WSD 430

DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin 53706

TECHNICAL REPORT NO. 430

September 1975

PRACTICAL APPROXIMATE SOLUTIONS TO LINEAR OPERATOR EQUATIONS WHEN THE DATA ARE NOISY

by

Grace Wahba

Department of Statistics
University of Wisconsin-Madison

MAR 18 1976

This work was supported by the U. S. Air Force Office of Scientific Research under Grant AF-AFOSR-2363-C.

Abstract

We consider approximate solutions $f_{n,\lambda}$ to linear operator equations $\mathcal{K}f = g$, of the form: $f_{n,\lambda}$ is the minimizer in $\mathcal{H}$ of $\frac{1}{n}\sum_{j=1}^{n}[(\mathcal{K}h)(t_j)-y(t_j)]^2 + \lambda\|h\|^2$, where $\mathcal{H}$ is a Hilbert space, and the data $\{y(t_j)\}$ satisfy $y(t_j) = g(t_j) + \epsilon(t_j)$, the $\{\epsilon(t_j)\}$ being measurement errors. $f_{n,\lambda}$ is the so-called regularized solution, and $\lambda > 0$ is the regularization parameter, to be chosen. It is important to choose λ correctly. The purpose of this paper is to propose the method of weighted cross-validation for choosing λ from the data. We suppose that g is very smooth and the errors are white noise. It is shown that the weighted cross-validation estimate $\hat{\lambda}$ estimates the value of λ which minimizes $\frac{1}{n} E \sum_{j=1}^{n}[(\mathcal{K}f_{n,\lambda})(t_j)-(\mathcal{K}f)(t_j)]^2$. Results related to the convergence of $\|f-f_{n,\hat{\lambda}}\|$, including rates, are obtained.

PRACTICAL APPROXIMATE SOLUTIONS TO LINEAR OPERATOR EQUATIONS WHEN THE DATA ARE NOISY

Grace Wahba

1. Introduction.

We consider the approximate solution of the Fredholm integral equation of the first kind: 1/

$$(1.1) \qquad \int_0^1 K(t,s)f(s)ds = g(t), \quad t \in [0,1]$$

where g is measured discretely and with errors. That is, the data are $y(t) = g(t) + \epsilon(t)$, $t = t_1, t_2, \ldots, t_n$, where the $\{\epsilon(t_j)\}$ are random errors. We suppose that $f \in \mathcal{H}_R$, where $\mathcal{H}_R$ is either $\mathcal{L}_2[0,1]$ or a reproducing kernel Hilbert Space (r.k.h.s.) with reproducing kernel (r.k.) $R(s,t)$, $s,t \in [0,1]$. We assume that the kernel $Q(s,t)$ given by

$$(1.2) \qquad \begin{aligned} Q(s,t) &= \int_0^1 K(s,u)K(t,u)du && \text{if } \mathcal{H}_R = \mathcal{L}_2[0,1] \\ Q(s,t) &= \int_0^1\int_0^1 K(s,u)K(t,v)R(u,v)dudv && \text{otherwise} \end{aligned}$$

is continuous on $[0,1] \times [0,1]$. We use a form of the <u>method of</u>

1/ Generalizations are given in Section 4.

regularization: The regularized solution $f_{n,\lambda}$ to (1.1), given data $y(t_1), y(t_2),\ldots,y(t_n)$, is taken as the solution to: Find $h \in \mathcal{H}_R$ to minimize

$$\frac{1}{n}\sum_{j=1}^{n} [(\mathcal{K}h)(t_j) - y(t_j)]^2 + \lambda\|h\|_R^2 , \quad (\mathcal{K}h)(t) = \int_0^1 K(t,s)h(s)ds . \tag{1.3}$$

Here $\|\cdot\|_R$ is the norm in $\mathcal{H}_R$, and $\lambda > 0$ is the regularization or smoothing parameter, to be chosen.

The purpose of this paper is to present the method of **weighted cross validation** for obtaining a good value of λ from the data and to give some of its properties.

Since $\mathcal{K}$ is a compact operator from $\mathcal{L}_2$ into $\mathcal{L}_2$, its inverse is unbounded, in the $\mathcal{L}_2$ topology. This is manifest in the frequently observed fact that simple discretization or collocation methods do not give satisfactory approximate solutions to (1.1); the linear system to be solved has an ill conditioned matrix; as $n \to \infty$, any norm of the approximate solution typically becomes large. This phenomenon of exploding norm is especially pronounced when the data are observed with errors. The method of regularization controls the norm $\|f_{n,\lambda}\|_R$ of the approximate solution through the choice of the parameter λ. As λ increases, $\|f_{n,\lambda}\|_R^2$ decreases, simultaneously however the infidelity of the approximate solution to the data, as measured by

$$\frac{1}{n}\sum_{j=1}^{n} [(\mathcal{K}f_{n,\lambda})(t_j) - y(t_j)]^2 ,$$

becomes large.

It is important to choose λ correctly, if λ is too large the approximate solution does not correspond to the data, if λ is too small, the norm of the approximate solution will be unduly large.

There are a number of studies concerned with the choice of λ, typically involving either numerical experiments [5, 8, 11] or a priori knowledge concerning **f as well as** an assumed structure for $\{\epsilon(t)\}$ [7,9,14,15,23,29], to mention a few. Some other of the many works in regularization are [6,10,16,18,25,26,27]. However, to our knowledge, there is no published practical method for choosing a good value of the regularization parameter λ from the data, with the exception of the interesting work of Anderssen and Bloomfield (AB) [1], [2] which we discuss later.

In this paper we introduce the method of weighted cross-validation for choosing λ from the data. Our method is designed for the situation when g is a very smooth function ("very smooth" to be defined precisely later) and the errors can be considered as uncorrelated zero mean random variables with common variance σ^2 ("white noise"). The idea of cross-validation is extremely simple: Let $f_{n,\lambda}^{(k)}$ be the minimizer of

$$\frac{1}{n}\sum_{\substack{j=1\\ j\neq k}}^{n} [(\mathcal{K}h)(t_j) - y(t_j)]^2 + \lambda\|h\|_R^2 .$$

If $\hat{\lambda}$ is a good choice for the regularization parameter, then $(\mathcal{K}f_{n,\hat{\lambda}}^{(k)})(t_k)$ should be closer to $y(t_k)$, on the average, than $(\mathcal{K}f_{n,\lambda}^{(k)})(t_k)$ for other values of λ. For each λ, we measure this closeness by the weighted mean square data prediction error $V(\lambda)$,

$$(1.4) \qquad V(\lambda) = \frac{1}{n}\sum_{k=1}^{n} [(\mathcal{K}f_{n,\lambda}^{(k)})(t_k) - y(t_k))]^2 \, w_k(\lambda)$$

where the $w_k(\lambda)$ are weights to be given in Section 2. Our choice $\hat{\lambda}$ for λ is the minimizer of $V(\lambda)$.

The weights $\{w_k(\lambda)\}$ we have chosen can be justified in several ways, the most intuitive being as follows: Let $T(\lambda)$ be the mean square true prediction error when λ is used: -

$$(1.5) \qquad T(\lambda) = \frac{1}{n}\sum_{j=1}^{n} [(\mathcal{K}f_{n,\lambda})(t_j) - (\mathcal{K}f)(t_j)]^2 .$$

$$= \frac{1}{n}\sum_{j=1}^{n} [(\mathcal{K}f_{n,\lambda})(t_j) - g(t_j)]^2 .$$

Then the minimizer $\hat{\lambda}$ of $V(\lambda)$ with the weights $w_k(\lambda)$ is an estimate of the minimizer of $T(\lambda)$. More precisely, letting E be mathematical expectation, we show (loosely stated),

<u>Theorem 1(a)</u>: If g is very smooth, the noise is "white", and $t_i = i/n$, then

$$\min_{\lambda} E\, V(\lambda) = \min_{\lambda} E\, T(\lambda)(1+o(1)) ,$$

where $o(1) \to 0$ as $n \to \infty$.

Furthermore, convergence of $f_{n,\lambda}$ seems assured. Let $\mathcal{K}^{+}g$ be that element f of minimal $\mathcal{H}_R$ norm satisfying $\mathcal{K}f = g$, and let $\lambda^* = \min_{\lambda} E\, V(\lambda)$. We show

<u>Theorem 2(a)</u>: Under the hypotheses of Theorem 1(a),

$$(1.6) \qquad \lim_{n\to\infty} E\|\mathcal{K}^{+}g - f_{n,\lambda^*}\|_R^2 = 0 .$$

We are even able to give convergence rates for the quantity in Theorem 2(a). These rates depend on the rate of decay of the eigenvalues $\{\lambda_\nu\}$ of the kernel $Q(s,t)$ of (1.2). More precisely, let the Mercer-Hilbert-Schmidt expansion [22], of Q be

$$Q(s,t) = \sum_{\nu} \lambda_\nu \Phi_\nu(s)\Phi_\nu(t) ,$$

where the $\{\Phi_\nu\}$ are orthonormal, and we include only non-zero eigenvalues. A generalization of Picard's Theorem given in [17] says that $g \in \mathcal{K}(\mathcal{H}_R)$ if and only if $g \in \text{span}\{\Phi_\nu\}$ and

$$\sum_{\nu} \frac{g_\nu^2}{\lambda_\nu} < \infty$$

where

$$g_\nu = \int_0^1 g(u)\Phi_\nu(u)\,du .$$

Definition:

$$g \text{ is very smooth if } \||g\||^2 \equiv \sum_{\nu} \frac{g_\nu^2}{\lambda_\nu^2} < \infty .$$

We have

Theorems 1(b) and 2(b). Let $\lambda_\nu = c_\nu^{-1}\nu^{-2m}$, for some $m \geq 1$ and c_ν satisfy: $0 < \alpha \leq c_\nu^{-1} \leq \beta < \infty$, $\nu = 1,2,\ldots$. Under the hypotheses of Theorem 1(a),

$$\lambda^* = \left[\frac{k_m}{4m} \frac{\sigma^2}{|||g|||^2} \frac{1}{n}\right]^{2m/(4m+1)} (1+o(1)) ,$$

where k_m is a constant and $o(1) \to 0$ as $n \to \infty$. Furthermore,

$$E\|\mathcal{K}^+ g - f_{n,\lambda^*}\|_R^2 \leq O(n^{-2m/(4m+1)}) .$$

Equation (1.4) is unsuitable for computing $V(\lambda)$. We are able to give a simple approximate expression for $V(\lambda)$ which involves the first n eigenfunctions and eigenvalues of Q. If $Q(s,t)$ is a periodic function of $s-t$, then the computations are vastly simplified and the fast Fourier transform can be used.

We note that cross-validation in essentially the form given here was first suggested by Wahba and Wold [35,36] in the context of smoothing spline functions, although the idea of cross-validation goes back much earlier, see Stone [24] for a bibliography. A series of numerical experiments have been carried out in [35] for the special case $\mathcal{K}f = f$, with $m = 2$, with impressive results.

2. The regularized solution. The weighted cross-validation estimate for λ.

To obtain a solution $f_{n,\lambda}$ to the minimization problem of (1.3) it is generally necessary that the linear functional L_t which maps $f \in \mathcal{H}_R \to (\mathcal{K}f)(t)$ is continuous, for each $t = t_1, t_2, \ldots, t_n$. Then, by the Riesz representation theorem, there exists $\eta_t \in \mathcal{H}_R$ such that

$$(\mathcal{K}f)(t) = \langle \eta_t, f \rangle_R , \quad t=t_1,t_2,\ldots,t_n,$$

where $\langle \cdot,\cdot \rangle_R$ is the inner product in $\mathcal{H}_R$.

The representer of a continuous linear functional in an r.k.h.s. is found by applying the linear functional to the r.k., and so if $\mathcal{H}_R$ is an r.k.h.s., then,

$$\eta_t(s) = \int_0^1 K(t,u)R(s,u)du .$$

If $\mathcal{H}_R = \mathcal{L}_2[0,1]$, then

$$\eta_t(s) = K(t,s) .$$

See [3], [17], [29] for the properties of r.k.h.s. that we use here. Furthermore

$$\langle \eta_t, \eta_s \rangle_R = Q(s,t) ,$$

where $Q(s,t)$ is given in (1.2). Since $Q(t,t)$ is well defined and finite for $t \in [0,1]$ by assumption, all the L_t are continuous. The solution $f_{n,\lambda}$ to the problem: Find $h \in \mathcal{H}_R$ to minimize

$$\frac{1}{n} \sum_{j=1}^{n} [(\mathcal{K}h)(t_j) - y(t_j)]^2 + \lambda \|h\|_R^2$$

is given by

(2.1) $$f_{n,\lambda} = (\eta_{t_1}, \eta_{t_2}, \ldots, \eta_{t_n})(Q_n + n\lambda I)^{-1}(y(t_1), y(t_2), \ldots, y(t_n))' ,$$

where Q_n is the $n \times n$ matrix with $j\,k^{th}$ entry

$$\langle \eta_{t_j}, \eta_{t_k} \rangle_R = Q(t_j, t_k) .$$

It will be useful to observe that

$$g_{n,\lambda} \equiv \mathcal{K}f_{n,\lambda}$$

is given by

(2.2) $$g_{n,\lambda} = (Q_{t_1}, Q_{t_2}, \ldots, Q_{t_n})(Q_n + n\lambda I)^{-1}(y(t_1), y(t_2), \ldots, y(t_n))' ,$$

where Q_t is given by

$$Q_t(s) \equiv Q(t,s) .$$

Our estimate for λ is the minimizer of

(2.3a) $$V(\lambda) = \frac{1}{n} \sum_{k=1}^{n} [g_{n,\lambda}^{(k)}(t_k) - y(t_k)]^2 w_k(\lambda) ,$$

where $g_{n,\lambda}^{(k)} = \mathcal{K}f_{n,\lambda}^{(k)}$. The $w_k(\lambda)$ are weighting functions given by

(2.3b) $$w_k(\lambda) = [m^{kk}(\lambda) / \frac{1}{n} \sum_{j=1}^{n} m^{jj}(\lambda)]^2$$

where $m^{jj} = m^{jj}(\lambda)$ is the jj^{th} entry of the matrix $M^{-1} = M^{-1}(\lambda)$ given by

$$M(\lambda) = Q_n + n\lambda I .$$

If Q_n is a circulant matrix then $(Q_n + n\lambda I)^{-1}$ is also a circulant and hence constant down the main diagonal, and in this case the $m_k(\lambda)$ are all the same functions of λ and $w_k(\lambda) \equiv 1$. If $(Q_n + n\lambda I)^{-1}$ is not constant down the main diagonal then the weights express the fact (loosely speaking) that the collection of g satisfying $g = \mathcal{K}f$, $\|f\|_R \leq 1$, say, will tend to be more "wiggly" for some values of t than others.

The formula (2.3) for $V(\lambda)$ is unsuited for computation and for study of the properties of its minimizer. We can obtain a more useful expression as follows:

Let

$$\underset{\sim}{y} = (y(t_1), \ldots, y(t_n))'$$

and let $\underset{\sim}{y}_k$ be the $n-1$ dimensional column vector formed from $\underset{\sim}{y}$ by deleting the k^{th} entry. Let M_k be the $(n-1) \times (n-1)$ matrix obtained from M by deleting the k^{th} row and column. Let $\underset{\sim}{q}_k$ be the $n-1$ dimensional vector obtained from the k^{th} column of M by deleting the k^{th} entry, equivalently $\underset{\sim}{q}_k$ is the k^{th} column of Q_n with the k^{th} entry deleted. Then

$$g^{(k)}_{n,\lambda}(t_k) = \underset{\sim}{q}_k' M_k^{-1} \underset{\sim}{y}_k .$$

Next, note that for any symmetric matrix of full rank partitioned as follows:

$$M = \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{12}' & M_{22} \end{array}\right)$$

where the diagonal blocks are square, then

$$M^{-1} = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{12}' & B_{22} \end{array}\right)$$

with

$$B_{12} = -B_{11} M_{12} M_{22}^{-1} .$$

It follows by rearranging and partitioning so that $M_{12} = \underset{\sim}{q}_k'$, that

$$[(Q_n + n\lambda I)^{-1} \underset{\sim}{y}]_k = m^{kk}(g^{(k)}_{n,\lambda}(t_k) - y(t_k)) ,$$

where the left hand side is the k^{th} entry of $(Q_n + n\lambda I)^{-1} \underset{\sim}{y}$. Thus

$$(2.4) \quad V(\lambda) = \frac{1}{n} \sum_{k=1}^{n} [(Q_n + n\lambda I)^{-1} \underset{\sim}{y}]_k^2 / [\tfrac{1}{n} \mathrm{Trace}(Q_n + n\lambda I)^{-1}]^2$$

$$= \frac{1}{n} \underset{\sim}{y}' (Q_n + n\lambda I)^{-2} \underset{\sim}{y} / [\tfrac{1}{n} \mathrm{Trace}(Q_n + n\lambda I)^{-1}]^2 .$$

Since Q_n is symmetric positive definite we can write

$$Q_n = \Gamma D \Gamma'$$

where Γ is orthogonal and D diagonal. Letting $\phi_{\nu n}(j)$ be the νj^{th} entry of Γ and $\lambda_{\nu n}$ be the $\nu\nu^{th}$ entry of D, one obtains

$$(2.5) \qquad V(\lambda) = \frac{1}{n} \sum_{\nu=1}^{n} \left(\frac{\tilde{y}_\nu}{\lambda_{\nu n} + n\lambda} \right)^2 \Bigg/ \left[\frac{1}{n} \sum_{\nu=1}^{n} \frac{1}{(\lambda_{\nu n} + n\lambda)} \right]^2$$

where the $\tilde{y}_\nu$ are the discrete generalized Fourier coefficients for $\underset{\sim}{y}$ with respect to the eigenvectors $\underset{\sim}{\phi}_{\nu n}$, $\underset{\sim}{\phi}_{\nu n} = (\phi_{\nu n}(1), \phi_{\nu n}(2), \ldots, \phi_{\nu n}(n))'$, that is,

$$\tilde{y}_\nu = \sum_{j=1}^{n} \phi_{\nu n}(j) y(t_j) .$$

We remark that we could also have obtained $V(\lambda)$ of (2.5) by first making the rotation of the coordinate system which maps $\underset{\sim}{y}$ into $\underset{\sim}{z} = W \Gamma' \underset{\sim}{y}$, where W is the rotation matrix which diagonalizes the circulant matrices, and then doing the equivalent of (unweighted) cross validation in that system.

So far, no particular conditions have been placed on the node sequence $t_1, t_2, \ldots, t_n$, and we believe that the procedure of choosing λ to minimize $V(\lambda)$ of (2.5) is quite reasonable for any node sequence which is not too "bunched up" locally, and possibly even then. However, in the remainder of this paper, we suppose that the $\{t_j\}$ are equally spaced, $t_j = j/n$, $j = 1,2,\ldots,n$. The reason for this is as follows:- The behavior of the minimizer of (2.5) depends on the eigenvalues $\{\lambda_{\nu n}\}$. When the $\{t_j\}$ are equally spaced there is a simple approximate relation between $\{\lambda_{\nu n}\}$ and the eigenvalues associated with the kernel Q, to be described next. This allows our analysis to proceed relatively reasily.

Since $Q(t,s)$ is continuous on $[0,1] \times [0,1]$, the operator $\mathfrak{Q}$ defined by

$$(\mathfrak{Q}g)(t) = \int_0^1 Q(t,s)g(s)ds$$

is a Hilbert-Schmidt operator in $\mathcal{L}_2[0,1]$. (See Riesz-Nagy [22], for details concerning these operators). The operator $\mathfrak{Q}$ possess a system of $\mathcal{L}_2$-orthonormal eigenfunctions $\{\phi_\nu\}_{\nu=1}^{\infty}$ and eigenvalues $\{\lambda_\nu\}_{\nu=1}^{\infty}$ satisfying

$$\lambda_\nu \phi_\nu(t) = \int_0^1 Q(t,s)\phi_\nu(s)ds, \quad \nu = 1,2,\ldots .$$

The ϕ_ν are always continuous. Furthermore Q possess the Mercer-Hilbert-Schmidt expansion

$$Q(s,t) = \sum_{\nu=1}^{\infty} \lambda_\nu \phi_\nu(s)\phi_\nu(t) , \tag{2.6}$$

which converges pointwise, and

$$\int_0^1 Q(s,s)ds = \text{Trace } \mathfrak{Q} = \sum_{\nu=1}^{\infty} \lambda_\nu < \infty .$$

For $t_j = j/n$, $j = 1,2,\ldots,n$ and n large, we have

$$Q_n(\tfrac{j}{n}, \tfrac{k}{n}) = \sum_{\nu=1}^{\infty} \lambda_\nu \phi_\nu(\tfrac{j}{n})\phi_\nu(\tfrac{k}{n}) \simeq \sum_{\nu=1}^{n} n\lambda_\nu \frac{\phi_\nu(\frac{j}{n})}{\sqrt{n}} \frac{\phi_\nu(\frac{k}{n})}{\sqrt{n}} .$$

However, by definition of $\{\lambda_{\nu n}, \phi_{\nu n}\}$,

$$Q_n(\tfrac{j}{n}, \tfrac{k}{n}) = \sum_{\nu=1}^{n} \lambda_{\nu n}\phi_{\nu n}(j)\phi_{\nu n}(k) .$$

Since

$$\frac{1}{n}\sum_{j=1}^{n} \phi_\mu(\tfrac{j}{n})\phi_\nu(\tfrac{j}{n}) \simeq \int_0^1 \phi_\mu(s)\phi_\nu(s)ds = 1, \quad \mu = \nu$$
$$= 0, \quad \mu \neq \nu \tag{2.7}$$

it is reasonable to make the approximations

$$\lambda_{\nu n} \simeq n\lambda_\nu, \qquad \nu = 1,2,\ldots,n .$$

$$\phi_{\nu n}(j) \simeq \frac{1}{\sqrt{n}}\phi_\nu(\tfrac{j}{n}), \quad \nu, j = 1,2,\ldots,n .$$

With these approximations, $V(\lambda)$ of (2.5) becomes

$$V(\lambda) \simeq \frac{\frac{1}{n}\sum_{\nu=1}^{n} \frac{|\tilde{Y}_\nu|^2}{(\lambda_\nu+\lambda)^2}}{\left[\frac{1}{n}\sum_{\nu=1}^{n}\frac{1}{(\lambda_\nu+\lambda)}\right]^2} \tag{2.8a}$$

where

$$(2.8b)\qquad \tilde{Y}_\nu = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \phi_\nu(\tfrac{j}{n}) y(\tfrac{j}{n}) .$$

The symbol $\simeq$ is to be understood as meaning "approximately". In any case, we shall henceforth assume that $V(\lambda)$ is computed using (2.8), and our claims concerning the goodness of the minimizer $\hat{\lambda}$ of $V(\lambda)$ as an estimate of λ will be based on the use of (2.8) as the definition of $V(\lambda)$.

Suppose $Q(s,t)$ is of the form $Q(s,t) = q(s-t)$, where $q(\tau)$ is periodic with period 1. Then

$$Q(s,t) = \sum_{\nu=-\infty}^{\infty} \lambda_\nu e^{2\pi i\nu(s-t)} \equiv \sum_{\nu=-\infty}^{\infty} \lambda_\nu \phi_\nu(s)\phi_\nu^*(t)$$

for some $\{\lambda_\nu\}$ with $\lambda_\nu = \lambda_{-\nu}^*$, $\sum |\lambda_\nu| < \infty$ and

$$\phi_\nu(s) = e^{2\pi i\nu s}, \quad \nu = 0, \pm 1, \pm 2, \ldots .$$

(* is complex conjugate here). Then the λ_ν of course satisfy

$$\lambda_\nu = \int_0^1 q(\tau)\phi_\nu(\tau)d\tau , \quad \nu = 0, \pm 1, \pm 2, \ldots ,$$

and

$$\tilde{Y}_\nu = \frac{1}{\sqrt{n}} \sum_{\nu=1}^{n} e^{2\pi i\nu j/n} y(\tfrac{j}{n})$$

can be computed easily using the fast Fourier transform. $|\tilde{Y}_\nu|^2$ is the

periodogram of the data. Furthermore (2.7) (modified appropriately with *) is exact for $\mu, \nu = 1,2,\ldots,n$.

In the periodic case it is easy to interpret regularization in the "frequency" or eigenfunction domain. The discrete Fourier coefficients of the data,

$$\sqrt{n}\ \tilde{Y}_\nu \equiv \sum_{j=1}^{n} \phi_\nu(\tfrac{j}{n}) y(\tfrac{j}{n}) \equiv \sum_{j=1}^{n} e^{2\pi i\nu j/n} y(\tfrac{j}{n})$$

are related to the discrete Fourier coefficients of the smoothing function $g_{n,\lambda}$ by

$$(2.9)\qquad \sum_{j=1}^{n} e^{-2\pi i\nu j/n} g_{n,\lambda}(\tfrac{j}{n}) = \frac{\lambda_{\nu n}}{(\lambda_{\nu n}+n\lambda)} \sqrt{n}\ \tilde{Y}_\nu \simeq \frac{\sqrt{n}\ \tilde{Y}_\nu}{(1+\lambda/\lambda_\nu)} .$$

Thus, regularization is equivalent to decomposing the (time) data into its frequency components, multiplying by the (low pass) filter function $\psi(\nu) \simeq \frac{1}{(1+\lambda/\lambda_\nu)}$, reconstructing the time function, and then solving the operator equation. If the λ_ν are monotone decreasing, then the half power point of the filter is at frequency ν_0 where $\lambda_{\nu_0} = \lambda$. We note that others (Baker, Fox, Mayers and Wright [4], Lanczos [12], Reinsch [21]), have suggested a filter function of the form

$$\psi(\nu) = 1, \quad \nu \le \nu_0$$
$$= 0, \quad \nu > \nu_0 ,$$

where ν_0 is to be chosen.

3. Asymptotic properties of the weighted cross-validation estimate of λ.

We prove the following

Theorem 1.

Suppose

(1) g is very smooth, that is $|||g|||^2 < \infty$

(2) $E\epsilon(t)=0$; $E\epsilon(s)\epsilon(t)=\sigma^2$, $s=t$, $=0, s\neq t$, ("white noise")

(3) $\lambda_\nu = c_\nu^{-1}\nu^{-2m}$, with $0 < \alpha \leq c_\nu^{-1} \leq \beta < \infty$, for some $m \geq 1$

(4) $t_i = i/n$, $i = 1,2,\ldots,n$

Further, suppose the following approximations are valid:

i) $g_{\nu n}^2/n \simeq g_\nu^2$, $\quad g_{\nu n} = \sum_{j=1}^n \phi_{\nu n}(j) g(\frac{j}{n})$

ii) $\frac{1}{n}\sum_{j=1}^n \phi_\nu(\frac{j}{n})\phi_\mu(\frac{j}{n}) \simeq 1, \quad \mu = \nu$

$\simeq 0, \quad \mu \neq \nu$

iii) $\lambda_{\nu n} \simeq n\lambda_\nu$

then

$$\min_\lambda E\,T(\lambda) = \left[\frac{k_m}{4m}\frac{\sigma^2}{|||g|||^2}\frac{1}{n}\right]^{2m/(4m+1)}(1+o(1))$$

$$= \min_\lambda E\,V(\lambda)(1+o(1)) ,$$

where $T(\lambda)$ and $V(\lambda)$ are defined by (1.5) and (2.8) respectively,

$$k_m = \frac{1}{c^{1/2m}}\int_0^\infty \frac{dx}{(1+x^{2m})^2}$$

and c is a co[illegible]nt satisfying $\alpha \leq c^{-1} \leq \beta$.

Before beginning the proof, we note that the smoothness condition on g is equivalent to the existence of $\rho \in \mathcal{L}_2$ such that

$$(3.1) \qquad g(t) = \int_0^1 Q(t,s)\rho(s)ds ,$$

if g is very smooth, then

$$|||g|||^2 \equiv \sum_\nu \frac{g_\nu^2}{\lambda_\nu^2} = \inf_{\rho \text{ satisfies } (3.1)} \int_0^1 \rho^2(u)du .$$

Proof of Theorem 1.

Using the fact that

$$\underset{\sim}{g}_{n,\lambda} \equiv (g_{n,\lambda}(\tfrac{1}{n}), g_{n,\lambda}(\tfrac{2}{n}),\ldots,g_{n,\lambda}(\tfrac{n}{n}))'$$

is given by

$$(3.2) \qquad \underset{\sim}{g}_{n,\lambda} = Q_n(Q_n + n\lambda I)^{-1}\underset{\sim}{y}$$

one obtains, after some calculation using (2), that

$$ET(\lambda) \equiv E\sum_{\nu=1}^n (g_{n,\lambda}(\tfrac{j}{n}) - g(\tfrac{j}{n}))^2$$

is equal to

$$n\lambda^2 \sum_{\nu=1}^n \frac{g_{\nu n}^2}{(\lambda_{\nu n}+n\lambda)^2} + \frac{\sigma^2}{n}\sum_{\nu=1}^n \frac{\lambda_{\nu n}^2}{(\lambda_{\nu n}+n\lambda)^2} ,$$

and the approximations i) and iii) give

$$(3.3) \qquad ET(\lambda) \approx \lambda^2 \sum_{\nu=1}^n \frac{g_\nu^2}{(\lambda_\nu+\lambda)^2} + \frac{\sigma^2}{n}\sum_{\nu=1}^n \frac{\lambda_\nu^2}{(\lambda_\nu+\lambda)^2} .$$

Defining

$$A(\lambda) = \frac{1}{n}\sum_{\nu=1}^{n} \frac{\lambda_\nu^2}{(\lambda_\nu+\lambda)^2} \quad , \quad G(\lambda) = \sum_{\nu=1}^{n} \frac{g_\nu^2}{(\lambda_\nu+\lambda)^2}$$

we have that the right hand side of (3.3) is minimized for

(3.4) $$2\lambda G_1(\lambda) + \sigma^2 A'(\lambda) = 0$$

where

$$G_1(\lambda) = \frac{1}{2\lambda}\frac{d}{d\lambda}[\lambda^2 G(\lambda)] = \sum_{\nu=1}^{n} \frac{\lambda_\nu g_\nu^2}{(\lambda_\nu+\lambda)^3} .$$

The expression (3.4) can be rewritten

(3.5) $$\lambda = \frac{\sigma^2 A'(\lambda)}{2G_1(\lambda)} .$$

Using (3), it can be shown that

(3.6) $$A(\lambda) \simeq \frac{1}{n}\int_0^\infty \frac{dx}{(1+c\lambda x^{2m})^2} = \frac{k_m}{n\lambda^{1/2m}} \quad , \quad A'(\lambda) \simeq -\frac{k_m}{2mn\lambda^{(2m+1)/2m}}$$

where

$$k_m = \frac{1}{c^{1/2m}}\int_0^\infty \frac{dx}{(1+x^{2m})^2} ,$$

c is a constant satisfying $\alpha \le c^{-1} \le \beta$, and "$\simeq$" means "$=(1+O(1/n\lambda^{1/2m}))$", as $n \to \infty$, $\lambda \to 0$ in such a way that $n\lambda^{1/2m} \to \infty$.

Substituting (3.6) into (3.5), multiplying thru by $\lambda^{(2m+1)/2m}$ and taking the $2m/(4m+1)$-th root, we have, that the minimizer $\tilde{\lambda}$ of $ET(\lambda)$ satisfies

$$\tilde{\lambda} = \left[\sigma^2 \left(\sum_{\nu=1}^{n} \frac{\lambda_\nu g_\nu^2}{(\lambda_\nu+\tilde{\lambda})^3}\right)^{-1} \frac{k_m}{4m}\frac{1}{n}\right]^{2m/(4m+1)} (1+o(1)) .$$

Upon showing that $\tilde{\lambda}$ must tend to 0 as n tends to ∞, one then has

$$\tilde{\lambda} = \left[\frac{k_m}{4m}\frac{\sigma^2}{|||g|||^2}\frac{1}{n}\right]^{2m/(4m+1)} (1+o(1))$$

where $o(1) \to 0$ as $n \to \infty$.

Next,

$$EV(\lambda) = \left\{\frac{1}{n}\sum_{\nu=1}^{n} \frac{g_{\nu n}^2}{(\lambda_\nu+\lambda)^2} + \frac{\sigma^2}{n}\sum_{\nu=1}^{n}\sum_{\mu=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}\phi_\nu(\tfrac{j}{n})\phi_\mu(\tfrac{j}{n})\right] \times \frac{1}{(\lambda_\nu+\lambda)(\lambda_\mu+\lambda)}\right\} \div \left[\frac{1}{n}\sum_{\nu=1}^{n}\frac{1}{(\lambda_\nu+\lambda)}\right]^2$$

$$\simeq \left[\lambda^2 \sum_{\nu=1}^{n} \frac{g_\nu^2}{(\lambda_\nu+\lambda)^2} + \frac{\sigma^2}{n}\sum_{\nu=1}^{n}\left(1 - \frac{\lambda_\nu}{(\lambda_\nu+\lambda)}\right)^2\right] \div \left[\frac{1}{n}\sum_{\nu=1}^{n}\left(1 - \frac{\lambda_\nu}{(\lambda_\nu+\lambda)}\right)\right]^2 ,$$

where $\simeq$ indicates we have used the approximations i)-iii).

Letting

$$B(\lambda) = \frac{1}{n}\sum_{\nu=1}^{n} \frac{\lambda_\nu}{\lambda_\nu+\lambda}$$

we have

$$EV(\lambda) = \frac{\lambda^2 G(\lambda) + \sigma^2[1-2B(\lambda)+A(\lambda)]}{[1-B(\lambda)]^2}$$

and $EV(\lambda)' = 0$ when

(3.7) $$\{\lambda^2 G(\lambda) + \sigma^2[1-2B(\lambda) + A(\lambda)]\}\{-2B'(\lambda)[1-B(\lambda)]\}$$
$$= [1-B(\lambda)]^2\{2\lambda G_1(\lambda) - 2\sigma^2 B'(\lambda) + \sigma^2 A'(\lambda)\}.$$

Rearranging the expression (3.7) gives

$$-\sigma^2[\,1-B(\lambda) + A(\lambda) - B(\lambda) + \frac{\lambda^2}{\sigma^2}G(\lambda)]\,\frac{2B'(\lambda)}{[1-B(\lambda)]} + 2\sigma^2 B'(\lambda) - \sigma^2 A'(\lambda) = 2\lambda G_1(\lambda)$$

which is equivalent to

(3.8) $$-\frac{\sigma^2}{2G_1(\lambda)}A'(\lambda)\left\{1 + \frac{2B'(\lambda)}{A'(\lambda)[1-B(\lambda)]}\left[A(\lambda) - B(\lambda) + \frac{\lambda^2}{\sigma^2}G(\lambda)\right]\right\} = \lambda .$$

or

$$\lambda = -\frac{\sigma^2 A'(\lambda)}{2G_1(\lambda)}\{1+H(\lambda)\}$$

where

$$H(\lambda) = \frac{2B'(\lambda)}{A'(\lambda)[1-B(\lambda)]}\left[A(\lambda) - B(\lambda) + \frac{\lambda^2}{\sigma^2}G(\lambda)\right].$$

Now

$$B(\lambda) \propto \frac{\tilde{k}_m}{n\lambda^{1/2m}}, \quad B'(\lambda) \propto -\frac{\tilde{k}_m}{2mn\lambda^{(2m+1)/2m}}$$

where

$$\tilde{k}_m = \frac{1}{\tilde{c}^{1/2m}}\int_0^\infty \frac{dx}{(1+x^{2m})},$$

with $\alpha \leq \tilde{c}^{-1} \leq \beta$.

Thus

$$H(\lambda) = O(1/n\lambda^{1/2m}) + O(\lambda^2) .$$

It can be shown that the minimizer λ^* of $EV(\lambda)$ must satisfy $\lambda^* \to 0$, $n\lambda^{*1/2m} \to \infty$, so that λ^* **is the solution to**

$$\lambda = -\frac{\sigma^2 A'(\lambda)}{2G_1(\lambda)}(1+o(1))$$

and so

$$\lambda^* = \tilde{\lambda}(1+o(1)) ,$$

and the proof is completed.

We note that this result holds for large n and <u>fixed</u> σ^2, it is not claimed to be true for $\sigma^2 \to 0$, due to the presence of the factor $1/\sigma^2$ in the **expression for** $H(\lambda)$.

We now consider the convergence properties of $\|\mathcal{K}^+ g - f_{n,\lambda}\|_R^2$. We prove the following

<u>Theorem 2.</u>

Under the hypotheses of Theorem 1, as $\lambda \to 0$, $n \to \infty$ in such a way that $n\lambda^{1/2m} \to \infty$,

$$E\|\mathcal{K}^+g-f_{n,\lambda}\|_R^2 \le \left[\lambda|||g|||^2 + \frac{\sigma^2}{n}\frac{\ell_m}{\lambda^{(2m+1)/2m}}\right](1+o(1))$$

where

$$\ell_m = d^{-(2m+1)/2m}\int_0^\infty \frac{x^{2m}dx}{(1+x^{2m})^2}$$

with $\alpha \le d^{-1} \le \beta$.

Corollary.

$$E\|\mathcal{K}^+g-f_{n,\lambda^*}\|_R^2 \le \theta n^{-2m/(4m+1)}(1+o(1))$$

where

$$\theta = (\sigma^2)^{2m/(4m+1)}|||g|||^{(2m+1)/(4m+1)}\left[\frac{k_m}{4m} + \ell_m\left(\frac{4m}{k_m}\right)^{(2m+1)/(4m+1)}\right]$$

and

$$\lambda^* = \min_\lambda EV(\lambda) = \left[\frac{k_m}{4m}\frac{\sigma^2}{|||g|||^2}\frac{1}{n}\right]^{2m/(4m+1)}(1+o(1)) .$$

Proof of Theorem 2.

We first note that $\mathcal{K}(\mathcal{H}_R) = \mathcal{H}_Q$, where $\mathcal{H}_Q$ is the r.k.h.s. with r.k. Q given in (1.2). See [17] for details. Furthermore, it is shown there that

$$\|\mathcal{K}^+g-f_{n,\lambda}\|_R^2 \equiv \|\mathcal{K}(\mathcal{K}^+g)-\mathcal{K}f_{n,\lambda}\|_Q^2 ,$$

where $\|\cdot\|_Q$ is the norm in $\mathcal{H}_Q$.

This last result follows from the isometric isomorphism between $\mathcal{H}_Q$ and $\mathcal{K}^+(\mathcal{H}_Q)$ generated by the correspondance $g \in \mathcal{H}_Q \sim f \in \mathcal{K}^+(\mathcal{H}_Q)$ iff $g = \mathcal{K}f$, See [17] for details. Thus, since $\mathcal{K}(\mathcal{K}^+g) = g$, we have

$$E\|\mathcal{K}^+g-f_{n,\lambda}\|_R^2 = E\|g-g_{n,\lambda}\|_Q^2 .$$

Let P_ng be the orthogonal projection in $\mathcal{H}_Q$ onto the subspace spanned by $\{Q_{t_i}\}_{i=1}^n$. (Recall that $Q_t(s) = Q(s,t)$). Provided that Q is strictly positive definite, then

$$P_ng = (Q_{t_1},Q_{t_2},\ldots,Q_{t_n})Q_n^{-1}(g(t_1),g(t_2),\ldots,g(t_n))' . \tag{3.9}$$

In any case, since P_ng and $g_{n,\lambda}$ are in the span of $\{Q_{t_i}\}_{i=1}^n$, and $g-P_ng$ is in the orthogonal complement of this subspace,

$$\|g-g_{n,\lambda}\|_Q^2 = \|g-P_ng\|_Q^2 + \|P_ng-g_{n,\lambda}\|_Q^2 .$$

The term $\|g-P_ng\|_Q^2$ is independent of λ and the data vector $\underset{\sim}{y} = (y(t_1),y(t_2),\ldots,y(t_n))'$. Convergence properties of $\|g-P_ng\|_Q^2$ have been given in Wahba [31]. It is shown in [31] that if $|||g|||^2 < \infty$ and Q has the continuity properties of a Green's function for a $2m^{th}$ order self-adjoint linear differential operator (which entails the eigenvalue decay rate of (3)) then $\|g-P_ng\|_Q^2 = O(n^{-2m})$. $O(n^{-2m})$ is negligible compared to $\|P_ng-g_{n,\lambda}\|_Q^2$ and we will henceforth ignore $\|g-P_ng\|_Q^2$. Our task therefore, is to examine the behavior of

$\|P_n g - g_{n,\lambda}\|_Q^2$.

Let

$$\underset{\sim}{g} = (g(t_1), g(t_2), \ldots, g(t_n))'$$

$$\underset{\sim}{\epsilon} = (\epsilon(t_1), \ldots, \epsilon(t_n))' .$$

Then $\underset{\sim}{y} = \underset{\sim}{g} + \underset{\sim}{\epsilon}$. Using (3.2) and (3.9),

$$P_n g - g_{n,\lambda} = (Q_{t_1}, Q_{t_2}, \ldots, Q_{t_n})[n\lambda(Q_n + n\lambda I)^{-1} Q_n^{-1} \underset{\sim}{g} - (Q_n + n\lambda I)^{-1} \underset{\sim}{\epsilon}] .$$

We can obtain an expression for $E\|P_n g - g_{n,\lambda}\|_Q^2$

as follows: Using $\langle Q_{t_i}, Q_{t_j} \rangle_Q = Q(t_i, t_j)$ and the properties of the $\{\epsilon(\frac{i}{n})\}$ gives

$$\begin{aligned} E\|P_n g - g_{n,\lambda}\|_Q^2 &= (n\lambda)^2 \underset{\sim}{g}'(Q_n + n\lambda I)^{-1} Q_n^{-1}(Q_n + n\lambda I)^{-1} \underset{\sim}{g} \\ &\quad + E\,\underset{\sim}{\epsilon}'(Q_n + n\lambda I)^{-1}(Q_n + n\lambda I)^{-1} \underset{\sim}{\epsilon} \\ &= (n\lambda)^2 \underset{\sim}{g}'(Q_n + n\lambda I)^{-1} Q_n^{-1}(Q_n + n\lambda I)^{-1} \underset{\sim}{g} \\ &\quad + \sigma^2 \operatorname{Trace}(Q_n + n\lambda I)^{-2} Q_n \\ &= (n\lambda)^2 \sum_{\nu=1}^{n} \frac{g_{\nu n}^2}{\lambda_{\nu n}(\lambda_{\nu n} + n\lambda)^2} \\ &\quad + \sigma^2 \sum_{\nu=1}^{n} \frac{\lambda_{\nu n}}{(\lambda_{\nu n} + n\lambda)^2} \end{aligned}$$

$$\begin{aligned} &\simeq \lambda^2 \sum_{\nu=1}^{n} \frac{g_\nu^2}{\lambda_\nu(\lambda_\nu + \lambda)^2} + \frac{\sigma^2}{n} \sum_{\nu=1}^{n} \frac{\lambda_\nu}{(\lambda_\nu + \lambda)^2} \\ &\leq \lambda \sum_\nu \frac{g_\nu^2}{\lambda_\nu^2} + \frac{\sigma^2}{n} \sum_{\nu=1}^{n} \frac{\lambda_\nu}{(\lambda_\nu + \lambda)^2} . \end{aligned}$$

We have

(3.10) $$\frac{1}{n} \sum_{\nu=1}^{n} \frac{\lambda_\nu}{(\lambda_\nu + \lambda)^2} \simeq \frac{\ell_m}{n\lambda^{(2m+1)/2m}}$$

and so the Theorem is proved. The Corollary follows by direct substitution.

4. Generalizations

The validity of the method of weighted cross validation is not restricted to first kind integral equations. Let $\mathcal{K}$ be an otherwise arbitrary linear operator with domain an r.k.h.s. $\mathcal{H}_R$ and satisfying

$$|(\mathcal{K}f)(t)| \le M\|f\|_R\,, \quad f \in \mathcal{H}_R\,, \quad t \in [0,1]\,.$$

Then there exists a family $\{\eta_t,\ t \in [0,1]\} \subset \mathcal{H}_R$ with the property

$$\langle \eta_t, f\rangle_R = (\mathcal{K}f)(t), \quad f \in \mathcal{H}_R\,, \quad t \in [0,1]\,.$$

Let

$$Q(s,t) \equiv \langle \eta_s, \eta_t\rangle_R\,.$$

be continuous. **Then Theorems 1 and 2 follow.**

Concerning spaces of real-valued functions on $[0,1]$, the only place that $t \in [0,1]$ was used was in quoting the result $\|g-P_n g\|_Q^2 = O(n^{-2m})$ from [31]. Apparently, the results here generalize to $Q(s,t)$ a trace-class kernel defined for $s,t \in T$ an otherwise arbitrary index set, provided the analogue of i), ii) and iii) hold. See [34].

It is easy to prove the analogues of Theorems 1 and 2 if $\lambda_\nu = O(e^{-\gamma\nu})$, see [33].

5. Other methods.

Generalizing a suggestion of Reinsch [19], if σ^2 were known, one might choose λ so that

$$S(\lambda) \equiv \frac{1}{n}\sum_{j=1}^{n} [(\mathcal{K}f_{n,\lambda})(t_j)-y(t_j)]^2 = \sigma^2\,.$$

It can be shown (provided Q_n is of full rank) that $S(\lambda)$ is a monotone function of λ and that a unique solution for λ always exists. It is shown in [32] (for the special case $\mathcal{K} = I$, but the result generalizes) that to minimize $T(\lambda)$ one wants to set $S(\lambda) = \sigma^2(1-k)$ where k is a positive (unknown in practice) "fudge factor" tending to zero as $n \to \infty$. If σ^2 is known, g is smooth and n is large, however, setting $S(\lambda) = \sigma^2$ is probably reasonable.

Anderson and Bloomfield (AB) [1][2] in two interesting papers concerned with numerical differentiation, suggest a different method of choosing λ from the data (σ^2 unknown). Their idea can immediately be considered in the generality of the present work. Our assumption $\|g\| < \infty$ is replaced in AB by the model that $g(t)$, $t \in [0,1]$ is a Gaussian stochastic process with $Eg(t) = 0$, $Eg(s)g(t) = bQ(s,t)$, where b is unknown. **(and $\{\epsilon(t)\}$ Gaussian white noise),** With this model the conditional expectation of $f(s)$ given $y(t)$, $t = t_1,\dots,t_n$ is

$$E(f(s)|y(t_1),\dots,y(t_n)) = (\eta_{t_1}(s),\dots,\eta_{t_n}(s))(Q_n+n\alpha I)^{-1}y\,,$$

where $\alpha = \sigma^2/nb$. A maximum likelihood estimate for α is found by AB and is equal to the minimizer of $L(\alpha)$ given approximately by

$$L(\alpha) \propto \frac{\frac{1}{n}\sum_{\nu=1}^{n} \frac{|\tilde{Y}_\nu|^2}{(\lambda_\nu+\alpha)}}{\left[\prod_{\nu=1}^{n} \frac{1}{(\lambda_\nu+\alpha)}\right]^{1/n}}$$

where $\tilde{Y}_\nu$ is as in (2.8 b)

If g is a stochastic process as described, then the $\{g_\nu\}$ are well known (see [13]) to be independent, zero mean random variables with variances $Eg_\nu^2 = b\lambda_\nu$ so that $E\sum_{\nu=1}^{n} g_\nu^2/\lambda_\nu = bn$. Thus the AB assumption is that g is "rough" rather than "very smooth". It is shown in [33], that, if n is large and g "smooth", that is $\sum g_\nu^2/\lambda_\nu^{1+\epsilon} < \infty$ for some $\epsilon > 0$, then

$$\alpha^* = \min_{\alpha} EL(\alpha) = O(n^{-2m/(2m+1)}) \ll \lambda^* = O(n^{-2m/(4m+1)})$$

and so the AB choice of λ asymptotically regularizes less than the cross-validation choice. This is not surprising, considering the differing assumptions on g. Furthermore it can be seen using (3.10) that $E\|\mathcal{K}^+ g - f_{n,\alpha^*}\|_R^2$ does not converge. On the other hand, AB have some very interesting numerical results for an example in which n is medium sized but too small for our asymptotics to be reliable. Further study is needed to determine the conditions under which each method is preferable.

ACKNOWLEDGEMENTS

We wish to thank Professor R.S. Anderssen for generously providir unpublished experimental data, and the Weizmann Institute, the Oxford University Mathematical Institute and the Statistics Department, Stanford University for their hospitality while this work was performe

REFERENCES

[1] R. S. Anderssen and P. Bloomfield, A time series approach to numerical differentiation, Technometrics 16 (1974), 69-75.

[2] R.S. Anderssen and P. Bloomfield, Numerical differentiation procedures for non-exact data, Numer. Math. 22 (1974), 157-182.

[3] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950), 337-404.

[4] C.T.H. Baker, L. Fox, D. F. Mayers, and K. Wright, Numerical solution of Fredholm integral equations of the first kind, Comput. J. 7 (1964), 141-148.

[5] Jane Cullum, Numerical differentiation and regularization, SIAM J. Numer. Anal. 8 (1971), 254-265.

[6] J. N. Franklin, Well posed extensions of ill posed linear problems, J. Math. Anal. Appl. 31 (1970), 682-716.

[7] V. I. Gorodnova and V. A. Morozov, Numerical parameter selection algorithms in the regularization method, Z. Vycisl. Mat. i Mat. Fiz. 13 (1973), 539-545.

[8] John Hilgers, Non-iterative methods for solving operator equations of the first kind, University of Wisconsin-Madison, Mathematics Research Center, TSR #1413, (1974).

[9] I. I. Ievlev, The approximate solution of equations of the first kind, Z. Vycisl. Mat. i. Mat. Fiz. 13 (1973), 1049-1056.

[10] Y. Ikebe, and E. Williams, The regularization method for the numerical solution of Fredholm integral equations of the first kind, Report CNA 72, Center for Numerical Analysis, University of Texas, Austin (1973).

[11] L. S. Jennings, An improperly posed quadratic programming problem, in Optimization, R. S. Anderssen, L. S. Jennings, and D. M. Ryan, Eds. University of Queensland Press, St. Lucia, Queensland, Australia (1972), 152-171.

[12] C. Lanczos, Evaluation of noisy data, SIAM J. Numer Anal. Ser. B. 1 (1964), 76-85.

[13] Michael Loeve, Probability Theory, D. Van Nostrand Co. (1955).

[14] V. A. Morozov, The choice of parameter when solving functional equations by regularization, Dokl. Akad. Nauk SSSR 175 (1967), 1225-1228.

[15] V. A. Morozov, The error principle in the solution of incompatible equations by Tihonov regularization, Z. Vycisl. Mat. i Mat. Fiz. 13 (1973), 1099-1111.

[16] M. Z. Nashed and G. Wahba, Convergence rates of approximate least square solutions of linear integral and operator equations of the first kind, Math. Comp. 28 (1974), 69-80.

[17] M. Z. Nashed and Grace Wahba, Generalized inverses in reproducing kernel spaces, SIAM J. Math. Anal. 5 (1974), 974-987.

[18] D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. Assoc. Comput. Mach. 9 (1962), 54-97.

[19] C. Reinsch, Smoothing by spline functions, Numer. Math. 10 (1967), 177-183.

[20] C. Reinsch, Smoothing by spline functions II, Numer. Math. 16 (1971), 451-454.

[21] C. Reinsch. Personal communication.

[22] F. Riesz and B. Sz.-Nagy, Functional Analysis. Ungar, New York (1955).

[23] T. I. Savelova and V. V. Tikhomirov, Solution of convolution-type integral equations of the first kind in the multi-dimensional case, Z. Vycisl. Mat. i Mat. Fiz. 3 (1973), 555-563.

[24] Mervyn Stone, Cross-validatory choice and assessment of statistical prediction (with discussion), J. Roy. Statist. Soc. Ser. B 36 (1974), 111-147.

[25] O. N. Strand and E. R. Westwater, Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind, J. Assoc. Comput. Mach. 15 (1968), 100-114.

[26] A. N. Tihonov, Solution of incorrectly formulated problems and the method of regularization, Soviet Math. Dokl. 4 (1963), 1035-1038.

[27] A. N. Tihonov, Regularization of incorrectly posed problems, Soviet Math. Dokl. 4 (1963), 1624-1627.

[28] A. N. Tihonov and V. B. Glasko, An approximate solution of Fredholm integral equations of the first kind, Z. Vycisl. Mat. i Mat. Fiz. 4 (1964), 564-571.

[29] Grace Wahba, Convergence properties of the method of regularization for noisy linear operator equations, University of Wisconsin-Madison, Mathematics Research Center TSR #1132, (1973).

[30] Grace Wahba, A class of approximate solutions to linear operator equations, J. Approximation Theory 9 (1973), 61-77.

[31] Grace Wahba, Convergence rates of certain approximate solutions to Fredholm integral equations of the first kind, J. Approximation Theory 7 (1973), 167-185.

[32] Grace Wahba, Smoothing noisy data by spline functions, Statistics Dept. University of Wisconsin-Madison TR #380, (1974), to appear, Numer. Math.

[33] Grace Wahba, Practical approximate solutions to linear operator equations when the data are noisy, to appear as University of Wisconsin-Madison, Mathematics Research Center TSR #1567, (1975).

[34] Grace Wahba, A canonical form for the problem of estimating smooth surfaces, Statistics Dept. University of Wisconsin-Madison TR #420, (1975).

[35] Grace Wahba and Svante Wold, A completely automatic French curve: Fitting spline functions by cross-validation, Comm. Statist. 4 (1975), 1-17.

[36] Grace Wahba and Svante Wold, Periodic splines for spectral density estimation: The use of cross-validation for determining the degree of smoothing, Comm. Statist. 4 (1975), 125-141.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #430	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): PRACTICAL APPROXIMATE SOLUTIONS TO LINEAR OPERATOR EQUATIONS WHEN THE DATA ARE NOISY	5. TYPE OF REPORT & PERIOD COVERED: Scientific Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Grace Wahba	8. CONTRACT OR GRANT NUMBER(s): AFOSR 72-2363 C
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, University of Wisconsin, 1210 W. Dayton Street	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research, Building 410, Bolling AFB, D. C. 20332	12. REPORT DATE: September 1975
	13. NUMBER OF PAGES: 33
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): This document has been approved for public release and sale; release and sale; its distribution	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is 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)

cross validation
linear operator equations
smooth approximation

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

See reverse side

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

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

Abstract

We consider approximate solutions $f_{n,\lambda}$ to linear operator equations $\mathcal{K}f = g$, of the form: $f_{n,\lambda}$ is the minimizer in $\mathcal{H}$ of $\frac{1}{n}\sum_{j=1}^{n}[(\mathcal{K}h)(t_j)-y(t_j)]^2 + \lambda\|h\|^2$, where $\mathcal{H}$ is a Hilbert space, and the data $\{y(t_j)\}$ satisfy $y(t_j) = g(t_j) + \epsilon(t_j)$, the $\{\epsilon(t_j)\}$ being measurement errors. $f_{n,\lambda}$ is the so-called regularized solution, and $\lambda > 0$ is the regularization parameter, to be chosen. It is important to choose λ correctly. The purpose of this paper is to propose the method of weighted cross-validation for choosing λ from the data. We suppose that g is very smooth and the errors are white noise. It is shown that the weighted cross-validation estimate $\hat{\lambda}$ estimates the value of λ which minimizes $\frac{1}{n}E\sum_{j=1}^{n}[(\mathcal{K}f_{n,\lambda})(t_j)-(\mathcal{K}f)(t_j)]^2$. Results related to the convergence of $\|f-f_{n,\hat{\lambda}}\|$, including rates, are obtained.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)