} c_j E_m(t,t_j) + \sum_{\nu=1}^{M} d_\nu \phi_\nu(t) \ . \tag{5}$$

where

$$E_m(s,t) = \theta_m |s-t|^{2m-2} \log|s-t| \ , \quad \theta_m = (2^{2m-1}\pi[(m-1)!]^2)^{-1}$$

$$\phi_\nu(t) = x^\alpha y^\beta \qquad \nu = 1,2,\ldots,M$$

where α, β run over all the M combinations of non-negative integers with $\alpha+\beta \leq m-1$, provided the $n\times M$ matrix T with $i\nu^{th}$ entry $\phi_\nu(t_i)$ is of rank M. The coefficients $\underset{\sim}{c} = (c_1,\ldots,c_n)'$ and $\underset{\sim}{d} = (d_1,\ldots,d_M)'$ are determined by

$$(K+\rho I)\,\underset{\sim}{c} + T\underset{\sim}{d} = \underset{\sim}{z} \tag{6}$$

$$T'\underset{\sim}{c} = 0 \tag{7}$$

where K is the $n\times n$ matrix with jk^{th} entry $E_m(t_j,t_k)$, and $\rho = n\lambda$.

This theorem is essentially due to Duchon (1976a, 1976b). Meinguet (1978)) has also proved very similar results in a reproducing kernel Hilbert space setting. For completeness, in the Appendix we outline a proof which roughly follows Meinguet's argument. By putting the minimization problems of (1) and (4) in a reproducing kernel Hilbert space setting, $f_{n,m,\lambda}$ and $u_{n,m,\lambda}$ can be shown to be Bayes estimates with a certain (partially improper) prior on f or u, see Wahba (1978a).

4. An Algorithm for Computation of the Smoothing Surface

We now want to compute $u_{n,m,\lambda}$ efficiently, and choose λ (and possibly m) by GCV. Our algorithm below has benefited from the algorithmic work of Paihua (1978). However it is different and seems especially well adapted to determining the generalized cross validation function $V(\lambda)$ for this case. We next derive the equations behind our computational approach.

Let R be any $n\times(n-M)$ dimensional matrix of rank $n-M$ satisfying $R'T = 0_{(n-M)\times M}$. Subscripts indicate the dimensions of the subscripted matrix. Since $T'\underline{c} = 0$, we have

$$\underline{c} = R\underline{\gamma} \tag{8}$$

for $\underline{\gamma}$ a unique $n-M$ dimensional vector. Left multiplying (6) by R' and substituting (8) into (6) gives

$$R'(K+\rho I)R\gamma = R'\underline{z} \tag{9}$$

$$\gamma = (R'(K+\rho I)R)^{-1}R'\underline{z} \tag{10}$$

$$\underline{c} = R(R'KR+\rho R'R)^{-1}R'\underline{z}\ . \tag{11}$$

The vector $\underline{d}$ is then given by $\underline{d} = (T'T)^{-1}T'(\underline{z}-K\underline{c})$, obtained by left multiplying (6) by T'. To estimate λ (equivalently ρ) by GCV, we want to choose λ to minimize

$$V(\lambda) = \frac{\frac{1}{n}\|(I-A(\lambda))\underline{z}\|^2}{(\frac{1}{n}\,\mathrm{Trace}(I-A(\lambda)))^2}\ .$$

where $A(\lambda)$ is the $n\times n$ matrix determined by

$$\begin{pmatrix} u_{n,m,\lambda}(t_1) \\ \vdots \\ u_{n,m,\lambda}(t_n) \end{pmatrix} = A(\lambda)\underline{z}\ .$$

To talk about good properties of GCV here, we suppose the $\{t_i\}$ will be in a bounded region of the plane R_2 (even though the minimization is over functions in R_2). The basic property (2) of GCV can then be shown to hold as the t_i become dense in this region - the proof (CW,GHW) is independent of the nature of the region.

To obtain a convenient representation for $A(\lambda)$, we see from (5) that

$$\underline{z} - \begin{pmatrix} u_{n,m,\lambda}(t_1) \\ \vdots \\ u_{n,m,\lambda}(t_n) \end{pmatrix} = \underline{z} - K\underline{c} - T\underline{d}\ . \tag{11}$$

From (6), we have

$$\underline{z}-T\underline{d} = (K+\rho I)\underline{c}$$

so that the right hand side of (11) equals $\rho\underline{c}$. Thus,

$$(I-A(\lambda))\underline{z} = \rho\underline{c} = \rho R(R'KR+\rho R'R)^{-1}R'\underline{z} \tag{12}$$

We need to compute $\underline{c}$, $\|(I-A(\lambda))\underline{z}\|^2$, and $\mathrm{Tr}(I-A(\lambda))$. Any $R_{n\times(n-M)}$ will have a singular value decomposition

$$R = U_{n\times(n-M)}D_{(n-M)\times(n-M)}V'_{(n-M)\times(n-M)} \tag{13}$$

where $U'U = V'V = I_{n-M}$ and D is diagonal. Then

$$\begin{aligned} R(R'KR+\rho R'R)^{-1}R' &= UDV'(VDU'KUDV'+\rho VDDV')^{-1}VDU' \\ &= U(U'KU+\rho I)^{-1}U' \end{aligned} \tag{14}$$

Define

$$B_{(n-M)\times(n-M)} = U'KU\ .$$

and define Γ and δ by

$$B = \Gamma\delta\Gamma' ,$$

where Γ and δ are the orthogonal and diagonal matrices in the eigenvalue decomposition of B. Then the right hand side of (14) becomes

$$U(\Gamma\delta\Gamma'+\rho)^{-1}U' \qquad (15)$$

$$= U\Gamma'(\delta+\rho I)^{-1}\Gamma U' \qquad (16)$$

$$= U\Gamma'\begin{pmatrix} \frac{1}{b_1+\rho} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{b_{n-M}+\rho} \end{pmatrix}\Gamma U' \qquad (17)$$

where $b_1,\ldots,b_{n-M}$ are the diagonal entries in δ (i.e. the eigenvalues of B). Given $U\Gamma'$, $\{b_i\}$ we compute

$$\underline{c} = U\Gamma'\begin{pmatrix} \frac{1}{b_1+\rho} & & \\ & \ddots & \\ & & \frac{1}{b_{n-M}+\rho} \end{pmatrix}\Gamma U'\underline{z} ,$$

$$||(I-A(\lambda))\underline{z}||^2 = \rho^2||(\delta+\rho I)^{-1}\Gamma U'\underline{z}||^2$$

and

$$(\frac{1}{n}\,\mathrm{Tr}(I-A(\lambda)))^2 = (\frac{1}{n}\sum_{i=1}^{n-M}\frac{\rho}{b_i+\rho})^2 .$$

We now discuss the determination of U. It can be seen that U is any matrix whose $n-M$ columns are orthonormal and perpendicular to the M columns of T, $U'_{(n-M)\times n}\, T = 0_{(n-M)\times M}$. We obtained U as follows. Let

$$I - T(T'T)^{-1}\,T' = \tilde{U}\,\Delta\,\tilde{U}' . \qquad (18)$$

where $\tilde{U}$ is orthogonal and Δ is diagonal. Since $I-T(T'T)^{-1}T'$ is a projection matrix of rank $n-M$, Δ is a matrix with M zeroes and $n-M$ ones on the diagonal. We used EISPACK (Smith et.al. (1975)) to perform the eigenvalue decomposition $\tilde{U}\,\Delta\,\tilde{U}'$ and the $n-M$ columns of U are taken as the columns of $\tilde{U}$ corresponding to the $n-M$ ones in Δ. Each such vector is perpendicular to the columns of T,

as can be seen by right multiplying (18) by T. The EISPACK computation of the entries of Δ was good to seven figures. Given U, B is computed and Γ and δ are also computed using the eigenvalue decomposition routines in EISPACK.

5. Numerical Results

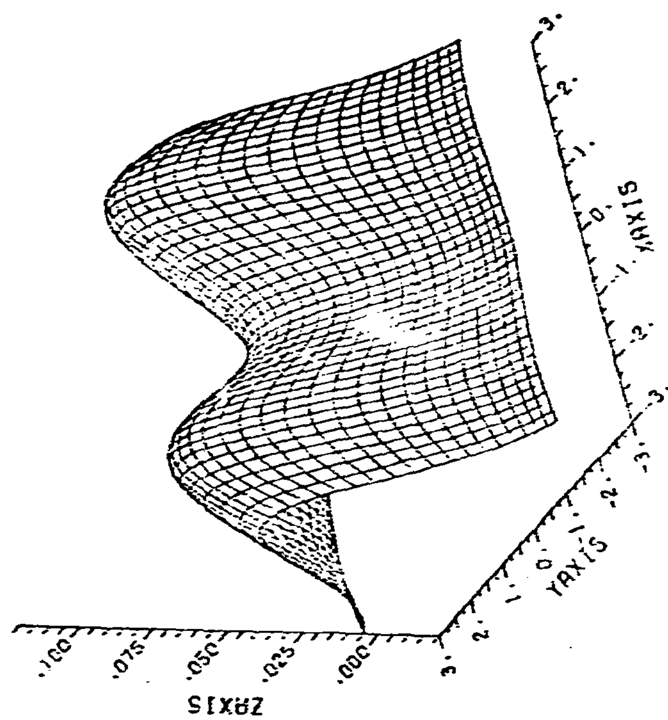
We present the results of a single Monte Carlo experiment, with $m=2$. Figure 1 gives a picture of the true function u that was the subject of the first experiment,

$$u(x,y) = \frac{1}{2\pi(1.3)^2}\left[e^{-\frac{1}{2(1.3)^2}((x-2)^2+y^2)} + e^{-\frac{1}{2(1.3)^2}((x+2)^2+y^2)}\right]$$

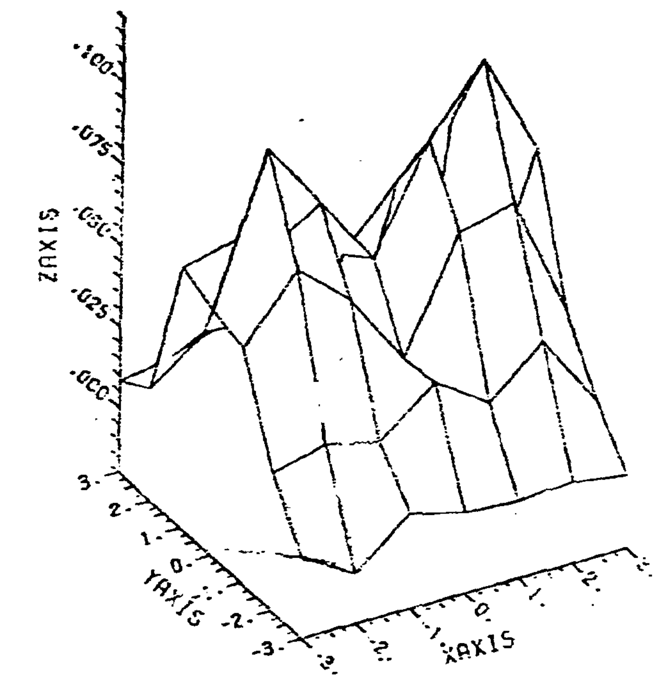
A regular 7×7 square array of 49 points t_i, $i = 1,2,\ldots,49$ was selected, with the middle point being (0,0) and the point spacing being 1.0. Data y_i were generated as

$$y_i = u(t_i) + \epsilon_i , \quad t_i = (x_i,y_i) , \quad i = 1,2,\ldots,49$$

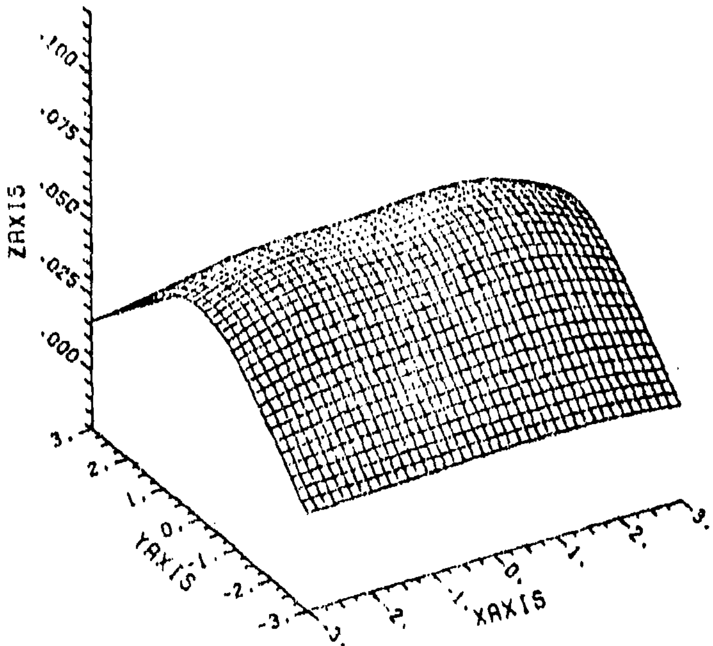
where the ϵ_i were pseudorandom $N(0,\sigma^2)$ random variables with $\sigma = .01$. σ is about 1/8 of the maximum height of u. Figure 2 presents a picture of the data points, which have been joined by straight lines. Figures 3 and 4 give $u_{n,2,\lambda}$ for two values of λ, in Figure 3, λ is too large, and in Figure 4, λ is too small. Figure 5 gives $u_{n,2,\hat{\lambda}}$, where $\hat{\lambda}$ is the minimizer of $V(\lambda)$. Figure 6 gives a plot of $R(\lambda)$ and $V(\lambda)$ against $\log\lambda$. It is seen that, in the neighborhood of the minimizer of $R(\lambda)$, $V(\lambda)$ roughly follows $R(\lambda)$. Theoretically, we have $\min_\lambda EV(\lambda) \simeq \min_\lambda ER(\lambda) + \sigma^2$, for large n, see CW, and this relationship is roughly approximated here. The achieved inefficiency, defined by $R(\hat{\lambda})/\min_\lambda R(\lambda)$, where $\hat{\lambda}$ is the minimizer of $V(\lambda)$, was 1.54. Note that $\min_\lambda R(\lambda) = .25\,\sigma^2$. If we were fitting a surface which is known to be a linear combination of given functions by regression we would expect the mean square error to be proportional to $\frac{\sigma^2}{n}$. Here numerical and theoretical results in the one



THE ACTUAL SURFACE
Figure 1

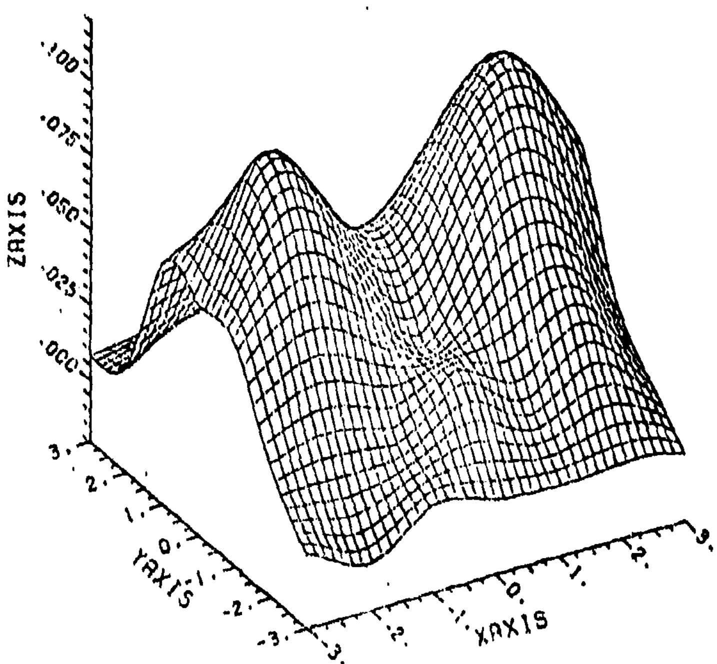


THE DATA
Figure 2



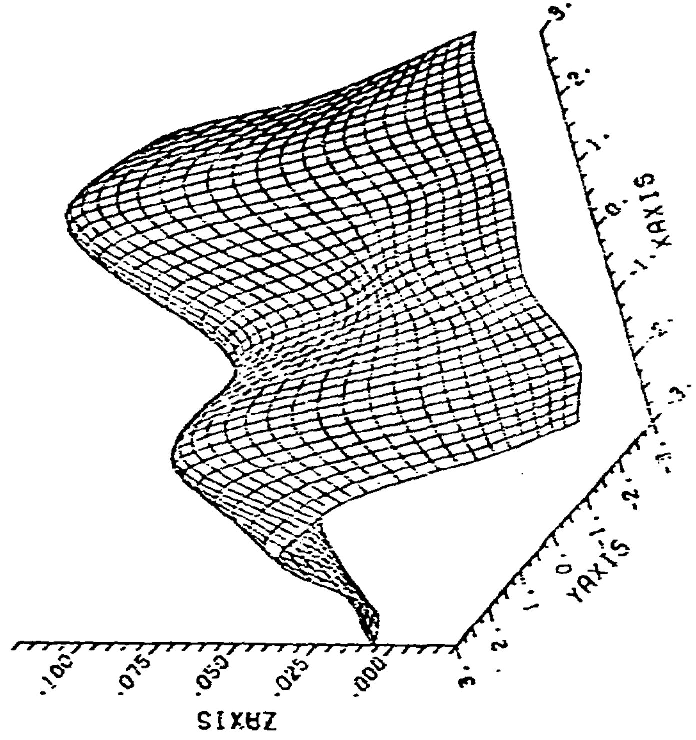
THE FITTED SURFACE $\lambda = 100\ \hat{\lambda}$.

Figure 3



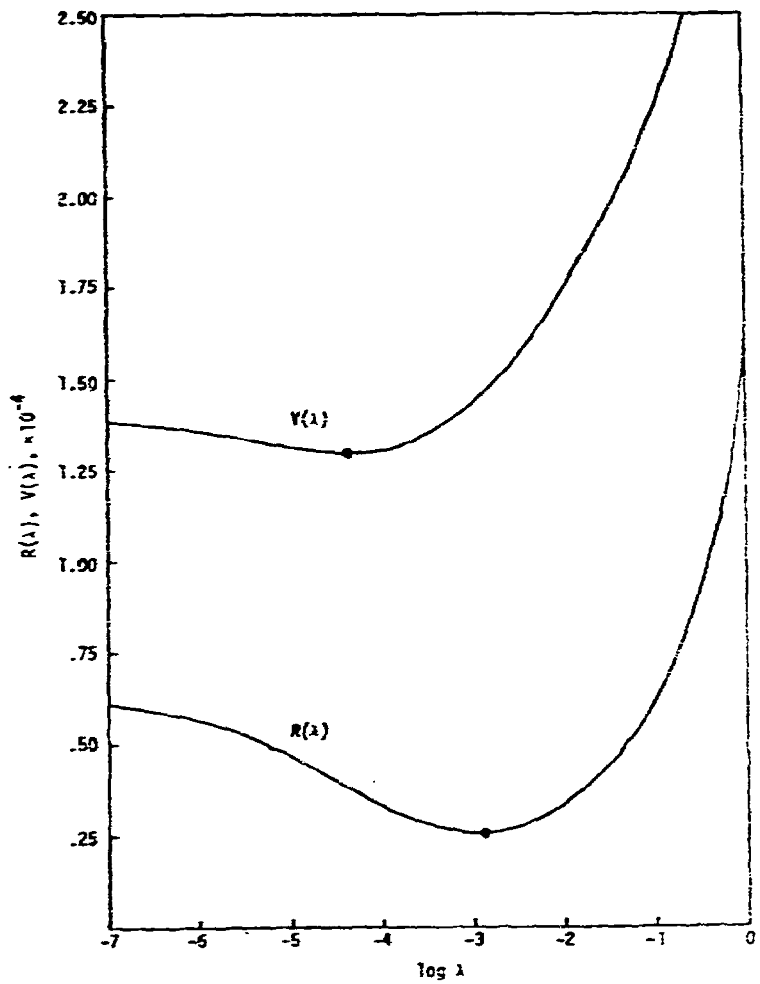
THE FITTED SURFACE $\lambda = .01\ \hat{\lambda}$.

Figure 4



THE FITTED SURFACE $\lambda = \hat{\lambda}$.

Figure 5



THE MEAN SQUARE ERROR $R(\lambda)$, AND THE CROSS-VALIDATION FUNCTION $V(\lambda)$.

Figure 6

dimensional case for reasonably regular arrangements of data points indicate that $\min_\lambda R(\lambda) \simeq \text{const.}(\sigma^2/n)^p$ where p is some power slightly less than one. p depends on the rate of decay of eigenvalues of an appropriate reproducing kernel. See Wahba and Wold (1975), CW, and Wahba (1975b, 1977). If $u \in H^{2m}(\mathbb{R}^2)$, $p = 2m/(2m+1)$. (In preparation).

Mr. Wendelberger's program is running for $n = 120$ and quite reasonable results have been obtained for this case, with randomly chosen points $\{t_i\}$. One cannot increase n with impunity, however. In the $n = 49$ case reported here the condition number of B, namely $\max_i b_i/\min_i b_i$ was around 200, and in the irregularly spaced $n = 120$ case this condition number was of the order of 4×10^5. (Irregularly spaced points will increase the condition number.) For large n and a condition number somewhere around (we guess) 10^6 or 10^7, the computation errors will begin to take over. Thus, in theory, a plot of $\log(\min R(\lambda))$ vs. n should be approximately linear with slope $-p$, however, as roundoff error gets large, this plot will flatten out. Laurent and colleagues (1978) have developed a procedure for patching together surfaces of this type so that groups of points may be handled separately.

The cost of running a program designed just to produce Figure 5 from data, we estimate to be about $4.00 at weekend rates at our computing center. To produce Figure 5 from a second set of data at the <u>same</u> points $\{t_i\}$, one would retain U, F and $\{b_i\}$, which depend on the $\{t_i\}$ but not $\underline{z}$, and then the cost would be very small.

6. Miscellaneous Remarks

We hope to implement the $m = 3$ and $m = 4$ cases. m can then be selected from the data by comparing $V(\hat\lambda)$ for each of the $m = 2, 3$ and 4 cases. For $m = 2$, the roughness penalty

$$\sum_{j=0}^{2} \iint \left(\frac{\partial^2 u}{\partial x^j \partial y^{2-j}}\right)^2 dxdy$$

is the bending energy of a thin plate. For this reason, Duchon christened the solutions "plaque mince" splines. We have reason to believe that the $m = 3$ case will be appropriate for the smoothing of certain meteorological data. In some cases the nature of the physical phenomena being smoothed may provide insight into a choice of m.

We note that the solutions $u_{n,m,\lambda}$ satisfy

$$\Delta^m u_{n,m,\lambda}\big|_t = 0\,, \qquad t \neq t_1,\dots,t_n\,,$$

where Δ is the Laplacian operator $\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$. The smoothing splines $f_{n,m,\lambda}$ satisfy $f^{(2m)}_{n,m,\lambda}(t) = 0$, $t \neq t_1,\dots,t_n$. For this reason, Prof. Iso Schoenberg has suggested to us that the functions $u_{n,m,\lambda}$ be called "Laplacian Smoothing Splines".

We have recently obtained what might be called the Laplacian histosplines, by analogy with Boneva, Kendall, and Stefanov (1971). These are functions which minimize the roughness penalty $\sum_{j=0}^{m} \iint_A \binom{m}{j} \left(\frac{\partial^m u}{\partial x^j \partial y^{m-j}}\right)^2 dxdy$ subject to volume matching conditions of the form

$$\iint_{A_i} u\, dxdy = p_i\,, \qquad i = 1,2,\dots,n$$

where the A_i are n bounded areas in $\mathbb{R}^n$ whose union is A. These functions satisfy

$$\Delta^m u = \text{constant on each } A_i\,.$$

(Dyn, Wong, and Wahba (1979)), and, in preparation).

Various optimality properties of smoothing splines and histosplines in one dimension are known. For example, it can be shown from CW and Wahba (1975b) that

$$E\int_0^1 (f_{n,m,\hat\lambda}(t)-f(t))^2 dt = O(n^{-(2m)/(2m+1)}),\ f \in H^m(0,1)$$
$$= O(n^{-(4m)/(4m+1)}),\ f \in H^{2m}(0,1)$$

and f satisfies some boundary conditions. It is part of the folklore that these rates cannot be improved upon. Density estimates determined by the minimizer of $\int_0^1 (f^{(m)}(t))^2 dt$ subject to the area-matching conditions

$$\int_{t_i}^{t_{i+1}} f(t)dt = \text{fraction of observations in } [t_i, t_{i+1}]$$

are known to achieve the best possible convergence rates over $f \in H^m$ provided the t_i are chosen properly. See Wahba (1975c, 1976). Stone (1978) has given some results on best possible pointwise convergence rates in d dimensions. We conjecture that all the nice convergence properties of polynomial splines can be extended to the Laplacian smoothing splines and Laplacian histosplines.

APPENDIX

Outline of Proof of Theorem

Let $r_1, r_2, \ldots, r_M$ be a subset of M points selected from $t_1, \ldots, t_n$ with the property that the M×M matrix T with $j\nu^{th}$ entry $\phi_\nu(r_j)$ is of full rank. The space $H = \{u: \; u \in D', \frac{\partial^m u}{\partial x^j \partial y^{m-j}} \in L_2, \; j = 0,1,\ldots,m-1\}$ can be decomposed into the direct sum of two spaces:

$$H = \pi_{m-1} \oplus \underline{X} \; .$$

where π_{m-1} is the M dimensional space of polynomials of total degree m-1 or less and $\underline{X} = \{u: \; u \in H, \; u(r_\nu) = 0, \; \nu = 1,2,\ldots,M\}$. It can then be shown that

$$\langle u,v \rangle_{\underline{X}} \equiv \iint \sum_{j=0}^{m} \binom{m}{j} \frac{\partial^m u}{\partial x^j \partial y^{m-j}} \; \frac{\partial^m v}{\partial x^j \partial y^{m-j}} \, dxdy$$

defines an inner product on $\underline{X}$. If an inner product is defined on π_{m-1} by $\langle u,v \rangle_{\pi_{m-1}} = \sum_{\nu=1}^{M} u(r_\nu)v(r_\nu)$, then π_{m-1} and $\underline{X}$ are orthogonal subspaces. $\underline{X}$ (and π_{m-1}, and hence H) are reproducing kernel spaces.

If the reproducing kernel K(s,t) for $\underline{X}$ can be found, then the solution $u_{n,m,\lambda}$ to the minimization problem of (4) will have a representation

$$u_{n,m,\lambda}(t) = \sum_{j=1}^{n} \bar{c}_j K(t,t_j) + \sum_{\nu=1}^{M} \bar{d}_\nu \phi_\nu(t) \; . \qquad (A.1)$$

(See, e.g. Kimeldorf and Wahba (1971)). $u_{n,m,\lambda}$, will, of course, be independent of the choice of $r_1, \ldots, r_M$. The reproducing kernel K has been found by Meinguet (1978, 1979) and is given by

$$K(s,t) = E_m(s,t) - \sum_{\mu=1}^{M} p_\mu(s)\, E_m(t,r_\mu)$$
$$- \sum_{\nu=1}^{M} p_\nu(t)\, E_m(s,r_\nu)$$
$$+ \sum_{\mu,\nu=1}^{M} p_\mu(s)\, p_\nu(t)\, E_m(r_\mu,r_\nu) \ , \qquad \text{(A.2)}$$

where $\{p_\mu\}_{\mu=1}^{M}$ span π_{m-1} and are chosen so that $p_\mu(r_\nu) = 1$, $\mu = \nu$, $= 0$, $\mu \neq \nu$. Substituting (A.2) into (A.1), it is seen that a representation of the form (5) for $u_{n,m,\lambda}$ holds.

To show that K is the reproducing kernel for $\underline{\overline{X}}$, it is necessary to show that

i) $K(s,\cdot) \in \underline{\overline{X}}$, each s

ii) $\langle K(s,\cdot), K(t,\cdot)\rangle_{\underline{\overline{X}}} = K(s,t)$, $\qquad$ (A.3)

where

$$\langle u,v\rangle_{\underline{\overline{X}}} = \sum_{j=1}^{M} \iint \binom{m}{j} \frac{\partial^m u}{\partial x^j \partial y^{m-j}} \frac{\partial^m v}{\partial x^j \partial y^{m-j}}\, dxdy \ . \qquad \text{(A.4)}$$

Define

$$H_s(t) = E_m(s,t) - \sum_{\mu=1}^{M} p_\mu(s)\, E_m(r_\mu,t) \ .$$

Then

$$K(s,t) = H_s(t) - \sum_{\nu=1}^{M} p_\nu(t)\, H_s(r_\nu) \ . \qquad \text{(A.5)}$$

The hard part is to show that $H_s \in H$. (Note that $E_m \notin H$.) Meinguet shows that $H_s \in H$, for each s, and we omit the proof. It then follows that $K(s,\cdot) \in H$, and, since $\sum_{\nu=1}^{M} p_\nu(\cdot)\, H_s(r_\nu)$ is the polynomial interpolating to H_s at $r_1,\ldots,r_M$, $K(s,r_\nu) = 0$, $\nu = 0,1,\ldots,M$, and so $K(s,\cdot) \in \underline{\overline{X}}$.

To establish (A.3), first note that

$$\frac{\partial^m}{\partial x^j \partial y^{m-j}} K(s,\cdot) = \frac{\partial^m}{\partial x^j \partial y^{m-j}} H_s(\cdot) \ . \qquad \text{(A.6)}$$

Consider the Green's formula

$$(-)^m \sum_{j=0}^{m} \binom{m}{j} \iint \frac{\partial^m u}{\partial x^j \partial y^{m-j}} \frac{\partial^m v}{\partial x^j \partial y^{m-j}}\, dxdy = \iint \Delta^m u \cdot v\, dxdy \qquad \text{(A.7)}$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. This formula holds provided, e.g. $v \in H \cap L_2$ and $u \in \mathcal{D}$.

If $u \in \mathcal{D}$, then the potential formula

$$\iint(\Delta^m u)(t)\, E_m(s,t)dt = u(s)$$

holds (see Schwartz (1966)) and in particular

$$\iint \Delta^m\, u \cdot H_s = u(s) - \sum_{\nu}^{M} p_\nu(s)\, u(r_\nu) \ . \qquad \text{(A.8)}$$

Meinguet argues that, in fact (A.7) and (A.8) hold for $u = H_t$, $v = H_s$, giving

$$(-1)^m \sum_{j=0}^{m} \binom{m}{j} \iint \frac{\partial^m}{\partial x^j \partial y^{m-j}} H_t \frac{\partial^m}{\partial x^j \partial y^{m-j}} H_s$$
$$= H_t(s) - \sum_{\nu=1}^{M} p_\nu(s)\, H_s(r_\nu) \equiv K(s,t) \ ,$$

which, combined with (A.6), gives (A.3).

Equation (7) can be obtained as follows: Considering $K(t,t_j)$ as a function of t,

$$K(t,t_j) = E_m(t,t_j) - \sum_{\nu=1}^{M} p_\nu(t_j)\, E_m(t,r_\nu) + \text{a polynomial of degree } m-1 \text{ or less.} \qquad \text{(A.9)}$$

Now, if ϕ is any element of π_{m-1}, we have

$$\phi(t) - \sum_{\nu=1}^{M} p_\nu(t)\, \phi(r_\nu) \equiv 0 \ . \qquad \text{(A.10)}$$

Letting $\alpha_1(j), \alpha_2(j), \ldots, \alpha_n(j)$ be the coefficients of $E_m(\cdot,t_1)$, $E_m(\cdot,t_2)$, ..., $E_m(\cdot,t_n)$, in (A.9), it can be verified from (A.10) that

$$\sum_{k=1}^{n} \alpha_k(j)\phi(t_k) \equiv 0, \quad j = 1,2,\ldots,m ,$$

which results directly in the conditions (7) on the coefficient vector $\underset{\sim}{c}$ in (5), namely, $T\underset{\sim}{c} = 0$. Equation (6) is obtained as follows: One substitutes (A.1) into (4), and then uses (A.3) to evaluate the expression (4) to be minimized. By repeatedly using $T'\underset{\sim}{c} = 0$, one obtains that $\underset{\sim}{c}$ and $\underset{\sim}{d}$ are chosen subject to $T'\underset{\sim}{c} = 0$, to minimize

$$||\underset{\sim}{z}-K\underset{\sim}{c}-T\underset{\sim}{d}||^2 + n\lambda \underset{\sim}{c}' K \underset{\sim}{c} .$$

Differentiating this expression with respect to $\underset{\sim}{c}$ and setting the result equal to zero, and using $T'\underset{\sim}{c} = 0$, gives (6).

REFERENCES

Athavale, M.L., and Wahba, G., (1978). Determination of an optimal mesh for a collocation - projection method for solving two point boundary value problems, University of Wisconsin-Madison, Statistics Department Technical Report #530, to appear, J. Approx. Theory.

Benedetti, J., (1977). On the nonparametric estimation of regression functions, J. Roy. Stat. Soc. B., 39, 2, 248-253.

Boneva, L., Kendall, D., and Stefanov, I., (1971). Spline transformations: Three new diagnostic aids for the statistical data-analyst (with Discussion), J. Roy. Stat. Soc. B., 33, 1-70.

Craven, P., and Wahba, G., (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math. 31, 377-403.

Duchon, Jean, (1976a). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. R.A.I.R.O. Analyse numerique, 10, 12, pp.5-12.

Duchon, Jean, (1976b). Fonctions spline du type "Plaque mince" en dimension 2. No. 231. Seminaire d'analyse numerique, Mathematiques Appliqueés, Universite Scientifique et Medicale de Grenoble.

Dyn, N., Wahba, G., and Wong, W.H., (1979). To be submitted to J. Am. Stat. Soc. as comments on a paper by Tobler.

Fleisher, J., (1979). Spline smoothing routines, Reference Manual for the 1110. Academic Computing Center, The University of Wisconsin-Madison.

Golub, G., Heath, M., and Wahba, G., (1977). Generalized cross-validation as a method for choosing a good ridge parameter, University of Wisconsin-Madison, Department of Statistics Technical Report #491, to appear, Technometrics.

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

Laurent, P.J., (1978). Seminar presented at the University of Wisconsin-Madison, Statistics Department.

Lucas, H., (1978). Estimation of smoothing parameters to smooth noisy data and confidence regions for the underlying function, Ph.D. Thesis, University of Wisconsin-Madison, Department of Statistics.

Meinguet, Jean, (1978). Multivariate interpolation at arbitrary points made simple. Report No. 118, Institute de Mathematique Pure et Appliqueé, Université Catholique de Louvain, to appear, ZAMP.

Meinguet, Jean, (1979). An intrinsic approach to multivariate spline interpolation at arbitrary points. To appear in the Proceedings of the NATO Advanced Study Institute on Polynomial and Spline Approximation, Calgary 1978, B. Sahney, Ed.

Merz, P., (1978a). Spline smoothing by generalized cross-validation, a technique for data smoothing. Chevron Research Company, Richmond, CA.

Merz, P.M., (1978b). Determination of absorption energy distribution by regularization and a characterization of certain absorption isotherms, Chevron Research Company, Richmond, CA.

Paihua Montes, Luis, (1978). Quelques methodes numeriques pour le calcul de fonctions splines a une et plusiers variables. Thesis, Analyse Numerique Université Scientifique et Medicale de Grenoble.

Schwartz, Laurent (1966). Theorie des distributions. Hermann, Paris.

Smith, B.T., et al, (1976). Matrix Eigensystem Routines - EISPACK Guide. Springer-Verlag.

Stone, C.J., (1978). Optimal rates of convergence for nonparametric estimates, Manuscript, Department of Mathematics, University of California, Los Angeles.

Stutzle, Werner, (1977). Estimation and Parameterization of Growth Curves. Thesis, Swiss Federal Institute of Technology Zurich. Department of Mathematics.

Utreras Diaz, F., (1978a). Sur le choix du parametre d'ajustement dans le lissage par fonctions spline, No. 296, Seminaire d'analyse numerique, Mathematiques Appliqueés, Université Scientifique et Medicale de Grenoble.

Utreras Dias, F., (1978b). Utilisation des programmes de calcul du parametre d'ajustement dans le lissage par fonctions spline. R.R. No. 121, Matheratiqueés Appliqueés, Université Scientifique et Medicale de Grenoble.

Wahba, G., (1971). On the regression design problem of Sacks and Ylvisaker, Ann. Math. Statist., 42, 3, 1035-1053.

Wahba, G., (1974). Regression design for some equivalence classes of kernels, Ann. Statist., 2, 5, 924-934.

Wahba, G., (1975a). A canonical form for the problem of estimating smooth surfaces, University of Wisconsin-Madison, Department of Statistics Technical Report #420.

Wahba, G., (1975b). Smoothing noisy data with spline functions, Numer. Math., 24, 383-393.

Wahba, G., (1975c). Interpolating spline methods for density estimation. I. Equispaced knots. Ann. Statist. 3, 30-48.

Wahba, G., (1976). On the optimal choice of nodes in the collocation-projection method for solving linear operator equations, J. Approx. Theory, 16, 2, 175-186.

Wahba, G., (1976). Histosplines with knots which are order statistics, J. Roy. Stat. Soc., Series B., 38, 2, 140-151.

Wahba, G., (1977). Discussion to "Consistent nonparametric regression", by Charles J. Stone, Ann. Statist. 5, 637-640.

Wahba, G., (1978a). Improper priors, spline smoothing and the problem of guarding against model errors in regression, University of Wisconsin-Madison, Department of Statistics Technical Report #508, to appear, J. Roy. Stat. Soc. B.

Wahba, G., (1978b). Discussion to curve fitting and optimal design for prediction, A. O'Hagan, J. Roy. Stat. Soc. B., 40, 1, 1-42.

Wahba, G., (1978c). Interpolating surfaces: High order convergence rates and their associated designs, with applications to X-ray image reconstruction, University of Wisconsin-Madison, Department of Statistics Technical Report #523, to appear, SIAM J. Numer. Anal.

Wahba, G., and Wold, S., (1975). A completely automatic French curve: Fitting spline functions by cross-validation, Communications in Statistics, 4, 1, 1-17.

Welch, J. (1979). Ph.D. dissertation, Dep't. of Chemical Engineering, University of Wisconsin, Madison, in preparation.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER TR#555	2. GOVT ACCESSION NO. ~~3. RECIPIENT'S CATALOG NUMBER~~ (9) Technical rept.
4. TITLE (and Subtitle) (6) HOW TO SMOOTH CURVES AND SURFACES WITH SPLINES AND CROSS-VALIDATION.	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER (15)
7. AUTHOR(s) (10) Grace Wahba	8. CONTRACT OR GRANT NUMBER(s) ONR N00014-77-C-0675, ARMY DAAG29-77-G-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (∅)
11. CONTROLLING OFFICE NAME AND ADDRESS U.S.Army Research Off., P.O. Box 12211, Res. Triangle Pk., N.C. 27709; Off. of Naval Res., 800 N. Quincy Street, Arlington, VA 22217 (11)	12. REPORT DATE March 1979
	13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 15p.	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited

(14) UWIS-DS-555

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

DDC RECEIVED MAY 8 1979 C

18. SUPPLEMENTARY NOTES

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

splines, smoothing, estimation of smooth surfaces, multidimensional splines, Laplacian splines

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) ~~We briefly review~~ the use of smoothing splines and the method of generalized cross validation (GCV) for smoothing discrete noisy data from an unknown but smooth curve is reviewed. ~~Then we describe~~ the use of "plaque mince" or Laplacian smoothing splines with GCV for smoothing discrete noisy data from an unknown but smooth surface is described. A numerical algorithm for this (nontrivial!) computational problem is described, and an example from a Monte Carlo study is presented to show how the method works on simulated data. The results are extremely promising. Some design problems are briefly mentioned. Some conjectures are made concerning optimality properties of Laplacian smoothing splines and Laplacian histosplines.

400 243 Gu

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)