In "Inverse and Ill-Posed Problems" M. Engl and G. Groetsch, Eds. Academic Press 1987 pp 37-50

THREE TOPICS IN ILL-POSED PROBLEMS

Grace Wahba

Department of Statistics
University of Wisconsin-Madison
Madison, Wisconsin

1. INTRODUCTION

In this paper we discuss three topics in ill posed problems. The first topic concerns the imposition of specified types of discontinuities on otherwise smooth two and three dimensional solutions of ill posed problems. The work has application to the recovery of the three dimensional atmospheric temperature distribution from satellite observed radiances and is joint work with Jyh-Jen Shiau of the University of Columbia-Missouri Computer Sciences Department and Donald R. Johnson of the University of Wisconsin-Madison Space Sciences and Engineering Center. (Shiau, Wahba, and Johnson (Dec. 1985)). The main new idea appears in Shiau's thesis (Shiau (June, 1985)). The work also relies upon the penalized likelihood approach to Tihonov regularization in Finbarr O'Sullivan's thesis (O'Sullivan (1983)) and O'Sullivan and Wahba (1985), which generalizes the point of view in Kimeldorf and Wahba (1971) and Wahba (1980b). A closely related reference is Svensson (February, 1985). The method proposed here for imposing specified discontinuities on the solution given discrete, noisy values on functionals of the solution can be expected to have application to a number of problems in the atmospheric and geological sciences, for example to the estimation of the three dimensional temperature distribution of the ocean, including the thermocline, and to the estimation of certain geological properties, e. g. density, transmittivity, etc.

The second topic concerns the Richardson/Landweber/Fridman/Picard/ Cimino iterative method(s) for iterative solutions of large linear systems. It has been observed by Prof. Brakhage at this conference and by other workers, that, when using certain iterative methods for solving large ill conditioned linear systems, the approximation to the solution appears to improve with iteration up to a certain point, when further iteration begins to

[1] *This research supported by the ONR under Grant No. N00014-77-C-0675 and by NASA under Grant No. NAG5-316.*

ISBN 0-12-239040-7

degrade the solution. See Strand (1974), Anderssen and Prenter (1981), Natterer (1986). A rough explanation for this phenomena is as follows: The first few iterations are recovering the projection of the solution onto singular vectors corresponding to large singular values. As the iteration proceeds, recovery of projections on singular vectors with small singular values begins, but recovery of these (high frequency) components is increasingly sensitive to noise in the data, either measurement error, or roundoff, as the singular values become smaller. A quantitative analysis of this phenomena appears in a report Wahba (1980b) prepared for the proceedings of the 1979 Delaware Conference, where it is observed that generalized cross validation (GCV) can be used to provide a data-based rule for choosing the point at which to stop the iteration, when the source of noise is random measurement error on the right hand side. Since those proceedings have not been published and there appears to be some interest in the topic at the present time, we review that result here.

The third topic concerns the parameter estimation problem for p. d. e.'s with particular application to the resevoir modelling problem. One has the equation (for example)

$$\frac{\partial u}{\partial t} = \sum_{l=1}^{3} \frac{\partial}{\partial x_l}(\alpha(x)\frac{\partial u}{\partial x_l}) + q(x,t), \quad x=(x_1,x_2,x_3)\varepsilon\Omega,\ t\varepsilon[0,T] \tag{1.1}$$

where the forcing function q is known exactly and u is measured with noise at a finite set of points in x and t. The problem is to estimate the transmittivity α, which is not necessarily a constant. Recently Kravaris and Seinfeld (1985) proposed a penalized likelihood approach to the estimation of α. This problem is both nonlinear and in general ill posed. The degree of ill posedness (see Wahba (1980b) for this concept) is a particularly tricky question since it can be very sensitive to the "true" u. We make a few conjectural remarks on the possiblity of solving the penalized likelihood equations subject to linear inequality constraints based on physical information (see Ewing, these proceedings) and using the GCV to choose the smoothing parameter in the resulting constrained implicit optimization problem.

Before beginning the substantive part of this paper we make a few remarks concerning ill posed experimental problems.

(1) An "ill posed experimental problem" is one for which there is not as much information in your experimental data as you really need to find out what you want to know.

(2) Good data analysis methods for ill posed experimental problems:

a) squeeze as much information as possible out of the data at hand

b) recognize the limitations of the available data

c) introduce whatever (valid) outside information is available, of diverse types. For example, prior belief concerning "smoothness", which leads to statistical and regularization methods (they may be the same, see Kimeldorf and Wahba (1971)), physical constraints, imposed both exactly and approximately, data from other experiments, qualitative information such as the location of discontinuities, etc.

The penalized likelikhood (also known as cross validated spline) methodology that is common to much of our work reads as follows: The model for the data is:

$$y_i = N_i(g) + \varepsilon_i, \qquad i=1,...,n, \tag{1.2}$$

where the ε_i are independent, identically distributed Gaussian random variables. g is assumed to be in some appropriately chosen Hilbert space H (a reproducing kernel space, if estimates of point values of g are going to be made), and the N_i are bounded linear or mildly nonlinear functionals on H. We will not here be considering modelling errors (errors in the specification of N_i), or other models for the ε_i, both of which can be important. Side information of the form $g \varepsilon C$, where C is a closed convex set in H (for example, $g(x) \geq 0$) may be imposed. The estimate g_λ for g is then the minimizer in H of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - N_i(g))^2 + \lambda J(g) \tag{1.3}$$

subject to $g \varepsilon C$, where $J(.)$ is a quadratic penalty functional.

II. PARTIAL SPLINE METHODS FOR INCLUDING DISCONTINUITIES IN OTHERWISE SMOOTH REGULARIZED SOLUTIONS OF ILL POSED PROBLEMS WITH NOISY DATA

The general abstract idea is as follows: It is desired to solve a variational problem in a space of functions which are smooth except for a specified discontinuity. One does this by solving the problem in a Hilbert space H which is constructed as the direct sum of a reproducing kernel Hilbert space of smooth functions (typically a Sobolev space) and a finite dimensional space of functions which possess the specified discontinuity. The abstract mathematical foundation is based on some geometry in reproducing kernel spaces (for proofs of the lemmas we use, see Kimeldorf and Wahba (1971).

Let $H = H_1 \oplus H_0$ where H_0 is of dimension $M+p$ and is spanned by $\Phi_1, \ldots, \Phi_M; \Psi_1, \ldots, \Psi_p$, the reason for the distinction between the Φ's and the Ψ's will become evident shortly. For the moment suppose that N_i is a bounded linear functional on H, thus it possess a representer η_i with $N_i(g) = \langle \eta_i, g \rangle$. Given

$$y_i = \langle \eta_i, g \rangle + \varepsilon_i, \quad i=1,\ldots,n, \tag{2.1}$$

we can find g in H to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \langle \eta_i, g \rangle)^2 + \lambda \| P_1 g \|, \tag{2.2}$$

where P_1 is the orthogonal projection onto H_1. Let T be the $n \times (M+P)$ "design matrix", for least squares regression on H_0, that is,

$$T = \begin{bmatrix} \langle \eta_1, \Phi_1 \rangle & \cdots & \langle \eta_1, \Phi_M \rangle & \langle \eta_1, \Psi_1 \rangle & \cdots & \langle \eta_1, \Psi_p \rangle \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \langle \eta_n, \Phi_1 \rangle & \cdots & \langle \eta_n, \Phi_M \rangle & \langle \eta_n, \Psi_1 \rangle & \cdots & \langle \eta_n, \Psi_p \rangle \end{bmatrix} \tag{2.3}$$

If T is of rank $M+p$ then the minimizer g_λ of

$$\frac{1}{n}\sum(y_i - \langle \eta_i, g \rangle)^2 + \lambda \| P_1 g \|^2$$

exists uniquely for every $\lambda > 0$ and

$$g_\lambda = \sum_{i=1}^{n} c_i (P_1 \eta_i) + \sum_{\nu=1}^{M} d_\nu \Phi_\nu + \sum_{j=1}^{p} \theta_j \Psi_j. \tag{2.4}$$

If ξ_i is the representer of N_i in H_1 then $P_1 \eta_i = P_1 \xi_i$ so it is not necessary to know (or define) η_i . In theory, one substitutes (2.4) into (2.2) and solves for $c = (c_1,\ldots,c_n)'$, $d = (d_1,\ldots,d_M)'$, and $\theta = (\theta_1, \ldots, \theta_p)'$.

A simple one dimensional example is

$$H = H_1 \oplus span\{\Phi_1, \ldots, \Phi_m\} \oplus span\{\Psi_1, \ldots, \Psi_p\}$$

where

$$H_1 \oplus span\{\Phi_1, \ldots, \Phi_m\}$$

is the Sobolev space $W_2^m[0,1]$ of functions with absolutely continuous $m-1$st derivatives and square integrable mth derivative with seminorm $J_m(f)$ given by

$$J_m(f) = \int_0^1 (f^{(m)}(x))^2 dx. \tag{2.5}$$

The $\Phi_1, \ldots, \Phi_m$ are a basis for the null space of J_m in W_2^m, namely, the polynomials of degree less than m, and $H_1 = N(J_m)^\perp$ in W_2^m. Suppose it is desired to allow $g \varepsilon H$ to have a jump in the first derivative at x^*. Then we let $p=1$ and $\Psi_1(x) = |x - x^*|$. Then, if $f \varepsilon W_2^m$ with $m \geq 2$ and $g = f + \theta \Psi_1$,

$$\left.\frac{\partial g}{\partial x}\right]_{x=x^*_+} - \left.\frac{\partial g}{\partial x}\right]_{x=x^*_-} = 2\theta.$$

More generally, to include a jump in the kth derivative, we can let the Ψ_q's be of the form $\Psi_q = \dfrac{(x-x^*_q)_+^k}{k!}$, for $k<m$. However, unless a very high quality data set is available, it will probably be difficult to estimate jumps in derivatives beyond the first.

Functions of the form of g_λ of (2.4) are known as partial splines since, in the case the N_i are evaluation functionals then $\sum c_i(P_1\eta_i) + \sum d_\nu \Phi_\nu$ is a spline, irrespective of the character of the Ψ_j's. For a more general discussion of partial splines, see Wahba (1984)

We now briefly describe the problem of estimating the three dimensional atmospheric temperature distribution from satellite-observed radiances, which was the motivation for some of this work.

A bank of narrow field-of view radiometers sits on a polar orbiting satellite looking down and scanning from side to side. The observed radiances $y_{i,\nu}$ from the νth channnel looking along the ith line of sight can be modelled as

$$y_{i,\nu} = \int_{line\ of\ sight\ i} K_\nu(g(x), x)dx + \varepsilon_{i,\nu} = N_{i,\nu}(g) + \varepsilon_{i,\nu}. \tag{2.6}$$

where $g(x)$ is the temperature at point x (= latitude, longitude, height). The K_ν are mildly nonlinear functions of g, further details may be found in Fritz et al. (1972), Susskind et al. (1984). The problem is to estimate g, given observations from the different channels along many lines of sight. The atmospheric temperature distribution tends to be a smooth function of location except at the tropopause (and at fronts or inversions) where a sharp minimum occurs. The solid line in Figure 1 is a hypothetical vertical temperature profile, which resembles a typical vertical temperature profile as observed by a radiosonde (weather balloon) over Green Bay, Wisconsin, in the winter. In acordance with the custom among meteorologists the figure is tipped on its side. A sharp minimum at the tropopause is clearly visible. The existence of a minimum is more or less typical of vertical temperture profiles, although the vertical coordinate of the minimum tends to vary slowly with latitude and longitude. Figure 2 gives a hypothetical two dimensional plot of the

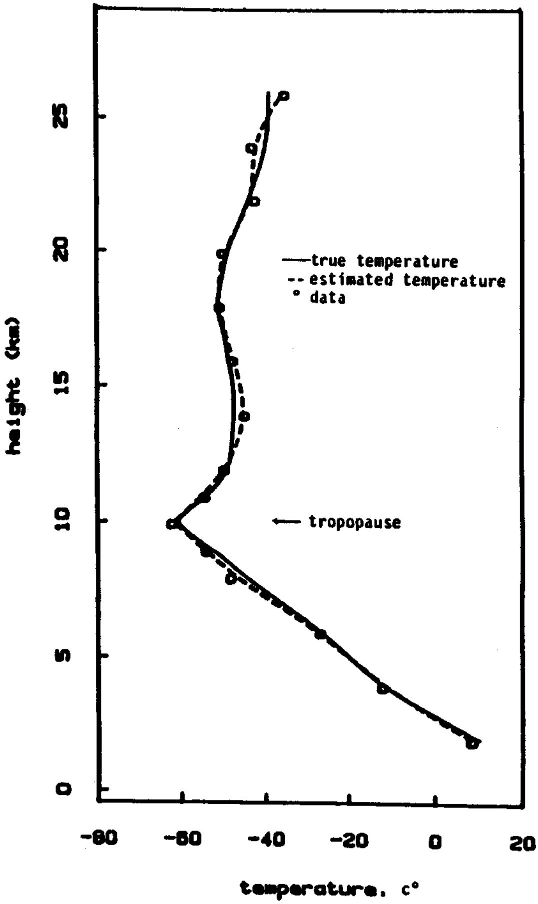

Fig. 1. Hypothetical vertical temperature profile.

location of the tropopause $x_2^*(x_1)$ along a circle of constant latitude, here x_1 is longitude and x_2^* is tropopause height. Information external to the radiance data $y_{i,\nu}$ of (2.6) may be available concerning the location of the tropopause, so that a curve like that in Figure 2 or a three dimensional analogue $x_2^*(lat., long.)$ of it may may be constructed. It is desired to include this information concerning the tropopause location in the estimate of g given the $y_{i,\nu}$.

To carry out the partial spline approach in two or three dimensions in practice we need to choose an appropriate penalty functional and Ψ's which carry the appropriate discontintuities. For a general discussion of the choice of penalty functionals in the problem of recovery of the three dimensional atmospheric temperature distribution, see Wahba (1985b). A convenient and simple penalty functional which generalizes J_m of (2.5) to several dimensions and has various favorable properties is the so called "thin plate" penalty functional. For $d = 2$ and $m = 2$ it is

$$J_m(f) = \int\int f_{x_1x_1}^2 + 2f_{x_1x_2}^2 + f_{x_2x_2}^2 dx_1 dx_2. \tag{2.7}$$

and in d dimensions it is

$$J_m(f) = \sum_{\alpha_1+\ldots+\alpha_d=m} \frac{m!}{\alpha_1! \cdots \alpha_d!} \int \cdots \int \left[\frac{\partial^m f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right]^2 dx_1 \cdots dx_d. \tag{2.8}$$

J_m is a seminorm in the space X of functions whose derivatives of total order m are square integrable (see Duchon (1976) , Meinguet (1979)). In order for evaluations to be bounded with respect to J_m it is necessary that $2m > d$. The span of the null space of J_m is the $M = \binom{m+d-1}{d}$ monomials of total degree less than m, call them $\Phi_1, \ldots, \Phi_M$. A reproducing kernel associated with the seminorm J_m from which can be obtained explicit formulae for the $P_1\xi_i$ can be found in Wahba and Wendelberger (1980).

Again, we can build in various types of discontinuities by our construction of the Ψ_j's. The magnitudes of the discontinuities are determined by θ. For example, letting d=2, to model a jump in the first derivative with respect to x_2 along a curve $x_2^*(x_1)$, let

$$\gamma(x) = \gamma(x_1,x_2) = |x_2 - x_2^*(x_1)|, \tag{2.9}$$

and

$$g(x) = f(x) + \theta(x_1)\gamma(x) \tag{2.10}$$

where $f \varepsilon X$ and θ may depend on x_1. Then

$$\left.\frac{\partial g}{\partial x_2}\right]_{x_2=x_2^*(x_1)_-} - \left.\frac{\partial g}{\partial x_2}\right]_{x_2=x_2^*(x_1)_+} = 2\theta(x_1). \tag{2.11}$$

If, for example

$$\theta(x_1) = \sum_{j=1}^{p} \theta_j q_j(x_1) \tag{2.12}$$

where the q_j's are given, then

$$\Psi_j(x) = q_j(x_1)\gamma(x). \tag{2.13}$$

Figure 3 gives the tropopause break function $\gamma(x_1,x_2)$ of (2.9) using the tropopause height $x_2^*(x_1)$ of Figure 2. Figure 4 gives a series of vertical temperature profiles as might be observed by a series of hopothetical radiosondes equally spaced along a latitude circle. The leftmost solid line corresponds to Figure 1 , and the tropopause heights (sharp minima) correspond to the curve $x_2^*(x_1)$ of Figure 2. The x_1 scale corresponds to the left most curve, the other curves have been shifted to the right for visual purposes. Figure 5 represents an estimate of the two dimensional temperature distribution given the hypothetical data (circles) of Figure 4 and was obtained as the solution to the problem:

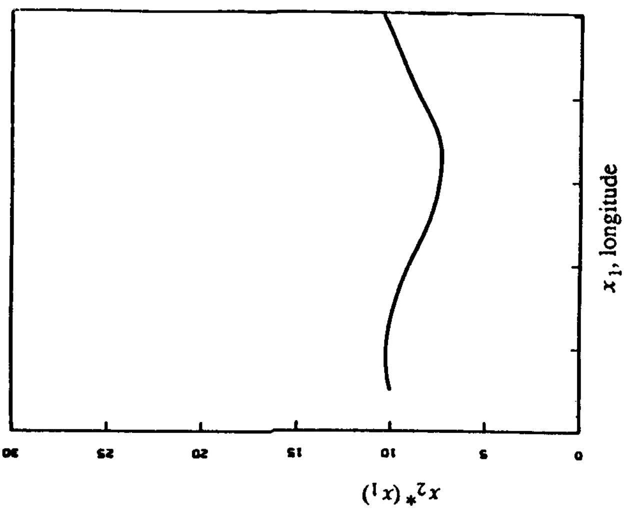

Fig. 2. Hypothetical tropopause height along a latitude circle.

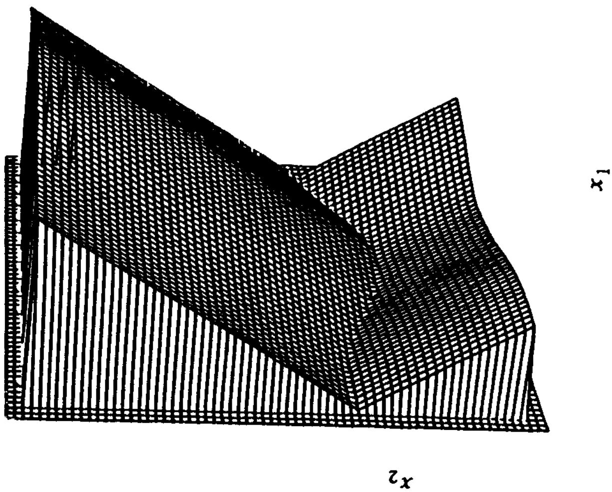

Fig. 3. Tropopause break function $\gamma(x_1,x_2)$.

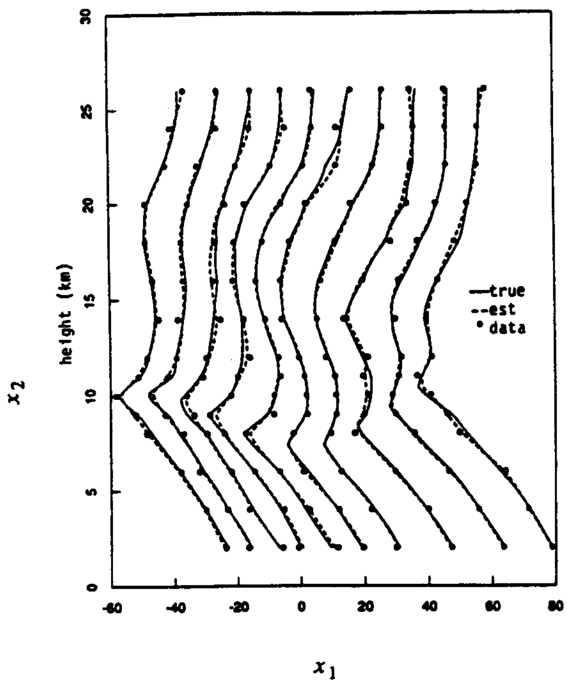

Fig. 4. Hypothetical vertical temperature profiles.

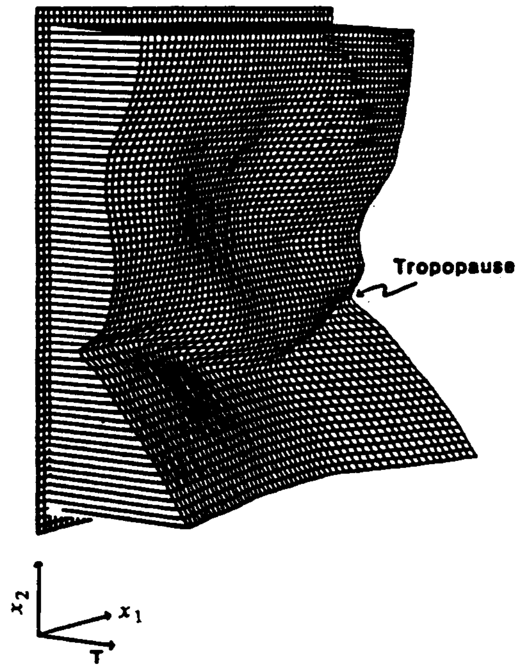

Fig. 5. $g_\lambda(x_1,x_2)$, estimated two dimensional temperature distribution.

Find $f \varepsilon X$ and $\theta = (\theta_1, \theta_2)$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x(i)) - \theta_1\Psi_1(x(i)) - \theta_2\Psi_2(x(i)))^2 + \lambda J_2(f), \tag{2.14}$$

where Ψ_1 and Ψ_2 are defined by (2.13) with $q_1(x)=1$ and $q_2(x)=x_1$. For further details, see Shiau, Wahba, and Johnson (Dec. 1985).

If, instead of the direct data y_i of Figure 4, one were given satellite radiance data

$$y_{i,\nu} = N_{i,\nu} + \varepsilon_{i,\nu}, \tag{2.15a}$$

where

$$N_{i,\nu}(g) = \int_{line\ of\ sight\ i} K_\nu(g(x), x)dx \tag{2.15b}$$

the problem becomes: find $f \varepsilon X$ and θ to minimize

$$\frac{1}{n}\sum_{i,\nu}(y_{i,\nu} - N_{i,\nu}(g))^2 + \lambda J_2(f) \tag{2.16}$$

where

$$g = f + \theta_1\Psi_1 + \theta_2\Psi_2.$$

III. COMPUTATIONAL PROBLEMS COMMON TO PARTIAL SPLINE MODELS

The minimizer of (2.2) can be computed directly, using the representation of g_λ in (2.4) (and the work depends on n) or, if n is large, f can be approximated as

$$f \sim \sum_{j=1}^{N} c_j B_j + \sum_{\nu=1}^{M} d_\nu \Phi_\nu \tag{3.1}$$

where the B_j's are some appropriately chosen basis functions. See Wahba (1980a) for some details in the thin plate spline case.

Substituting (3.1) back into (2.2) results in the problem of finding c, d, θ to minimize

$$\frac{1}{n}\| y - Xc - Td - S\theta \| + \lambda c'\Sigma c. \tag{3.2}$$

where X, T, S, Σ are given matrices. To find λ by GCV, minimize

$$V(\lambda) = \frac{\frac{1}{n}\|(I - A(\lambda))y\|^2}{\left[\frac{1}{n}Tr(I - A(\lambda))\right]^2} \tag{3.3}$$

where $A(\lambda)$ is the influence or hat matrix. $A(\lambda)$ satisfies

$$\begin{pmatrix} N_1 g_\lambda \\ \cdot \\ \cdot \\ N_n g_\lambda \end{pmatrix} = A(\lambda)\begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \tag{3.4}$$

where g_λ is the minimizer. (For properties of the GCV estimate of λ, see Wahba (1985a)). This problem can be put in a standard form for which transportable code has been developed. Let

$$Z = (X : T : S) \tag{3.5}$$

$$\beta = \begin{pmatrix} c \\ d \\ \theta \end{pmatrix} \tag{3.6}$$

$$J = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \tag{3.7}$$

then the problem has the form: Find β to min

$$\frac{1}{n} \| y - Z\beta \|^2 + \lambda \beta' J \beta, \tag{3.8}$$

choose λ by GCV. Transportable code Bates et al. (November 1985) is available for minimizing (3.8) with λ the minimizer of (3.3). The code uses Cholesky decomposition and a truncated singular value decomposition.

In the nonlinear case of (2.15) and (2.16), let

$$f = \sum_{j=1}^{N} c_j B_j + \sum_{\nu=1}^{n} d_\nu \Phi_\nu, \tag{3.9}$$

$$N_i(f + \sum \theta_j \Psi_j) = N_i(\beta), \quad \beta = \begin{pmatrix} c \\ d \\ \theta \end{pmatrix} \tag{3.10}$$

Then we want to minimize

$$I_\lambda = \frac{1}{n} \sum (y_i - N_i(\beta))^2 + \lambda \beta' J \beta \tag{3.11}$$

$$= Q(\beta) + \lambda \beta' J \beta, \ \textit{say}.$$

The Gauss Newton iteration for this problem is

$$\beta_\lambda^{k+1} = (\nabla^2 Q + n\lambda J)^{-1} [\nabla^2 Q \, \beta_\lambda^k - \nabla Q] \tag{3.12}$$

$$\nabla = \textit{gradient } I_\lambda, \ \nabla^2 = \textit{Hessian } I_\lambda .$$

β_λ^{k+1} is the minimizer of

$$\| y^k - Z^k \beta \|^2 + n\lambda \beta' J \beta \tag{3.13}$$

with

$$y^k = y - N(\beta_\lambda^k) + Z^k \beta_\lambda^k$$

$$Z^{k\prime} Z^k = \nabla^2 Q$$

$$Z^{k\prime} y^k = \nabla^2 Q \, \beta_\lambda^k - \nabla Q$$

To choose λ by GCV for nonlinear problems, fix λ. When this iteration converges, find $V(\lambda)$ as a function of the influence matrix $A^k(\lambda)$ and y^k for the kth iterate, that is, $A^k(\lambda)$ satisfies $Z^k \beta_\lambda^{k+1} = A^k(\lambda) y^k$. Step to a new λ. Find $\hat{\lambda}$ which minimizes $V(\lambda)$ this way.

See O'Sullivan and Wahba (1985).)

IV. THE USE OF GCV AS A STOPPING RULE IN THE ITERATIVE SOLUTION OF LARGE LINEAR SYSTEMS.

Consider the matrix model

$$y = Kg + \varepsilon \tag{4.1}$$

where K is an $n \times p$ matrix and ε is an error vector, presumed to behave like "white noise". Let Q be a strictly positive definite $p \times p$ matrix with symmetric square root $Q^{1/2}$, and let the singular value decomposition of $KQ^{1/2}$ be

$$KQ^{1/2} = UDV' \tag{4.2}$$

where D is $p \times p$ with diagonal entries d_i and let $u_1, \ldots, u_p$ and $v_1, \ldots, v_p$ be the p columns of U and V respectively. The Q-generalized inverse solution K_Q^{+} of the equation $y = Kg$ is defined as that vector g which minimizes $g'Q^{-1}g$ subject to $Kg = \hat{y}$ where $\hat{y}$ is the orthogonal projection of y onto the range of K. K_Q^{+} is given by

$$K_Q^{+}y = QK'(KQK')^{+}y = \sum_{j:d_j \neq 0} \frac{(y,u_j)}{d_j} Q^{1/2} v_j \tag{4.3}$$

Now consider the generalized Richardson/Landweber/Fridman/Picard/Cimino iteration

$$g^k = g^{k-1} + \beta QK'(y - Kg^{k-1}),\ k = 1,2,\ldots \tag{4.4}$$

$$= (I - \beta QK'K)g^{k-1} + \beta QK'y$$

with $g^0 = 0$. It is necessary that β satisfies $\beta < \frac{2}{d_1^2}$. Then

$$Q^{-1/2}g^k = (I - \beta Q^{1/2}K'KQ^{1/2})Q^{-1/2}g^{k-1} + \beta Q^{1/2}K'y. \tag{4.5}$$

Let

$$\tilde{g} = Q^{-1/2}g^k,\ \ \tilde{K} = KQ^{1/2}. \tag{4.6}$$

Then

$$\tilde{g}^k = (I - \beta\tilde{K}'\tilde{K})\tilde{g}^{k-1} + \beta\tilde{K}'y. \tag{4.7}$$

It is not hard to show, using the identity $[I + (I-B) + \ldots + (I-B)^{k-1}]B = I - (I-B)^k$ for symmetric matrices, that

$$\tilde{g}^k = \sum_{j:d_j>0} (1-(1-\beta d_j^2)^{k-1}) \frac{(y,u_j)}{d_j} v_j \tag{4.8}$$

hence

$$g^k = \sum_{j:d_j>0} (1-(1-\beta d_j^2)^{k-1}) \frac{(y,u_j)}{d_j} Q^{1/2} v_j. \tag{4.9}$$

g^k is related to $K_Q^{+}y$ as follows: To get g^k from $K_Q^{+}y$ the component of $K_Q^{+}y$ in the direction of $Q^{1/2}v_j$ is damped by the factor $(1-(1-\beta d_j^2)^{k-1}$. This factor decreases to 0 as d_j^2

decreases and reduces the influence of the factor $\frac{(y,u_j)}{d_j}$ which is sensitive to ε for d_j small. For comparison, the minimizer of

$$\frac{1}{n}\|y-Kg\|^2+\lambda g'Q^{-1}g \tag{4.10}$$

is

$$g_\lambda=\sum_{j:d_j^2>0}\frac{1}{1+(\frac{n\lambda}{d_j^2})}\frac{(y,u_j)}{d_j}Q^{1/2}v_j \tag{4.11}$$

so that in the Richardson iteration the damping factor $(1-(1-\beta d_j^2)^{k-1}$ is analogous to the damping factor $1/(1+\frac{n\lambda}{d_j^2})$ in regularization, and the pair (k,β) play the role of λ.

One can obtain the cross validation function $V(k,\beta)$ analogous to (3.3) as follows: since $KQ^{1/2}v_i=d_iu_i$ we have

$$Kg^k=\sum_{j:d_j>0}(1-(1-\beta d_j^2)^{k-1})(y,u_j)u_j. \tag{4.12}$$

This equation defines the $n\times n$ matrix $A(k,\beta)$ which plays the role of $A(\lambda)$ in the cross validation function (3.3) via $Kg^k=A(k,\beta)y$. Here $A(k,\beta)=U\Lambda U'$, where Λ is the $p\times p$ diagonal matrix with jjth entry $(1-(1-\beta d_j^2)^{k-1})$. Thus $V(k,\beta)$ is given by

$$V(k,\beta)=\frac{\frac{1}{n}\|y-Kg^k\|^2}{[\frac{1}{n}Tr(I-\beta KQK')^{k-1}]^2}. \tag{4.13}$$

One computes $V(1,\beta),V(2,\beta),\ldots,$ until a minimum is found, and then stops the iteration. If ε behaves like a white noise vector, then this choice of k should share the optimality properties of the GCV method.

V. GCV AND CONSTRAINED REGULARIZATION FOR THE PARAMETER ESTIMATION PROBLEM

The remarks in this section are at the present time mostly conjectural as to the actual feasibility of carrying out the program suggested. Some hope that the program is feasible can be gleaned from some ideas in O'Sullivan (to appear). See also Kravaris and Seinfeld (1985). It is, of course to be remembered that that what may seem like a great expense for data analysis can look quite small when compared to the cost of collecting the data.

Let

$$\frac{\partial u}{\partial t}=\sum_{l=1}^{3}\frac{\partial}{\partial x_l}(\alpha(x)\frac{\partial u}{\partial x_l})+q(x,t),\ x=(x_1,x_2,x_3)\,\varepsilon\,\Omega,\ t\,\varepsilon[0,T] \tag{5.1a}$$

$$u(x,0)=0,\ \ \frac{\partial u}{\partial n}=0, \tag{5.1b}$$

where n is the direction normal to the boundary. Here u is pressure and q accounts for the withdrawal or injection of fluid in the region Ω. We will suppose that q is known as accurately as desired over $\Omega\times[0,T]$. Noisy observations y_{ij} of u are obtained,

$$y_{ij} = u(x(i),t(j)) + \varepsilon_{ij}, \tag{5.2}$$

and it is desired to estimate α.

First, approximate $\alpha(x)$ by

$$\alpha(x) = \sum_{k=1}^{K} \alpha_k B_k(x) \tag{5.3}$$

where for notational convenience we have incorporated all of the desired basis functions into the set of B_k's. Given q, $\alpha(x)$ of the form (5.3), and the boundary conditions (5.1b), we define u_α as the solution to the p. d. e. of (5.1), and $\alpha_\lambda = (\alpha_{1,\lambda}, \ldots, \alpha_{K,\lambda})$ as the minimizer of

$$I_\lambda(\alpha) = \sum_{ij}(u_\alpha(x(i),t(j)) - y_{ij})^2 + \lambda\alpha'\Sigma\alpha \tag{5.4}$$

where Σ is some non negative definite matrix derived from a reasonable penalty functional, with a sufficiently small null space. For fixed λ one needs a numerical method for finding α_λ to minimize (5.4). See O'Sullivan (to appear) for some ideas along these lines.

In principle one could seek the minimizer of (5.4) subject to linear inequality constraints on the vector α which reflected known physical constraints on the function $\alpha(x)$, for example, (a discretized version of) $\alpha_{min} \le \alpha(x) \le \alpha_{max}$, $x \in \Omega$.

A GCV function for constrained problems is defined in Wahba (1982), eq. (3.7), see also Villalobos and Wahba (March 1985). In principle this method for choosing λ should be applicable to the present problem. An approximation to the cross validation function $V(\alpha)$ would be an approximation to

$$V(\alpha) = \frac{\sum_{ij}(u_\alpha(x(i),t(j)) - y_{ij})^2}{(1 - \frac{1}{n}\sum_{ij} a_{ij}(\lambda))^2} \tag{5.5a}$$

where $a_{ij}(\lambda)$ is an approximation to

$$\frac{\partial u_\alpha(x(i),t(j))}{\partial y_{ij}}, \tag{5.5b}$$

and n is the number of data points.

REFERENCES

Anderssen, R. S., and Prenter, P. M. (1981), "A Formal Comparison of Methods Proposed for the Numerical Solution of First Kind Integral Equations," *J. Australian Math. Soc. (Ser. B)*, 22, 488-500.

Bates, D. M., Lindstrom, M. J., Wahba, G., and Yandell, B. (November 1985) "GCVPACK - Routines for Generalized Cross Validation." Technical Report 775, Dept. of Statistics, University of Wisconsin-Madison.

Duchon, J. (1976), "Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces," in *Constructive Theory of Functions of Several Variables*, eds. W. Schempp, and K. Zeller, 85-100.

Fritz, S., Wark, D., Fleming, J., Smith, W., Jacobowitz, H., Hilleary, D., and Alishouse, J. (1972) "Temperature Sounding from Satellites." NOAA Tech. Rept. NESS 59, Washington, D. C.: National Oceanic and Atmkospheric Administration.

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

Kravaris, C., and Seinfeld, J.H. (1985), "Identification of Parameters in Distributed Parameter Systems by Regularization," *SIAM J. Control Opt.*, 23, 217-241.

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

Natterer, F. (1986) "Numerical Treatment of Ill Posed Problems." *manuscript*, .

O'Sullivan, F. (1983), *The Analysis of Some Penalized Likelihood Estimation Schemes*, University of Wisconsin-Madison, Statistics Dept.. (Technical Report #726)

O'Sullivan, F., and Wahba, G. (1985), "A Cross Validated Bayesian Retrieval Algorithm for Non-Linear Remote Sensing Experiments," *J. Comput. Physcs*, 59, 441-455.

O'Sullivan, F. (to appear) "A Statistical Perspective on Ill-Posed Inverse Problems." *Statistical Science*, .

Shiau, J., Wahba, G., and Johnson, D.R. (Dec. 1985) "Partial Spline Models for the Inclusion of Tropopause and Frontal Boundary Information in Otherwise Smooth Two and Three Dimensional Objective Analysis." Technical Report # 777 University of Wisconsin-Madison Statistics Dept.

Shiau, J. (June, 1985) "Smoothing Spline Estimation of Functions with Discontinuities." Technical Report # 768 University of Wisconsin-Madison Statistics Dept, Madison, WI:.

Strand, O.N. (1974), "Theory and Methods Related to the Singular Value Expansion and Landweber'S Iteration for Integral Equations of the First Kind.," *SIAM J. Numer. Anal*, 11, 798-825.

Susskind, J., Rosenfield, J., Reuter, D., and Chahine, M.T. (1984), "Remote Sensing of Weather and Climate Parameters from HIRS@/MSU on TIROS-N," *J. Geophysical research*, 89, 4677-4697.

Svensson, J. (February, 1985), *A Nonlinear Inversion Method for Derivation of Temperature Profiles from TOVS Data*, Norrkoping, Sweden: The Swedish Meteorological and Hydrological Institute. (manuscript, presented at The Second International TOVS Study Conference)

Villalobos, M., and Wahba, G. (March 1985) "Inequality Constrained Multivariate Smoothing Splines with Application to the Estimation of Posterior Probabilities." *J. Am. Statist. Assoc., to appear.* .

Wahba, G., and Wendelberger, J. (1980), "Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross-Validation," *Monthly Weather Review*, 108, 36-57.

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

Wahba, G. (1980b) "Ill Posed Problems: Numerical and Statistical Methods for Mildly, Moderately, and Severely Ill Posed Problems with Noisy Data." University of Wisconsin-Madison Statistics Department Technical Report No. 595. (To appear in the Proceedings of the International Conference on Ill Posed Problems, M.Z. Nashed, ed.)

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

Wahba, G. (1984), "Cross Validated Spline Methods for the Estimation of Multivariate Functions from Data on Functionals," in *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory*, eds. H. A. David, and H. T. David Iowa State University Press.

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

Wahba, G. (1985b), "Variational Methods for Multidimensional Inverse Problems," in *Advances in Remote Sensing Retrieval Methods*, eds. A. Deepak, H. E. Fleming, and M. Chahine, Hampton, Va: A. Deepak Publishing, 385-410.