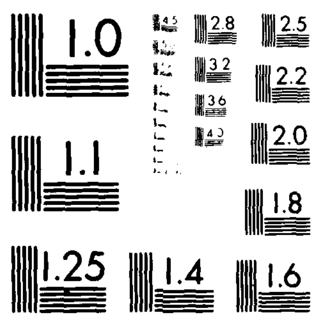
AD-A110 769 WISCONSIN UNIV-MADISON DEPT OF STATISTICS F/G 12/1
CONSTRAINED REGULARIZATION FOR ILL POSED LINEAR OPERATOR EQUATI--ETC(U)
AUG 81 G WAHBA N00014-77-C-0675
UNCLASSIFIED TR-646 NL

1 of 1
AD A
110769

END
DATE
FILMED
3-82
DTIC



MICROCOPY RESOLUTION TEST CHART

(1) LEVEL

DEPARTMENT OF STATISTICS

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

AD A110769

TECHNICAL REPORT NO. 646

August 1981

CONSTRAINED REGULARIZATION FOR ILL POSED LINEAR OPERATOR EQUATIONS, WITH APPLICATIONS IN METEOROLOGY AND MEDICINE

by

Grace Wahba

This work was supported by the Office of Naval Research under Contract No. N00014-77-C-0675.

Abstract

The relationship between certain regularization methods for solving ill posed linear operator equations and ridge methods in regression problems is described. The regularization estimates we describe may be viewed as ridge estimates in a (reproducing kernel) Hilbert space H. When the solution is known *a priori* to be in some closed, convex set in H, for example, the set of nonnegative functions, or the set of monotone functions, then one can propose regularized estimates subject to side conditions such as nonnegativity, monotonicity, etc. Some applications in medicine and meteorology are described. We describe the method of generalized cross validation for choosing the smoothing (or ridge) parameter in the presence of a family of linear inequality constraints. Some successful numerical examples, solving ill posed convolution equations with noisy data, subject to nonnegativity constraints, are presented. The technique appears to be quite successful in adding information, doing nearly the optimal amount of smoothing, and resolving distinct peaks in the solution which have been blurred by the convolution operation.

Prepared for the Proceedings of the Third Purdue Symposium on Statistical Decision Theory, S. S. Gupta and J. O. Berger, eds.

1. Introduction

We are interested in the Hilbert space version of constrained ridge regression, which we will show has many interesting applications.

The (ridge) regression setup is:

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1} \tag{1.1}$$

$$\varepsilon \sim N(0,\sigma^2 I)$$

$$\beta \sim N(0,b\Sigma)$$

where X and Σ are known, σ^2, b are unknown. A "ridge-Stein" estimate of β, call it β_λ, is given by the minimizer of $Q_\lambda(\beta)$,

$$Q_\lambda(\beta) = \frac{1}{n}||y-X\beta||^2 + \lambda\beta'\Sigma^{-1}\beta,$$

where $||\cdot||$ is the Euclidean mean. If λ is taken as σ^2/nb, then it is not hard to show that

$$\beta_\lambda = E(\beta|y). \tag{1.2}$$

If it is known that β is in some closed convex set C in E_p, then one may estimate β as the minimizer of $Q_\lambda(\beta)$ subject to the constraint $\beta\varepsilon C$. Some interesting C are those determined by a finite number of linear inequality constraints, for example $\beta_i \geq 0$, $i = 1,2,\ldots,p$, or $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_p$. M.E. Bock discusses a related setup in these proceedings.

We particularly want to allow β to have a partially improper prior, for example, $\sigma_{11} = \infty$. Then Σ^{-1} is defined in the natural way and will

---

This work was supported by the Office of Naval Research under Contract No. N00014-77-C-0675.

then not be of full rank. This causes no problem provided X and Σ^{-1} are such that

$$\frac{1}{n}\beta' X' X\beta + \lambda\beta'\Sigma^{-1}\beta = 0 \Rightarrow \beta = 0. \tag{1.3}$$

An example of a Hilbert space version of this problem (an indirect sensing experiment) is

$$y(t_i) = \int_0^1 K(t_i,s)f(s)ds + \varepsilon_i,\ i = 1,2,\ldots,n,\ 0 \le t_1 <\ldots< t_n \le 1 \tag{1.4}$$

$$\varepsilon \sim N(0,\sigma^2 I)$$

where K is known, f is known to be in the Sobolev space W_2^m(W_2^m= f:f,f',...,$f^{(m-1)}$ abs.cont., $f^{(m)} \varepsilon L_2[0,1]$),see Adams (1975)), and σ^2 is unknown. A so called "regularized" estimate f_λ of f is given by the minimizer in W_2^m of

$$Q_\lambda(f) = \frac{1}{n}\sum_{i=1}^{n}(y(t_i) - \int_0^1 K(t_i,s)f(s)ds)^2 + \lambda\int_0^1 (f^{(m)}(s))^2 ds. \tag{1.5}$$

$Q_\lambda(f)$ is analogous to

$$Q_\lambda(\beta) = \frac{1}{n}||y-X\beta||^2 + \lambda\beta'\Sigma^{-1}\beta.$$

If the linear functionals $f \rightarrow \int_0^1 K(t_i,s)f(s)ds$ are bounded in W_2^m for each i = 1,2,...,n, and

$$\frac{1}{n}\sum_{i=1}^{n}(\int_0^1 K(t_i,s)f(s)ds)^2 + \lambda\int_0^1 (f^{(m)}(s))^2 ds = 0 \Rightarrow f = 0 \tag{1.6}$$

then $Q_\lambda(f)$ will have a unique minimizer, call it f_λ,in W_2^m.

If f is endowed with the zero mean Gaussian prior defined by: f is $\sqrt{b}$ times an unpinned m-fold integrated Weiner process (Shepp (1966)), with a diffuse prior on the initial conditions, then it can be shown (Kimeldorf and Wahba (1971), Wahba (1978)), that

$$f_\lambda(t) = E\{f(t)|y(t_1),\ldots,y(t_n)\}, \tag{1.7}$$

where $\lambda = \sigma^2/nb$. This prior may be colloquially described as "$f^{(m)}$=white noise". However, with this prior $E\int_0^1(f^{(m)}(s))^2ds$ is not finite, and the meaning of b as a process parameter becomes unclear for $f\varepsilon W_2^m$. If it is assumed that $f\varepsilon W_2^m$, then it appears to be more appropriate to view λ as the "bandwidth parameter" which governs the square bias-variance tradeoff.

If (1.6) holds, then $Q_\lambda(f)$ *will have a unique minimizer in any* closed convex set $C\subseteq H$ (see Wong (1980), Gorenflo and Hilpert (1980). The set of non-negative functions $\{f:f(s)\geq 0, 0\leq s\leq 1\}$ is closed and convex in W_2^m for m = 1,2,..., and the set of monotone increasing functions $\{f:f'(s)\geq 0,\ 0\leq s\leq 1\}$ is closed convex in W_2^m for m = 2,3,... . See also Wright and Wegman (1980).

We are interested in the general formulation of the above problem. The model is

$$y_i = L_{t_i}f + \varepsilon_i,\ i = 1,2,\ldots,n$$

where it is known that $f\varepsilon C\subseteq H$, where H is a given Hilbert space, C is a closed, convex set in H, and $L_{t_i}\ldots L_{t_n}$ are n continuous linear functions on H. $J(\cdot)$ is a seminorm on H with an m dimensional null space, and it is "believed" that J(f) is not too large. We propose estimating f as the minimizer of

$$Q_\lambda(f) = \frac{1}{n}\sum_{i=1}^{n}(L_{t_i}f-y_i)^2 + \lambda J(f) \tag{1.8}$$

subject to $f\varepsilon C$.

If

$$\frac{1}{n}\sum_{i=1}^{n}(L_{t_i}f)^2 + \lambda J(f) = 0$$

$\Rightarrow f = 0$, then there will be a unique solution, call it f_λ^C. We will refer to this solution as the constrained regularized estimate, sometimes dropping the superscript C.

There are now two problems. One, given λ, how does one compute a good approximation to f_λ^C, and two, how does one estimate a good value of λ. In many interesting cases, when H is a reproducing kernel space, the constraint set C can be discretized in a convergent way, see Wahba (1973). For example, the minimizer of $Q_\lambda(f)$ subject to $f \varepsilon C = \{f: f(s) \geq 0, 0 \leq s \leq 1\}$ is well approximated by the minimizer of $Q_\lambda(f)$ subject to $f \varepsilon C_L = \{f: f(\frac{i}{L}) \geq 0, i = 1,2,\ldots,L\}$ for $H = W_2^m$, $J(\cdot) = \int_0^1 (f^{(m)}(s))^2 ds$, L large. If C_L is any (closed) set defined by L linear inequality constraints, the problem of minimizing $Q_\lambda(f)$ subject to $f \varepsilon C_L$ can be reduced to a quadratic programming problem with linear inequality constraints in at most n + m + L variables. See Kimeldorf and Wahba (1971). The researcher interested in numerical methods for this and related problems may consult Anselone and Laurent (1968), Utreras (1979), Wahba (1978, 1980a, 1980b, 1981), Wahba and Wendelberger (1980). (The formulae in Kimeldorf and Wahba are inappropriate for computational purposes.) Remarks concerning the effect of quadrature in this setting may be found in Wahba (1981). Library software for solving the quadratic programming problem by the principal pivoting method is available, for moderate n + m + L, see MACC (1979). We will go through a relatively simple example in Section 4.

Our main interest in this paper is the development of a method for choosing λ which is suitable for the constrained problem.

In this paper we propose an extension of the generalized cross validation (GCV) method, to the constrained case. This method was proposed in the unconstrained case in Craven and Wahba (1979), Golub, Heath and Wahba (1979), and Wahba (1977). The GCV estimate of λ we propose in the constrained case can be expensive to compute. Thus we propose a first order approximation to it which is very much cheaper to compute, and appears to be satisfactory in the examples we tried.

We experimentally tested the constrained regularization method with the approximate GCV estimate of λ on a convolution equation with several simulated data sets generated according to the model (1.4) with non-negative f's. For comparison, we first estimated f by minimizing $Q_\lambda(f)$ in W_2^2 and using the (usual) unconstrained GCV estimate $\hat{\lambda}$ for λ. We then estimated f by minimizing $Q_\lambda(f)$ in C_n where $C_n = \{f : f(\frac{i}{n}) \geq 0, i=1,2,\ldots,n\}$, and choosing λ by the approximate GCV method for constrained problems. The constrained estimates with the approximate GCV choice of λ were all dramatic improvements over the unconstrained estimates. As a practical matter, they displayed a remarkable ability to resolve closely spaced peaks in the solution that have been blurred in the data by the convolution operation. The convolution equation is ill posed, and the positivity constraints are apparently supplying much needed information. Three cases of the exact GCV method for constrained problems were tried for choosing λ. It gave a very slightly better (and possibly more stable) estimate of the optimal λ. However it's much more expensive to compute.

2. Some Applications

i) Meteorology

In recent years several satellites have been put in orbit which carry detectors which measure the upwelling radiation at selected frequencies. The observed radiation at frequency ν, when the subsatellite point is P, may be modelled (after some linearization and approximation) as

$$I_\nu(P) = \int_{\Omega_P} K_\nu(P,P')T(P')dP',$$

where P' is a point in the atmosphere, Ω_P is the volume within the detector field of view when the subsatellite point is P, T(P') is the atmospheric temperature at point P' and K_ν is determined from the equations of radiative transfer. See for example Fritz et al (1972), Smith et al (1979), Westwater (1979). It is desired to estimate T(P) to use as initial conditions in numerical weather forecasting. Occasionally, outside information, such as the existence of a temperature inversion, is available, thus providing some inequality conditions on the derivative of T(P) in the vertical direction.

ii) Computerized Tomography

Computerized tomography machines are in most well equipped hospitals. Computerized tomography machines observe line (or more accurately, strip) integrals of the X-ray density f of parts of the human body, and from this data

$$y_i = \int_{\ell_i} f(P)dP + \varepsilon_i, \; i = 1,2,\ldots,n,$$

estimates of f(P) are made. Algorithms for estimating f must be capable of dealing with $n \approx 10^5$, see Herman and Natterer (1981), Shepp and Kruskal (1978). The true f is non-negative.

iii) Stereology

Scientists studying tumor growth feed laboratory mice a carcinogen, sacrifice the mice, and then freeze and slice the livers. Images of the liver slices are magnified and areas of tumor cross sections are measured. It is expensive to examine the liver slices, thus it is desired to take a sample of the possible slices and from the resulting data infer numbers and (three dimensional) size distributions of tumors in the entire liver from data from a few slices. In the "random spheres" model, the tumors are assumed to be spherical with the radii density f(s). If the slices are "random" then the cross sectional (two dimensional) density g(t) is related to f by

$$g(t) = \frac{t}{\mu}\int_t^{\infty} \frac{f(s)}{\sqrt{s^2-t^2}}\, ds, \quad \mu = \int_0^{\infty} s f(s)\, ds.$$

See Anderssen and Jakeman (1975), Watson (1971), Wicksell (1926). This setup does not fit into the model (1.4) because i) in theory a random sample from the population with density g is observed (not $g(t_i)+\varepsilon_i$) and ii) in practice the liver is embedded in a paraffin block and sliced systematically perpendicular to an axis which (roughly) maximizes the cross sectional area of the liver being sliced. Nonetheless, it is fruitful to think of this problem in the context of ill posed integral equations (see Anderssen and Jakeman (1975), Nychka (1981)).

iv) Convolution Equations

Convolution equations in one and higher dimensions arise in many areas of physics. See, for example Chambless (1980), Davies (1979). These equations can be surprisingly ill posed.

v) Other applications

Other applications may be found in the books of Anderssen, DeHoog and Lukas (1980), Golberg (1978), Tihonov and Arsenin (1977), Twomey (1977), Nashed (1981).

3. Cross validation for constrained problems

We first define the ordinary cross validation (OCV) or "leaving out one" method of choosing λ.

Let $L_i = L_{t_i}$, and let $f_\lambda^{[k]}$ be the minimizer of

$$\frac{1}{n}\sum_{\substack{i=1\\ i\neq k}}^{n}(L_i f - y_i)^2 + \lambda J(f) \tag{3.1}$$

subject to $f \in C \subseteq H$, where we assume sufficient conditions on the $\{L_i\}$ and $J(\cdot)$ for existence and uniqueness. A figure of merit can be defined for λ by

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n}(L_k f_\lambda^{[k]} - y_k)^2, \tag{3.2}$$

where $L_k f_\lambda^{[k]}$ is the prediction of y_k given the data $y_1,\ldots,y_{k-1},y_{k+1},\ldots,y_n$, and using λ. The OCV estimate of λ is the minimizer of $V_0(\lambda)$. In the unconstrained ridge regression case this estimate is known as Allen's PRESS (see Hocking's discussion to Stone (1974)). The names of Mosteller and Tukey (1968) Geisser (1975), M. Stone (1974) and others are associated with early work on ordinary cross validation. See also Wahba and Wold (1975). In the ridge regression case the OCV or Allen's PRESS has the undesireable property of not being invariant under arbitrary rotations $y \to \Gamma y$ of the data space. If one observed Γy instead of y the OCV estimate of λ may be different. GCV (to be defined below) may be thought of as a rotation invariant version of OCV, for which some good theoretical properties may be obtained. For further discussion see Craven and Wahba (1979), Golub, Heath and Wahba (1979), Wahba (1977), Utreras (1978), Speckman (1981).

To extend the definition of the GCV estimate of λ to constrained problems, we will use the Theorem given below.

Theorem: Let H be a Hilbert space, $J(\cdot)$ a semi norm on H and $L_1,\ldots,L_n$ be n continuous linear functionals on H, with the property, that for any fixed $\lambda > 0$,

$$\frac{1}{n}\sum_{\substack{i=1\\ i\neq k}}^{n}(L_i f)^2 + \lambda J(f) = 0 \Rightarrow f = 0 \qquad k = 1,2,\ldots,n.$$

Let C be a closed convex set in H and let $f_\lambda^{[k]}[z]$ and $f_\lambda[z]$ be the minimizers in C of

$$\frac{1}{n}\sum_{\substack{i=1\\ i\neq k}}^{n}(L_i f - z_i)^2 + \lambda J(f)$$

and

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

respectively, where $z = (z_1,\ldots,z_n)'$. Then

$$f_\lambda[y+\underset{\sim}{\delta}_k] = f_\lambda^{[k]}[y],\ k = 1,2,\ldots,n \tag{3.3}$$

where $\underset{\sim}{\delta}_k = (0,\ldots,0,L_k f_\lambda^{[k]}[y]-y_k,0,\ldots,0)'$, (the non 0 entry is in the kth position.

Remark: This theorem says, that given data

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{k-1} \\ L_k f_\lambda^{[k]}[y] \\ y_{k+1} \\ \vdots \\ y_n \end{pmatrix}$$

the minimizer of $Q_\lambda(f)$ in C is $f_\lambda^{[k]}$.

Proof: Proofs in special cases may be found in Craven and Wahba (1979) and Golub, Heath and Wahba (1979). A proof in the generality cited here is in Wahba (1980c), although no doubt the result is a special case of classic optimization theory results.

Now define the "differential influence" of y_k when λ is used, by $a_{kk}^*(\lambda)$,

$$a_{kk}^*(\lambda) = \frac{L_k f_\lambda[y+\underset{\sim}{\delta}_k] - L_k f_\lambda[y]}{\delta_k} \tag{3.4}$$

where

$$\delta_k = L_k f_\lambda^{[k]}[y] - y_k. \tag{3.5}$$

$a_{kk}^*(\lambda)$ is a divided difference of $L_k f_\lambda$ considered as a function of the kth data point (and is well defined).

Applying L_k to both sides of (3.3) and substituting the result into (3.4) and (3.4) into (3.2) gives the identity

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n} \frac{(L_k f_\lambda - y_k)^2}{(1-a_{kk}^*(\lambda))^2} \quad . \tag{3.6}$$

The GCV estimate of λ is obtained by replacing $a_{kk}^*(\lambda)$ in (3.6) by the "average differential influence" $\frac{1}{n}\sum_{k=1}^{n} a_{kk}^*(\lambda)$, that is, the GCV estimate of λ is obtained by minimizing $V(\lambda) = V^C(\lambda)$ defined by

$$V^C(\lambda) = \frac{\frac{1}{n}\sum_{i=1}^{n}(L_k f_\lambda - y_k)^2}{(1-\frac{1}{n}\sum_{i=1}^{n} a_{kk}^*(\lambda))^2} \tag{3.7}$$

Some properties of this estimate in the unconstrained case are known. First, in the unconstrained ($C=H$) case, $L_k f\ [y]$ is linear in y, and there exists an influence matrix $A(\lambda)$ with the property

$$\begin{pmatrix} L_1 f_\lambda \\ \vdots \\ L_n f_\lambda \end{pmatrix} = A(\lambda)y.$$

In this case $a^*_{kk}(\lambda)$, the divided difference of $L_k f_\lambda$ with respect to $y_k + \delta_k$ and y_k, is also the first derivative

$$a^*_{kk}(\lambda) = \frac{\partial L_k f_\lambda}{\partial y_k} = a_{kk}(\lambda)$$

where $a_{kk}(\lambda)$ is the kkth entry of $A(\lambda)$. Then $V(\lambda)$ can be written

$$V(\lambda) = \frac{\frac{1}{n}||I-A(\lambda)y||^2}{(\frac{1}{n}Tr(I-A(\lambda)))^2} \tag{3.8}$$

To understand the known (and potentially obtainable) properties of the GCV estimate of λ we will first compare it with the unbiassed risk estimates of Stein (see Hudson (1974), Mallows (1973)).

Let $L(f,\lambda)$ be the predictive mean square error when λ is used

$$\begin{aligned} L(f,\lambda) &= \frac{1}{n}\sum_{i=1}^{n}(L_k f_\lambda - L_k f)^2 \\ &= \frac{1}{n}||A(\lambda)y-g||^2 \end{aligned}$$

where $g = (L_1 f,\ldots,L_n f)' = E_f y$.

If σ^2 is known (or an unbiassed estimate of it is available) then an unbiassed estimate $\hat{R}(\lambda)$ of $R(\lambda) = E_f L(f,\lambda) = \frac{1}{n}||(I-A(\lambda))g||^2 + \frac{\sigma^2}{n} TrA^2(\lambda)$ is available and is given by

$$\hat{R}(\lambda) = \frac{1}{n}||(I-A(\lambda))y||^2 - \frac{\sigma^2}{n}Tr(I-A(\lambda))^2 + \frac{\sigma^2}{n}TrA^2(\lambda),$$

this corresponds to Mallows' C_L, see Mallows (1973), Craven and Wahba (1979).

To talk about convergence, consider a family L_t, $t\varepsilon[0,1]$ of continuous linear functionals on H, with $L_{t_1},\ldots,L_{t_n}$ a subset. Let K be the operator which maps H into the real valued functions on $[0,1]$ by $(Kf)(t) = L_t f$. Loosely speaking, if $K(H)$ is a reproducing kernel space with sufficiently smooth reproducing kernel, then as $t_1,\ldots,t_n$ become dense in $[0,1]$,

$$E_f V(\lambda) \approx E_f L(f,\lambda) + \sigma^2$$

for λ in the neighborhood of the minimizer of $E_f L(f,\lambda)$. See Wahba (1977). Under various circumstances it can be shown (Craven and Wahba (1979)), that

$$\frac{E_f L(f,\hat{\lambda})}{\min_{\lambda} E_f L(f,\lambda)} \downarrow 1 \text{ as } n\to\infty,\ f\varepsilon H \tag{3.9}$$

where $\hat{\lambda}$ is the minimizer of $E_f V(\lambda)$. Utreras (1978) and Speckman (1981) have recently rigorized and strengthened these results.

In general for (3.9) to be true one appears to need that $\mu_1(\lambda)\to 0$ and $\mu_1^2(\lambda)/\mu_2(\lambda)\to 0$ for λ in the neighborhood of λ^* where $\mu_i(\lambda) = \frac{1}{n}\mathrm{Tr}A^i(\lambda)$ and λ^* is the minimizer of $E_f L(f,\lambda)$. Intuitively, this means that the signal must be concentrated in a small "corner" of the data space E_n. Optimal rates of convergence for $f_{\hat{\lambda}}$ corresponding to those given by C. Stone (1980) can be obtained in some cases Craven and Wahba (1979), Wahba (1977a, 1977b, 1979b), Lukas (1981).

We now return to the constrained case, $f\varepsilon C$. We consider only the case where C is (or is well approximated by) the intersection of a finite number of half-spaces,

$$C_L = \{f: N_\ell f \geq \alpha(\ell),\ \ell = 1,2,\ldots,L\},$$

where the N_ℓ are continuous linear functions on H. Even in this special case it appears that to evaluate $V(\lambda)$ of (3.7) for a single λ one must solve n quadratic programming problems in as many as n + m + L variables. *To avoid this we propose the following approximation:*

Replace the divided difference

$$a^*_{kk}(\lambda) = \frac{L_k f_\lambda[y+\delta_k]-L_k f_\lambda[y]}{\delta_k} \tag{3.10}$$

by the derivative

$$a_{kk}(\lambda) = \frac{\partial}{\partial y_k} L_k f_\lambda[y]|_y. \tag{3.11}$$

Thus $V(\lambda)$ of (3.7) is replaced by $V^C_{approx}(\lambda) = V_{approx}(\lambda)$ defined by

$$V_{approx}(\lambda) = \frac{\frac{1}{n}\sum_{k=1}^n (L_k f_\lambda - y_k)^2}{(1-\frac{1}{n}\sum_{k=1}^n \frac{\partial}{\partial y_k} L_k f_\lambda|_y)^2}. \tag{3.12}$$

For each λ, $V_{approx}(\lambda)$ can be obtained by solving one quadratic optimization problem. We outline the procedure, for more details, see Wahba (1980b) and the example in Section 4. First, solve the quadratic optimization problem to obtain f_λ and determine which constraints are active. Suppose these correspond to $N_{\ell_1}, N_{\ell_2}, \ldots, N_{\ell_{L'}}$. f_λ is then also the solution to the quadratic optimization problem: Minimize $Q_\lambda(f)$ subject to $N_{\ell_i} f = \alpha(\ell_i)$, $i = 1,2,\ldots,L'$,. The solution to this latter problem is linear in y and is related to the data through an influence matrix, call it $A_{L'}(\lambda)$. Then

$$\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial y_k} L_k f_\lambda|_y = \frac{1}{n}\mathrm{Tr}A_{L'}(\lambda). \tag{3.13}$$

$A_{L'}(\lambda)$ is given explicitly in Wahba (1980b), see also below. The ingredients for computing $TrA_{L'}(\lambda)$ will generally have been obtained in the process of setting up and solving the quadratic optimization problem.

Unfortunately $\frac{\partial}{\partial y_k} L_k f_\lambda |_y$ may be only piecewise well defined and continuous in λ. If a change in λ causes a change in the active constraint set, then one or more of the $\frac{\partial}{\partial y_k} L_k f_\lambda |_y$ may have a jump. This can be seen in the examples in Section 4 and is the major drawback of the method. The exact cross validation function $V(\lambda)$ of (3.7) appears to be a continuous function of λ for $\lambda > 0$.

4. Numerical Experiments

We numerically studied convolution equations with the model

$$y_i = \int_0^1 k(\tfrac{i}{n}-s)f(s)ds + \varepsilon_i, \; i = 1,2,\ldots,n, \; n \text{ even.}$$

$$f(s) \geq 0, 0 \leq s \leq 1,$$

With $J(f) = \int_0^1 (f^{(m)}(s))^2 ds$. The constraints will be discretized to $f(\frac{i}{n}) \geq 0$, $i = 1,2,\ldots,n$. To simplify the calculations while retaining many of the features of the original problem we assumed that $k(\cdot)$ and $f(\cdot)$ were both in the n dimensional subspace F of W_2^m spanned by

$$\{1, \cos 2\pi\nu t, \; \nu=1,2,\ldots,n/2, \; \sin 2\pi\nu t, \; \nu=1,2,\ldots,n/2-1\}.$$

Thus all functions in F_n are periodic and the null space of $J(\cdot)$ in F_n is spanned by the single function "1". Also, f and k are of the form

$$f(t) = \alpha_0 + 2\sum_{\nu=1}^{n/2-1} \alpha_\nu \cos 2\pi\nu t + 2\sum_{\nu=1}^{n/2-1} \beta_\nu \sin 2\pi\nu t + \alpha_{n/2}\cos \pi n t \tag{4.1}$$

$$k(t) = \xi_0 + 2\sum_{\nu=1}^{n/2-1} \xi_\nu \cos 2\pi\nu t + 2\sum_{\nu=1}^{n/2-1} \eta_\nu \sin 2\pi\nu t + \xi_{n/2}\cos \pi n t \tag{4.2}$$

where

$$\alpha_\nu = \frac{1}{n}\sum_{i=1}^{n} \cos 2\pi\nu\frac{i}{n} f(\tfrac{i}{n}), \quad \beta_\nu = \frac{1}{n}\sum_{i=1}^{n} \sin 2\pi\nu\frac{i}{n} f(\tfrac{i}{n}) \tag{4.3}$$

$$\xi_\nu = \frac{1}{n}\sum_{i=1}^{n} \cos 2\pi\nu\frac{i}{n} k(\tfrac{i}{n}), \quad \eta_\nu = \frac{1}{n}\sum_{i=1}^{n} \sin 2\pi\nu\frac{i}{n} k(\tfrac{i}{n}). \tag{4.4}$$

We have

$$g(t) = \int_0^1 k(t-s)f(s)ds$$

$$= \xi_0\alpha_0 + 2\sum_{\nu=1}^{n/2-1} (\alpha_\nu\xi_\nu - \beta_\nu\eta_\nu)\cos 2\pi\nu t$$

$$+ 2\sum_{\nu=1}^{n/2-1} (\alpha_\nu\eta_\nu + \beta_\nu\xi_\nu)\sin 2\pi\nu t$$

$$+\tfrac{1}{2}\alpha_{n/2}\xi_{n/2}\cos\pi nt, \tag{4.5}$$

and

$$J(f) = 2\sum_{\nu=1}^{n/2-1} (\alpha_\nu^2+\beta_\nu^2)(2\pi\nu)^{2m} + (1/2)\alpha_{n/2}^2(\pi n)^{2m}. \tag{4.6}$$

f_λ, the minimizer in F_n of

$$Q_\lambda(f) = \frac{1}{n}\sum_{i=1}^{n}\left(\int_0^1 k\left(\frac{i}{n}-s\right)f(s)ds - y_i\right)^2 + \lambda\int_0^1 (f^{(m)}(s))^2 ds \tag{4.7}$$

is given by

$$f_\lambda(t) = \hat{\alpha}_0 + 2\sum_{\nu=1}^{n/2-1}\hat{\alpha}_\nu\cos 2\pi\nu t + 2\sum_{\nu=1}^{n/2-1}\hat{\beta}_\nu\sin 2\pi\nu t$$

$$+ \hat{\alpha}_{n/2}\cos\pi nt \tag{4.8}$$

where

$$\hat{\alpha}_0 = a_0/\xi_0$$

$$\hat{\alpha}_\nu = \frac{1}{\xi_\nu^2+\eta_\nu^2+\lambda\lambda_\nu}(a_\nu\xi_\nu - b_\nu\eta_\nu) \qquad \nu = 1,2,\ldots,n/2-1$$

$$\hat{\beta}_\nu = \frac{1}{\xi_\nu^2+\eta_\nu^2+\lambda\lambda_\nu}(a_\nu\eta_\nu + b_\nu\xi_\nu) \tag{4.9}$$

$$\hat{\alpha}_{n/2} = \frac{1}{\frac{1}{2}\xi_{n/2}^2+\lambda\lambda_{n/2}}a_{n/2}\xi_{n/2}$$

with

$$\lambda_\nu = (2\pi\nu)^{2m} \tag{4.10}$$

$$\begin{aligned} a_\nu &= \frac{1}{n}\sum_{j=1}^{n} \cos 2\pi\nu\frac{j}{n}y_j \quad \nu=0,1,\dots,n/2 \\ b_\nu &= \frac{1}{n}\sum_{j=1}^{n} \sin 2\pi\nu\frac{j}{n}y_j \quad \nu=1,2,\dots,n/2-1, \end{aligned} \tag{4.11}$$

The cross validation function $V(\lambda)$ of (3.8) in the unconstrained case becomes

$$V(\lambda) = \frac{2\sum_{\nu=1}^{n/2-1}\left[\frac{\lambda\lambda_\nu}{\xi_\nu^2+\eta_\nu^2+\lambda\lambda_\nu}\right]^2(a_\nu^2+b_\nu^2)+\left[\frac{\lambda\lambda_{n/2}}{\frac{1}{2}\xi_\nu^2+\lambda\lambda_{n/2}}\right]^2 a_{n/2}^2}{\left[\frac{2}{n}\sum_{\nu=1}^{n/2-1}\frac{\lambda\lambda_\nu}{\xi_\nu^2+\eta_\nu^2+\lambda\lambda_\nu}+\frac{1}{n}\frac{\lambda\lambda_{n/2}}{\frac{1}{2}\xi_\nu^2+\lambda\lambda_{n/2}}\right]^2} . \tag{4.12}$$

In principle m can be chosen by cross validation (see Gamber (1979), Wahba and Wendelberger (1980). In these experiments we have (arbitrarily) set m = 2.

To study the constrained case we write this problem as follows: Letting $x = (f(\frac{1}{n}),\dots,f(\frac{n}{n}))'$, we have

$$Q_\lambda(f) = ||KWx-Wy||^2 + \lambda x'W'JWx \tag{4.13}$$

where the n×n matrices K, J and W are given by

$$K = \left(\begin{array}{c|ccc|c|ccc}
\xi_0 & 0 & & & 0 & 0 & & \\ \hline
 & \xi_1 & & & & -\eta_1 & & \\
0 & & \ddots & & 0 & & \ddots & \\
 & & & \xi_{n/2-1} & & & & -\eta_{n/2-1} \\ \hline
0 & 0 & & & \text{[illegible]} & 0 & & \\ \hline
 & \eta_1 & & & & \xi_1 & & \\
0 & & \ddots & & 0 & & \ddots & \\
 & & & \eta_{n/2-1} & & & & \xi_{n/2-1}
\end{array}\right)$$

$$J = \left(\begin{array}{c|ccc|c|ccc}
0 & 0 & & & 0 & 0 & & \\ \hline
 & \lambda_1 & & & & & & \\
0 & & \ddots & & 0 & & 0 & \\
 & & & \lambda_{n/2-1} & & & & \\ \hline
0 & 0 & & & \frac{1}{2}\lambda_{n/2} & 0 & & \\ \hline
 & & & & & \lambda_1 & & \\
0 & 0 & & & 0 & & \ddots & \\
 & & & & & & & \lambda_{n/2-1}
\end{array}\right)$$

$$W = \begin{pmatrix} - & c_0 & - \\ - & \sqrt{2}\, c_1 & - \\ & \vdots & \\ - & \sqrt{2}\, c_{n/2-1} & - \\ - & c_{n/2} & - \\ - & \sqrt{2}\, s_1 & - \\ & \vdots & \\ - & \sqrt{2}\, s_{n/2-1} & - \end{pmatrix}$$

where

$$c_0 = \frac{1}{n}(1,\ldots,1)$$

$$c_\nu = \frac{1}{n}(\cos 2\pi\nu\frac{1}{n}\ ,\ \cos 2\pi\nu\frac{2}{n},\ldots,\cos 2\pi\nu\frac{n}{n})$$

$$s_\nu = \frac{1}{n}(\sin 2\pi\nu\frac{1}{n}\ ,\ \sin 2\pi\nu\frac{2}{n},\ldots,\sin 2\pi\nu\frac{n}{n})\ .$$

Note that $WW' = \frac{1}{n}I$.

We let f_λ^C be the minimizer of (4.13) subject to $x \geq 0$. The program QUADPR in the Madison Academic Computing Center Library (MACC, 1977) was used to find x to minimize the right hand side of (4.13) subject to $x \geq 0$. This code employs the principal pivoting method of Cottle (1968). Call the minimizer x_λ. Letting the ith component of x_λ be $x_\lambda(i)$, the indices $i_1,\ldots,i_L$, for which $x_\lambda(i) > 0$ are determined.

Let E be the $n \times L'$ indicator matrix of these indices, that is, E has a 1 in the ith row and jth column if $i = i_j$, $j = 1,2,\ldots,L'$, and zeroes elsewhere. The solution to the problem: minimize

$$||KWx-Wy||^2 + \lambda x'W'JWx$$

subject to $x(i) = 0$ for i not one of $i_1,\ldots,i_{L'}$ is

$$x_\lambda = E(E'W'K'KWE+\lambda E'W'JWE)^{-1}E'W'K'Wy \tag{4.14}$$

Defining g_λ^C by

$$g_\lambda^C(t) = \int_0^1 k(t-s)f_\lambda^C(s)ds$$

where $f_\lambda^C \varepsilon F_n$ satisfies $(f_\lambda^C(\frac{1}{n}),\ldots,f_\lambda^C(\frac{n}{n})) = x_\lambda$, we have $L_i f_\lambda^C = g_\lambda^C(\frac{i}{n})$, and

$$\begin{pmatrix} L_1 f_\lambda^C \\ \vdots \\ L_n f_\lambda^C \end{pmatrix} = nW'KW\, x_\lambda \doteq A_{L'}(\lambda)y \tag{4.15}$$

where

$$A_{L'}(\lambda) = nW'KWE(\textstyle\sum_K+\lambda\sum_J)^{-1}E'W'K'W,$$

with

$$\textstyle\sum_K = E'W'K'KWE,\ \sum_J = E'W'JWE.$$

Therefore (provided all i for which $x_\lambda(i) = 0$ are active constraints!) we have

$$\begin{aligned} n - \sum_{i=1}^{n} \frac{\partial L_k f_\lambda}{\partial y_k} &= Tr(I-A_{L'}(\lambda)) \\ &= n-L'+\lambda TrB \end{aligned}$$

where

$$B = \Sigma_J(\Sigma_K+\lambda\Sigma_J)^{-1},$$

and the approximate cross validation function $V_{approx}(\lambda) = V^C_{approx}(\lambda)$ is

$$V^C_{approx}(\lambda) = \frac{||KWx_\lambda - Wy||^2}{(\frac{1}{n}(n-L'+\lambda TrB))^2} \quad . \tag{4.16}$$

$TrB = Tr\Sigma_J(\Sigma_K+\lambda\Sigma_J)^{-1}$ is computed by first using LINPACK (Dongarra et al (1979)) to solve L' linear systems for B defined by

$$(\Sigma_K+\lambda\Sigma_J)B = \Sigma_J$$

and then computing TrB.

We pause to caution the reader that roundoff error lurks everywhere in calculating with ill posed problems (as this will be if k is at all "smooth"), all calculations must be done in double precision and care must be taken with such simple quantities as $||u-v||^2$ (don't compute $(u,u)-2(u,v)+(v,v)$!).

To get a nice example function h in F_n for our Monte Carlo study, we began with a convenient analytically defined function $h_{oo}(t)$ with $h_{oo}(0) \simeq h_{oo}(1)$, constructed a function $h_o(t)$ satisfying $h_o(0) - h_o(1)$ by setting

$$h_o(t) = h_{oo}(t) + (h_{oo}(0)-h_{oo}(1))t + \tfrac{1}{2}(h_{oo}(1)-h_{oo}(0)).$$

Then we took as our example function h the trigonometric interpolant to h_o via (4.1)-(4.4). For n = 64 the h_{oo} and h we used as example functions cannot be distinguished visually on a $8\frac{1}{2} \times 11$ plot. For our examples we constructed k and several $f's \varepsilon F_n$ from k_{oo} and the f_{oo}'s given below:

$$k_{oo}(t) = \frac{1}{\sqrt{2\pi}s} e^{-t^2/2s^2} + e^{-(1-t)^2/2s^2} \quad , \; s = .043$$

$$f_{oo}(t) = \frac{1}{3} \frac{1}{\sqrt{2\pi}s_1} e^{-(t-.3)^2/2s_1^2} + \frac{2}{3} \frac{1}{\sqrt{2\pi}s_2} e^{-(t-\mu)^2/2s_2^2}$$

where

$$s_1 = .015, \; s_2 = .045$$

and four different f's were generated by letting the peak separation μ-.3 be as in Table 1. In each example $g(t) = \int_0^1 k(t-s)f(s)ds$ is computed from (4.3)-(4.5) given $k(\frac{i}{n})$, $f(\frac{i}{n})$ for $i = 1,2,\ldots,n$. Figure 1 gives a plot of k(t).

Table 1

Example	Peak separation	I_{DOMAIN}	I_{RANGE}
1	.2	1.005	1.002
2	.15	1.016	1.081
3	.10	1.224	1.081
4	.05	6.650	1.318

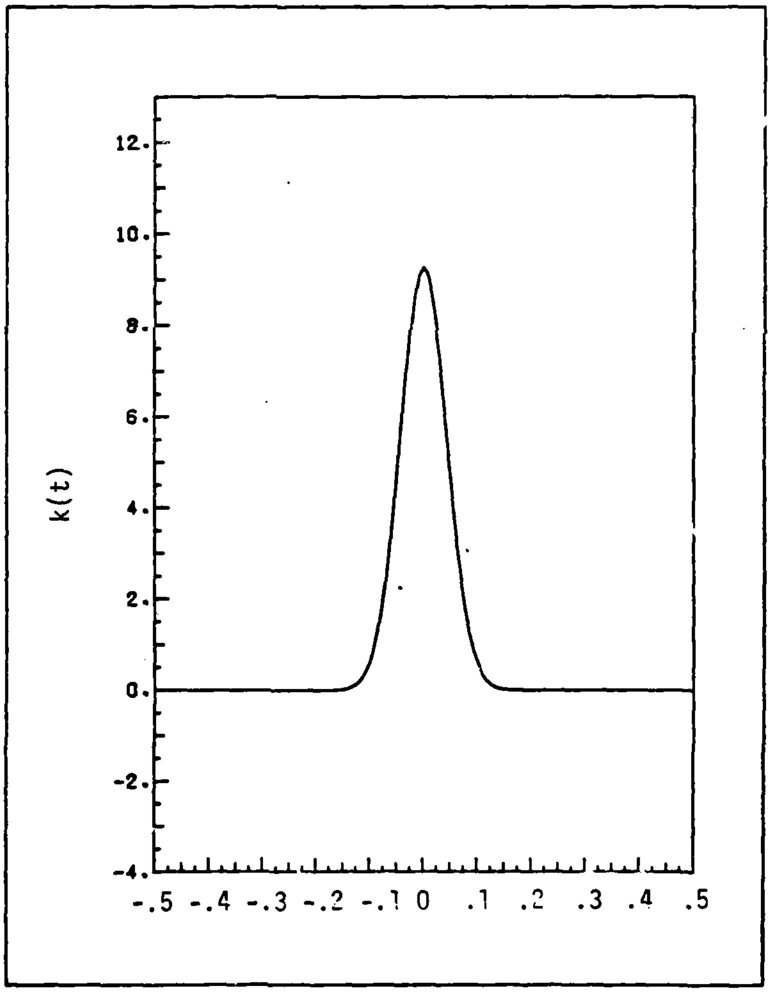
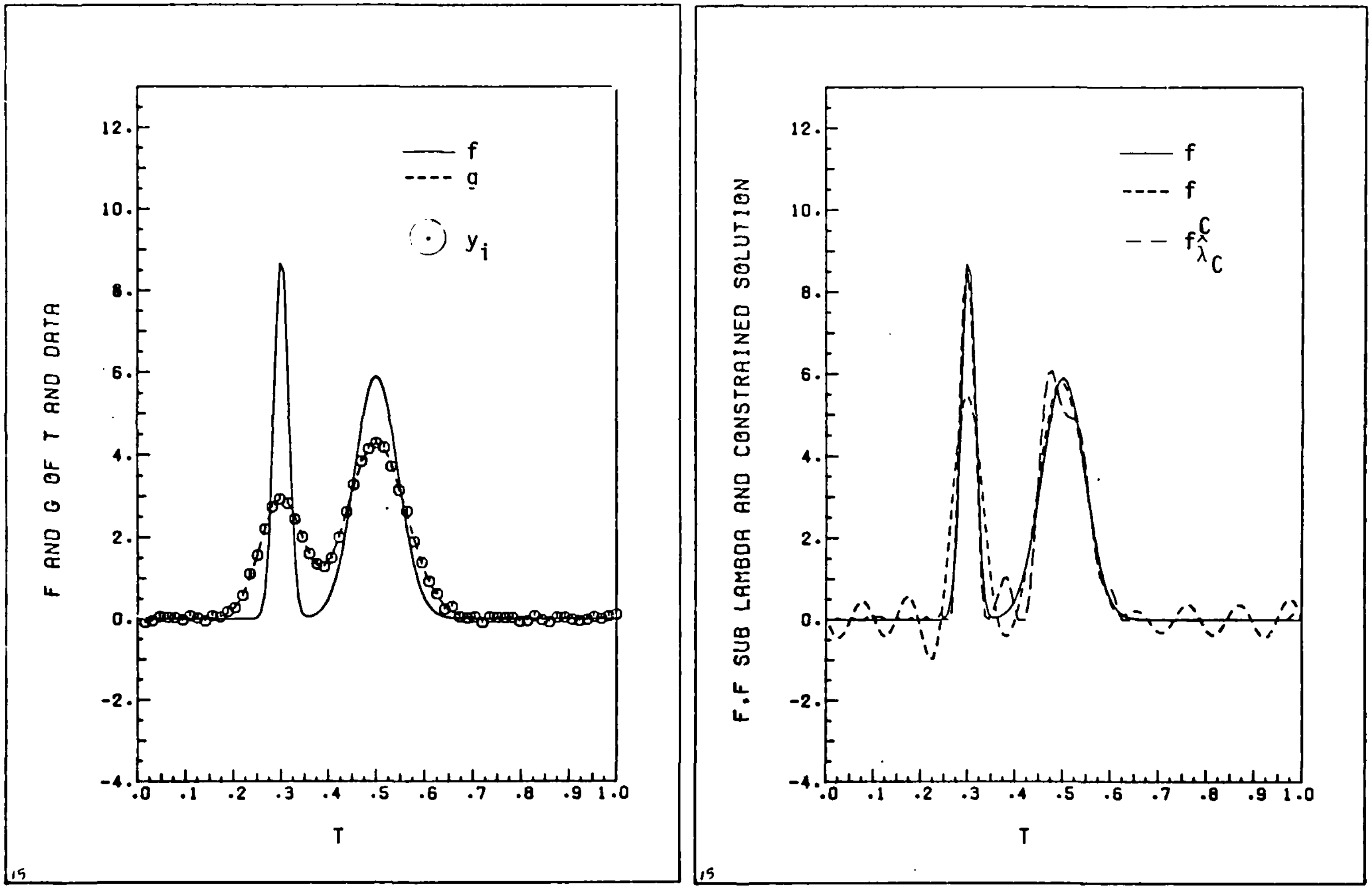


Figure 1. The convolution kernel k(t).

Figures 2a, 3a, 4a and 5a give f(t), $g(t) = \int_0^1 k(t-s)f(s)ds$, and $y_i = g(\frac{i}{n}) + \varepsilon_i$, for examples 1-4, where the ε_i were i.i.d. $N(0,\sigma^2)$ pseudo random variables with $\sigma = .05$. Figures 2b, 3b, 4b and 5b give f, $f_{\hat\lambda}$ and $f^C_{\hat\lambda_C}$ for these same 4 examples. $\hat\lambda$ is the minimizer of $V(\lambda)$ for unconstrained problems given by (4.12) and computed by evaluating $V(\lambda)$ at equally spaced increments in $\log_{10}\lambda$, performing a global search, evaluating $V(\lambda)$ at a finer set of equally spaced increments centered at the previous minimum etc. The final search is performed on $V(\lambda)$ evaluated at increments of $\frac{1}{9}$ in $\log\lambda$. $\hat\lambda$ is the minimizer of $V^C_{approx}(\lambda)$ of (4.16). In these examples the minimum was found by evaluating $V^C_{approx}(\lambda)$ at values of λ satisfying $\log\lambda - \log\hat\lambda = j(.1)$ for $j = 0, \pm 1, \ldots$, etc. The possible perils of this process will be discussed later.

In each example, a "ringing" phenomena in the unconstrained solution is very evident. Intuitively, the approximate solution retains some high frequency components in an attempt to capture the two narrow peaks. In each of the four examples the imposition of positivity constraints provided a dramatic improvement in the solution. Anyone who has attempted a numerical solution of an ill posed problem knows that the visual character of the solution can vary significantly with λ (and to a lesser extent with m, given the optimal λ for that m.) In the unconstrained solutions, the cross validation estimate of λ was near optimal in Examples 1 and 2, good in Example 3 and poor (from the point of mean square error of the solution) in Example 4. The data behind this remark are given in Table 1. The inefficiencies I_{DOMAIN} and I_{RANGE} in that table are defined by



(a) (b)

Figure 2. f, g, data, $f_{\hat{\lambda}}$ and $f^C_{\hat{\lambda}_C}$ for Example 1, peak separation = .2.

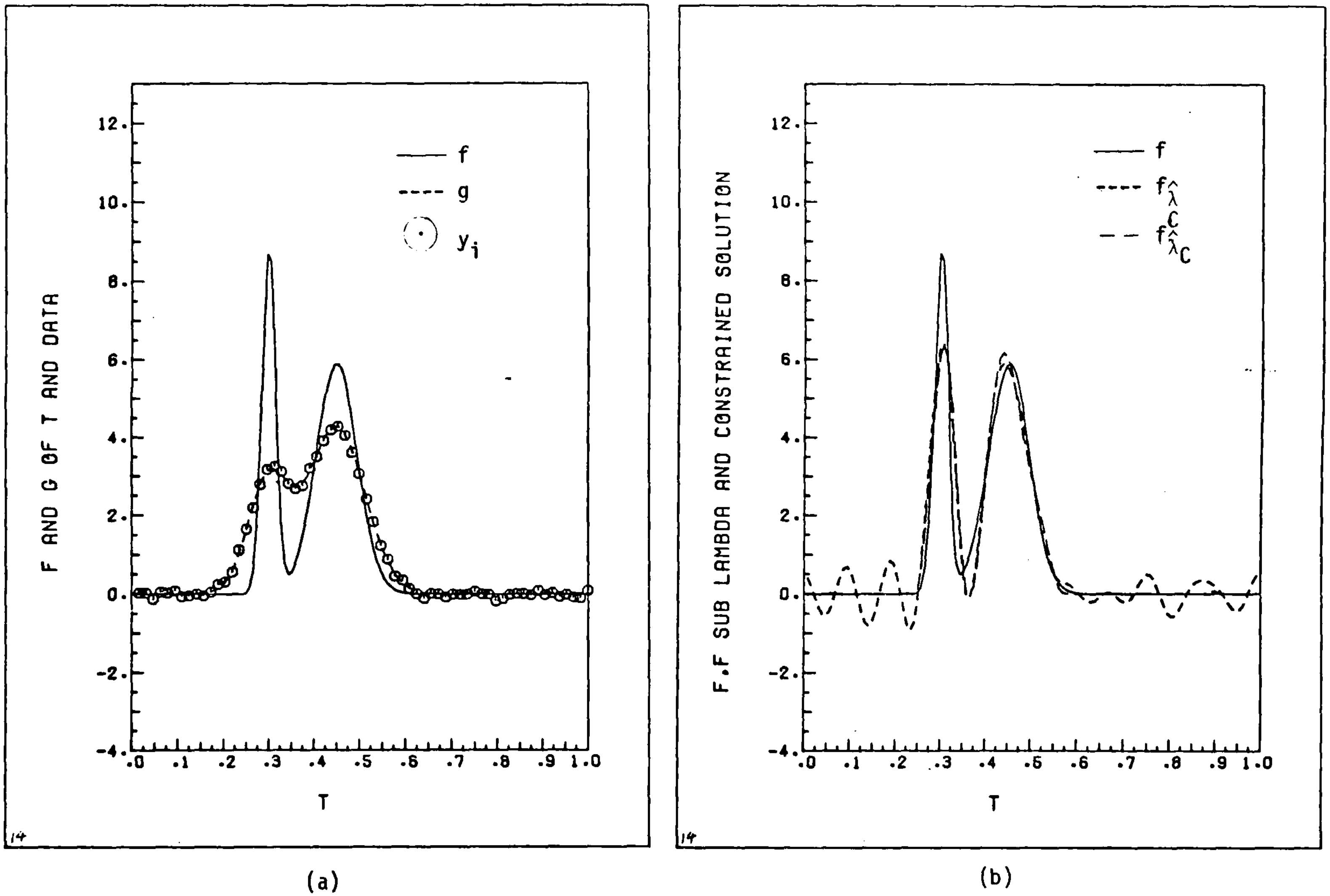


Figure 3. f, g, data, $f_{\hat\lambda}$ and $f^C_{\hat\lambda_C}$ for Example 2, peak separation = .15.

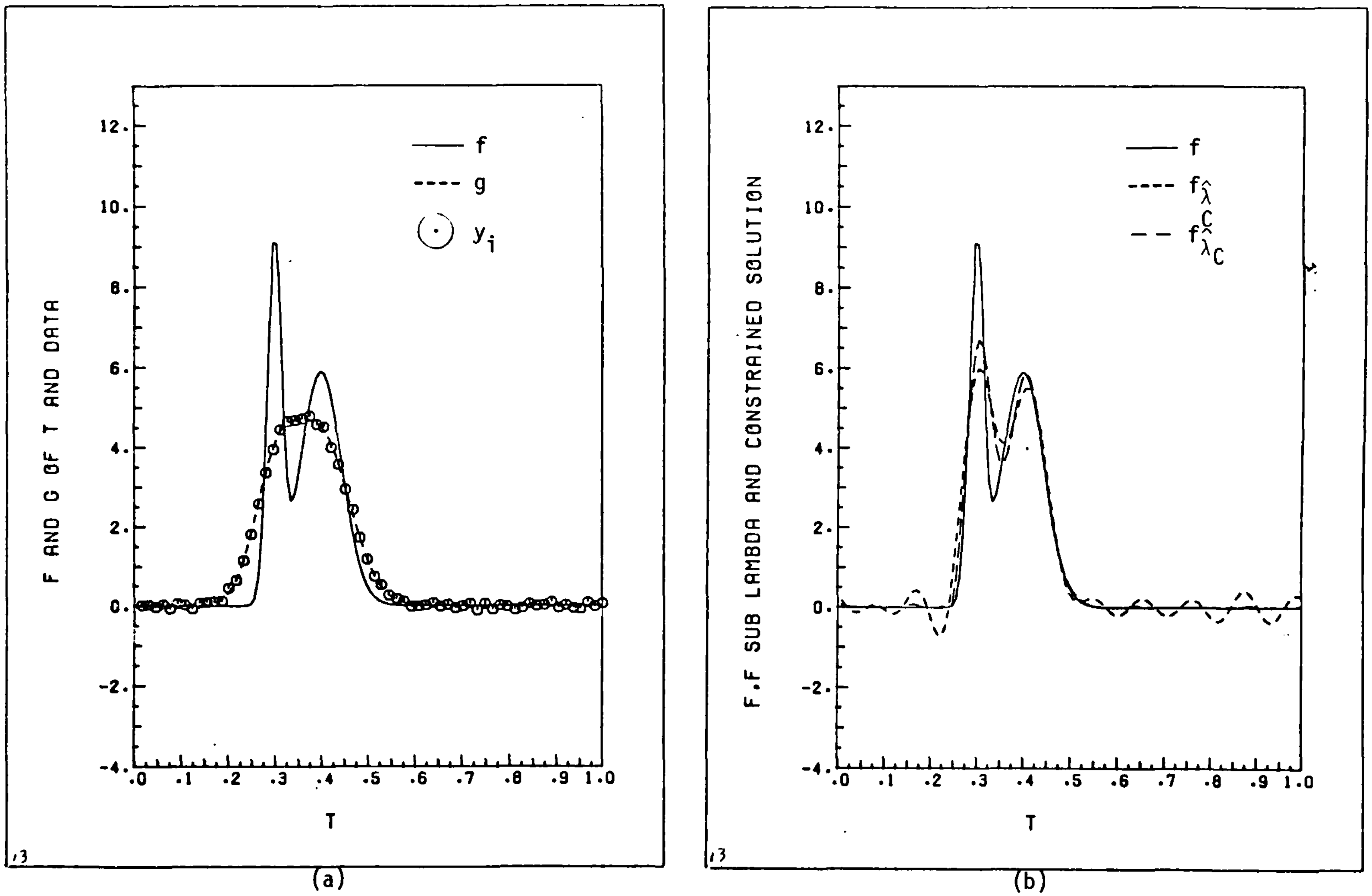


Figure 4. f, g, data, $f_{\hat\lambda}$ and $f^C_{\hat\lambda_C}$ for Example 3, peak separation = .10.

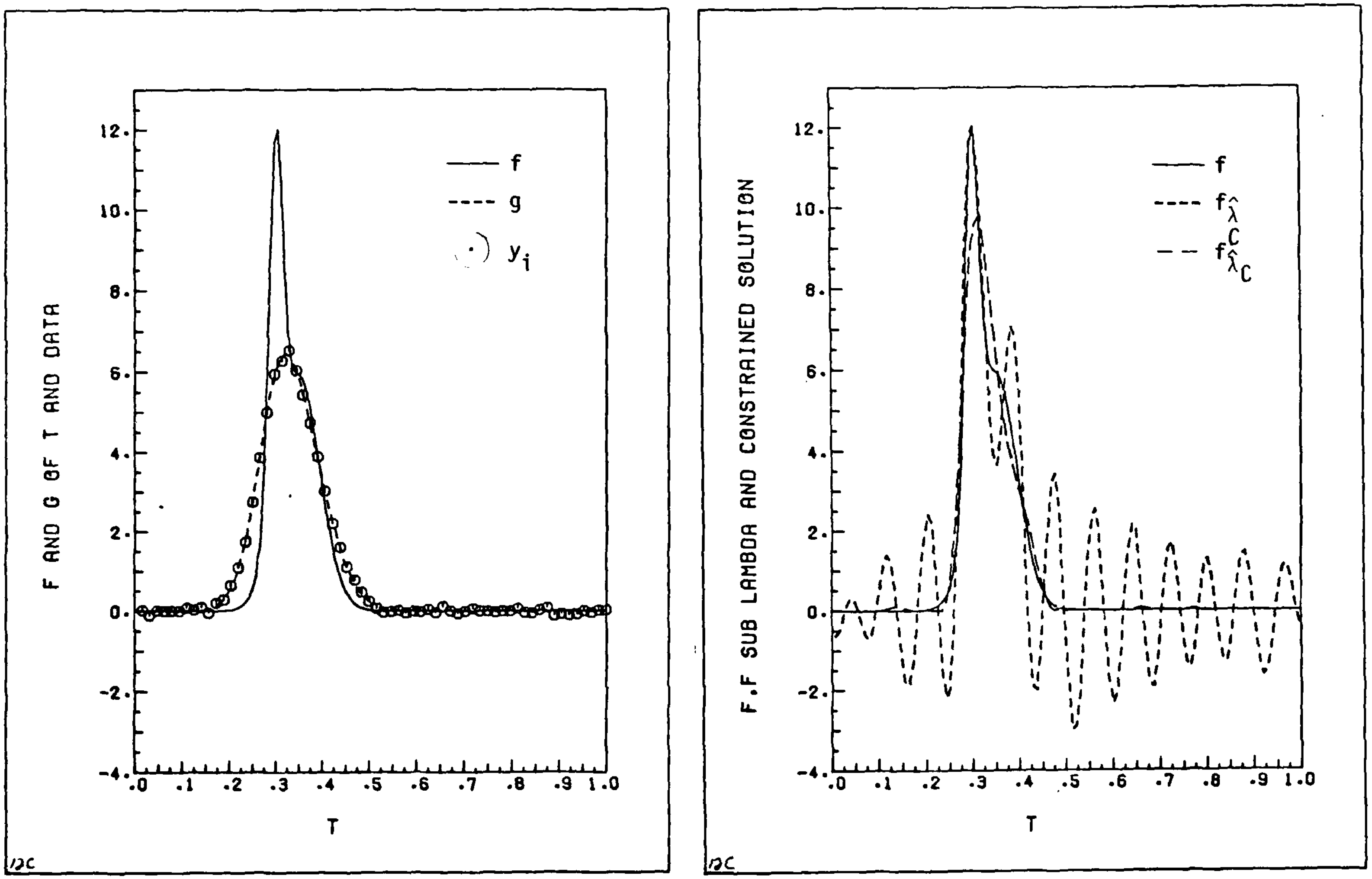


Figure 5. f, g, $f_{\hat{\lambda}}$ and $f^C_{\hat{\lambda}_C}$ for Example 3, peak separation = .05.

$$I_{DOMAIN} = \frac{\frac{1}{n}\sum_{i=1}^{n}(f_{\hat{\lambda}}(\frac{i}{n})-f(\frac{i}{n}))^2}{\min_{\lambda} \frac{1}{n}\sum_{i=1}^{n}(f_{\lambda}(\frac{i}{n})-f(\frac{i}{n}))^2}$$

$$I_{RANGE} = \frac{\frac{1}{n}\sum_{i=1}^{n}(g_{\hat{\lambda}}(\frac{i}{n})-g(\frac{i}{n}))^2}{\min_{\lambda} \frac{1}{n}\sum_{i=1}^{n}(g_{\lambda}(\frac{i}{n})-g(\frac{i}{n}))^2}$$

The theory (Equation (3.9)) concerning the GCV estimate $\hat{\lambda}$ says (roughly) that $I_{RANGE} = (1+o(1))$ as $n\to\infty$.

We now discuss Example 3 in greater detail. Figure 6 gives the mean square error of f_λ, f_λ^C, g_λ and g_λ^C as a function of λ. ($MSE(f_\lambda) = \frac{1}{n}\sum_{i=1}^{n}(f_\lambda(\frac{i}{n})-f(\frac{i}{n}))^2$, etc.). We have taken the origin as $\log\hat{\lambda}(\log\hat{\lambda}=-9.889)$. Since the GCV estimate of λ estimates the minimizer of $MSE(g_\lambda)$ or $MSE(g_\lambda^C)$, it will generally be a good estimate of the minimizer of $MSE(f_\lambda)$ or $MSE(f_\lambda^C)$ to the extent that $MSE(f_\lambda)$ and $MSE(g_\lambda)$, or $MSE(f_\lambda^C)$ and $MSE(g_\lambda^C)$ have the same minimizer. The minimizers of the four curves are marked by arrows. In these and other cases we have tried ($n\varepsilon[30,100]$, smooth f, σ a few percent of $\max_t|g(t)|$), the optimal λ for $MSE(f_\lambda)$ and $MSE(g_\lambda)$ appear to be close, as a practical matter. As a theoretical phenomena for large n it may or may not be true, see Lukas (1981) for some asymptotic results on the optimal λ for different loss functions in the unconstrained case.

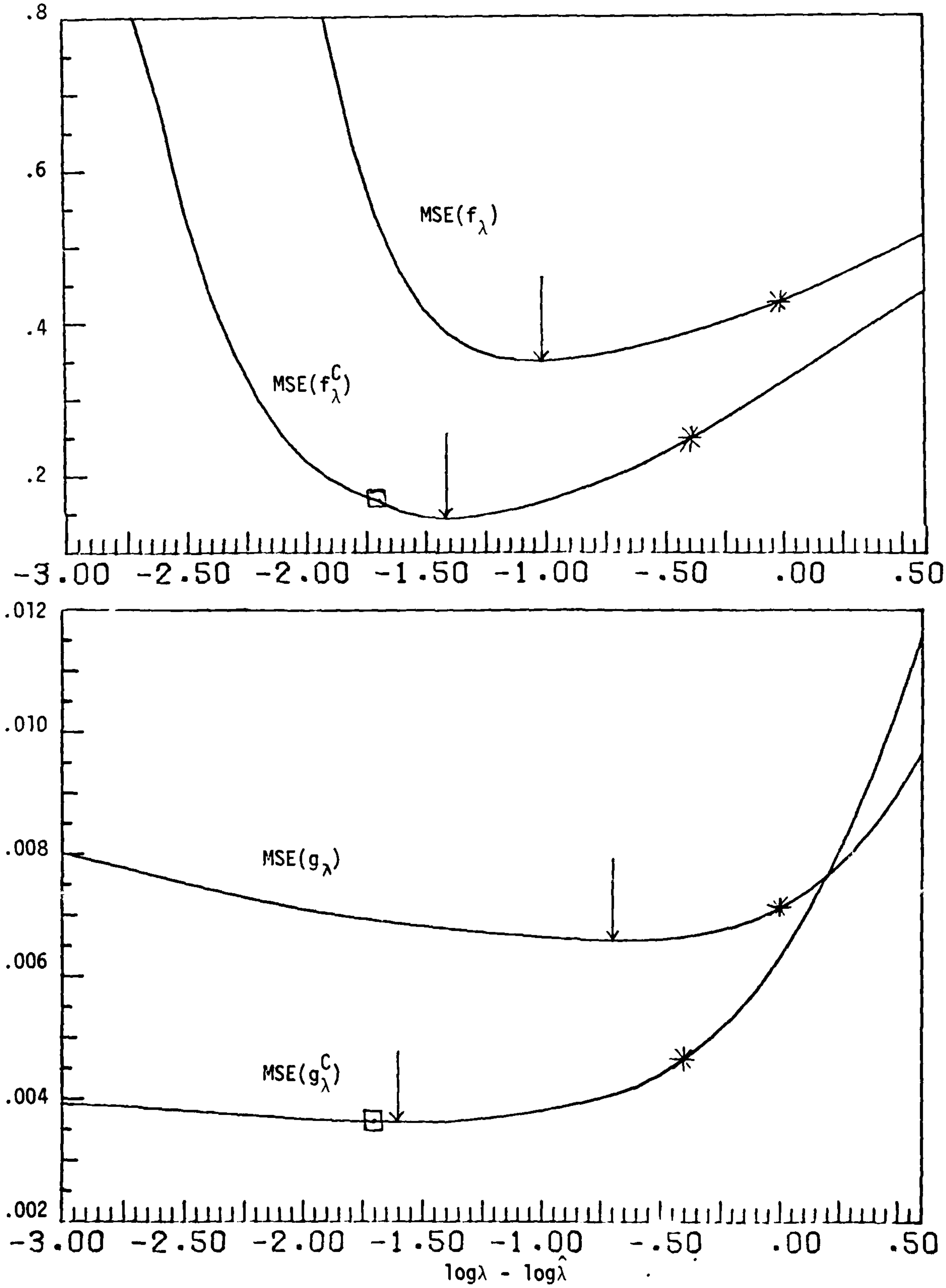


Figure 6. *Comparison of mean square error of estimates of f and g, as a function of* λ.

Figure 7 gives $V(\lambda)$ of (4.12), $V^C_{approx}(\lambda)$ of (4.16) and $V^C(\lambda)$ of (3.7) for Example 3. $V(\lambda)$ and $V^C_{approx}(\lambda)$ were computed at increments of .1 in $\log\lambda$. $\hat{\lambda}_C$ was taken as the global minimizer of the computed V^C_{approx} values. V and V^C_{approx} at their respective minimizers $\hat{\lambda}$ and $\hat{\lambda}_C$ are marked by a large *. In Figure 6, the corresponding MSE values at $\hat{\lambda}$ and $\hat{\lambda}_C$ are also marked by a large *. In Figure 7, some of the computed values of V^C_{approx} have been connected by a smooth curve. Two adjacent points have not been connected if the set of active constraints is different for the two corresponding values of λ. V^C_{approx} can be expected to have at least one discontinuity somewhere between the two corresponding values of λ, (including the end points). Although the estimates $\hat{\lambda}_C$ worked well in this and the other three examples tried, there are obvious pitfalls in minimizing a discontinuous function, e.g. sensitivity to the increment in $\log\lambda$.

We decided to invest a fair amount of computer time to compute $V^C(\lambda)$ for this one example. The computed values are indicated by $\square$ in Figure 7. The computation was attempted for $\log\lambda-\log\hat{\lambda}$ from -3.00 to .6 in steps of .1. There are missing values whenever the quadratic optimization routine QUADPR terminated with an error message. This happened during the constrained minimization of the leaving out one version of (4.13) in the process of calculating a^*_{kk} of (3.4), for some k (typical error message: "no complement variable found"). Nevertheless it appears possible to connect the computed values by a smooth curve and find the minimum by a global search in a neighborhood about or below $\hat{\lambda}$.

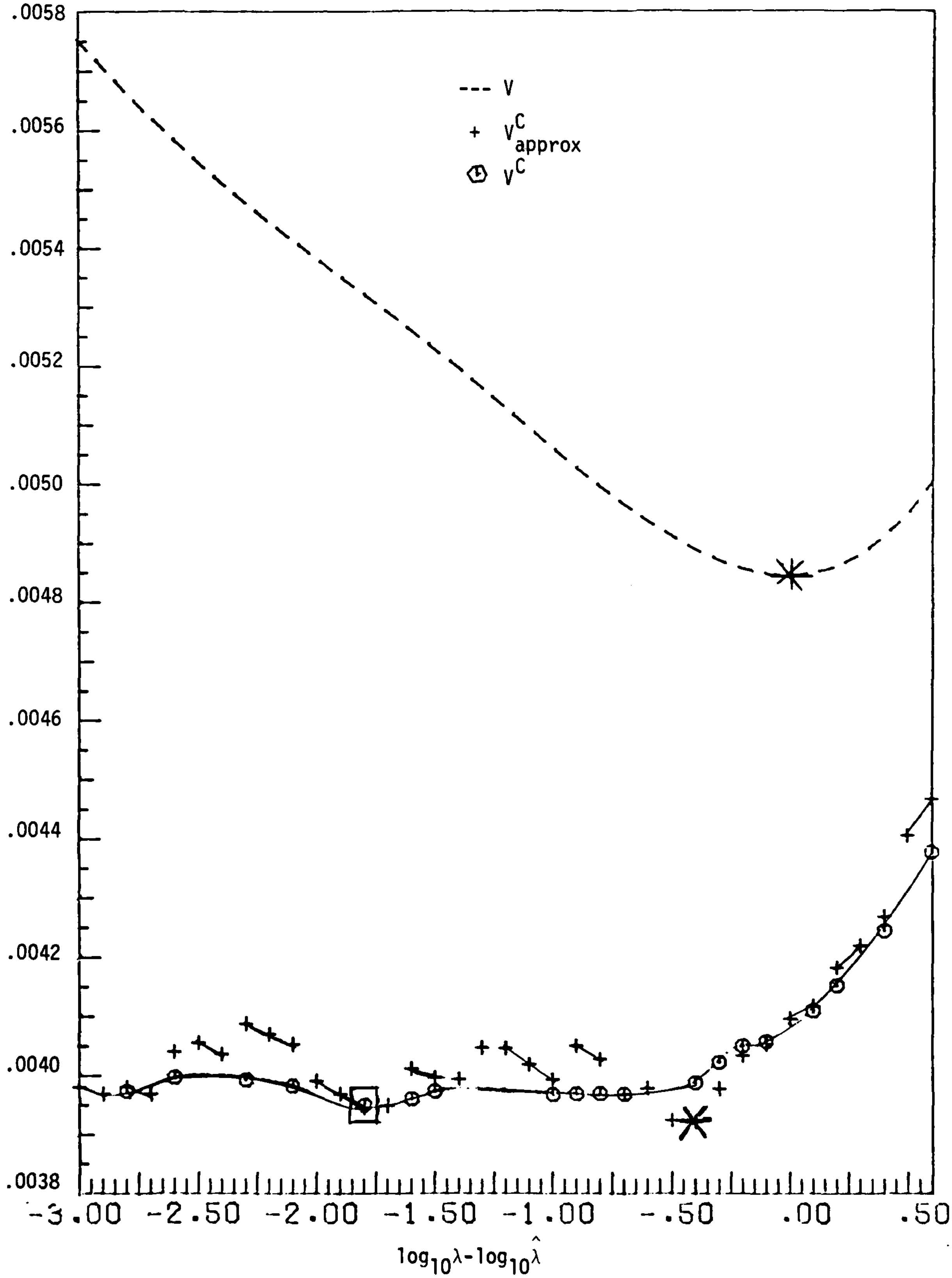


Figure 7. V, V^C_{approx} and V^C

V^C at its global minimizer is marked by $\square$ in Figure 7, and the MSE curves for f_λ^C and g_λ^C in Figure 6 are also marked by a $\square$ at the minimizer of V^C. Out of concern for the computational failures with QUADPR noted above, it was decided to try this example for n = 50. The difficulty of the quadratic program increases with n. Two replications were tried. In the first, $V^C(\lambda)$ as well as $V^C_{approx}(\lambda))$ was successfully computed for $\log\lambda - \log\hat{\lambda}$ in steps of .1 from -2.4 to .6. The CPU time for n = 50 was around $\frac{1}{2}(\sim(\frac{50}{64})^3)$ times that for n = 64. $V^C(\lambda)$ was visually smooth and convex near its minimum when plotted to the same scale as Figure 7 (equivalently, to 3 but not 4 significant figures). V^C_{approx} showed the same apparently piecewise continuous behavior as in the example for n = 64. Both functions had their global minimizers at $\log\lambda - \log\hat{\lambda} = -.7$ while $MSE(f_\lambda^C)$ was minimized at $\log\lambda - \log\hat{\lambda} = -.8$, for an I^C_{DOMAIN} of 1.009 (I^C_{DOMAIN} is defined analogously to I_{DOMAIN} with f replaced by f^C, etc.) In the second replication the computation of a $V^C(\lambda)$ for a few scattered values of λ terminated in an error message but nevertheless a minimum of $V^C(\lambda)$ was easily found, and resulted in I^C_{DOMAIN} of 1.02.

The innocuous-looking convolution equation we have studied here is very ill posed, a phenomena surprisingly common in many experiments. We may write

$$y = nW'KWx + \varepsilon,$$

thus the design matrix X is nW'KW. If k is symmetric (as it is here), then the η_i's are all 0 and K is diagonal. Table 2 gives the ξ_ν's of (4.2) and (4.13), which are also the singular values of the design matrix. $\xi_1, \ldots, \xi_{n/2-1}$ are of multiplicity 2. Also given in Table 2 are the α_ν, β_ν, $\hat{\alpha}_\nu$ and $\hat{\beta}_\nu$ defined by (4.3) and (4.9), with $\lambda = \hat{\lambda}$. If ξ_ν is

	Fourier coefficients of f		Fourier coefficients of $f_{\hat{\lambda}}$		Singular values of X (Eigenvalues of K)
ν	α_ν	β_ν	$\hat{\alpha}_\nu$	$\hat{\beta}_\nu$	ξ_ν
0	1.0000000	--	1.0056062		1.0000000
1	-0.6207604	0.6921165	-0.6215352	0.6961828	0.9641602
2	-0.0893528	-0.7328304	-0.0848551	-0.7344837	0.8641653
3	0.4028712	0.2542137	0.4029176	0.2429338	0.7200172
4	-0.1885802	0.0365562	-0.1962951	0.0955639	0.5576829
5	-0.0528776	0.0200001	-0.0537360	-0.0060144	0.4215413
6	-0.0401296	-0.1772403	-0.0061505	-0.1516747	0.2687643
7	0.2459903	0.2681774	0.2405176	0.0936229	0.1672289
8	-0.1869963	0.1965549	-0.1173723	0.1667934	0.0967274
9	-0.0930543	-0.2366141	-0.0024572	-0.1934057	0.0522099
10	0.2260386	-0.0000008	0.0572545	-0.0546176	0.0259969
11	-0.0644608	0.1883329	0.0276649	0.0190262	0.0120796
12	-0.1416100	-0.1053629	-0.0047038	-0.0089655	0.0052175
13	0.1275489	-0.0917606	0.0215188	0.0017124	0.0020952
14	0.0429244	0.1325941	0.0000653	-0.0001728	0.0007821
15	-0.1226323	-0.0000074	-0.0000249	-0.0000563	0.0002714
16	0.0330138	-0.1016864	-0.0000495	0.0000217	0.0000876
17	0.0747306	0.0542659	-0.0000043	-0.0000022	0.0000263
18	-0.0639485	0.0464957	-0.0000029	-0.0000012	0.0000073
19	-0.0207693	-0.0637932	0.0000007	-0.0000003	0.0000019
20	0.0564273	-0.0000585	0.0000000	0.0000000	0.0000005
21	-0.0144560	0.2447695	0.0000000	0.0000000	0.0000001
22	-0.0315319	-0.0227541	-0.0000000	0.0000000	0.0000000
23	0.0256734	-0.0188771	0.0000000	-0.0000000	-0.0000000
24	0.0082505	0.0245512	0.0000000	0.0000000	0.0000000
25	-0.0208860	0.0003692	0.0000000	0.0000000	0.0000000
26	0.0046742	-0.0160679	-0.0000000	-0.0000000	0.0000000
27	0.0112024	0.0072426	-0.0000000	0.0000000	0.0000000
28	-0.0079654	0.0070270	0.0000000	-0.0000000	0.0000000
29	-0.0039043	-0.0075671	-0.0000000	0.0000000	-0.0000000
30	0.0067293	-0.0018679	-0.0000000	-0.0000000	-0.0000000
31	0.0006966	0.0059954	0.0000000	-0.0000000	-0.0000000
32	-0.0057113		-0.0000000		-0.0000000

Eigenvalues of the design matrix and true and (unconstrained) estimates of Fourier coefficients of the solution, Example 3.

Table 2.

sufficiently small then α_ν, β_ν are not estimable with double precision arithmetic and it is seen that $\hat{\alpha}_\nu$ and $\hat{\beta}_\nu$ are 0 (to as many figures as we have printed). Although XX' is theoretically of full rank (64), the 40th largest eigenvalue is around 10^{-14} times the largest.

From the examples we have studied, it appears that the imposition of positivity constraints can be an important source of information in very ill posed problems, and that the GCV estimate for λ for constrained problems, and its approximate version appear to do a good job of estimating λ. Of course not all problems will show such a dramatic improvement, with the imposition of constraints, since, if no constraints are active, then no information has been added. In some sense the examples tried here were chosen in anticipation of negative unconstrained solutions (and, we must admit, with some subjective hunches on the part of the author concerning the type of problem the method is likely to do well on).

The evaluation of $V^C(\lambda)$ required n + 1 calls to QUADPR at a cost per call for n = 64 of around 5 to 8 seconds CPU time on the Madison UNIVAC 1110 while the computation of $V^C_{approx}(\lambda)$ requires one such call. It is possible that a clever search procedure utilizing information from $V(\lambda)$ or $V^C_{approx}(\lambda)$ could be used to obtain the minimizer of $V^C(\lambda)$ with a small number of functional evaluations, particularly with an improved quadratic optimation routine. On the other hand the minimizer of V^C_{approx} may be adequate in many situations. It is clear that both the exact and the approximate GCV method warrants further study, both theoretically and numerically.

5. Acknowledgments

We thank D. Chambless for providing us early drafts of his work involving positivity constraints, A.R. Davies and P. Merz for stimulating conversations concerning problems with nonnegative solutions, E. Wegman for encouragement, J. Fleisher for help in using QUADPR and last but not least C. Sheridan, who wrote the computer program.

REFERENCES

Adams, R.A. (1975). Sobolev Spaces. Academic Press, New York.

Anderssen, R.S., and Jakeman, A.J. (1975). Abel type integral equations in stereology,II. Computational methods of solution and the random spheres approximation. J. Microscopy 105, 2, 135-153.

Anderssen, R.S., de Hoog, F.R., and Lukas, M.A., eds. (1980). "The application and numerical solution of integral equations". Sijthoff and Noordhoff.

Anselone, P.M. and Laurent, P.J. (1968). A general method for the construction of interpolating or smoothing spline-functions. Numerische Mathematik 12, 66-82.

Aronszajn, N. (1950). Theory of reproducing kernels. Transactions of the American Mathematical Society 68, 337-404.

Chambless, D.A. (1980). Radiological data analysis in the time and frequency domain II. Auburn University, Department of Mathematics, Montgomery, AL, report.

Cottle, R.W. (1968). The principal pivoting method of quadratic programming. Mathematics of the Decision Sciences, 1, 144-162.

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.

Davies, A.R. (1979). The numerical inversion of integral transforms in laser anemometry and photon correlation. To appear, Proceedings of the International Conference on Ill Posed Problems, M.Z. Nashed, ed.

Dongarra, J.J., Moler, C.B., Bunch, J.R., and Stewart, G.W. (1979). LINPACK User's Guide. SIAM, Philadelphia, PA.

Fritz, S., Wark, D.Q., Fleming, J.E., Smith, W.P., Jacobowitz, H., Hilleary, D.T. and Alishouse, J.C. (1972). Temperature sounding from satellites, NOAA Technical Report NESS 59, National Oceanic and Atmospheric Administration, Washington, D.C.

Gamber, H. (1979). Choice of an optimal shape parameter when smoothing noisy data. Commun. Statist. A8, 14, 1425-1436.

Geisser, S. (1975). The predictive sample reuse method with applications. J. Amer. Statist. Assoc., 70, 320-328.

Golberg, M.A. (1978), ed. "Solution methods for integral equations, Theory and Applications". Plenum Press, New York.

Golub, G., Heath, M. and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21, 215-223.

Gorenflo, P. and Hilpert, M. (1980). On the continuity of convexly constrained interpolation, in "Approximation Theory III", E.W. Cheney, ed., Academic Press, 449-454.

Herman, G.T., and Natterer, F. (1981). "Mathematical aspects of computerized tomography". Springer-Verlag, New York.

Hudson, H.M. (1974). Empirical Bayes Estimation, Technical Report No. 58, Stanford University, Department of Statistics, Stanford, CA.

IMSL (International Mathematical and Statistical Library)(1980). Version 8, Subroutine ICSSCV.

Kimeldorf, G., and Wahba, G. (1971). Some results on Tchebycheffian spline functions, J. Math. Anal. and Applic., 33, 1, 82-95.

Lukas, M. (1981). Regularization of linear operator equations, Thesis, Department of Pure Mathematics, Australian National University, Canberra.

MACC (Madison Academic Computing Center), University of Wisconsin-Madison (1977). QUADPR/QUADMP Quadratic Programming Subroutines. Madison, WI.

Mallows, C.L. (1973). Some comments on C_p. Technometrics 14, 661-675.

Merz, P.H. (1980). Determination of adsorption energy distribution by regularization and a characterization of certain adsorption isotherms. J. Comput. Physics 38, 64-85.

Mosteller, F., and Tukey, J.W. (1968). Data analysis including Statistics, in "Handbook of Social Psychology", Vol. 2. Addison-Wesley, Reading Madd. 80-203.

Nashed, M.A., ed. (1981). Proceeding of the International Conference on Ill-Posed Problems held at Newark, Delaware, November 2-6, 1979, to appear.

Nychka, D. In preparation.

Shepp, L.A. (1966). Radon-Nikodym derivatives of Gaussian measures. Ann. Math. Statist. 37, 2, 321-354.

Shepp, L.A. and Kruskal, J.B. (1978). Computerized tomography: The new medical x-ray technology. Amer. Math. Monthly 85, 420-439.

Smith, W.L., Woolf, H.M., Hayden, C.M., Wark, D.Q., and McMillin, L.M. (1979). The TIROS-N Operational vertical sounder. Bull. American Meteorological Society, 60, 10, 1177-1187.

Speckman, P. (1981). Spline smoothing and optimal rates of convergence in nonparametric regression models, UNiversity of Oregon, manuscript.

Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. Ann. Statist. 8, 6, 1348-1360.

Stone, M. (1974). Cross-validitory choice and assessment of statistical prediction, JRSS, Series B, 36, 2, 111-147.

Tihonov, A.N. and Arsenin, V.Y. (1977). "Solutions of ill-posed problems". Translation editor Fritz John, V.H. Winston and Sons, Washington, D.C.

Twomey, S. (1977). "Introduction to the mathematics of inversion in remote sensing and indirect measurements." Elsevier, New York.

Utreras, F. (1979). Cross validation techniques for smoothing spline functions in one or two dimensions. In "Smoothing Techniques for Curve Estimation". T. Gasser and M. Rosenblatt, eds. Lecture Notes in Mathematics, No. 757, Springer-Verlag, Verlin.

Utreras, F. (1978). Quelques resultats d'optimalite pour la methode de validation crossee. Seminaire d'Analyse Numerique No. 301, Universite Scientifique et Medicale de Grenoble, Grenoble, France.

Wahba, G. (1973). On the minimization of a quadratic functional subject to a continuous family of linear inequality constraints, SIAM J. Control, 11, 1.

Wahba, G. (1977a). Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numerical Analaysis, 14, 4,651-667.

Wahba, G. (1977b). Comments to "Consistent nonparametric regression, by C.J. Stone, Ann. Statist., 5, 4, 647-640.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. J. Roy. Stat. Soc. Ser. B., 40, 3.

Wahba, G. (1979a). Smoothing and ill posed problems, in "Solution Methods for Integral Equations with Applications". Michael Golberg, ed., Plenum Press, 183-194.

Wahba, G. (1979b). Convergence rates of "thin plate" smoothing splines when the data are noisy in "Smoothing Techniques for Curve Estimation". T. Gasser and M. Rosenblatt, eds. Springer-Verlag, Heidelberg, 232-245.

Wahba, G. (1980a). Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data, in"Approximation Theory III",E.W. Cheney, ed. 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 Department of Statistics Technical Report No. 595.

Wahba, G. (1980c). Cross validation and constrained regularization methods for mildly ill posed problems. University of Wisconsin-Madison Technical Report No. 629, to appear, Proceedings of the International Conference on Ill Posed Problems. M.Z. Nashed, ed. Academic Press.

Wahba, G. (1981). Numerical experiments with the thin plate histospline. University of Wisconsin-Madison, Department of Statistics Technical Report No. 638, to appear, Commun. Statist. A.

Wahba, G. and Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross validation. Monthly Weather Review, 108, 8, 1122-1143.

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

Westwater, E.D. (1979). Ill posed problems in remote sensing of the earth's atmosphere by microwave radiometry. Manuscript, to appear, Proceedings of the International Confernece on Ill-Posed Problems, M.Z. Nashed, ed., Academic Press.

Watson, G.S. (1971). Estimating functionals of particle size distributions, Biometrika 58, 3, 483-490.

Wicksell, S.D. (1925). The corpuscle problem, Part I. Biometrika 17, 87-97.

Wong, W.H. (1980). An analysis of the volume-matching problem and related topics in smooth density estimation. Ph.D. thesis, University of Wisconsin-Madison.

Wright, I.W. and Wegman, E.J., (1980). Isotonic, convex and related splines. Ann. Statist. 8, 5, 1023-1035.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 646	2. GOVT ACCESSION NO. A110 769	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CONSTRAINED REGULARIZATION FOR ILL POSED LINEAR OPERATOR EQUATIONS, WITH APPLICATIONS IN METEOROLOGY AND MEDICINE		5. TYPE OF REPORT & PERIOD COVERED Scientific Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Grace Wahba		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0675
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics 1210 W. Dayton St. Madison, WI 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA		12. REPORT DATE August 1981 13. NUMBER OF PAGES 41
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

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

Ill posed problems, constrained regularization, cross validation in the presence of constraints.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see reverse sheet)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

The relationship between certain regularization methods for solving ill posed linear operator equations and ridge methods in regression problems is described. The regularization estimates we describe may be viewed as ridge estimates in a (reproducing kernel) Hilbert space H. When the solution is known a priori to be in some closed, convex set in H, for example, the set of nonnegative functions, or the set of monotone functions, then one can propose regularized estimates subject to side conditions such as nonnegativity, monotonicity, etc. Some applications in medicine and meteorology are described. We describe the method of generalized cross validation for choosing the smoothing (or ridge) parameter in the presence of a family of linear inequality constraints. Some successful numerical examples, solving ill posed convolution equations with noisy data, subject to nonnegativity constraints, are presented. The technique appears to be quite successful in adding information, doing nearly the optimal amount of smoothing, and resolving distinct peaks in the solution which have been blurred by the convolution operation.