The Annals of Statistics
1985, Vol. 13, No. 4, 1378–1402

A COMPARISON OF GCV AND GML FOR CHOOSING THE SMOOTHING PARAMETER IN THE GENERALIZED SPLINE SMOOTHING PROBLEM[1]

BY GRACE WAHBA

University of Wisconsin

The partially improper prior behind the smoothing spline model is used to obtain a generalization of the maximum likelihood (GML) estimate for the smoothing parameter. Then this estimate is compared with the generalized cross validation (GCV) estimate both analytically and by Monte Carlo methods. The comparison is based on a predictive mean square error criteria. It is shown that if the true, unknown function being estimated is smooth in a sense to be defined then the GML estimate undersmooths relative to the GCV estimate and the predictive mean square error using the GML estimate goes to zero at a slower rate than the mean square error using the GCV estimate. If the true function is "rough" then the GCV and GML estimates have asymptotically similar behavior. A Monte Carlo experiment was designed to see if the the asymptotic results in the smooth case were evident in small sample sizes. Mixed results were obtained for $n = 32$, GCV was somewhat better than GML for $n = 64$, and GCV was decidedly superior for $n = 128$. In the $n = 32$ case GCV was better for smaller σ^2 and the comparison close for larger σ^2. The theoretical results are shown to extend to the generalized spline smoothing model, which includes the estimate of functions given noisy values of various integrals of them.

1. Introduction. We consider the same smoothing spline procedures as in Wahba (1978b, 1983) and elsewhere, and their extension to the solution of linear operator equations with noisy data. The (special) spline smoothing model is

$$y_i = f(t_i) + \varepsilon_i, \qquad i = 1, 2, \dots, n,\ t_i \in [0,1],$$

where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \sim \mathcal{N}(0, \sigma^2 I_{n\times n})$, σ^2 is unknown, and $f(\cdot)$ is some function in the Sobolev space $W_2^m[0,1]$,

$$W_2^m[0,1]\{f\colon f, f', \dots, f^{(m-1)} \text{ abs. cont., } f^{(m)} \in \mathscr{L}_2[0,1]\}.$$

The smoothing spline estimate $f_{n,\lambda}$ of f is the minimizer in $W_2^m[0,1]$ of

$$\frac{1}{n}\sum_{i=1}^{n} (f(t_i) - y_i)^2 + \lambda \int_0^1 (f^{(m)}(t))^2\, dt. \tag{1.1}$$

$f_{n,\lambda}$ is the celebrated polynomial smoothing spline of degree $2m-1$. The bandwidth parameter λ controls the tradeoff between the infidelity to the data as measured by $1/n\sum_{i=1}^n (f(t_i) - y_i)^2$ and the roughness $\int_0^1 (f^{(m)}(u))^2\, du$ of the estimated solution.

Received December 1983; revised July 1985.

[1] This research support by the Office of Naval Research under Contract No. N00014-77-C-0675.

AMS 1980 *subject classifications.* 65D07, 65D10, 62J02, 65R20

Key words and phrases. Spline smoothing, cross validation, maximum likelihood, integral equations.

The generalized cross validation (GCV) estimate of λ is the minimizer of $V(\lambda)$,

$$V(\lambda) = \frac{(1/n)\|(I - A(\lambda))y\|^2}{[(1/n)\,\mathrm{tr}(I - A(\lambda))]^2}, \tag{1.2}$$

where $A(\lambda)$ is the $n \times n$ influence matrix, which satisfies

$$\begin{pmatrix} f_{n,\lambda}(t_1) \\ \vdots \\ f_{n,\lambda}(t_n) \end{pmatrix} = A(\lambda)y, \qquad y = (y_1, \dots, y_n)'. \tag{1.3}$$

The GCV estimate of λ estimates the λ which minimizes the predictive mean square error $R(\lambda)$ defined by

$$R(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(f(t_i) - f_{n,\lambda}(t_i)\right)^2. \tag{1.4}$$

$f_{n,\lambda}(t)$, $t \in [0,1]$ is also a Bayes estimate of $f(t)$, if f is endowed with a certain zero mean Gaussian prior, which is partially improper.

The purpose of this paper is to derive a maximum likelihood (ML) estimate for λ, based on this prior, which generalizes the usual notion of ML estimates to the case of improper distributions, and then to compare the properties of this estimate of λ (called the GML estimate) with the behavior of the GCV estimate of λ. We decided to make this comparison at this time because of recent interest in related ML estimators.

The GML estimate we derive is the minimizer of $M(\lambda)$ given by

$$M(\lambda) = \frac{y'(I - A(\lambda))y}{[\det{}^{+}(I - A(\lambda))]^{1/n-m}} \tag{1.5}$$

where $\det^{+}(I - A(\lambda))$ is the product of the $n - m$ nonzero eigenvalues of $(I - A(\lambda))$. The GML estimate reduces the usual ML estimate, as first given by Anderssen and Bloomfield (1974) when the prior is "proper," and is an extension of an estimate recently given by Barry (1983). The comparisons we make between GCV and GML also hold for the proper prior case.

Our comparison of the GCV and GML estimates is based on the criterion of minimizing predictive mean square error $R(\lambda)$ defined in (1.4). Although this might appear to be a somewhat special criterion, under certain conditions other loss functions (for example, mean square error in the derivative) turn out to be minimized by a λ close to the minimizer of $R(\lambda)$. Some references are given below.

Let λ_{opt} be the minimizer of $ER(\lambda)$, where the expectation is taken over ε. The asymptotic behavior of λ_{opt} and $ER(\lambda_{\mathrm{opt}})$ has been studied by a number of authors, under mild regularity conditions on the data points. [See Cox (1983a,b 1984), Craven and Wahba (1979), Ragozin (1983), Rice and Rosenblatt (1983), Speckman (1985), Utreras (1981), and Wahba (1975).] The results include

$$f \in \pi^{m-1} \Rightarrow \lambda_{\text{opt}} = \infty, \quad ER(\lambda_{\text{opt}}) = O\left(\frac{1}{n}\right).$$

$$f \in W_2^m \Rightarrow ER(\lambda_{\text{opt}}) = O(n^{-2m/(2m+1)})$$

and this rate is achieved with

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

$$f \in \mathscr{C}_p \Rightarrow ER(\lambda_{\text{opt}}) = O(n^{-2mp/(2mp+1)})$$

and this rate is achieved with

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

Here π^{m-1} are the polynomials of degree $m-1$ or less and $\mathscr{C}_p = \mathscr{C}_p(W_2^m)$ is the class of functions in W_2^m with $0 < \int_0^1 (f^{(m)}(u))^2\,du < \infty$, and satisfying certain additional smoothness conditions indexed by $p \in (1,2]$, to be defined more precisely later. If mp is an integer, then it is conjectured that $f \in \mathscr{C}_p$ entails that $f \in W_2^{mp}$ and f satisfies the homogeneous boundary conditions

$$f^{(j)}(0) - f^{(j)}(1) = 0, \qquad j = m, m+1, \dots, mp-1.$$

Let f be fixed and let λ_{GML} and λ_{GCV} be the minimizers of $EM(\lambda)$ and $EV(\lambda)$, respectively. Let the "expectation inefficiency" of λ_X relative to λ_Y be $I_{X/Y}$ defined by

$$I_{X/Y} = ER(\lambda_X)/ER(\lambda_Y).$$

In this paper, we obtain information concerning $I_{\text{GML/opt}}$ as $n \to \infty$ under three (distinct) "smoothness" assumptions on f, namely

(1) $f \in \pi^{m-1}$,
(2) $f \in \mathscr{C}_p$ for some $p \in (1,2]$,
(3) f behaves like a "sample function" from a stochastic process with the given prior.

The results are

$$(1) \Rightarrow I_{\text{GML/opt}} = 1,$$
$$(1.6) \qquad (2) \Rightarrow I_{\text{GML/opt}} \to \infty,$$
$$(3) \Rightarrow I_{\text{GML/opt}} = 1 + o(1).$$

The "borderline" case $f \in W_2^m$ with $0 < \int_0^1 (f^{(m)}(u))^2\,du$ and $f \notin \mathscr{C}_p$ for any $p > 1$ is unresolved at this time. (We call this the borderline case because $\mathscr{C}_1 \oplus \text{span}\,\pi^{m-1} = W_2^m$.)

It is well known that if f satisfies (1), (2), (3), or is a borderline case

$$(1.7) \qquad I_{\text{GCV/opt}} = 1 + o(1).$$

For numerical and theoretical results, see Craven and Wahba (1979), Erdal (1983), Golub, Heath, and Wahba (1979), Utreras (1979, 1980, 1981, 1983), Wahba (1977b), and Wahba and Wendelberger (1980). Speckman (1982) has recently obtained stronger theoretical results, without the "expectation," and Li (1983) has recently related GCV and Stein's unbiased risk estimate. All of the results in this paper relate to ER, EV and EM rather than R, V and M. We believe that the "E" can be removed, possibly without strengthening the hypotheses, but that is not done here.

In the light of (1.7), it follows that

$$\text{(1.8)} \qquad \begin{aligned} &(1)\ I_{\mathrm{GML/GCV}} = 1,\\ &(2)\ I_{\mathrm{GML/GCV}} = \infty,\\ &(3)\ I_{\mathrm{GML/GCV}} = 1 + o(1). \end{aligned}$$

Following these theoretical results we present a small Monte Carlo study with three example fs satisfying (2). The inferiority of the GML estimate is perceptibly evident at $n = 64$ and strongly evident at $n = 128$.

Cross validated spline methods have been used with some success in the solution of linear integral equations with noisy data. [See Crump and Seinfeld (1982), Halem and Kalnay (1983), Merz (1980), O'Sullivan and Wahba (1984), and Wahba (1977b, 1979, 1982b).] One of the reasons for this success is that under various circumstances the λ that minimizes $R(\lambda)$ also minimizes (or nearly minimizes) other, possibly more interesting loss functions, for example $R_D(\lambda) = \int_0^1 (f'_{n,\lambda}(u) - f'(u))^2\,du$. [For theoretical results see Cox (1983a), Lukas (1981), Nychka (1983), and Ragozin (1983).] Special cases of this may be obtained by, e.g., comparing the optimal λ for $ER(\lambda)$ and $ER_D(\lambda)$ using the results from Theorems 1–4 in Rice and Rosenblatt (1983), or by comparing the optimal λs in Theorems 1 and 2 of Wahba (1977b). Numerical evidence supporting this result may be found in Craven and Wahba (1979) and Wahba (1979b, 1982b).

It is fairly straightforward to state and prove most of our results comparing GML and GCV in the context of the generalized smoothing spline model, which includes spline smoothing on the plane and in several dimensions, and on the sphere [Cox (1982), Utreras (1979), Wahba (1979a, 1981a, 1982a), and Wahba and Wendelberger (1980)], as well as the integral equation case discussed above. We will do that here.

The generalized smoothing spline model (of which the special spline smoothing model is the most widely known special case) is

$$\text{(1.9)} \qquad y_i = L_i f + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where ε is as before, f is assumed to be in some reproducing kernel Hilbert space $\mathscr{H}_Q$ of real valued functions on some index set $\mathscr{T}$, and the L_i are bounded linear functionals on $\mathscr{H}_Q$. The (generalized) smoothing spline estimate $f_{n,\lambda}$ of f is the minimizer in $\mathscr{H}_Q$ of

$$\text{(1.10)} \qquad \frac{1}{n}\sum_{i=1}^{n} (L_i f - y_i)^2 + \lambda \|Pf\|_Q^2,$$

where $\|\cdot\|_Q$ is the norm in $\mathscr{H}_Q$, and P is the orthogonal projection operator in $\mathscr{H}_Q$ onto the orthocomplement of the span of m given linearly independent basis functions $\{\phi_\nu\}_{\nu=1}^m$. [See Kimeldorf and Wahba (1971), Wahba (1984), and references cited therein.] The reader only interested in the special spline smoothing model may make the associations: $\mathscr{T} = [0,1]$, $\phi_\nu(t) = t^{\nu-1}/(\nu-1)!$, $\nu = 1,2,\dots,m$, $\mathscr{H}_Q = W_2^m[0,1]$, $L_i f = f(t_i)$, and $\|Pf\|_Q^2 = \int_0^1 (f^{(m)}(u))^2\,du$.

When solving (first kind) integral equations, we have

$$L_i f = \int_{\mathscr{T}} K(s_i, t) f(t)\,dt, \qquad i = 1,2,\dots,n,$$

where $K(\cdot,\cdot)$ is known.

In the general case $V(\lambda)$ and $M(\lambda)$ are still defined by (1.2) and (1.5), respectively, where now $A(\lambda)$ satisfies

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

and $R(\lambda)$ becomes

$$R(\lambda) = \frac{1}{n} \sum_{i=1}^{n} (L_i f - L_i f_{n,\lambda})^2. \tag{1.11}$$

The truth of (1.6) and (1.8) will actually be argued in this more general setting, with the extra smoothness condition $f \in \mathscr{C}_p$ appropriately generalized.

In Section 2 we derive the GML estimate of λ for the model of (1.9) and discuss the related maximum likelihood estimates of Barry (1983) and of Wecker and Ansley (1982). In Section 3 we obtain the asymptotic behavior of λ_{GML} under conditions (1)–(3). In Section 4 we compare λ_{GML}, λ_{GCV}, and λ_{opt}. Section 5 presents the Monte Carlo results and Section 6 discusses the extension to the model of (1.9).

For the results under (2) (1.6) we have given very general hypotheses under which the conclusions hold. A limitation of this general approach is that verification of the hypotheses in many interesting cases requires further work.

We briefly indicate both the generality and the limitations of the analytical results of (2). First we note (see details in Section 3) that in both the special and generalized smoothing spline model $I - A(\lambda)$ has a representation

$$I - A(\lambda) = n\lambda W(D + n\lambda I)^{-1} W',$$

where $W_{n\times n-m}$ satisfies $W'W = I_{n-m}$ and $D = \operatorname{diag}(\lambda_{1n},\dots,\lambda_{n-m,n})$ with $\lambda_{\nu n} \geq 0$. Here both W and D depend on $(t_1,\dots,t_n)$ or $(L_1,\dots,L_n)$. Writing $(t_{1n},\dots,t_{nn})$ or $(L_{1n},\dots,L_{nn})$ instead of $(t_1,\dots,t_n)$ or $(L_1,\dots,L_n)$ to denote the emphasis on n, define

$$\begin{pmatrix} g_{1,n} \\ \vdots \\ g_{n-m,n} \end{pmatrix} = W' \begin{pmatrix} f(t_{1n}) \\ \vdots \\ f(t_{nn}) \end{pmatrix} \quad \text{or} \quad W' \begin{pmatrix} L_{1n} f \\ \vdots \\ L_{nn} f \end{pmatrix}.$$

Our hypotheses for (2) are stated in terms of the conditions (as $n \to \infty$, and

$\lambda \to 0$)

$$\sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2/n}{(\lambda_{\nu n}/n)p} \le J_p \quad \text{independent of } n, \text{ for some } p \in (1,2], \tag{1.12}$$

$$\begin{aligned} \frac{1}{n}\operatorname{tr} A(\lambda)^j &\doteq \frac{1}{n}\sum_{\nu=1}^{n-m}\left(\frac{\lambda_{\nu n}}{n\lambda + \lambda_{\nu n}}\right)^j \\ &= \frac{\text{const}(j)}{n\lambda^{1/r}}(1 + o(1)) \quad \text{for } j = 1,2 \text{ and some } r > 1. \end{aligned} \tag{1.13}$$

The expressions on the left depend on $(t_{1n}, \ldots, t_{nn})$ or $(L_{1n}, \ldots, L_{nn})$ as well as f and a certain reproducing kernel. In certain very special cases, for example, the special spline smoothing model with $t_{in} = (i/n)$ these conditions can be rigorously related to readily understandable smoothness conditions on f and the (known) eigenvalues of Q_1, the reproducing kernel for $H_Q \setminus \{\operatorname{span} \phi_1, \ldots, \phi_m\}$. [See Rice and Rosenblatt (1983) and Utreras (1983).] Conjectures relating to the general spline smoothing model may be found in Wahba (1977b, 1977c) and in Sections 4 and 6 to follow. Since this paper was written, Nychka and Cox (1984) have provided further information on convergence properties of the solution to the generalized spline smoothing problem.

Throughout the paper we assume "some regularity conditions" on the $\{t_i\} = \{t_{in}\}$. We believe that in the case $t \in [0,1]$ sufficient regularity conditions for the results in (2) and (3) of (1.6) are: the $\{t_{in}\}$ satisfy

$$\frac{i}{n}(1 + o(1)) = \int_0^{t_{in}} w(t)\, dt$$

for some strictly positive bounded density w; and that the arguments do not always hold if the $\{t_{in}\}$ accumulate to a fixed finite number (independent of n) of accumulation points.

2. The GML estimate of λ. The Bayesian model behind the estimate $f_{n,\lambda}$ goes as follows:

$$y_i = L_i f + \varepsilon_i, \qquad i = 1,2,\ldots,n, \tag{2.1}$$

where the ε_i are as before but $f(t)$, $t \in \mathscr{T}$ is supposed to be a certain zero mean Gaussian stochastic process with a partially improper prior. The meaning of L_i will be given shortly. Let $Q(s,t)$ be the reproducing kernel for $\mathscr{H}_Q$ and let $Q_1(s,t)$ be the reproducing kernel for the orthocomplement of span$\{\phi_\nu\}$ in $\mathscr{H}_Q$, that is,

$$Q_1(s,t) = Q(s,t) - \sum_{\mu,\nu=1}^{m} \phi_\mu(s)\phi_\nu(t)k^{\mu\nu}, \tag{2.2}$$

where $k^{\mu\nu}$ is the $\mu\nu$th entry of the inverse of the Gram matrix $\{\langle \phi_\mu, \phi_\nu \rangle_Q\}$ of the $\{\phi_\nu\}$.

Let

$$X_\xi(t) = \sum_{j=1}^{m} \theta_j \phi_j(t) + b^{1/2} Z(t), \tag{2.3}$$

where $\theta = (\theta_1, \dots, \theta_m) \sim \mathcal{N}(0, \xi I_{m \times m})$, b is some constant, and $Z(t)$, $t \in \mathcal{T}$ is a zero mean Gaussian stochastic process independent of θ with

$$EZ(s)Z(t) = Q_1(s, t).$$

In the polynomial spline case, Q_1 is the covariance of the m-fold integrated Weiner process. We let f have the prior distribution of X_ξ, as $\xi \to \infty$. It is shown in Wahba (1978b), in the case $L_i f = f(t_i)$, that

$$f_{n,\lambda}(t) = \lim_{\xi \to \infty} E_\xi\{f(t)|y_1, \dots, y_n\}. \tag{2.4}$$

with $\lambda = \sigma^2/nb$.

Now, the sample functions X_ξ are not in $\mathcal{H}_Q$ (see below), so that the exact meaning of L_i must be clarified. According to Parzen (1962) [for further details, see Wahba (1982a)], L_i is a bounded linear functional on $\mathcal{H}_Q$ if and only if $L_i X_\xi$ is a zero mean Gaussian random variable well defined in quadratic mean. Then the covariances will be

$$E(L_i X_\xi)(L_j X_\xi) = L_{i(s)} L_{j(t)}\{EX_\xi(s) X_\xi(t)\}, \tag{2.5}$$

where $L_{i(s)}$ means L_i is applied to the operand (in braces) considered as a function of s.

Letting T be the $n \times m$ matrix with $i\nu$th entry $L_i \phi_\nu$ and Σ the $n \times n$ matrix with ijth entry $L_{i(s)} L_{j(t)} Q_1(s, t)$, we have, using (2.3) and (2.5),

$$\{E(L_i X_\xi)(L_j X_\xi)\} = \xi TT' + b\Sigma. \tag{2.6}$$

Using this fact, straightforward substitution in Wahba (1978b) [see also Wahba (1983)] can be used to show that (2.4) holds for the $\{L_i\}$ any set of bounded linear functions on $\mathcal{H}_Q$ such that rank T is m. If $L_i f = f(t_i)$, etc., then $L_{i(s)} L_{j(t)} Q_1(s, t) = Q_1(t_i, t_j)$.

Using (2.1) and (2.6), it follows that

$$y \sim \mathcal{N}(0, \xi TT' + b\Sigma + \sigma^2 I).$$

Setting $\lambda = \sigma^2/nb$ and $\eta = \xi/b$, we have

$$y \sim \mathcal{N}(0, b(\eta TT' + \Sigma + n\lambda I)). \tag{2.7}$$

We find the GML estimate of λ by letting $\eta \to \infty$ in (2.7) in an appropriate manner. We do this by letting $R_{n-m \times n}$ be any $n - m \times n$ matrix satisfying $RR' = I_{n-m}$ and $RT = 0_{n-m \times m}$. Let

$$\begin{pmatrix} x \\ \cdots \\ u \end{pmatrix} = \begin{pmatrix} R \\ \cdots \\ \frac{1}{\sqrt{\eta}} T' \end{pmatrix} y.$$

Then

$$\begin{aligned} Exx' &= b(R\Sigma R' + n\lambda I), \\ \lim_{\eta \to \infty} Exu' &= 0, \\ \lim_{\eta \to \infty} Euu' &= b(T'T)(T'T). \end{aligned} \tag{2.8}$$

Since in the limit the distribution of u does not depend on λ, we claim it is appropriate to define the GML estimate of λ as the (usual) ML estimate based on the distribution of x. Peter Green has kindly pointed out to us that this argument has previously been used by Patterson and Thompson (1971) in a different context. A straightforward calculation gives, that the ML estimate of λ based on $x \sim \mathcal{N}(0, b(R\Sigma R' + n\lambda I))$ is the minimizer of $M(\lambda)$ defined by

$$M(\lambda) = \frac{x'(R\Sigma R' + n\lambda I)^{-1}x}{\left[\det(R\Sigma R' + n\lambda I)^{-1}\right]^{1/n-m}}. \tag{2.9}$$

Substituting in $x = Ry$ and $\det(R\Sigma R' + n\lambda I)^{-1} = \det{}^{+}R'(R\Sigma R' + n\lambda I)^{-1}R$ gives

$$M(\lambda) = \frac{y'R'(R\Sigma R' + n\lambda I)^{-1}Ry}{\left[\det{}^{+}R'(R\Sigma R' + n\lambda I)^{-1}R\right]^{1/n-m}}. \tag{2.10}$$

To put $M(\lambda)$ in final form for further study, we observe that

$$\begin{aligned} R'(R\Sigma R' + n\lambda I)^{-1}R &= (\Sigma + n\lambda I)^{-1} - (\Sigma + n\lambda I)^{-1} \\ &\quad \times T\left(T'(\Sigma + n\lambda I)^{-1}T\right)^{-1}T'(\Sigma + n\lambda I)^{-1}. \end{aligned} \tag{2.11}$$

To see this, note that both sides of (2.11) have the same action on the m columns of T and the $n - m$ columns of $(\Sigma + n\lambda I)R'$. It can be shown from, e.g., Kimeldorf and Wahba (1971) that $I - A(\lambda)$ is equal to $n\lambda$ times the right-hand side of (2.11). Thus for $\lambda > 0$, $M(\lambda)$ can be rewritten

$$M(\lambda) = \frac{y'(I - A(\lambda))y}{[\det{}^{+}(I - A(\lambda))]^{1/(n-m)}}. \tag{2.12}$$

Anderssen and Bloomfield (1974) were the first to suggest the use of a maximum likelihood estimate for λ in a smoothing context, and (2.12) will reduce essentially to their estimate in the case of a proper prior, that is, when the set of $\{\phi_\nu\}$ is empty, equivalently $\|Pf\|_Q^2 = \|f\|_Q^2$. Barry (1983), in a forthcoming thesis, has recently obtained the equivalent of (2.12) in two cases where the dimension of the null space of P is one. In the two cases he studied, the joint distributions for the $n - 1$ variables $(y_2 - y_1, \ldots, y_n - y_{n-1})$ or $(y_1 - \bar{y}, \ldots, y_{n-1} - \bar{y})$ are proper, and he exploited this fact to obtain his estimate. Thus the GML estimate generalizes the estimate obtained by Barry.

We compare this result with a maximum likelihood estimate for λ given by Wecker and Ansley (1983), (4.5). By making the associations, their λ is our $1/n\lambda$ and their Λ is our $(n\lambda)^{-1}(\Sigma + n\lambda I)$, and using (2.11) it can be shown that their maximum likelihood estimate is the minimizer of $M_{\mathrm{WA}}(\lambda)$ given by

$$M_{\mathrm{WA}}(\lambda) = \frac{y'R(R\Sigma R + n\lambda I)^{-1}Ry}{[\det(\Sigma + n\lambda I)]^{1/n}}, \tag{2.13}$$

which is to be compared with (2.10) and (2.11). The difference results from the fact that they include the estimation of $(\theta_1, \ldots, \theta_m)$ of (2.3) as part of the likelihood equations while we do not. [See O'Hagan (1975) for more on the role of nuisance parameters in ML estimation.] We remark that Wecker and Ansley are in error in their claim that GCV cannot be done with repeated observations.

Neither the GCV nor the maximum likelihood estimates require Σ to be of full rank. The only condition on the observations is that the matrix T be of rank m.

3. Asymptotic behavior of λ_{GML}. Let $R_{n-m \times n}$ be defined as in the previous section and let the eigenvalue eigenvector decomposition of $R\Sigma R'$ be

$$R\Sigma R' = UDU' \tag{3.1}$$

where $UU' = I_{n-m}$ and D_{n-m} is diagonal with diagonal entries $\lambda_{\nu n}$. Let

$$W_{n \times n-m} = R'U'$$

$$(w_{1n}, \ldots, w_{n-m,n})' = w_{n-m} = W'y.$$

Then

$$\begin{aligned} I - A(\lambda) &= n\lambda R'(UDU' + n\lambda UU')^{-1}R \\ &= n\lambda R'U'(D + n\lambda I)^{-1}UR \\ &= n\lambda W(D + n\lambda I)^{-1}W' \end{aligned}$$

and

$$M(\lambda) = \frac{\sum_{\nu=1}^{n-m} w_{\nu n}^2 (n\lambda/(n\lambda + \lambda_{\nu n}))}{\left(\prod_{\nu=1}^{n-m}(1 - \lambda_{\nu n}/(n\lambda + \lambda_{\nu n}))\right)^{1/n-m}}. \tag{3.2}$$

Letting

$$\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} L_1 f \\ \vdots \\ L_n f \end{pmatrix}, \qquad \begin{pmatrix} g_{1,n} \\ \vdots \\ g_{n-m,n} \end{pmatrix} = W'h \tag{3.3}$$

we have

$$EM(\lambda) = \frac{\sum_{\nu=1}^{n-m}[n\lambda/(n\lambda + \lambda_{\nu n})] g_{\nu n}^2 + \sigma^2 \sum_{\nu=1}^{n-m} n\lambda/(n\lambda + \lambda_{\nu n})}{\prod_{\nu=1}^{n-m}(1 - \lambda_{\nu n}/(n\lambda + \lambda_{\nu n}))^{1/(n-m)}}. \tag{3.4}$$

Letting

$$G(\lambda) = \sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2}{(n\lambda + \lambda_{\nu n})}, \tag{3.5}$$

$$\mu_1(\lambda) = \frac{1}{n-m} \sum_{\nu=1}^{n-m} \frac{\lambda_{\nu n}}{n\lambda + \lambda_{\nu n}} = \frac{1}{n-m}[\operatorname{tr} A(\lambda) - m], \tag{3.6}$$

$$D(\lambda) = \prod_{\nu=1}^{n-m} \left(1 - \frac{\lambda_{\nu n}}{n\lambda + \lambda_{\nu n}}\right)^{1/(n-m)} \tag{3.7}$$

gives

$$\frac{1}{n} EM(\lambda) = \frac{\lambda G(\lambda) + \tilde{\sigma}^2(1 - \mu_1(\lambda))}{D(\lambda)}, \tag{3.8}$$

where $\tilde{\sigma}^2 = ((n-m)/n)\sigma^2$.

We first assume $f \in \mathscr{H}_Q$. We must consider the cases $\|Pf\|_Q^2 = 0$ and $\|Pf\|_Q^2 > 0$ separately. Now $\|Pf\|_Q^2 = 0$ if and only if $f = \sum_{\nu=1}^m \theta_\nu \phi_\nu$ for some $\theta = (\theta_1, \dots, \theta_m)$, then $h = T\theta$, $g = W'h = 0$, and $G(\lambda) = 0$, all λ. Then

$$\frac{1}{\sigma^2(n-m)} EM(\lambda) = \frac{(1/(n-m))\sum_{\nu=1}^{n-m} n\lambda/(n\lambda + \lambda_{\nu n})}{\left(\prod_{\nu=1}^{n-m} n\lambda/(n\lambda + \lambda_{\nu n})\right)^{1/(n-m)}}, \tag{3.9}$$

and the right-hand side, being the ratio of an arithmetic to a geometric mean, is bounded below by 1. Assuming that the $\lambda_{\nu n}$ are not all equal, this expression achieves its lower bound for $\lambda = \infty$.

We now return to the case $\|Pf\|_Q^2 > 0$. Differentiating the right-hand side of (3.8) with respect to λ, using the fact that

$$D'(\lambda) = \frac{D(\lambda)}{\lambda} \mu_1(\lambda), \tag{3.10}$$

and setting the result equal to 0 gives

$$\begin{gathered} \left[\lambda G(\lambda) + \tilde{\sigma}^2(1 - \mu_1(\lambda))\right]\left[\frac{D(\lambda)}{\lambda}\mu_1(\lambda)\right] \\ = D(\lambda)\left[\lambda G'(\lambda) + G(\lambda) - \tilde{\sigma}^2\mu_1'(\lambda)\right], \\ \tilde{\sigma}^2\left[\mu_1(\lambda) + \lambda\mu_1'(\lambda) - \mu_1^2(\lambda)\right] = \lambda G(\lambda) + \lambda^2 G'(\lambda) - \lambda G(\lambda)\mu_1(\lambda). \end{gathered} \tag{3.11}$$

Now

$$\begin{aligned} \mu_1(\lambda) + \lambda\mu_1'(\lambda) &= \frac{1}{n-m}\left[\sum_{\nu=1}^{n-m} \frac{\lambda_{\nu n}}{n\lambda + \lambda_{\nu n}} - \frac{n\lambda\lambda_{\nu n}}{(n\lambda + \lambda_{\nu n})^2}\right] \\ &= \frac{1}{n-m}\sum_{\nu=1}^{n-m}\left(\frac{\lambda_{\nu n}}{n\lambda + \lambda_{\nu n}}\right)^2 \\ &= \mu_2(\lambda), \quad \text{say}, \end{aligned} \tag{3.12}$$

and

$$\begin{aligned} \lambda G(\lambda) + \lambda^2 G'(\lambda) &= \sum_{\nu=1}^{n-m} \frac{\lambda g_{\nu n}^2}{(n\lambda + \lambda_{\nu n})} - \frac{n\lambda^2 g_{\nu n}^2}{(n\lambda + \lambda_{\nu n})^2} \\ &= \lambda \sum_{\nu=1}^{n-m} \frac{\lambda_{\nu n} g_{\nu n}^2}{(n\lambda + \lambda_{\nu n})^2} \\ &= \lambda G_1(\lambda), \quad \text{say}. \end{aligned} \tag{3.13}$$

Thus, (3.11) can be written

$$\tilde{\sigma}^2\mu_2(\lambda) - \lambda G_1(\lambda) = \mu_1(\lambda)\left[\tilde{\sigma}^2\mu_1(\lambda) - \lambda G(\lambda)\right]. \tag{3.14}$$

It is well known that if $0 < \|Pf\|_Q < \infty$ it is necessary that $\lambda \to 0$ and $\mu_2(\lambda) \to 0$ in order that $R(\lambda) \to 0$. Since $\mu_2(\lambda) \geq \mu_1^2(\lambda)$, we will only consider roots of this equation for which $\lambda \to 0$, $\mu_2(\lambda) \to 0$, and $\mu_1(\lambda) \to 0$.

We now want to impose a further "smoothness" condition. Further discussion of this condition will appear in Section 4. Define

$$J_p^n = \sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2/n}{(\lambda_{\nu n}/n)^p}$$

and suppose that $J_p^n \le J_p$ for some $p \in (1,2]$ independent of n. Then

$$G_1(0) - G_1(1) = \sum \left[\frac{n\lambda}{n\lambda + \lambda_{\nu n}} \right]^2 \frac{g_{\nu n}^2}{\lambda_{\nu n}} + \sum_{\nu=1}^{n-m} \frac{2n\lambda}{(n\lambda + \lambda_{\nu n})^2} g_{\nu n}^2$$

$$\le 3\lambda^{p-1} \sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2/n}{(\lambda_{\nu n}/n)^p} \le 3\lambda^{p-1} J_p,$$

$$G(0) - G(\lambda) = \sum_{\nu=1}^{n-m} \frac{n\lambda}{(n\lambda + \lambda_{\nu n})} \frac{g_{\nu n}^2}{\lambda_{\nu n}} \le \lambda^{p-1} \sum \frac{g_{\nu n}^2/n}{(\lambda_{\nu n}/n)^p} \le \lambda^{p-1} J_p$$

and so, as $\lambda \to 0$, $G(\lambda) = G(0) + o(1)$, $G_1(\lambda) = G_1(0) + o(1)$, independent of n. Now

$$G_1(0) = G(0) = \sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2}{\lambda_{\nu n}}.$$

It can be shown that

$$\sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2}{\lambda_{\nu n}} = \|Pf_n\|_Q^2 \le \|Pf\|_Q^2, \tag{3.15}$$

where f_n is that element in $\mathscr{H}_Q$ which minimizes $\|Pf\|_Q$ subject to $L_i f_n = h_i$, $i = 1, 2, \dots, n$. The demonstration proceeds by showing that

(3.16)

$$\|Pf_n\|_Q^2 = h'\left(\Sigma^{-1} - \Sigma^{-1}T(T'\Sigma^{-1}T)^{-1}T'\Sigma^{-1}\right)h = h'R'(R\Sigma R')^{-1}Rh.$$

Thus, as $n \to \infty$, $G(0)$ and $G_1(0)$ increase monotonically but are bounded above by $\|Pf\|_Q^2$. (Assuming, of course, that the set $\{L_1, \dots, L_{n+1}\}$ contains the set $\{L_1, \dots, L_n\}$.) We have the following

THEOREM. *Suppose*

$$\sum \frac{g_{\nu n}^2/n}{(\lambda_{\nu n}/n)^p} < J_p \quad \text{for some } p > 1, \text{independent of } n$$

and

$$\mu_1(\lambda) = \frac{l}{n\lambda^{1/r}}(1 + o(1)), \tag{3.17}$$

$$\mu_2(\lambda) = \frac{\tilde{l}}{n\lambda^{1/r}}(1 + o(1)), \tag{3.18}$$

as $n \to \infty$ for some $r > 1$. Then,

$$\lambda_{\text{GML}} = \left(\frac{\tilde{\sigma}^2 \tilde{l}}{\|Pf_n\|_Q^2} \right)^{r/(r+1)} \frac{1}{n^{r/(r+1)}}(1 + o(1)) \tag{3.19}$$

is a zero of (3.14).

PROOF. Substituting (3.17) and (3.18) into (3.14) gives

$$(3.20)\quad \left(\tilde{\sigma}^2 \frac{\tilde{l}}{n\lambda^{1/r}} - \lambda G_1(0)\right)(1 + o(1)) = \frac{l}{n\lambda^{1/r}}\left[\frac{\tilde{\sigma}^2 l}{n\lambda^{1/r}} - \lambda G(0)\right](1 + o(1)),$$

which is satisfied by (3.19). □

In the case $L_i f = f(t_{in})$, with $t_{in} = i/n$, it is generally believed that the asymptotic behavior of $\mu_1(\lambda)$ and $\mu_2(\lambda)$ can be related to the asymptotic behavior of the eigenvalues of the reproducing kernel Q_1. [See the heuristic argument in Wahba (1977b).] In the special spline smoothing case with $H_Q = W_2^m$, it is known that (3.17) and (3.18) are satisfied with $r = 2m$. [See Craven and Wahba (1979) and Utreras (1980, 1981, 1983).] Roughly, the $\lambda_{\nu n}$ behave like n times the eigenvalues of the reproducing kernel.

Let $w(s)$ be a strictly positive smooth density on [0, 1], let

$$F(t) = \int_0^t w(s)\, ds,\ K_1(s, t)$$

$$= Q_1(F^{-1}(s), F^{-1}(t)), \text{ and } t_{in} \text{ satisfy } i/n = \int_0^{t_{in}} w(s)\, ds.$$

Then by the same reasoning as in Wahba (1977b) one could argue that the behavior of $\mu_1(\lambda)$ and $\mu_2(\lambda)$ can be related to the behavior of the eigenvalues of K_1. Letting $\{\xi_\nu\}$ and $\{\psi_\nu\}$ be the eigenvalues and eigenfunctions of K_1, we have

$$\xi_\nu \psi_\nu(t) = \int_0^1 Q_1(F^{-1}(t), F^{-1}(s))\psi_\nu(s)\, ds,$$

and making the change of variables $y = F^{-1}(t)$, $x = F^{-1}(s)$ gives

$$\xi_\nu \psi_\nu(F(y))w^{1/2}(y) = \int w^{1/2}(y) Q_1(y, x) w^{1/2}(x)\psi_\nu(F(x))w^{1/2}(x)\, dx,$$

which shows that the eigenvalues of K_1 are the same as the eigenvalues of $\tilde{K}_1(y, x) = w^{1/2}(y)Q_1(y, x)w^{1/2}(x)$. Now if Q_1 is the Green's function for a $2m$th order self-adjoint differential operator, $\tilde{K}_1$ is also the Green's function for a $2m$th order self-adjoint differential operator, its eigenvalues are $\xi_\nu = O(\nu^{-2m})$ and the same heuristic argument gives (3.17) and (3.18) satisfied with $r = 2m$. In the case of thin plate splines in d dimensions with

$$\|Pf\|_Q^2 = \sum_{\alpha_1 + \cdots + \alpha_d = m} \int \cdots \int \left(\frac{\partial^m}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right)^2 dx_1 \cdots dx_d,$$

partial results are available to the effect that (3.17) and (3.18) are satisfied with $r = 2m/d$. [See Cox (1984) and Wahba (1979c).] For information concerning tensor product splines on the plane, see Barry (1983), Micchelli and Wahba (1981) and Wahba (1978a); for splines on the sphere, see Wahba (1981). Behavior of $\mu_1(\lambda)$ for other L_i is discussed in Section 6, and very recent results of Nychka and Cox (1984) shed light on this question.

Inspection of the proof of the Theorem reveals that if $0 < \|Pf\| < \infty$, and the L_i are such that $\|Pf_n\|^2 \uparrow \|Pf\|_Q^2$, further smoothness assumptions of f cannot change the asymptotic behavior of λ_{GML}. A simple example will be given in the next section.

To complete our study of the behavior of λ_{GML}, we now consider the case f "behaves like" a sample function from the original prior. If f is a random function from the original prior, it can be shown that

$$E_f g_{\nu n}^2 = b\lambda_{\nu n},$$

where E_f is expectation with respect to this prior. Then

$$E_f G(0) = b \sum_{\nu=1}^{n-m} \frac{\lambda_{\nu n}}{\lambda_{\nu n}} = b(n-m)$$

and

$$E_f \|Pf\|_Q^2 = \infty.$$

Of course $\|Pf\|_Q^2 = \infty$ entails $f \notin \mathscr{H}_Q$. To see quickly what will happen, set $g_{\nu n}^2 = b\lambda_{\nu n}$ in (3.4) giving

$$\frac{1}{b} E_f EM(\lambda) = \frac{\sum_{\nu=1}^{n-m}(n\lambda/(n\lambda+\lambda_{\nu n}))(\lambda_{\nu n}+\sigma^2/b)}{\left[\sum_{\nu=1}^{n-m}(1-\lambda_{\nu n}/(n\lambda+\lambda_{\nu n}))\right]^{1/(n-m)}}. \quad (3.21)$$

Differentiation of the right-hand side of (3.21) or substitution into (3.14) gives that (3.21) is minimized by

$$\lambda_{GML} = \frac{\sigma^2}{nb}.$$

It appears that Wecker and Ansley's ML estimate, call it λ_{WA} would be only approximately equal to σ^2/nb in this case as $n \to \infty$, and slightly suboptimal in the case $f \in \pi^{m-1}$. It appears that we will have $\lambda_{GML} = O(1/n)$, if f is a fixed function (not in $\mathscr{H}_{Q_1}$!) and the L_i are such that

$$\lim_{n\to\infty} \frac{1}{n-m} \sum_{\nu=1}^{n-m} \frac{g_{\nu n}}{\lambda_{\nu n}} = \text{const}.$$

Conditions under which this will occur are suggested in the next section.

4. Comparison of λ_{GML}, λ_{opt}, and λ_{GCV}. We first consider the case $f \in \mathscr{H}_Q$. The predictive mean square error $R(\lambda)$ is defined by

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

and

$$ER(\lambda) = \frac{1}{n}\|(I - A(\lambda))h\|^2 + \frac{\sigma^2}{n} \operatorname{tr} A^2(\lambda) \quad (4.1)$$

or

$$ER(\lambda) - \frac{\sigma^2 m}{n} = \frac{1}{n} \sum_{\nu=1}^{n-m} \left(\frac{n\lambda}{n\lambda+\lambda_{\nu n}}\right)^2 g_{\nu n}^2 + \tilde{\sigma}^2 \mu_2(\lambda). \quad (4.2)$$

If $\Sigma g_{\nu n}^2 = 0$, then

$$ER(\lambda) = \frac{\sigma^2}{n} \operatorname{tr} A^2(\lambda), \tag{4.3}$$

which is minimized for $\lambda = \lambda_{\mathrm{GML}} = \infty$, in which case $f_{n\lambda}$ is that element in $\operatorname{span}\{\phi_\nu\}$ best fitting the data in the least squares sense and

$$ER(\infty) = \sigma^2 \frac{m}{n}. \tag{4.4}$$

Thus, from (3.9)

$$(1)\ f \in \operatorname{span}\{\phi_\nu\} \Rightarrow I_{\mathrm{GML/opt}} = 1. \tag{4.5}$$

Returning to the general case we have

$$ER(\lambda) - \sigma^2 \frac{m}{n} = \frac{1}{n} \sum_{\nu=1}^{n-m} \left(\frac{n\lambda}{n\lambda + \lambda_{\nu n}} \right)^2 g_{\nu n}^2 + \tilde{\sigma}^2 \mu_2(\lambda) \tag{4.6}$$

$$\le \lambda \sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2}{\lambda_{\nu n}} + \tilde{\sigma}^2 \mu_2(\lambda) \tag{4.7}$$

$$\le \lambda \|Pf\|_Q^2 + \tilde{\sigma}^2 \mu_2(\lambda). \tag{4.8}$$

If $\mu_2(\lambda) = (\tilde{l}/n\lambda^{1/r})(1 + o(1))$ for some r, the right-hand side of (4.8) is minimized by setting

$$\lambda^* = \left(\frac{\tilde{\sigma}^2 \tilde{l}}{r\|Pf\|_Q^2} \right)^{r/(r+1)} \frac{1}{n^{r/(r+1)}} (1 + o(1)) \tag{4.9}$$

and $ER(\lambda_{\mathrm{opt}}) \le ER(\lambda^*) = O(n^{-r/(r+1)})$ ($= O(n^{-2m/(2m+1)})$ in the special spline case). If no further assumptions are made on f it appears that this rate cannot be improved upon.

However, it is well known that if f satisfies certain additional smoothness conditions then higher rates of convergence can be obtained by choosing λ to go to 0 more slowly. We always have, for any $p \in [1, 2]$,

$$\frac{1}{n} \sum_{\nu=1}^{n-m} \left(\frac{n\lambda}{n\lambda + \lambda_{\nu n}} \right)^2 g_{\nu n}^2 \le \lambda^p \sum_{\nu=1}^{n-m} \frac{(g_{\nu n}^2/n)}{(\lambda_{\nu n}/n)^p} = \lambda^p J_p^n. \tag{4.10}$$

If J_p^n is uniformly bounded by J_p independent of n, and $\mu_2(\lambda) = (\tilde{l}/n\lambda^{1/r})(1 + o(1))$, then

$$ER(\lambda) - \sigma^2 \frac{m}{n} \le \lambda^p J_p + \frac{\tilde{\sigma}^2 \tilde{l}}{n\lambda^{1/r}} (1 + o(1)). \tag{4.11}$$

The right-hand side of (4.11) is minimized by

$$\lambda_p^* = \left(\frac{\tilde{\sigma}^2 \tilde{l}}{rpJ_p} \right)^{r/(r+1)} \frac{1}{n^{r/(rp+1)}} (1 + o(1))$$

and

$$ER(\lambda_{\text{opt}}) \le ER(\lambda_p^*) = O\left(\frac{1}{n^{rp/(rp+1)}}\right). \tag{4.12}$$

Thus

$$I_{\text{GML/opt}} = \text{const}\; n^{rp/(rp+1) - r/(r+1)} \to \infty. \tag{4.13}$$

The condition $J_p^n \le J_p$ is satisfied in some interesting cases. For example, suppose $\mathscr{H}_Q = W_2^m(\text{per})$, the space of periodic functions in $W_2^m[0,1]$ satisfying the periodic boundary conditions $f^{(\nu)}(0) = f^{(\nu)}(1)$, $\nu = 0, 1, \dots, m-1$, and suppose $L_i f = f(i/n)$. Then, very roughly [for details, see Craven and Wahba (1979) and Utreras (1980)],

$$g_{\nu n}^2 \simeq n f_n^2, \qquad f_\nu = \int_0^1 f(s)\cos(2\pi\nu s)\,ds \quad \text{or} \quad \int_0^1 f(s)\sin(2\pi\nu s)\,ds,$$

$$\lambda_{\nu n} \simeq n\lambda_\nu, \qquad \lambda_\nu = (2\pi\nu)^{-2m}.$$

Then if $f \in W_2^{mp}(\text{per})$ for some $1 < p \le 2$, we have

$$\int_0^1 \left(f^{(mp)}(s)\right)^2 ds = \sum (2\pi\nu)^{2mp} f_\nu^2 \simeq \sum \frac{g_{\nu n}^2/n}{(\lambda_{\nu n}/n)^p}(1 + o(1)).$$

This example, with $f(t) = f_0 \cos 2\pi\nu_0 t$, say, can be used to show directly that $ER(\lambda_{\text{GML}})$ is still $O(n^{-2m/(2m+1)})$ while $ER(\lambda_{\text{opt}}) = O(n^{-4m(4m+1)})$ by observing that in this example

$$\lambda G_1(\lambda) \simeq \lambda \frac{(2\pi\nu_0)^{2m} f_0^2}{\left(1 + \lambda(2\pi\nu_0)^{2m}\right)^2}$$

and

$$\frac{1}{n}\sum_{\nu=1}^{n-m}\left(\frac{n\lambda}{n\lambda + \lambda_{\nu n}}\right)^2 g_{\nu n}^2 \simeq \lambda^2 \frac{(2\pi\nu_0)^{4m} f_0^2}{\left(1 + \lambda(2\pi\nu_0)^{2m}\right)^2},$$

and carrying through the minimizations directly.

For the case $L_i f = f(t_i)$, we state as a conjecture a general condition for $J_p^n = \sum_{\nu=1}^{n-m}(g_{\nu n}^2/n)/(\lambda_{\nu n}/n)^p$ to be uniformly bounded. Suppose $Q_1(s,t)$ possess the Mercer–Hilbert–Schmidt expansion

$$Q_1(s,t) = \sum_{\nu=1}^{\infty} \lambda_\nu u_\nu(s) u_\nu(t),$$

where the $\{\lambda_\nu, u_\nu\}$ are the eigenvalues and the eigenfunctions of Q_1. (For this it is sufficient that $\int\int_{\mathscr{T}\times\mathscr{T}} Q_1^2(s,t)\,ds\,dt = \sum_{\nu=1}^{\infty}\lambda_\nu < \infty$ [see Riesz and Sz. Nagy (1955)]; this condition is being implicitly assumed throughout this paper.) Let $J_p^*(f)$ be defined by

$$J_p^*(f) = \sum_{\nu=1}^{\infty} \frac{(Pf, u_\nu)^2}{\lambda_\nu^p}, \qquad (f, u_\nu) = \int f(t) u_\nu(t)\,dt. \tag{4.14}$$

We say that $f \in \mathscr{C}_p^*$ if $0 < J_p^*(f) < \infty$. It is conjectured that $f \in \mathscr{C}_p^*$ and some regularity conditions on the $\{t_i\} = \{t_{in}\}_{i=1}^n$, $n = 1, 2, \ldots$, imply that there is some constant c_p such that

$$0 < J_p^n(f) \le c_p J_p^*(f).$$

To see the behavior of $R(\lambda)$ when f behaves like a sample function, we only consider the case $g_{\nu n} = \pm b\lambda_{\nu n}$; we suggest that the results can be extended to functions f for which $\lim_{n\to\infty} 1/(n-m)\sum_{\nu=1}^{n-m} g_{\nu n}^2/\lambda_{\nu n} \to$ const, and we conjecture that it is sufficient for this that some regularity conditions hold in the L_i and

$$Pf \sim \sum_{\nu=1}^{\infty} f_\nu u_\nu, \qquad f_\nu = (f, u_\nu),$$

with $\lim_{n\to\infty}(1/n)\sum_{\nu=1}^n f_\nu^2/\lambda_\nu \to b$ for some $0 < b < \infty$. Letting $g_{\nu n}^2 = b\lambda_{\nu n}$ in (4.2) gives

$$nb\left[ER(\lambda) - m\sigma^2\right] = \sum_{\nu=1}^{n-m} \lambda_{\nu n}\left\{\frac{(n\lambda)^2 + (\sigma^2/b)\lambda_{\nu n}}{(n\lambda + \lambda_{\nu n})^2}\right\}.$$

Differentiating the expression in braces on the right with respect to λ and setting the result equal to 0 gives

$$\left[(n\lambda)^2 + \frac{\sigma^2}{b}\lambda_{\nu n}\right] n(n\lambda + \lambda_{\nu n}) = (n\lambda + \lambda_{\nu n})^2[n(n\lambda)],$$

which is satisfied for

$$\lambda_{\text{opt}} = \frac{\sigma^2}{nb}$$

for every $\lambda_{\nu n}$. Thus, in this example (assuming λ_{opt} and λ_{GML} are global minimizers), $I_{\text{GML/opt}} = 1$.

We can summarize the result of the last two sections as follows:

Let $\mathscr{H}_Q$ and $L_1, \ldots, L_n$ be such that

$$\mu_1(\lambda) \to \frac{l}{n\lambda^{1/r}},$$

$$\mu_2(\lambda) \to \frac{\tilde{l}}{n\lambda^{1/r}}$$

for some $r > 1$, as $n \to \infty$, $\lambda \to 0$, and let $\mathscr{C}_p$ be defined by

$$\mathscr{C}_p = \left\{ f\colon \|Pf\|_Q > 0 \text{ and } \sum_{\nu=1}^{n-m} \frac{g_{\nu n}^2/n}{(\lambda_{\nu n}/n)^p} \le J_p(f)(1 + o(1)) \right\}$$

for some constant J_p independent of n. Then we have the results in Table 1.

We remark that we do not prove, but merely state as a conjecture that in the special spline case with $t_i = i/n$, and mp an integer, that the definitions of $\mathscr{C}_p$ here and in the introduction are equivalent, and that the methods in, e.g., Cox (1983b) and Rice and Rosenblatt (1983), can be used to show it.

TABLE 1

(1)	$f \in \operatorname{span}\{\phi_\nu\} \Rightarrow I_{\mathrm{GML/opt}} = 1$
(2)	$f \in \mathscr{C}_p$ for some $p > 1 \Rightarrow I_{\mathrm{GML/opt}} \to \infty$
(3)	f "behaves like" a sample function $\Rightarrow I_{\mathrm{GML/opt}} = 1 + o(1)$.

To compare λ_{GML} and λ_{GCV}, we have, from Craven and Wahba, that

(1) $f \in \operatorname{span}\{\phi_\nu\} \Rightarrow I_{\mathrm{GCV/opt}} = 1$.
(2) $f \in \mathscr{H}_Q$, $\mu_1(\lambda_{\mathrm{opt}}) \to 0$ and $\mu_1^2(\lambda_{\mathrm{opt}})/\mu_2(\lambda_{\mathrm{opt}}) \to 0$ imply $I_{\mathrm{GCV/opt}} = 1 + o(1)$.

Although the arguments are carried out for a special case, it is seen by following them that the results hold in the generality of this paper.

Now, suppose f behaves like a sample function from the stochastic process. $V(\lambda)$ is given by

$$nV(\lambda) = \frac{\sum_{\nu=1}^{n-m} w_{\nu n}^2 (n\lambda/(n\lambda + \lambda_{\nu n}))^2}{(\sum_{\nu=1}^{n-m}(1 - \lambda_{\nu n}/(n\lambda + \lambda_{\nu n})))^2}$$

and

$$(4.15)\quad nEV(\lambda) = \frac{\sum_{\nu=1}^{n-m}(n\lambda/(n\lambda + \lambda_{\nu n}))^2 g_{\nu n}^2 + \sigma^2 \sum_{\nu=1}^{n-m}(n\lambda/(n\lambda + \lambda_{\nu n}))}{(\sum_{\nu=1}^{n-m}(1 - \lambda_{\nu n}/(n\lambda + \lambda_{\nu n})))^2}.$$

Replacing $g_{\nu n}^2$ by its expected value $b\lambda_{\nu n}$ under the prior in (4.5) gives

$$(4.16)\quad \frac{1}{b} E_f nEV(\lambda) = \frac{\sum_{\nu=1}^{n-m}(n\lambda/(n\lambda + \lambda_{\nu n}))^2(\lambda_{\nu n} + \sigma^2/b)}{(\sum_{\nu=1}^{n-m} n\lambda/(n\lambda + \lambda_{\nu n}))^2}$$

and a straightforward calculation [which appears in Wahba (1977a)] shows that the right-hand side of this expression is minimized for $\lambda = \sigma^2/nb$. It appears that, for $g_{\nu n}^2/\lambda_{\nu n} \sim$ const, we have $\lambda_{\mathrm{GCV}} = O(1/n)$. The proof of Theorem 4.2 in Craven and Wahba shows that $\lambda_{\mathrm{opt}} = O(1/n)$, $\lambda_{\mathrm{GCV}} = O(1/n)$ and $\mu_1(\lambda) \simeq l/n\lambda^{1/r}$, $\mu_2(\lambda) = \tilde{l}/n\lambda^{1/r}$ for some $r > 1$ entails that $I_{\mathrm{GCV/opt}} = 1 + o(1)$. We conclude that

(3) f "behaves like" a sample function $\Rightarrow I_{\mathrm{GCV/opt}} = 1 + o(1)$.

Thus, in each of the three entries in Table 1, we may replace $I_{\mathrm{GML/opt}}$ by $I_{\mathrm{GML/GCV}}$.

5. Monte Carlo results. A Monte Carlo study was carried out to see whether some of the preceding asymptotic results would be manifest in small to medium sized samples. Three experimental test functions were used, given in Cases 1, 2, and 3 below.

CASE 1. $f(t) = \frac{1}{3}\beta_{10,5}(t) + \frac{1}{3}\beta_{7,7}(t) + \frac{1}{3}\beta_{5,10}(t)$.

CASE 2. $f(t) = \frac{6}{10}\beta_{30,17}(t) + \frac{4}{10}\beta_{3,11}(t)$.

CASE 3. $f(t) = \frac{1}{3}\beta_{20,5}(t) + \frac{1}{3}\beta_{12,12}(t) + \frac{1}{3}\beta_{7,30}(t)$, where

$$\beta_{p,q}(t) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1}(1-t)^{q-1}, \qquad 0 \le t \le 1.$$

We considered only $m = 2$, all of these functions are in W_2^4(per), and the periodic smoothing spline in W_2^2 was implemented. The study reported here was done simultaneously with the Monte Carlo study in Wahba (1983) but not published at that time. Plots of f of Cases 1–3 above and sample Monte Carlo data appear there, and some of the values of $I_{\mathrm{GCV/opt}}$ appearing here for comparison purposes are also reported there. We considered only $L_i f = f(i/n)$, $i = 1, 2, \ldots, n$, and $n = 32, 64, 128$. (Some examples with $n = 16$ were also tried but the results were erratic.) Five values of σ, $\sigma = 0.0125, 0.0250, 0.05, 0.1$, and 0.2, were tried. Since $\int_0^1 |f(t)|\, dt = 1$, the smallest two values of σ represent "engineering accuracy," or two-figure data, while a σ of 0.2 is one-figure data. For each of the 3 cases $\times$ 3 ns $\times$ 5 σs, 10 replicates were generated from the model

$$y_i = f\left(\frac{i}{n}\right) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \qquad i = 1, 2, \ldots, n.$$

For each replicate the GML and GCV estimates $\hat\lambda_{\mathrm{GML}}$ and $\hat\lambda_{\mathrm{GCV}}$, the minimizers of $V(\lambda)$ and $M(\lambda)$, were computed, along with $\tilde\lambda_{\mathrm{opt}}$, the minimizer of $R(\lambda)$. Then the inefficiencies $\hat I_{\mathrm{GML/opt}}$ and $\hat I_{\mathrm{GCV/opt}}$ defined by

$$\hat I_{\mathrm{GML}} = \frac{R(\hat\lambda_{\mathrm{GML}})}{R(\tilde\lambda_{\mathrm{opt}})}, \qquad \hat I_{\mathrm{GCV}} = \frac{R(\hat\lambda_{\mathrm{GCV}})}{R(\tilde\lambda_{\mathrm{opt}})}$$

were computed. The Appendix gives a complete table of $\hat I_{\mathrm{GML}}$ and $\hat I_{\mathrm{GCV}}$ for each

TABLE 2
$\bar I_{\mathrm{GML}}$, $\bar I_{\mathrm{GCV}}$, and the GML score.

		$\sigma = 0.0125$			$\sigma = 0.0250$			$\sigma = 0.05$			$\sigma = 0.10$			$\sigma = 0.20$		
n	Case	$\bar I_{\mathrm{GML}}$	$\bar I_{\mathrm{GCV}}$	GML score out of 10	$\bar I_{\mathrm{GML}}$	$\bar I_{\mathrm{GCV}}$	GML score	$\bar I_{\mathrm{GML}}$	$\bar I_{\mathrm{GCV}}$	GML score	$\bar I_{\mathrm{GML}}$	$\bar I_{\mathrm{GCV}}$	GML score	$\bar I_{\mathrm{GML}}$	$\bar I_{\mathrm{GCV}}$	GML score
	1	1.49	1.41	1	1.32	1.33	3	1.22	1.94	8	1.25	1.20	2	1.39	1.40	5
32	2	1.38	1.24	3	1.83	1.27	1	1.41	1.09	$1\frac{1}{2}$	1.23	1.07	$2\frac{1}{2}$	1.05	1.07	6
	3	1.50	1.45	4	1.51	1.13	0	1.62	1.43	1	1.17	1.11	3	1.12	2.02	$5\frac{1}{2}$
	1	1.40	1.07	2	1.40	1.09	0	1.23	1.05	0	1.48	1.83	$5\frac{1}{2}$	1.22	1.32	8
64	2	2.09	1.31	1	1.49	1.33	2	1.43	1.16	1	1.24	1.05	2	1.18	1.06	2
	3	1.94	1.10	0	1.51	1.06	0	1.30	1.14	1	1.20	1.21	1	1.12	1.45	4
	1	1.67	1.06	1	1.39	1.09	1	1.29	1.06	2	1.32	1.16	2	1.07	1.50	3
128	2	1.75	1.03	0	1.59	1.07	0	1.38	1.03	0	1.26	1.06	1	1.30	1.18	0
	3	1.69	1.09	1	1.34	1.07	0	1.28	1.07	2	1.20	1.04	0	1.23	1.19	3

of the $3 \times 3 \times 5$ sets of 10 replicates. A summary of this data appears in Table 2. For each n, σ, and case, Table 2 gives $\bar{I}_{\mathrm{GML}}$ and $\bar{I}_{\mathrm{GCV}}$, where I_{GML} is the average of the 10 replicated values of $\hat{I}_{\mathrm{GML}}$, and similarly for $\bar{I}_{\mathrm{GCV}}$. Table 2 also gives the GML score, defined as the number of times out of 10 replicates, that $\hat{I}_{\mathrm{GML}} < \hat{I}_{\mathrm{GCV}}$. A tie is counted as $\frac{1}{2}$. It can be seen that in the extreme NE corner of the table, $n = 32$, $\sigma = 0.2$, GML appears to have a modest edge over GCV (perhaps not "significant"), and mixed results obtain in other entries towards the NE. For smaller σ and all of the $n = 128$ entries except the $\sigma = 0.20$ case, the GCV edge is fairly striking. The results of this experiment, with $n = 32$ and 64, are in rough agreement with the Monte Carlo results of Barry (1983) and Davies et al. (1983). Barry considered $n = 20$ and 40, and Davies et al., considered $n = 50$.

6. The case of general L_i. We may study the asymptotic behavior of λ_{GML}, λ_{GCV}, and λ_{opt} with general L_i if the $\{L_i\}$ can be imbedded in a nice family L_s, $s \in \mathscr{S}$, of bounded linear functionals on $\mathscr{H}_Q$. This generally can be done if one is trying to solve a so-called Fredholm integral equation of the first kind. To show how this study proceeds, we first review some relevant facts from Nashed and Wahba (1974).

Let $\mathscr{S}$ be an index set and, for each $s \in \mathscr{S}$, let L_s be a bounded linear functional on $\mathscr{H}_Q$. Later we shall let $L_i = L_{s_i}$. We can define a linear operator $\mathscr{K}$ with domain $\mathscr{H}_Q$ and range contained in the real valued functions on $\mathscr{S}$ by

$$\mathscr{K}f = g, \qquad g(s) = (\mathscr{K}f)(s) = L_s f, \qquad f \in \mathscr{H}_Q, \qquad s \in \mathscr{S}.$$

The most interesting case concerns $\mathscr{K}$ an integral operator,

$$(\mathscr{K}f)(s) = \int K(s,t) f(t)\, dt, \qquad s \in \mathscr{S},$$

for some known K. It was shown by Nashed and Wahba (1974), that

$$\mathscr{K}(\mathscr{H}_Q) = \mathscr{H}_R,$$

where $\mathscr{H}_R$ is the reproducing kernel space with reproducing kernel $R(u, v)$ with

$$R(u,v) = L_{u(s)} L_{v(t)} Q(s,t),$$

which, if $\mathscr{K}$ is an integral operator, becomes

$$R(u,v) = \iint K(u,s) Q(s,t) K(v,t)\, ds\, dt.$$

We also have $\mathscr{K}(\mathscr{H}_{Q_1}) = \mathscr{H}_{R_1}$ where

$$R_1(u,v) = L_{u(s)} L_{v(t)} Q_1(s,t).$$

The null space of $\mathscr{K}$ in $\mathscr{H}_Q$ consists of all $f \in \mathscr{H}_Q$ with $L_s f = 0$, $s \in \mathscr{S}$. Let $\mathscr{V}$ be the null space perpendicular of $\mathscr{K}$ in $\mathscr{H}_Q$. If we endow $\mathscr{H}_R$ with its reproducing kernel space topology, then there is a 1 : 1 inner product perserving map between $\mathscr{V}$ and $\mathscr{H}_R = \mathscr{K}(\mathscr{V})$ under which

$$f \in \mathscr{V} \sim g = \mathscr{K}f \in \mathscr{H}_R \tag{6.1}$$

and

$$\langle f_1, f_2 \rangle_Q = \langle g_1, g_2 \rangle_R \tag{6.2}$$

whenever $f_1, f_2 \in \mathscr{V}$, $\mathscr{K}f_1 = g_1$, $\mathscr{K}f_2 = g_2$. Thus the geometry of $\mathscr{V}$ and $\mathscr{H}_R$ are the same under the 1 : 1 correspondence " $\sim$ " given in (6.1).

We assume that the dimension of the span of $\{\mathscr{K}\phi_\nu\}$ is m. (If it is not, T cannot be of rank m.) Let $\tilde{P}$ be the orthogonal projection in $\mathscr{H}_R$ onto $\mathscr{H}_{R_1}$ (which is the orthocomplement of span $\{\mathscr{K}\phi_\nu\}$). Letting $\mathscr{K}^+$ be defined, for g in $\mathscr{H}_R$, as that element in $\mathscr{H}_R$ of minimal norm which satisfies $\mathscr{K}f = g$, we have $\mathscr{K}^+(\mathscr{H}_R) = \mathscr{V} = \mathscr{K}^+(\mathscr{H}_{R_1}) \oplus \{\phi_\nu\}$, $P\mathscr{K}^+ g \sim \tilde{P}g$, and so $\|P\mathscr{K}^+ g\|_Q^2 = \|\tilde{P}g\|_R^2$. Let $g_{n,\lambda}$ be that element in $\mathscr{H}_R$ which minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\left(g(s_i) - y_i\right)^2 + \lambda\|\tilde{P}g\|_R^2.$$

Using $g(s_i) = L_i f$, $\|\tilde{P}g\|_R^2 = \|P\mathscr{K}^+ g\|_Q^2$, and the fact that $f_{n,\lambda}$ must be in $\mathscr{V}$, it can be shown that $\mathscr{K}f_{n,\lambda} = g_{n,\lambda}$ and $\mathscr{K}^+ g_{n,\lambda} = f_{n,\lambda}$. Furthermore,

$$\|f - f_{n,\lambda}\|_Q^2 = \|\mathscr{K}^+\mathscr{K}f - f_{n,\lambda}\|_Q^2 + \|f - \mathscr{K}^+\mathscr{K}f\|_Q^2$$

and

$$\|\mathscr{K}^+\mathscr{K}f - f_{n,\lambda}\|_Q^2 = \|g - g_{n,\lambda}\|_R^2$$

by (6.2). Further details may be found in Nashed and Wahba (1974).

Now consider the problem of studying the behavior of λ_{GML}, λ_{GCV}, and λ_{opt} for the case of general L_i, and suppose the L_i can be embedded in a family L_s, $s \in \mathscr{S}$, by $L_i = L_{s_i}$. The problem then reduces to examining the properties of g and $g_{n,\lambda}$ in $\mathscr{H}_R$, with the loss function of (1.11) becoming $R(\lambda) = (1/n)\sum_{i=1}^{n}(g(s_i) - g_{n,\lambda}(s_i))^2$. The entries of Σ are $R_1(s_i, s_j)$. Thus if the s_i are regularly distributed, the behavior of the $\{\lambda_{\nu n}\}$ will be related to the eigenvalues of R_1 (instead of Q_1). This can be used in some cases to establish the asymptotic behavior of $\mu_1(\lambda)$ and $\mu_2(\lambda)$ [see, e.g., Lukas (1981), Rice and Rosenblatt (1983), and Wahba (1977b)].

Conditions (1)–(3) on f can now be transported to conditions on $g = \mathscr{K}f$ and we have

(1) $g \in \text{span}\, \mathscr{K}\{\phi_\nu\} \Rightarrow I_{\text{GML/opt}} = 1$.
(2) $g \in \mathscr{C}_p$ for some $p > 1 \Rightarrow I_{\text{GML/opt}} \to \infty$.
(3) g "behaves like" a sample function $\Rightarrow I_{\text{GML/opt}} = 1 + o(1)$.

With some abuse of notation, we are letting $\mathscr{C}_p$ be defined by

$$\mathscr{C}_p = \left\{g\colon \|Pg\|_R^2 > 0 \text{ and } \sum_{\nu=1}^{n-m}\frac{g_{\lambda n}^2/n}{(\lambda_{\nu n}/n)^p} \le J_p(g)(1 + o(1))\right\},$$

for some constant J_p independent of n.

Let $\tilde{J}_p^*(g)$ be defined by

$$\tilde{J}_p^*(g) = \sum_{\nu=1}^{\infty}\frac{(\tilde{P}g, \tilde{u}_\nu)}{\tilde{\lambda}_\nu^p},$$

where $\{\tilde{\lambda}_\nu$ and $\tilde{u}_\nu\}$ are the eigenvalues and eigenvectors of R_1, and say that $g \in \mathscr{C}_p^*$ if $0 < J_p^*(g) < \infty$. It is now conjectured that $g \in \mathscr{C}_p^*$ and some regularity conditions on the $\{s_i\}$, $i = 1, \ldots, n$, imply that $g \in \mathscr{C}_p$. Since

$$\|g - g_{n,\lambda}\|_R^2 = \|\mathscr{K}^+\mathscr{K}f - f_{n,\lambda}\|_Q^2,$$

the study of the optimal λ with certain other loss functions referenced to f can be studied by examining the problem in $\mathscr{H}_R$. For example, compare the methods and conclusions of Theorems 1 and 2 of Wahba (1977b). There is some continuing research in this area [see Cox (1983a) and Nychka (1983)]. We remark that the generalized spline smoothing problem has recently been extended to nonlinear functionals and non-Gaussian errors by O'Sullivan (1983).

7. Acknowledgments. We wish to thank J. A. Hartigan for providing us with an early version of Barry's thesis, and we would like to acknowledge many stimulating discussions during the Madison miniseminar, especially with Dennis Cox, Tom Leonard, Mark Lukas and Doug Nychka. Special thanks go to Christopher Sheridan, presently with IBM, who wrote the computer program for the Monte Carlo studies.

Appendix. Values $\hat{I}_{\text{GML}}$ and $\hat{I}_{\text{GCV}}$ for all replicates. ISUBM $= \hat{I}_{\text{GML}}$; ISUBV $= \hat{I}_{\text{GCV}}$; REPL = replicate number.

CASE 1

		σ = 0.0125		σ = 0.025		σ = 0.050		σ = 0.10		σ = 0.20	
	REPL	ISUBM	ISUBV	ISUBM	ISUBV	ISUBM	ISUBV	ISUBM	ISUBV	ISUBM	ISUBV
$n = 32$	1	1.501	1.015	1.698	1.410	1.481	1.218	1.623	1.461	1.769	1.331
	2	1.317	1.018	1.066	1.013	1.016	1.195	1.040	1.625	1.030	1.307
	3	1.101	1.109	1.184	1.002	1.092	1.004	1.072	1.047	1.002	1.016
	4	1.282	1.087	2.071	1.442	1.210	4.219	1.961	1.340	1.084	1.021
	5	1.343	1.022	1.176	1.001	1.268	1.856	1.067	1.065	1.112	1.019
	6	1.440	1.034	1.054	1.108	1.027	1.069	1.262	1.129	1.010	1.243
	7	2.714	1.959	1.190	1.055	1.000	1.314	1.007	1.044	3.576	3.762
	8	1.637	1.200	1.117	1.124	1.792	2.953	1.109	1.048	1.026	1.241
	9	1.303	1.019	1.550	3.120	1.002	1.351	1.246	1.205	1.137	1.011
	10	1.226	1.000	1.094	1.003	1.310	3.317	1.096	1.014	1.239	1.065
$n = 64$	1	1.276	1.001	1.357	1.065	1.265	1.101	1.082	1.988	1.012	1.088
	2	1.351	1.010	1.156	1.058	1.068	1.027	2.597	1.811	1.000	1.239
	3	2.105	1.000	2.035	1.347	1.204	1.062	1.060	1.652	1.279	1.052
	4	1.230	1.004	1.155	1.069	1.107	1.000	1.187	2.100	1.045	1.095
	5	1.426	1.074	1.068	1.064	1.117	1.049	3.363	3.363	1.000	1.277
	6	1.879	1.087	1.730	1.178	1.175	1.008	1.083	1.024	1.002	1.136
	7	1.096	1.236	1.320	1.002	1.450	1.076	1.078	1.003	1.031	1.021
	8	1.118	1.073	1.152	1.007	1.422	1.096	1.223	3.262	1.012	1.181
	9	1.357	1.010	1.648	1.139	1.281	1.014	1.075	1.004	1.220	1.288
	10	1.114	1.235	1.396	1.005	1.210	1.019	1.000	1.109	2.556	2.865
$n = 128$	1	2.040	1.127	2.310	1.428	1.051	1.129	1.000	1.217	1.094	1.010
	2	1.511	1.000	1.394	1.179	1.344	1.027	1.269	1.027	1.186	1.020
	3	1.286	1.062	1.211	1.009	1.168	1.000	1.749	1.440	1.058	1.001
	4	2.124	1.091	1.419	1.020	1.422	1.039	1.159	1.009	1.053	1.676
	5	2.063	1.115	1.132	1.026	1.689	1.124	1.020	1.091	1.036	1.009
	6	1.208	1.007	1.072	1.112	1.586	1.155	1.113	1.010	1.177	1.023
	7	1.918	1.035	1.375	1.000	1.286	1.015	1.198	1.023	1.013	1.007
	8	1.933	1.085	1.191	1.000	1.093	1.041	2.492	1.810	1.000	1.192
	9	1.101	1.116	1.689	1.141	1.154	1.003	1.075	1.003	1.003	5.036
	10	1.513	1.000	1.145	1.023	1.077	1.078	1.094	1.007	1.088	1.023

Case 2

	REPL	σ = 0.125 ISUBM	σ = 0.125 ISUBV	σ = 0.025 ISUBM	σ = 0.025 ISUBV	σ = 0.050 ISUBM	σ = 0.050 ISUBV	σ = 0.10 ISUBM	σ = 0.10 ISUBV	σ = 0.20 ISUBM	σ = 0.20 ISUBV
$n = 32$	1	1.674	1.389	1.600	1.600	1.318	1.318	1.369	1.173	1.046	1.293
	2	1.234	1.234	1.860	1.860	1.503	1.000	1.114	1.004	1.003	1.007
	3	1.989	1.504	2.053	1.018	1.207	1.029	1.084	1.001	1.224	1.171
	4	1.601	1.601	1.772	1.013	1.317	1.001	1.001	1.107	1.000	1.027
	5	1.382	1.382	1.487	1.232	1.468	1.035	1.774	1.205	1.093	1.045
	6	1.176	1.176	1.841	1.133	1.369	1.053	1.137	1.050	1.017	1.047
	7	1.000	1.000	1.430	1.042	1.082	1.030	1.253	1.063	1.075	1.052
	8	1.275	1.093	3.454	1.149	1.043	1.223	1.306	1.032	1.002	1.012
	9	1.509	1.039	1.599	1.289	2.186	1.240	1.220	1.017	1.001	1.009
	10	1.000	1.000	2.205	1.387	1.607	1.021	1.068	1.044	1.037	1.000
$n = 64$	1	2.280	1.008	1.567	1.000	1.437	1.041	1.243	1.023	1.042	1.005
	2	1.676	1.004	1.139	1.073	1.547	1.053	1.470	1.134	1.130	1.010
	3	1.673	1.021	1.729	1.088	2.162	1.235	1.104	1.000	1.003	1.082
	4	2.220	2.220	1.843	2.630	1.480	2.219	1.434	1.070	1.431	1.137
	5	2.219	1.005	1.820	1.143	1.473	1.047	1.091	1.001	1.039	1.004
	6	2.789	2.730	1.241	1.025	1.389	1.000	1.285	1.012	1.566	1.227
	7	2.237	1.004	1.730	2.209	1.215	1.000	1.172	1.024	1.219	1.002
	8	1.564	1.010	1.302	1.050	1.255	1.005	1.491	1.089	1.098	1.000
	9	2.002	1.034	1.122	1.051	1.291	1.013	1.036	1.057	1.227	1.047
	10	2.186	1.047	1.440	1.050	1.124	1.026	1.027	1.081	1.002	1.101
$n = 128$	1	1.767	1.024	1.573	1.006	1.336	1.000	1.070	1.025	1.499	1.285
	2	1.315	1.019	1.379	1.011	1.638	1.107	1.166	1.031	1.087	1.002
	3	1.974	1.185	1.618	1.022	1.257	1.000	1.151	1.023	1.778	1.719
	4	1.736	1.005	1.744	1.038	1.271	1.004	1.285	1.016	1.258	1.133
	5	2.036	1.107	1.503	1.022	1.711	1.052	1.015	1.144	1.184	1.044
	6	2.365	1.014	2.315	1.486	1.188	1.001	1.122	1.000	1.021	1.036
	7	1.518	1.000	1.346	1.012	1.334	1.001	1.278	1.051	1.153	1.029
	8	1.485	1.000	1.438	1.006	1.351	1.000	1.765	1.250	1.806	1.446
	9	1.769	1.000	1.560	1.080	1.547	1.026	1.291	1.012	1.052	1.002
	10	1.490	1.002	1.433	1.019	1.205	1.062	1.416	1.004	1.109	1.076

Case 3

	REPL	σ = 0.0125 ISUBM	σ = 0.0125 ISUBV	σ = 0.025 ISUBM	σ = 0.025 ISUBV	σ = 0.050 ISUBM	σ = 0.050 ISUBV	σ = 0.10 ISUBM	σ = 0.10 ISUBV	σ = 0.20 ISUBM	σ = 0.20 ISUBV
$n = 32$	1	1.061	1.061	1.963	1.249	1.081	1.023	1.135	1.046	1.005	1.055
	2	1.494	1.491	1.597	1.169	2.913	2.448	1.247	1.272	1.117	1.117
	3	1.329	1.329	1.453	1.069	1.165	1.005	1.000	1.014	1.699	1.479
	4	2.156	1.931	1.653	1.071	1.542	1.244	1.301	1.161	1.042	1.469
	5	1.380	1.096	1.203	1.154	1.331	1.069	1.292	1.244	1.126	1.223
	6	1.293	1.293	1.206	1.090	2.105	2.105	1.001	1.178	1.018	2.703
	7	1.543	1.543	1.387	1.002	1.042	1.083	1.112	1.018	1.021	1.001
	8	1.431	1.431	1.615	1.188	1.166	1.045	1.117	1.019	1.007	1.002
	9	1.654	1.654	1.376	1.287	2.643	2.643	1.110	1.001	1.025	1.039
	10	1.661	1.661	1.593	1.075	1.228	1.062	1.400	1.192	1.093	1.079
$n = 64$	1	1.882	1.052	2.298	1.348	1.283	1.008	1.094	1.002	1.082	1.006
	2	1.736	1.000	1.429	1.007	1.141	1.000	1.377	1.140	1.240	1.183
	3	1.992	1.209	1.446	1.023	1.070	1.015	1.022	1.004	1.091	4.089
	4	1.935	1.258	1.637	1.084	1.601	1.371	1.258	2.660	1.116	1.029
	5	1.635	1.028	1.432	1.046	1.605	1.337	1.280	1.072	1.087	1.746
	6	2.350	1.000	1.481	1.019	1.034	1.076	1.228	1.084	1.052	1.000
	7	1.804	1.000	1.153	1.014	1.143	1.004	1.444	1.141	1.001	1.153
	8	1.884	1.028	1.292	1.007	1.634	1.527	1.024	1.017	1.035	1.007
	9	1.768	1.090	1.356	1.018	1.426	1.018	1.152	1.010	1.000	1.050
	10	2.505	1.342	1.596	1.024	1.086	1.018	1.093	1.023	1.497	1.201
$n = 128$	1	1.447	1.017	2.024	1.274	1.256	1.000	1.539	1.236	1.000	1.121
	2	1.477	1.005	1.320	1.021	1.238	1.002	1.086	1.005	1.335	1.157
	3	1.646	1.001	1.230	1.058	1.536	1.126	1.366	1.049	1.031	1.022
	4	1.865	1.055	1.198	1.030	1.043	1.234	1.031	1.018	1.925	1.728
	5	1.757	1.029	1.291	1.011	1.214	1.007	1.044	1.019	1.004	1.047
	6	1.541	1.005	1.407	1.008	1.103	1.012	1.368	1.073	1.007	1.188
	7	1.243	1.064	1.466	1.044	1.340	1.000	1.127	1.006	1.563	1.284
	8	2.567	1.646	1.221	1.150	1.745	1.224	1.224	1.016	1.029	1.017
	9	1.578	1.020	1.110	1.048	1.262	1.024	1.135	1.008	1.038	1.000
	10	1.802	1.019	1.157	1.024	1.046	1.072	1.062	1.015	1.355	1.300

REFERENCES

ANDERSSEN, R. S. and BLOOMFIELD, P. (1974). A time series approach to numerical differentiation. *Technometrics* **16** 69–75.

BARRY, DANIEL (1983). Nonparametric Bayesian Regression. Thesis, Yale University.

COX, DENNIS, (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal*, 21, 789–813.

COX, DENNIS, (1983a). Approximation of method of regularization estimators, Technical Report No. 723, Statistics Department, University of Wisconsin-Madison.

COX, DENNIS, (1983b). Asymptotics for M type smoothing splines. *Ann. Math. Statist.* **11** 530–551.

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**, 317–403.

CRUMP, J. G. and SEINFELD, J. H. (1982). A new algorithm for inversion of aerosol size distribution data. *Aerosol Sci. Tech.* **1** 15–34.

DAVIES, A. R., IQBAL, M., MALEKNEJAD, K. and REDSHAW, T. C. (1983). A comparison of statistical regularization and fourier extrapolation methods for numerical deconvolution. To appear in Proceedings of the Heidelberg Workshop on Numerical Treatment of Inverse Problems in Differential and Integral Equations (P. Deuflhard and E. Hairer, eds.) Birkhauser, Boston.

ERDAL, AYTUL (1983). Cross validation for ridge regression and principal component analysis. Thesis, Div. of Applied Mathematics, Brown University.

GOLUB, G. H., HEATH, M. and WAHBA, G. (1979). Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics* **21** 215–224.

HALEM, M. and KALNAY, E. (1983). Variational global analysis of satellite temperature soundings. Preprints, Ninth Conference on Aerospace and Aeronautical Meteorology, American Meteorological Society.

KIMELDORF, G. S. and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95.

LI, K. C. (1983). From Stein's unbiassed risk estimates to the method of generalized cross validation, Purdue University Statistics Dept. TR 83–84.

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

MALEKNEJAD, K. (1983). Thesis, Aberystwyth.

MERZ, P. (1980). Determination of adsorption energy distribution by regularization and a characterization of certain adsorption isotherms. *J. Comput. Phys.* **38** 64–85.

MICCHELLI, C. and WAHBA, G. (1981). Design problems for optimal surface interpolation. In *Approximation Theory and Applicat.* (Z. Ziegler, ed.) 329–348, Academic, New York.

NASHED, M. Z. and WAHBA, G. (1974). Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations. *SIAM J. Math. Anal.* **5** 974–987.

NYCHKA, D. (1983). The solution of Abel-type integral equations with an application in stereology. Thesis, University of Wisconsin-Madison.

NYCHKA, D. and COX, D. (1984). Convergence rates for regularized solutions of integral equations from discrete noisy data. Technical Report No. 752, Statistics Department, University of Wisconsin-Madison.

O'HAGAN, A. (1976). On posterior joint and marginal modes. *Biometrika* **63** 329–333.

O'SULLIVAN, F. (1983). The analysis of some penalized likelihood estimation schemes. Technical Report No. 726, Statistics Department, University of Wisconsin-Madison.

O'SULLIVAN, F. and WAHBA, G. (1984). A cross validated Bayesian retrieval algorithm for non-linear remote sensing experiments. Technical Report No. 747, Statistics Department, University of Wisconsin-Madison (to appear in *J. Comput. Phys.*).

PARZEN, E. (1962). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58** 545–554.

RAGOZIN, D. C. (1983). Error bounds for derivative estimates based on spline smoothing of exact or noisy data. *J. Approx. Thy.* **37** 4 335–355.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

RIESZ, F. and B. SZ. NAGY (1955). Functional analysis. *Ungar.* 242–246.

SILVERMAN, B. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 3 795–810.

SPECKMAN, P. (1982). Efficient nonparametric regression with cross validated smoothing splines. Department of Statistics, University of Missouri-Columbia. Manuscript.

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

STEINBERG, D. (1983). Bayesian models for response surfaces of uncertain functional form. Mathematics Research Center, Report No. TSR 2474, University of Wisconsin-Madison.

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 #757, 196–231. Springer, Berlin.

UTRERAS, F. (1980). Sur le choix du parametre d'adjustment dan le lissage par fonctions spline. *Numer. Math.* **34** 15–28.

UTRERAS, F. (1981). Optimal smoothing of noisy data using spline functions. *SIAM J. Sci. Stat. Comput.* **2** 349–362.

UTRERAS, F. (1983). Natural spline functions: their associated eigenvalue problem. *Numer. Math.* **42** 107–117.

WAHBA, G. (1975). Smoothing noisy data by spline functions. *Numer. Math.* **24** 383–393.

WAHBA, G. (1977a). Optimal smoothing of density estimates. In *Classification and Clustering* (J. Van Ryzin, ed.) 423–458. Academic, New York.

WAHBA, G. (1977b). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14** 651–667.

WAHBA, G. (1977c). Discussion to consistent nonparametric regression, by Charles Stone. *Ann. Statist.* **5** 637–640.

WAHBA, G. (1978a). Interpolating surfaces: High order convergence rates and their associated designs, with application to X-ray image reconstruction. Technical Report No. 523, Statistics Department, University of Wisconsin-Madison.

WAHBA, G. (1978b). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Stat. Soc. Ser. B.* **40** 364–372.

WAHBA, G. (1979a). How to smooth curves and surfaces with splines and cross validation. In *Proceedings of the 24th Conference on the Design of Experiments*, U.S. Army Research Office, Report No. 79-2, 167–192.

WAHBA, G. (1979b). Smoothing and ill posed problems. In *Solution Methods for Integral Equations with Applications* (Michael Golberg, ed.) 183–194. Plenum, New York.

WAHBA, G. (1979c). Convergence rates of "thin plate" smoothing splines when the data are noisy. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.) Lecture Notes in Mathematics No. 757, 232–246. Springer, Berlin.

WAHBA, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* **2** 5–16.

WAHBA, G. (1982a). Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data. In *Multivariate Approximation Theory* **2** (W. Schempp and K. Zeller, eds.) 407–429. Birkhauser, Boston.

WAHBA, G. (1982b). Constrained regularization for ill posed linear operator equations, with applications in meteorology and medicine. In *Statistical Decision Theory and Related Topics III* **2** (S. S. Gupta and J. O. Berger, eds.) 383–418, Academic, New York.

WAHBA, G. (1983). Bayesian "confidence intervals" for the cross validated smoothing spline. *J. Roy. Stat. Soc. B* **45** 133–150.

WAHBA, G. (1984). Cross validated spline methods for direct and indirect sensing experiments. In *Statistical Signal Processing* (E. J. Wegman, and J. G. Smith eds). Marcel Dekker, 179–197.

WAHBA, G. and WENDELBERGER, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Rev.* **108** 36–57.

WECKER, W. and ANSLEY, C. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *J. Am. Stat. Assoc.* **78** 81–89.

DEPT. OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WI 53706