Numer. Math. 31, 377–403 (1979)

Numerische Mathematik

Smoothing Noisy Data with Spline Functions

Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation*

Peter Craven[1] and Grace Wahba[2]**

[1] The Computer Laboratory, The University of Liverpool, Liverpool, England

[2] Department of Statistics, University of Wisconsin, Madison, WI 53706, USA

Summary. Smoothing splines are well known to provide nice curves which smooth discrete, noisy data. We obtain a practical, effective method for estimating the optimum amount of smoothing from the data. Derivatives can be estimated from the data by differentiating the resulting (nearly) optimally smoothed spline.

We consider the model $y_i = g(t_i) + \varepsilon_i$, $i = 1, 2, \ldots, n$, $t_i \in [0, 1]$, where $g \in W_2^{(m)} = \{f : f, f', \ldots, f^{(m-1)}$ abs. cont., $f^{(m)} \in \mathcal{L}_2[0, 1]\}$, and the $\{\varepsilon_i\}$ are random errors with $E\varepsilon_i = 0$, $E\varepsilon_i\varepsilon_j = \sigma^2\delta_{ij}$. The error variance σ^2 may be unknown. As an estimate of g we take the solution $g_{n,\lambda}$ to the problem: Find $f \in W_2^{(m)}$ to minimize $\frac{1}{n}\sum_{j=1}^{n}(f(t_j) - y_j)^2 + \lambda\int_0^1 (f^{(m)}(u))^2\,du$. The function $g_{n,\lambda}$ is a smoothing polynomial spline of degree $2m-1$. The parameter λ controls the tradeoff between the "roughness" of the solution, as measured by $\int_0^1 [f^{(m)}(u)]^2\,du$, and the infidelity to the data as measured by $\frac{1}{n}\sum_{j=1}^{n}(f(t_j) - y_j)^2$, and so governs the average square error $R(\lambda; g) = R(\lambda)$ defined by

$$R(\lambda) = \frac{1}{n}\sum_{j=1}^{n}(g_{n,\lambda}(t_j) - g(t_j))^2.$$

We provide an estimate $\hat{\lambda}$, called the generalized cross-validation estimate, for the minimizer of $R(\lambda)$. The estimate $\hat{\lambda}$ is the minimizer of $V(\lambda)$ defined by $V(\lambda) = \frac{1}{n}\|(I - A(\lambda))y\|^2 \Big/ \left[\frac{1}{n}\operatorname{Trace}(I - A(\lambda))\right]^2$, where $y = (y_1, \ldots, y_n)'$ and $A(\lambda)$ is the $n \times n$ matrix satisfying $(g_{n,\lambda}(t_1), \ldots, g_{n,\lambda}(t_n))' = A(\lambda)y$. We prove that there exist a sequence of minimizers $\tilde{\lambda} = \tilde{\lambda}(n)$ of $EV(\lambda)$, such that as the (regular) mesh $\{t_i\}_{i=1}^{n}$

* Research supported in part under U.S. Air Force Grant AF-AFOSR-77-3272 and by the Science Research Council (GB)

** To whom offprint requests should be sent becomes finer, $\lim_{n\to\infty} ER(\tilde{\lambda})/\min_{\lambda} ER(\lambda)\downarrow 1$. A Monte Carlo experiment with several smooth g's was tried with $m=2$, $n=50$ and several values of σ^2, and typical values of $R(\hat{\lambda})/\min_{\lambda} R(\lambda)$ were found to be in the range 1.01–1.4. The derivative g' of g can be estimated by $g'_{n,\hat{\lambda}}(t)$. In the Monte Carlo examples tried, the minimizer of $R_D(\lambda)=\frac{1}{n}\sum_{j=1}^{n}(g'_{n,\lambda}(t_j)-g'(t_j))$ tended to be close to the minimizer of $R(\lambda)$, so that $\hat{\lambda}$ was also a good value of the smoothing parameter for estimating the derivative.

Subject Classifications. MOS: 65D10; CR: 5.17; MOS: 65D25.

1. Introduction

We consider the model

$$y(t)=g(t)+\varepsilon(t), \qquad t\in[0,1] \tag{1.1}$$

where $g(t)$ is a "smooth" curve, and $\varepsilon(t)$ is a white noise process, $E\,\varepsilon(t)=0$, $E\,\varepsilon(s)\,\varepsilon(t)=\sigma^2$, $s=t$, $=0$, otherwise (E is mathematical expectation). $y(t)$ is observed for $t=t_1, t_2, \ldots, t_n$, $0\leqq t_1<t_2<\cdots<t_n\leqq 1$. It is desired to reconstruct g from the data $y(t_j)\equiv y_j$, $j=1,2,\ldots,n$. We assume that $g\in W_2^{(m)}$, where

$$W_2^{(m)}=\{g: g^{(\nu)} \text{ abs. cont.}, \ \nu=0,1,\ldots,m-1, \ g^{(m)}\in\mathscr{L}_2[0,1]\}.$$

Our estimate of g is $g_{n,\lambda}$, where $g_{n,\lambda}$ is the solution to the problem: Find $f\in W_2^{(m)}$ to minimize

$$\frac{1}{n}\sum_{j=1}^{n}(f(t_j)-y_j)^2+\lambda\int_0^1(f^{(m)}(u))^2\,du. \tag{1.2}$$

The function $g_{n,\lambda}$ is well known to be a polynomial smoothing spline of degree $2m-1$. See Reinsch [9, 10], Schoenberg [11], Wahba [13] for properties of smoothing splines. A Bayesian argument that the use of smoothing splines is appropriate when a certain prior distribution is attached to the $\{g(t_j)\}_{i=1}^{n}$ may be obtained from the discussion in Kimeldorf and Wahba [7], see also [18].

The parameter λ, which must be chosen, controls the tradeoff between the "roughness" of the solution, as measured by

$$\int_0^1[f^{(m)}(u)]^2\,du$$

and the infidelity to the data as measured by

$$\frac{1}{n}\sum_{j=1}^{n}(f(t_j)-y_j)^2. \tag{1.3}$$

The problem is to obtain a good value of λ. Reinsch [9] suggests, roughly, that if σ^2

is known, then λ should be chosen so that the infidelity satisfies

$$\frac{1}{n}\sum_{j=1}^{n}(g_{n,\lambda}(t_j)-y_j)^2=\sigma^2. \tag{1.4}$$

Wahba [13] obtains theoretical results for the optimum choice of λ in the equally spaced data case when certain further smoothness and periodicity conditions are imposed. The optimum λ is defined as the λ which minimizes the true mean square error averaged over the data points. This true mean square error is defined as $R(\lambda)$, given by

$$R(\lambda)=\frac{1}{n}\sum_{j=1}^{n}(g_{n,\lambda}(t_j)-g(t_j))^2. \tag{1.5}$$

The results in [13] show that λ should be chosen so that the infidelity defined by the left-hand side of (1.4), is actually slightly less than σ^2. However, this result is not practical, in that how much less depends on n as well as on the unknown g and on σ^2, which may also be unknown.

If σ^2 is known, then a good value of λ may be obtained from the data as follows: Define $A(\lambda)$ as the $n\times n$ matrix depending on $\{t_i\}_{i=1}^n$ and λ satisfying

$$\begin{pmatrix} g_{n,\lambda}(t_1)\\ \vdots \\ g_{n,\lambda}(t_n)\end{pmatrix}=A(\lambda)\begin{pmatrix} y_1\\ \vdots \\ y_n\end{pmatrix}.$$

Since $g_{n,\lambda}(t)$ is a linear function of $y_1, y_2, \ldots, y_n$ for each t, such $A(\lambda)$ exists. Then

$$ER(\lambda)=E\frac{1}{n}\|A(\lambda)\,y-g\|^2 \tag{1.6}$$

where $y=(y_1,\ldots,y_n)'$, $g=(g(t_1),\ldots,g(t_n))'$, and "$t$" is transpose. The norm is the Euclidean norm. It follows from elementary calculations on (1.6) using the assumed mean and covariance properties of $\varepsilon=(\varepsilon(t_1),\ldots,\varepsilon(t_n))'$, that

$$ER(\lambda)=\frac{1}{n}\|(I-A(\lambda))\,g\|^2+\frac{\sigma^2}{n}\operatorname{Trace}A^2(\lambda). \tag{1.7}$$

It is then trivial to demonstrate the following

Theorem 1.1. An unbiassed estimate of $ER(\lambda)$ is given by $\hat{R}(\lambda)$ defined by

$$\hat{R}(\lambda)=\frac{1}{n}\|(I-A(\lambda))\,y\|^2-\frac{\sigma^2}{n}\operatorname{Tr}(I-A(\lambda))^2+\frac{\sigma^2}{n}\operatorname{Tr}A^2(\lambda), \tag{1.8}$$

that is

$$E\hat{R}(\lambda)=ER(\lambda).$$

Therefore, the minimizer of $\hat{R}(\lambda)$ can be taken as a good choice of λ. An estimate of this type has been proposed by Mallows [8] in the context of ridge regression, see also Hudson [6].

The main result of this paper is to obtain a good estimate of the minimizer of $ER(\lambda)$ from the data *which does not require knowledge of* σ^2. This estimate, to be called the generalized cross-validation (GCV) estimate, takes as the estimate of λ, the minimizer of $V(\lambda)$ defined by

$$V(\lambda)=\frac{1}{n}\|(I-A(\lambda))\,y\|^2\Big/\left[\frac{1}{n}\operatorname{Tr}(I-A(\lambda))\right]^2. \tag{1.9}$$

We will demonstrate, under general conditions (to be given) on g, and on the mesh sequence $\{t_i\}_{i=1}^n \equiv \{t_{in}\}_{i=1}^n$, $n=1,2,\ldots$, that, for large n, $EV(\lambda)-\sigma^2 \approx ER(\lambda)$ for λ in the neighborhood of the minimizer of $ER(\lambda)$.

As a consequence of this, we have the following:

Theorem 4.3. For $g\in W_2^{(m)}$ and mild conditions on the mesh sequence $\{t_{in}\}_{i=1}^n$, there exists a sequence $\tilde{\lambda}=\tilde{\lambda}(n)$ of minimizers of $EV(\lambda)$ with the property that

$$\lim_{n\to\infty}\frac{ER(\tilde{\lambda})}{\min_{\lambda} ER(\lambda)}=1. \tag{1.10}$$

This theorem says that the expected mean square error using $\tilde{\lambda}$ tends to the minimum possible expected mean square error, as $n\to\infty$.

We now describe the origin of the GCV estimate. The intuitive idea of cross-validation is quite simple and goes as follows: Let $g_{n,\lambda}^{[k]}$ be the smoothing spline using all the data points, except the kth. We take the ability of $g_{n,\lambda}^{[k]}$ to predict the missing data point y_k, as a measure of the goodness of λ. Formally, let $g_{n,\lambda}^{[k]}$ be the function $f\in W_2^{(m)}$ which minimizes

$$\frac{1}{n}\sum_{\substack{j=1\\ j\neq k}}^{n}(f(t_j)-y_j)^2+\lambda\int_0^1 (f^{(m)}(u))^2\,du,$$

and let

$$V_0(\lambda)=\frac{1}{n}\sum_{k=1}^{n}(g_{n,\lambda}^{[k]}(t_k)-y_k)^2. \tag{1.11}$$

The (ordinary) cross-validation estimate of λ is defined to be the minimizer of $V_0(\lambda)$. The equally spaced data points case was considered in Wahba and Wold [15, 16], where the (ordinary) cross-validation estimate of λ was introduced. Fairly extensive Monte Carlo experiments [15] showed that the minimizer of $V_0(\lambda)$ was an amazingly good estimate of the minimizer of $R(\lambda)$ over a variety of g and σ^2 tried. Theoretical results related to the optimality of the minimizer of $V_0(\lambda)$ were also obtained for a special case equivalent to constraining g and $g_{n,\lambda}$ to be periodic and requiring $t_j=j/n$, $j=1,2,\ldots,n$. We shall call this the symmetric case.

Note that in the symmetric case all data points are treated symmetrically. That is, the prediction error at t_k is weighted the same as at any other t_j. In the general case, we let

$$V(\lambda)=\frac{1}{n}\sum_{k=1}^{n}(g_{n,\lambda}^{[k]}(t_k)-y_k)^2\,w_k(\lambda), \tag{1.13}$$

where the weights $w_k(\lambda)$ are to compensate for nonequally spaced data points and the possible nonperiodicity of g. If

$$w_k(\lambda) = \left[(1 - a_{kk}(\lambda)) \Big/ \frac{1}{n} \operatorname{Tr}(I - A(\lambda))\right]^2, \tag{1.14}$$

$k = 1, 2, \ldots, n$, where the $\{a_{kk}(\lambda)\}$ are the diagonal elements of $A(\lambda)$, then $V(\lambda)$ of (1.13) becomes $V(\lambda)$ of (1.9), and then (1.10) holds.

That is, we have obtained $\{w_k\}$ so that (1.10) holds. A different intuitive argument for the choice of $V(\lambda)$ as in (1.9) is given in [17], and involves finding a rotation of Euclidean n-space which transforms the general problem into one equivalent to the symmetric problem and then doing ordinary cross validation. This point will be discussed further in Sect. 3.

In the process of proving (1.10) we have obtained a basis for the smoothing spline $g_{n,\lambda}$ in terms of n periodic functions which are piecewise shifted Bernoulli polynomials with one knot, plus $m+1$ polynomials of degree $\leq m$. (See Golomb [3] for earlier results on periodic splines.) An interesting fact about the n piecewise shifted Bernoulli polynomials is that, in the equally spaced data case their $n \times n$ Gram matrix is a circulant matrix. This representation will illuminate the remark that the smoothing spline for unequally spaced sampled non-periodic data is the natural generalization of the output of a low pass filter with the data as input.

In Sect. 2, we obtain the aforementioned representation of $g_{n,\lambda}$ in terms of polynomials plus periodic piecewise shifted Bernoulli polynomials, and we obtain the explicit formula for $A(\lambda)$ that will be used in the proof of (1.10). In Sect. 3 we obtain a simplified form of $V_0(\lambda)$ of (1.11) and show that $V(\lambda)$ of (1.13) with the weights $\{w_k(\lambda)\}$ given by (1.14) is equal to $V(\lambda)$ of (1.9). In Sect. 4, we prove the main theorem, namely (1.10). In Sect. 5, we present some Monte Carlo examples illustrating the effectiveness of the method. Data according to the model (1.1) was generated with several smooth g's and range of values of σ^2. Typical values of $R(\hat{\lambda})/\min_{\lambda} R(\lambda)$ are to be found in the range 1.01–1.4 where $\hat{\lambda}$ is the minimizer of $V(\lambda)$. The minimizer of $\hat{R}(\lambda)$ of (1.8) and the value of λ satisfying (1.4) were also computed. The use of the minimizer of $V(\lambda)$ was found to be roughly about as good as the minimizer of $\hat{R}(\lambda)$, while the use of (1.4) gave estimates of λ that were consistently too large.

We note that the method of generalized cross-validation is also applicable to choosing the regularization parameter in the method of regularization for solving Fredholm integral equations of the first kind, see [14].

2. Bernoulli Polynomials and Smoothing Splines

Let $B_r(t)$, $r = 0, 1, \ldots$ be the Bernoulli polynomials on $t \in [0, 1]$. The $\{B_r\}$ are defined by letting $B_0(t) \equiv 1$, $\frac{1}{(r+1)} \frac{d}{dt} B_{r+1}(t) = B_r(t)$, and choosing the constant of integration so that $\int_0^1 B_r(u)\, du = 0$, $r = 1, 2, \ldots$. Letting $[x]$ be the fractional part of x, we define

$$k_r(t) = B_r([t])/r!.$$

Let $L_k, k=0, 1, \ldots$ be the linear functionals

$$L_0 f = \int_0^1 f(u)\,du$$

$$L_k f = f^{(k-1)}(1) - f^{(k-1)}(0) \equiv \int_0^1 f^{(k)}(u)\,du, \qquad k=1,2,\ldots$$

Then

$$\begin{aligned} L_k(k_r) &= 1, \qquad k=r \\ &= 0, \qquad k \neq r,\ k, r = 0, 1, 2, \ldots . \end{aligned} \tag{2.1}$$

Define the "Bernoulli kernel" $k_r(s, t)$ by

$$k_r(s,t) = -\sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} \frac{1}{(2\pi i \nu)^r} e^{2\pi i \nu (s-t)}, \qquad r=1,2,\ldots \tag{2.2}$$

It is known (see Abramowitz and Stegun [1], p. 805), that

$$k_r(s,t) = \frac{1}{r!} B_r([s-t]) = k_r([s-t]), \tag{2.3}$$

and it can be verified from the definition of $k_r(s, t)$ that

$$\begin{aligned} &\frac{\partial^p}{\partial s^p} k_r(s,t) = k_{r-p}(s,t) \\ & \qquad\qquad\qquad\qquad\qquad\qquad p=1,2,\ldots,r-2 \\ &\frac{\partial^p}{\partial t^p} k_r(s,t) = (-)^p\, k_{r-p}(s,t), \qquad s,t \in [0,1] \end{aligned} \tag{2.4}$$

$$\begin{aligned} &\frac{\partial^{r-1}}{\partial s^{r-1}} k_r(s,t) = k_1(s,t) \\ & \qquad\qquad\qquad\qquad\qquad\qquad s,t \in [0,1],\ s \neq t. \\ &\frac{\partial^{r-1}}{\partial t^{r-1}} k_r(s,t) = (-1)^{r-1} k_1(s,t) \end{aligned} \tag{2.5}$$

$$\int_0^1 \frac{\partial^m}{\partial s^m} k_{2m}(s,u) \frac{\partial^m}{\partial t^m} k_{2m}(t,u)\,du = (-1)^{m-1} k_{2m}(s,t). \tag{2.6}$$

We are now ready to obtain a representation for $g_{n,\lambda}$ in terms of piecewise Bernoulli polynomials.

Theorem 2.1. The solution $g_{n,\lambda}$, to the problem: Find $f \in W_2^{(m)}$ to minimize

$$\frac{1}{n} \sum_{j=1}^{n} (f(t_j) - y_j)^2 + \lambda \int_0^1 (f^{(m)}(u))^2\,du \tag{2.7}$$

is, for $n \geqq m$, unique, and has the representation

$$g_{n,\lambda}(t) = \sum_{r=0}^{m} \theta_r k_r(t) + (-1)^{m-1} \sum_{j=1}^{n} \alpha_j k_{2m}(t, t_j), \tag{2.8a}$$

where $\theta = (\theta_0, \theta_1, \ldots, \theta_m)'$ and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$ are given by

$$\begin{aligned} \theta &= (T' M^{-1} T + \Delta)^{-1} T' M^{-1} y \\ \alpha &= M^{-1}(y - T\theta) \\ y &= (y_1, y_2, \ldots, y_n)', \end{aligned} \tag{2.8b}$$

T is the $n \times (m+1)$ dimensional matrix with jrth entry

$$T_{jr} = k_r(t_j), \qquad r = 0, 1, \ldots, m \qquad j = 1, 2, \ldots, n, \tag{2.8c}$$

Δ is the $(m+1) \times (m+1)$ dimensional matrix of all zeroes except 1 in the $(m+1)$, $(m+1)$ position, M is given by

$$M = K + n\lambda I$$

where K is the $n \times n$ matrix with jkth entry K_{jk},

$$K_{jk} = (-1)^{m-1} k_{2m}(t_j, t_k) \tag{2.8d}$$

and I is the $n \times n$ identity matrix. The matrix $A(\lambda)$ is given by

$$A(\lambda) = KM^{-1}[I - T(T' M^{-1} T + \Delta)^{-1} T' M^{-1}] + T(T' M^{-1} T + \Delta)^{-1} T' M^{-1}. \tag{2.9}$$

Proof. The expression for $A(\lambda)$ follows immediately from (2.8). We first show that $g_{n,\lambda} \in \text{span}\{\{k_r(\cdot)\}_{r=0}^{m} \cup \{k_{2m}(\cdot, t_j)\}_{j=1}^{n}\}$. This demonstration can be carried out a number of ways using known results on splines. We rely on the arguments in Kimeldorf and Wahba [7]. The reproducing kernel $Q(s, t)$ for $W_2^{(m)}$ endowed with the inner product

$$\langle f, g \rangle = \sum_{r=0}^{m-1} (L_r f)(L_r g) + \int_0^1 f^{(m)}(u)\, g^{(m)}(u)\, du$$

is shown in Lemma 2.1, of the Appendix, to be

$$Q(s, t) = \sum_{r=0}^{m} k_r(s)\, k_r(t) + (-1)^{m-1} k_{2m}(s, t). \tag{2.10}$$

It then follows from the arguments in [7] that $g_{n,\lambda}$ must lie in

$$\mathscr{S} = \text{span}\{\{k_r(\cdot)\}_{r=0}^{m-1} \cup \{Q_{t_j}\}_{j=1}^{n}\} \quad \text{where } Q_{t_j}(\cdot) \equiv Q(\cdot, t_j).$$

However $\mathscr{S}$ is contained in span $\{\{k_r(\cdot)\}_{r=0}^{m} \cup \{k_{2m}(\cdot, t_j)\}_{j=1}^{n}\}$ so that $g_{n,\lambda}$ has the representation (2.8a) for some θ, α. Substituting (2.8a) into (2.7), and using (2.6) gives

$$\sum_{j=1}^{n} (g_{n,\lambda}(t_j)-y_j)^2 + n\lambda \int_0^1 (g_{n,\lambda}^{(m)}(u))^2\, du$$
$$= \sum_{j=1}^{n} \left[\sum_{r=0}^{m} \theta_r k_r(t_j) + (-1)^{m-1} \sum_{k=1}^{n} \alpha_k k_{2m}(t_j, t_k) - y_j\right]^2$$
$$+ n\lambda \left[\sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \alpha_k (-1)^{m-1} k_{2m}(t_j, t_k) + \theta_m^2\right]$$
$$\equiv \|T\theta + K\alpha - y\|^2 + n\lambda(\alpha^t K\alpha + \theta_m^2). \tag{2.11}$$

The vectors θ and α are to be chosen to minimize this expression. By differentiating the right hand side of (2.11) with respect to θ and α and setting the result equal to 0 we obtain the Theorem.

We remark that $(-)^{m-1} k_{2m}(t, t_k)$, considered as a function of t is a monospline of degree $2m$, that is, the sum of the monomial t^{2m} plus a polynomial spline of degree $2m-1$ (with a single knot at t_k). However, it can be checked that $\sum_{j=1}^{n} \alpha_j = 0$, so that $g_{n,\lambda}$ is a polynomial spline of degree $2m-1$, as is well known.

When the knots $\{t_j\}$ are equally spaced, K, and hence M, are circulant matrices. The details are given in the (well-known)

Lemma 2.2. $\left\{(-1)^{m-1} k_{2m}\left(\frac{j}{n}, \frac{k}{n}\right)\right\}_{j,k=1,\ldots,n} = WDW^*$

where "*" denotes complex conjugate transpose and W is the $n \times n$ unitary matrix with rsth entry W_{rs} given by

$$W_{rs} = \frac{1}{\sqrt{n}} e^{2\pi i rs/n},$$

D is the diagonal matrix with $\nu\nu$th entry $D_{\nu\nu}$ given by

$$D_{\nu\nu} = \lambda_{\nu n}^{2m}$$

where

$$\lambda_{\nu n}^{r} = n \sum_{\xi=-\infty}^{\infty} \frac{1}{[2\pi(\nu+\xi n)]^r}, \qquad \nu = 1, 2, \ldots, n-1, \quad (\lambda_{\nu n}^{r} \equiv \lambda_{n-\nu, n}^{r})$$
$$\lambda_{nn}^{r} = n \sum_{\substack{\xi=-\infty \\ \xi \neq 0}}^{\infty} \frac{1}{[2\pi \xi n]^r}. \tag{2.12}$$

Proof.

$$(-1)^{m-1} k_{2m}\left(\frac{j}{n}, \frac{k}{n}\right) = \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} \frac{1}{(2\pi\nu)^{2m}} e^{2\pi i \nu (j-k)/n}$$
$$= \sum_{\nu=1}^{n} \sum_{\substack{\xi=-\infty \\ (\nu,\xi) \neq (n,-1)}}^{\infty} \frac{1}{[2\pi(\nu+\xi n)]^{2m}} e^{2\pi i \nu (j-k+\xi n)/n}$$
$$= \sum_{\nu=1}^{n} \sum_{\substack{\xi=-\infty \\ (\nu,\xi) \neq (n,-1)}}^{\infty} \frac{1}{[2\pi(\nu+\xi n)]^{2m}} e^{2\pi i \nu (j-k)/n}$$

The $\lambda_{\nu n}^{r}$ can be expressed in terms of the polygamma function, see Abramowitz and Stegun [1], Sect. 6.4. However, sufficient computational accuracy will usually be obtained with only a few terms in (2.12).

3. The Generalized Cross-Validation Function $V(\lambda)$

We first obtain a simplified representation for the (ordinary) cross-validation function $V_0(\lambda)$ defined by

$$V_0(\lambda)=\frac{1}{n}\sum_{k=1}^{n}(g_{n,\lambda}^{[k]}(t_k)-y_k)^2.$$

Recall that $g_{n,\lambda}^{[k]}$ is the solution to the problem: Find $f\in W_2^{(m)}$ to minimize

$$\frac{1}{n}\sum_{\substack{j=1\\j\neq k}}^{n}(f(t_j)-y_j)^2+\lambda\int_0^1(f^{(m)}(u))^2\,du.$$

It will be useful to know that if we replace the data point y_k by $g_{n,\lambda}^{[k]}(t_k)$ and solve the original (n-data point) minimization problem (1.2) with the data $y_1, y_2, \ldots, y_{k-1}$, $g_{n,\lambda}^{[k]}(t_k)$, $y_{k+1}, \ldots, y_n$, we get $g_{n,\lambda}^{[k]}$ for the solution. This is the content of

Lemma 3.1. Let $n\geqq m$ and let $g_{n,\lambda}(t;k,z_k)$ be the solution to the problem: Find $f\in W_2^{(m)}$ to minimize

$$\frac{1}{n}\left[(f(t_k)-z_k)^2+\sum_{\substack{j=1\\j\neq k}}^{n}(f(t_j)-y_j)^2\right]+\lambda\int_0^1(f^{(m)}(u))^2\,du.$$

Then

$$g_{n,\lambda}(t;k,g_{n,\lambda}^{[k]}(t_k))=g_{n,\lambda}^{[k]}(t).$$

Proof. Let $h=g_{n,\lambda}^{[k]}$, let $z_k=g_{n,\lambda}^{[k]}(t_k)$ and let f be any element of $W_2^{(m)}$ different from h. Then

$$\begin{aligned}
&\frac{1}{n}\left[\sum_{\substack{j=1\\j\neq k}}^{n}(h(t_j)-y_j)^2+(h(t_k)-z_k)^2\right]+\lambda\int_0^1(h^{(m)}(u))^2\,du\\
&=\frac{1}{n}\left[\sum_{\substack{j=1\\j\neq k}}^{n}(h(t_j)-y_j)^2+\lambda\int_0^1(h^{(m)}(u))^2\,du\right]\\
&<\frac{1}{n}\left[\sum_{\substack{j=1\\j\neq k}}^{n}(f(t_j)-y_j)^2+\lambda\int_0^1(f^{(m)}(u))^2\,du\right]\\
&\leqq\frac{1}{n}\left[\sum_{\substack{j=1\\j\neq k}}^{n}(f(t_j)-y_j)^2+(f(t_k)-z_k)^2\right]+\lambda\int_0^1(f^{(m)}(u))^2\,du.
\end{aligned}$$

Comparing the left and rightmost expressions, we see that h solves the n-data point minimization problem with y_k replaced by z_k.

The results of Lemma 3.1 allow us to prove

Lemma 3.2.

$$g_{n,\lambda}^{[k]}(t_k)-y_k=(g_{n,\lambda}(t_k)-y_k)\Big/\left(1-\frac{\partial}{\partial y_k}g_{n,\lambda}(t_k)\right).$$

Proof. Let $z_k=g_{n,\lambda}^{[k]}(t_k)$. Then Lemma 3.1 and the fact that for each t, $g_{n,\lambda}(t)$ depends linearly on y_k, gives

$$z_k=g_{n,\lambda}(t;k,z_k)=g_{n,\lambda}(t_k;k,y_k)+(z_k-y_k)\frac{\partial g_{n,\lambda}(t_k)}{\partial y_k}.$$

$$=g_{n,\lambda}(t_k)+(z_k-y_k)\frac{\partial g_{n,\lambda}(t_k)}{\partial y_k}$$

and the result follows after some algebraic manipulation.

Denoting the entries of $A(\lambda)$ by a_{jk}, we have

$$g_{n,\lambda}(t_k)\doteq\sum_{j=1}^{n}a_{kj}y_j$$

and so

$$\frac{\partial g_{n,\lambda}(t_k)}{\partial y_k}=a_{kk}$$

and it follows from Lemma 3.2 that

$$V_0(\lambda)=\frac{1}{n}\sum_{k=1}^{n}\left\{\left(\sum_{j=1}^{n}a_{kj}y_j-y_k\right)^2/(1-a_{kk})^2\right\}. \tag{3.1}$$

To motivate the definition of $V(\lambda)$ consider the periodic version of the smoothing problem: it is: Find $g\in W_2^{(m)}$, periodic and with integral 0, to minimize

$$\frac{1}{n}\sum_{j=1}^{n}(f(t_j)-y_j)^2+\lambda\int_0^1(f^{(m)}(u))^2\,du.$$

The function g periodic with integral 0 in this context means

$$L_k g=0,\qquad k=0,1,\ldots,m.$$

It can be shown that the solution $h_{n,\lambda}$ is given by

$$h_{n,\lambda}(t)=\sum_{j=1}^{n}\alpha_j(-1)^{m-1}k_{2m}(t,t_j) \tag{3.2}$$

where

$$\alpha=(K+n\lambda I)^{-1}y\equiv M^{-1}y.$$

Here the role of A is played by KM^{-1}. If $t_j = j/n$, $j = 1, 2, \ldots, n$, then KM^{-1} is circulant for every λ and hence constant down the diagonals, $a_{kk} \equiv \frac{1}{n} \sum_{j=1}^{n} a_{jj} \equiv \frac{1}{n} \operatorname{Trace} A$ and $V_0(\lambda)$ becomes

$$V_0(\lambda) = \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{j=1}^{n} a_{kj} y_j - y_k \right)^2 \Big/ \left(1 - \frac{1}{n} \sum_{k=1}^{n} a_{kk}\right)^2$$
$$= \frac{1}{n} \|(I - A)\, y\|^2 \Big/ \left[\frac{1}{n} \operatorname{Tr}(I - A)\right]^2. \tag{3.3}$$

(This expression is given in Wahba and Wold [16] for the periodic, equally spaced case considered there.)

To obtain generalized cross-validation from "ordinary" cross-validation in general, one rotates the coordinate system so the matrix, call it $\tilde{A}(\lambda)$, which plays the role in the new coordinate system of the prediction matrix $A(\lambda)$, is circulant. Since A is symmetric this can always be done by writing $A(\lambda) = UD^2(\lambda)\,U'$ where D^2 is diagonal and U is orthogonal. Then, letting $\Gamma = WU'$, $\tilde{A}(\lambda) = \Gamma A(\lambda)\,\Gamma'$ is circulant. Let $\tilde{y} = \Gamma y$. Then the "smoothed" $\tilde{y}$ is $\Gamma(g_{n,\lambda}(t_1), \ldots, g_{n,\lambda}(t_n))' = \Gamma A(\lambda)\, y = \tilde{A}(\lambda)\, \tilde{y}$, say. Now do "ordinary" cross-validation on the "data" $\tilde{y}$. The result is $V(\lambda)$.

We remark that inspection of $h_{n,\lambda}$ [Eq. (3.2)] reveals the "low pass filter" character of the smoothing spline in the periodic, equally spaced data case. From Lemma 2.2 and Eq. (3.2) we find that the sample Fourier coefficients $\{h_{n,\lambda,\nu}\}$ of $h_{n,\lambda}$,

$$h_{n,\lambda,\nu} \doteq \frac{1}{n} \sum_{j=1}^{n} h_{n,\lambda}\left(\frac{j}{n}\right) e^{-2\pi i \nu j/n}$$

are related to the Fourier coefficients $\{\hat{h}_\nu\}$ of the data

$$\hat{h}_\nu \doteq \frac{1}{n} \sum_{j=1}^{n} y\left(\frac{j}{n}\right) e^{-2\pi i \nu j/n}$$

by the equations

$$h_{n,\lambda,\nu} = f_\nu \hat{h}_\nu, \qquad \nu = 1, 2, \ldots, n,$$

where

$$f_\nu = \frac{1}{1 + n\lambda/\lambda_{\nu n}^{2m}}.$$

When $\nu \ll n$ we may approximate the summation for $\lambda_{\nu n}^{r}$ in (2.12) by the $\xi = 0$ term and so obtain

$$f_\nu \approx \frac{1}{1 + n\lambda(2\pi\nu)^{2m}} = B\left(\frac{\nu}{\nu_0}\right) B^*\left(\frac{\nu}{\nu_0}\right)$$

where $B\left(\frac{\nu}{\nu_0}\right)$ is the Butterworth filter, well known to electrical engineers, having half power point $\nu_0 = \dfrac{1}{2\pi(n\lambda)^{1/2m}}$.

4. Optimal Properties of the Generalized Cross-Validation Estimate of λ

Recall that the true mean square error is given by

$$R(\lambda)=\frac{1}{n}\sum_{i=1}^{n}(g_{n,\lambda}(t_i)-g(t_i))^2$$
$$=\frac{1}{n}\|A(\lambda)\,y-g\|^2$$

and the cross-validation function $V(\lambda)$ is given by

$$V(\lambda)=\frac{\frac{1}{n}\|(I-A(\lambda))\,y\|^2}{\left(1-\frac{1}{n}\operatorname{Tr}A(\lambda)\right)^2}.$$

The general idea is that one wishes to choose λ to minimize $R(\lambda)$. This cannot be done directly, of course, since $R(\lambda)$ involves the unknown g. If σ^2 is known, then the minimizer of $\hat{R}(\lambda)$ of (1.8) can be used to estimate the λ which minimizes $R(\lambda)$. If σ^2 is not known, we will show that the minimizer of $V(\lambda)$ can be used.

To demonstrate the usefulness of $V(\lambda)$, we must distinguish two cases. If $g(\cdot)\in\pi_{m-1}$, where π_{m-1} are the polynomials of degree $m-1$ or less, we shall first show that $ER(\lambda)$ and $EV(\lambda)$ are both minimized for $\lambda=\infty$. (Recall that $f_{n,\infty}$ is the $m-1^{\text{st}}$ degree polynomial best fitting the data in the least squares sense.) In general, we will show that if $\tilde{\lambda}$ is the minimizer of $EV(\lambda)$, then the inefficiency I^* of the method of generalized cross validation, defined by

$$I^*=\frac{ER(\tilde{\lambda})}{\min_{\lambda} ER(\lambda)} \qquad (4.1)$$

tends to 1 as $n\to\infty$. Thus, the mean square error when λ is estimated by minimizing V should be close to the minimum possible mean square error.

It follows immediately that $I^*=1$ if $g\in\pi_{m-1}$, since $ER(\cdot)$ and $EV(\cdot)$ have the same minimizer. In the general case $g\in W_2^{(m)}$, $g\notin\pi_{m-1}$, it will turn out that $\tilde{\lambda}$ and λ^*, the minimizers of $EV(\lambda)$ and $ER(\lambda)$ respectively, must satisfy $\tilde{\lambda}\to 0$, $\lambda^*\to 0$, $1/n\tilde{\lambda}^{1/2m}\to 0$, $1/n\lambda^{*1/2m}\to 0$. We will proceed to prove that $I^*\downarrow 1$ in several steps. First we show that

$$\left|\frac{ER(\lambda)+\sigma^2-EV(\lambda)}{ER(\lambda)}\right|\leqq h(\lambda) \qquad (4.2)$$

where $h(\lambda)$ is a small quantity to be defined. We will then show that

$$I^*\equiv\frac{R(\tilde{\lambda})}{R(\lambda^*)}\leq\frac{1+h(\lambda^*)}{1-h(\tilde{\lambda})}.$$

Finally we show that $h(\lambda)=O(1/n\lambda^{1/2m})$ and that λ^*, $\tilde{\lambda}$ must satisfy $1/n(\lambda^*)^{1/2m}\to 0$,

$1/n(\tilde{\lambda})^{1/2m} \to 0$, from which it will follow that $(1+h(\lambda^*))/(1-h(\tilde{\lambda}))\downarrow 1$ and hence $I^*\downarrow 1$.

Let

$$b^2(\lambda)=\frac{1}{n}g'(I-A(\lambda))^2 g=\frac{1}{n}\|(I-A(\lambda))g\|^2$$

$$\mu_1(\lambda)=\frac{1}{n}\operatorname{Tr}A(\lambda)$$

$$\mu_2(\lambda)=\frac{1}{n}\operatorname{Tr}A^2(\lambda).$$

Then

$$ER(\lambda)=b^2(\lambda)+\sigma^2\mu_2(\lambda)$$

$$EV(\lambda)=\frac{b^2(\lambda)+\sigma^2(1-2\mu_1(\lambda)+\mu_2(\lambda))}{[1-\mu_1(\lambda)]^2}.$$

We first consider the case $g(\cdot)\in\pi_{m-1}$. In this case $g=(g(t_1),g(t_2),\ldots,g(t_n))'$ is a linear combination of the first m columns of T and so $(I-A(\lambda))g=0$ for all λ and $b(\lambda)\equiv 0$. Thus the minimization of $ER(\lambda)$ reduces to the minimization of $\operatorname{Tr}A^2(\lambda)$, which is clearly minimized for $\lambda=\infty$. Similarly $EV(\lambda)$ becomes

$$EV(\lambda)=(1-2\mu_1(\lambda)+\mu_2(\lambda))/(1-\mu_1(\lambda))^2.$$

Now $I-A(\lambda)$ has m zero eigenvalues, and the remaining $n-m$ eigenvalues can be shown to be of the form $n\lambda(n\lambda+\xi_{vn})^{-1}$, $v=1,2,\ldots,n-m$ where ξ_{1n}, $\xi_{2n},\ldots,\xi_{n-m,n}$ are $n-m$ positive numbers, not all the same (see Sect. 5 for more details), and so the above expression for $EV(\lambda)$ becomes

$$EV(\lambda)=\frac{1}{n}\sum_{v=1}^{n-m}\left(\frac{n\lambda}{n\lambda+\xi_{vn}}\right)^2\Big/\left(\frac{1}{n}\sum_{v=1}^{n-m}\frac{n\lambda}{n\lambda+\xi_{vn}}\right)^2$$

$$=\frac{1}{\left(\frac{n-m}{n}\right)}\frac{\frac{1}{n-m}\sum_{v=1}^{n-m}\left(\frac{n\lambda}{n\lambda+\xi_{vn}}\right)^2}{\left[\frac{1}{n-m}\sum_{v=1}^{n-m}\left(\frac{n\lambda}{n\lambda+\xi_{vn}}\right)\right]^2}\geqq\frac{1}{\left(\frac{n-m}{n}\right)},$$

and the minimum is attained if and only if $\lambda=\infty$.

We now proceed to the general case. We have

Theorem 4.1.

$$\frac{ER(\lambda)+\sigma^2-EV(\lambda)}{ER(\lambda)}=\frac{-\mu_1(2-\mu_1)}{(1-\mu_1)^2}+\frac{\sigma^2}{b^2+\sigma^2\mu_2}\cdot\frac{\mu_1^2}{(1-\mu_1)^2}$$

and so

$$\frac{|ER(\lambda)+\sigma^2-EV(\lambda)|}{ER(\lambda)}<h(\lambda)$$

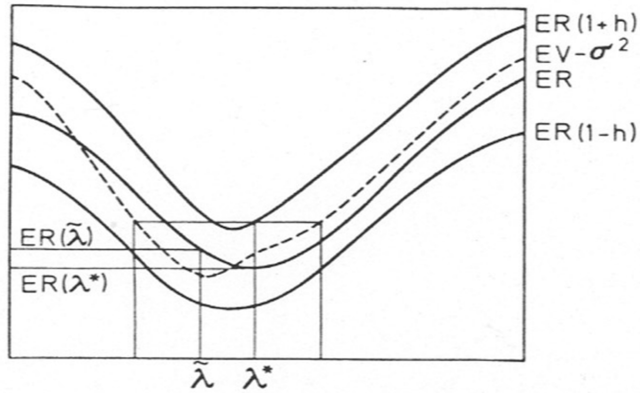

Fig. 1. Graphical suggestion of the proof of Theorem 4.2

where

$$h(\lambda)=\left[2\mu_1(\lambda)+\frac{\mu_1^2(\lambda)}{\mu_2(\lambda)}\right]\frac{1}{(1-\mu_1(\lambda))^2}.$$

Proof of Theorem. The result follows trivially from

$$ER(\lambda)+\sigma^2-EV(\lambda)=ER(\lambda)\left(1-\frac{1}{(1-\mu_1(\lambda))^2}\right)+\sigma^2\frac{\mu_1^2(\lambda)}{(1-\mu_1(\lambda))^2}.$$

From Theorem 4.1 one can deduce

Theorem 4.2. Let λ^* be the minimizer of $ER(\lambda)$. Then $EV(\lambda)$ has a minimum $\tilde{\lambda}$ so that the "expectation inefficiency" I^* defined by

$$I^*=\frac{ER(\tilde{\lambda})}{ER(\lambda^*)}$$

satisfies

$$I^*\leqq\frac{1+h(\lambda^*)}{1-h(\tilde{\lambda})}.$$

Proof. Let $\Lambda=\{\lambda\colon 0\leqq\lambda\leqq\infty,\ EV(\lambda)-\sigma^2\leqq R(\lambda^*)(1+h(\lambda^*))\}$.

Since

$$ER(\lambda)(1-h(\lambda))<EV(\lambda)-\sigma^2<ER(\lambda)(1+h(\lambda)),\qquad 0\leqq\lambda<\infty,$$

and ER, EV, and h are continuous functions of λ, then Λ is a non-empty closed set. If 0 is not a boundary point of Λ, then $EV(\lambda)$ has a minimum in the interior of Λ, (or possibly at ∞) call it $\tilde{\lambda}$ (see Fig. 1). Now by Theorem 4.1

$$ER(\tilde{\lambda})(1-h(\tilde{\lambda}))<EV(\tilde{\lambda})-\sigma^2<ER(\lambda^*)(1+h(\lambda^*))$$

If Λ includes 0, then $\tilde{\lambda}$ may be on the boundary of Λ, i.e., $\tilde{\lambda}=0$, but the above bound on I^* still holds. Our aim now is to prove that $h(\lambda^*)$ and $h(\tilde{\lambda})\to 0$ as $n\to\infty$. We will use several lemmas, whose proofs we relegate to the Appendix.

Lemma 4.1. If $g \in W_2^{(m)}$,

$$b^2(\lambda) \leqq \lambda \int_0^1 (g^{(m)}(u))^2\, du$$

Lemma 4.2. Let $\{t_i\}_{i=1}^n \equiv \{t_{in}\}_{i=1}^n$ satisfy

$$\int_0^{t_{in}} w(u)\, du = i/n, \qquad i=1,2,3,\ldots,n, \; n=1,2,\ldots$$

where $w(u)$ is a continuous strictly positive weight function. Then if $g \notin \pi_{m-1}$, (and not identically 0) and λ is bounded away from 0 as $n \to \infty$, then $b^2(\lambda)$ is also bounded away from 0.

Lemma 4.3. Let $\{t_{in}\}_{i=1}^n$ satisfy the hypothesis of Lemma 4.2 with $0 < \alpha \leqq w(t) \leqq \beta < \infty$. Then

$$\frac{k_m}{\beta^{1/2m}} + o(1) \leqq n\lambda^{1/2m} \mu_1(\lambda) \leqq \frac{k_m}{\alpha^{1/2m}} + o(1)$$

$$\frac{l_m}{\beta^{1/2m}} + o(1) \leqq n\lambda^{1/2m} \mu_2(\lambda) \leqq \frac{l_m}{\alpha^{1/2m}} + o(1),$$

where

$$o(1) = O(\lambda) + O(1/n\lambda^{1/2m}), \qquad \text{as } \lambda \to 0, \; n\lambda^{1/2m} \to \infty$$

and

$$k_m = \int_0^\infty \frac{dx}{(1+x^{2m})}, \qquad l_m = \int_0^\infty \frac{dx}{(1+x^{2m})^2}.$$

Conversely, if $n\lambda^{1/2m}$ is bounded away from 0, then so are $\mu_1(\lambda)$ and $\mu_2(\lambda)$. We remark that it is a consequence of Lemma 4.3 that $\mu_1^2(\lambda)/\mu_2(\lambda) \to 0$.
We conclude from Lemmas 4.1–4.3 that if $g(\cdot) \notin \pi_{m-1}$, then, as $\lambda \to 0$, $n\lambda^{1/2m} \to \infty$,

$$ER(\lambda) = b^2(\lambda) + \sigma^2 \mu_2(\lambda) = O(\lambda) + O(1/n\lambda^{1/2m}) \to 0,$$

and if either λ or $1/n\lambda^{1/2m}$ is bounded away from 0, $ER(\lambda)$ does not tend to 0. Thus, to minimize $ER(\lambda)$, we must have $\lambda^* \to 0$, $n(\lambda^*)^{1/2m} \to \infty$, so that $h(\lambda^*) \to 0$. Now it can be checked that $EV(\lambda) \geqq \sigma^2$. Furthermore $EV(\tilde{\lambda}) \downarrow \sigma^2$ since $EV(\tilde{\lambda}) - \sigma^2 \leqq ER(\lambda^*)(1 + h(\lambda^*)) \to 0$. If $g \notin \pi_{m-1}$, it is necessary that $\tilde{\lambda} \to 0$, $n(\tilde{\lambda})^{1/2m} \to \infty$ in order that $EV(\tilde{\lambda}) \downarrow \sigma^2$, and so it can be concluded that $h(\tilde{\lambda}) \to 0$ as $n \to \infty$.

Combining the above arguments with Theorem 4.2 gives the following main

Theorem 4.3. Let $g(\cdot) \in W_2^{(m)}$, and let $\{t_{in}\}_{i=1}^n$ satisfy $\frac{i}{n} \int_0^{t_{in}} w(u)\, du$, where $w(u)$ is a strictly positive continuous weight function. Then there exist a sequence $\tilde{\lambda} = \tilde{\lambda}(n)$ of minima of $EV(\lambda)$ such that

$$\lim_{n \to \infty} \frac{ER(\tilde{\lambda})}{ER(\lambda^*)} = 1.$$

5. Numerical Results

We have tried the method on artificial data of the form $y(t_i)=g(t_i)+\varepsilon_i$, where ε_i are normally distributed pseudo-random numbers with mean 0 and variance σ^2, and $m=2$. For $m=2$, $g_{n,\lambda}$ is a cubic smoothing spline.

In the $m=2$ case, it can be established from Reinsch [9], p. 179, that

$$I-A(\lambda)=\tilde{Q}(\tilde{Q}^t\tilde{Q}+p\tilde{T})^{-1}\tilde{Q}^t \tag{5.1}$$

where

$$p=1/n\lambda,$$

$\tilde{Q}$ is the $n\times(n-2)$ dimensional tridiagonal matrix with entries $\tilde{q}_{ij}$, $i=1,2,\ldots,n$, $j=1,2,\ldots,n-2$, given by

$$\tilde{q}_{i,i+1}=i/h_{i+1}, \qquad \tilde{q}_{ii}=-1/h_i-1/h_{i+1}, \qquad \tilde{q}_{i+1,i}=1/h_{i+1},$$

where $h_i=t_{i+1}-t_i$, and $\tilde{T}$ is the $(n-2)\times(n-2)$ dimensional tridiagonal matrix with entries $\tilde{t}_{ij}$, $i,j=1,2,\ldots,n-2$ given by

$$\tilde{t}_{ii}=2(h_i+h_{i+1})/3, \qquad \tilde{t}_{i,i+1}=\tilde{t}_{i+1,i}=h_{i+1}/3.$$

The matrix $\tilde{T}$ is strictly positive definite (assuming $h_i>0$). Let $F=\tilde{Q}\tilde{T}^{-1/2}$, where $\tilde{T}^{-1/2}$ is the symmetric square root of $\tilde{T}^{-1}$. When $h_i\equiv\frac{1}{n}$, $i=1,2,\ldots,n$, $\tilde{T}^{-1/2}$ can be found analytically from the formula

$$\begin{pmatrix} \alpha & \beta & & & 0 \\ \beta & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \beta \\ 0 & & & & \beta & \alpha \end{pmatrix} = \Gamma D \Gamma^t \tag{5.2}$$

where

$$\Gamma_{jk}=\sqrt{\frac{2}{n+1}}\sin\frac{jk\pi}{n+1}$$

and D is the diagonal matrix with jjth entry $\alpha+2\beta\cos\frac{j\pi}{n+1}$, thus

$$\tilde{T}^{-1/2}=\Gamma D^{-1/2}\Gamma^t.$$

Then

$$I-A=F(F^tF+pI)^{-1}F^t. \tag{5.3}$$

Let the singular value decomposition of F be (see [5])

$$F = UDV^T$$

where U and V are $n \times (n-2)$ and $(n-2) \times (n-2)$ orthogonal matrices and D has the (non-zero) singular values of F, call them $d_1, d_2, \ldots, d_{n-2}$ on the diagonal and zeroes elsewhere. Then

$$I - A = U \begin{pmatrix} \frac{d_1^2}{d_1^2 + p} & & 0 \\ & \ddots & \\ 0 & & \frac{d_{n-2}^2}{d_{n-2}^2 + p} \end{pmatrix} U^t \tag{5.4}$$

and

$$V(p) = \frac{1}{n} \sum_{j=1}^{n-2} \left(\frac{d_j^2}{d_j^2 + p} \right)^2 z_j^2 \Big/ \left[\frac{1}{n} \sum_{j=1}^{n-2} \left(\frac{d_j^2}{d_j^2 + p} \right) \right]^2 \tag{5.5}$$

where

$$z = (z_1, \ldots, z_{n-2})^t = U^t y$$

The numerical experiments were conducted as follows: To conform to Reinsch's formulae, λ is everywhere replaced by $p = 1/n\lambda$. For given g, σ^2, and n, data y_i, $i = 1, 2, \ldots, n$, were generated by

$$y_i = g\left(\frac{i-1}{n}\right) + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where the ε_i are pseudo-random variates with mean 0 and variance σ^2. $V(p)$ is computed using (5.5), for $\log_{10} p$ in increments of 1/9, and the minimizing p, call it $\hat{p}$, was obtained by global search. Then $g_{n,\hat{\lambda}}$ for $\hat{\lambda} = 1/n\hat{p}$ is computed using Reinsch [9], (Eqs. 8, 9, 13 and 14), and $R(p)$,

$$R(p) = \frac{1}{n} \sum_{i=1}^{n} (g(t_i) - g_{n,\lambda}(t_i))^2$$

is obtained for comparison.

Test functions of the form

$$g(t) = \sum_{j=1}^{r} w_j \beta_{p_j, q_j}(t),$$

where

$$\beta_{pq}(t) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} t^{p-1} (1-t)^{q-1}$$

and Γ is the gamma function were used.

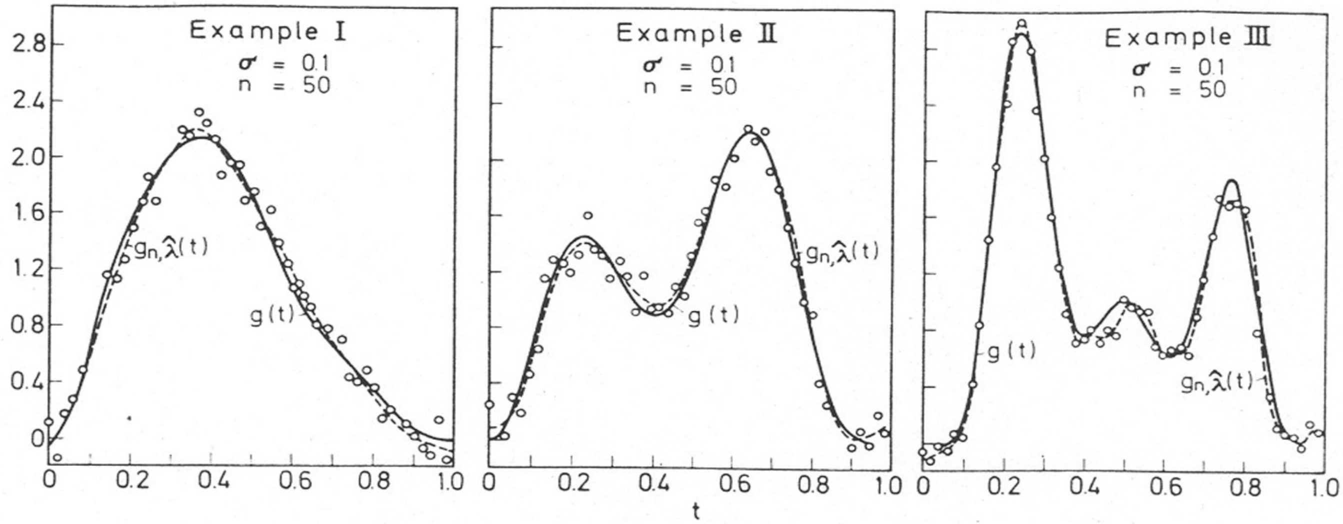

Fig. 2. Examples I, II, and III, g, $g_{n,\hat{\lambda}}$, and the data

The examples are

Example I	$r=3$	$w_1=0.2$	$p_1=4$	$q_1=15$
		$w_2=0.7$	$p_2=5$	$q_2=7$
		$w_3=0.1$	$p_3=12$	$q_3=5$
Example II	$r=2$	$w_1=0.4$	$p_1=12$	$q_1=7$
		$w_2=0.6$	$p_2=4$	$q_2=11$
Example III	$r=3$	$w_1=0.5$	$p_1=10$	$q_1=30$
		$w_2=0.2$	$p_2=20$	$q_2=20$
		$w_3=0.3$	$p_3=30$	$q_3=10$

Figure 2 gives plots of the original function $g(t)$, the data $y_i=g\left(\frac{i}{n}\right)+\varepsilon_i$, $i=1,2,\ldots,n$, and $g_{n,\hat{\lambda}}(t)$, $\hat{\lambda}=1/n\,\hat{p}$, and $\hat{p}$ is the minimizer of $V(p)$. Here $\sigma=0.1$ and the number of data points $n=50$. Figure 3 gives plots of $V(p)$, $\hat{R}(p)$, and $R(p)$.

$\hat{R}(p)$ is defined by (1.8) and is computed by

$$\hat{R}(p)=\frac{1}{n}\sum_{j=1}^{n-2}\left(\frac{d_j^2}{d_j^2+p}\right)^2 z_j^2+\frac{2\sigma^2}{n}\sum_{j=1}^{n-2}\left(\frac{d_j^2}{d_j^2+p}\right)-\sigma^2.$$

The minima of each of these curves is marked with a circle. Reinsch's suggestion for choosing p [Eq. (1.4)] when σ^2 is known, was also implemented. In our notation, his suggestion becomes: Choose p so that $S(p)/\sigma^2=1$. To evaluate this suggestion, $S(p)$ is also plotted, and the point $S(p)=\sigma^2$ is also marked with a circle.

In each example I, II, III it is seen that $\hat{R}(p)$ tracks $R(p)$, and in the neighborhood of the minimum of $R(p)$, $V(p)\approx R(p)+\text{constant}$, where the constant is around σ^2. (Note that $\Sigma\,\varepsilon_i^2/\sigma^2$ is a pseudo random X^2 variate.) It is seen

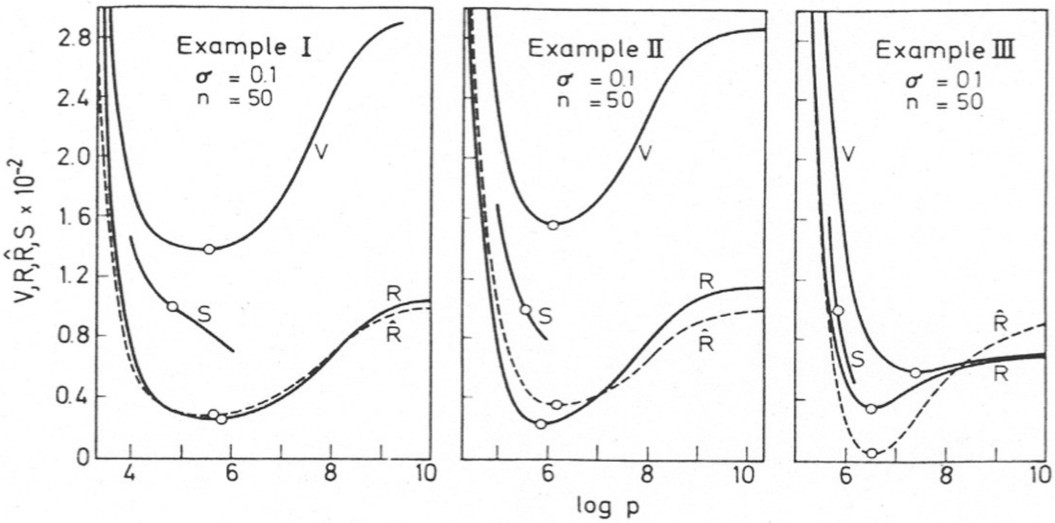

Fig. 3

Table 1. Inefficiencies associated with V, $\hat{R}$ and S

	$\sigma=0.1$			$\sigma=0.01$		
	$\dfrac{R(\hat{p})}{\min_p R(p)}$	$\dfrac{R(\hat{p}_R)}{\min_p R(p)}$	$\dfrac{R(\hat{p}_S)}{\min_p R(p)}$	$\dfrac{R(\hat{p})}{\min_p R(p)}$	$\dfrac{R(\hat{p}_R)}{\min_p R(p)}$	$\dfrac{R(\hat{p}_S)}{\min_p R(p)}$
Example I	1.01	1.00	1.21	1.02	1.06	2.38
Example II	1.04	1.10	1.14	1.01	1.04	1.07
Example III	1.42	1.01	2.02	1.22	1.00	2.06
	$\sigma=0.001$					
Example II	1.12	1.04	1.97			

that the p obtained by setting $S(p)=\sigma^2$ consistently results in p too small, confirming the theoretical results to this effect in [13]. We caution the reader that a good value of σ^2 is required in order that the minimizer of $\hat{R}(p)$ be near that of $R(p)$. In the computations, σ^2 is taken as the variance used to generate pseudo random numbers ε_i.

Letting $\hat{p}$, $\hat{p}_{\hat{R}}$ and $\hat{p}_S$ be the estimates of p using Generalized Cross-Validation, the minimizer of $\hat{R}(p)$, and Reinsch's suggestion respectively, the first three columns of the top of Table 1 gives the observed inefficiencies

$$\frac{R(\hat{p})}{\min_p R(p)}, \quad \frac{R(\hat{p}_R)}{\min_p R(p)} \quad \text{and} \quad \frac{R(\hat{p}_S)}{\min_p R(p)}.$$

These experiments were replicated for $\sigma=0.01$ and $\sigma=0.001$. Plots of V, $\hat{R}$, R and S for the $\sigma=0.01$ case appear in Fig. 4, and the inefficiencies appear in the

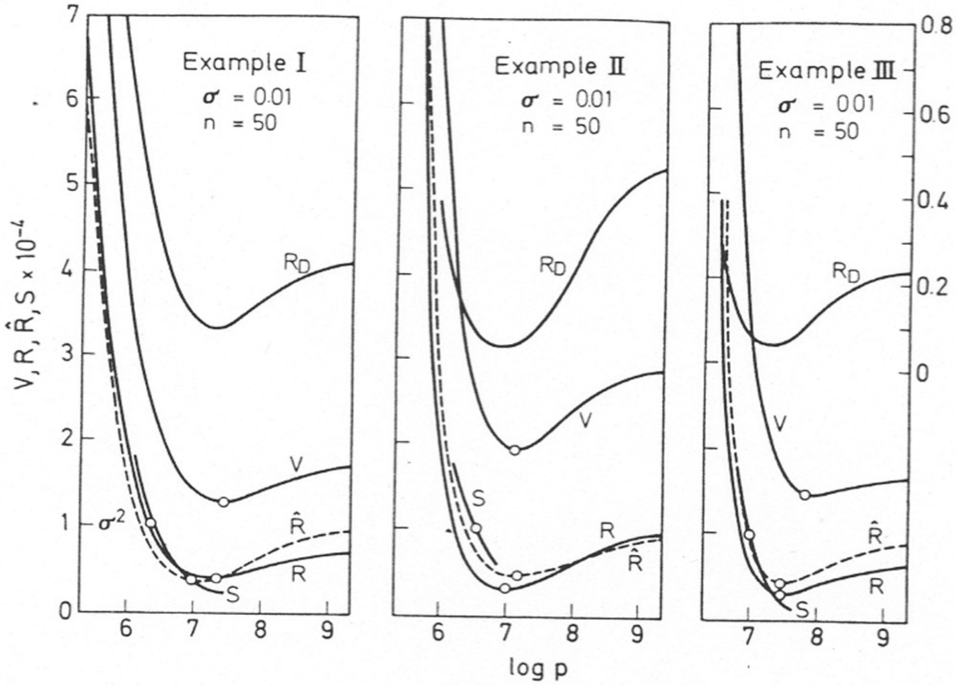

Fig. 4

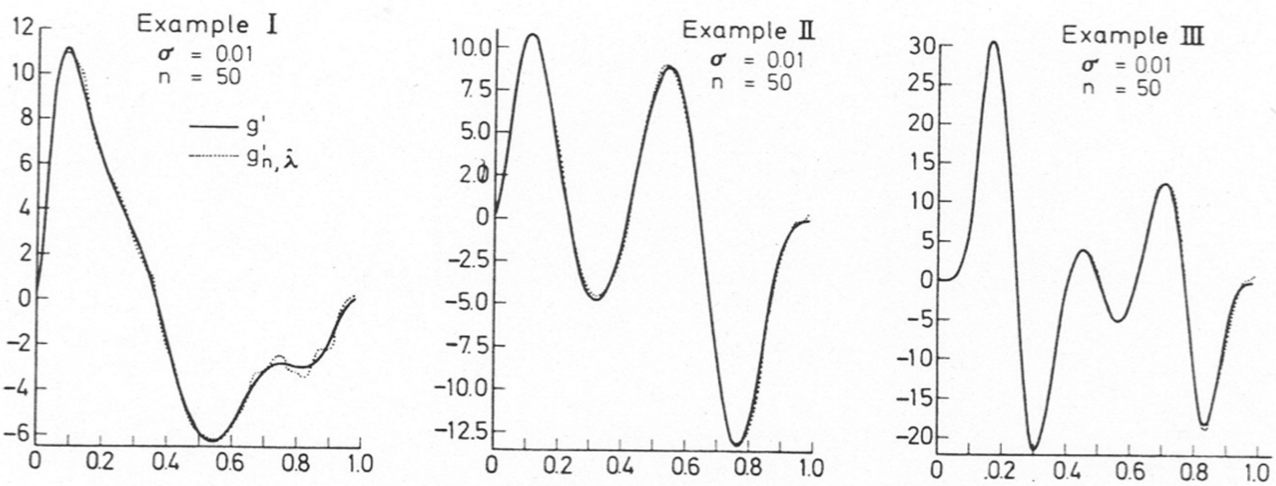

Fig. 5. The derivative g' and its estimate $g'_{n,\hat{\lambda}}$

third through sixth columns of Table 1. The functions g and $g_{n,\hat{\lambda}}$ in the $\sigma=0.01$ case (which is roughly 1 % of the average g) are nearly visually indistinguishable and are not plotted. Good estimates of the derivative of g can be obtained by differentiating $g_{n,\hat{\lambda}}$. The functions g' and $g'_{n,\hat{\lambda}}$ are plotted in Fig. 5, and it can be seen that at this signal to noise ratio the results are impressive. The mean square error $R_D(p)$ in estimating the derivative,

$$R_D(p)=\frac{1}{n}\sum_{j=1}^{n}\left(g'\left(\frac{j}{n}\right)-g'_{n,\lambda}\left(\frac{j}{n}\right)\right)^2$$

is also plotted in Fig. 4. Note that the minimum of $R_D(p)$ is close to the minimum of $R(p)$, so that in these examples both the GCV estimate and the

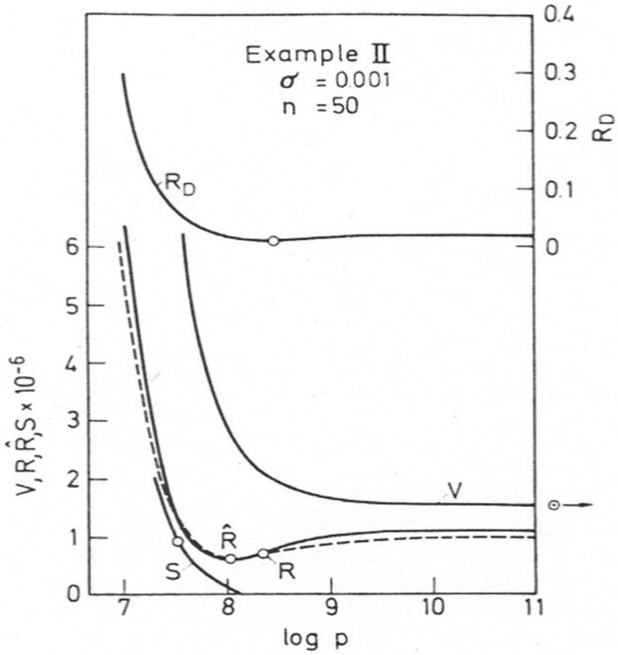

Fig. 6

minimizer of $\hat{R}(p)$ are good from the point of view of minimizing $R_D(p)$. This phenomena also obtained for the noisier data with $\sigma=0.1$, however the best derivative estimate with this 10% noise is fairly crude.

As $\sigma^2 \to 0$ $R(p)$ will flatten out so that the optimum $p \to \infty$, and $\hat{R}(p)$ and $V(p)$ also display this behavior. To illustrate what can happen as $\sigma^2 \to 0$ we present the $V(p)$, $\hat{R}(p)$, $R(p)$, $S(p)$ and $R_D(p)$ curves for Example II, $\sigma=0.001$ in Fig. 6. Note that, while $V(p)$ appears to have its minimum at $p=\infty$, R and $\hat{R}$ have finite minima. Judged from the point of view of inefficiency, however, $\hat{p}$ and $\hat{p}_R$ are not bad. The estimate $\hat{p}_S$ becomes very inefficient in the σ^2 small case (again agreeing with the theoretical results in [13], Eq. (1.3b) there says that as $\sigma^2 \to 0$, Reinsch's suggestion becomes progressively worse).

6. Conclusions

The method of generalized cross validation has been shown both theoretically, and by example, to be an effective method for estimating that value of the spline smoothing parameter which minimizes the mean square error. Excellent estimates of the derivative are also obtained in examples involving roughly 1% and $\frac{1}{10}$ of 1% noise.

Appendix

In this Appendix we give proofs of Lemmas 2.1 and 4.1 through 4.3.

Lemma 2.1. $Q(s,t)$ given by

$$Q(s,t)=\sum_{r=0}^{m} k_r(s)\,k_r(t)+(-1)^{m-1}\,k_{2m}(s,t)$$

is the reproducing kernel for $W_2^{(m)}$ endowed with the inner product

$$\langle f, g\rangle = \sum_{r=0}^{m-1} (L_r f)(L_r g) + \int_0^1 f^{(m)}(u)\, g^{(m)}(u)\, du.$$

Proof. Let $Q_t(\cdot) \equiv Q(t, \cdot)$. We have to show

i) $Q_t \in W_2^{(m)}$ for each t

ii) $\langle Q_t, f\rangle = f(t)$, $f \in W_2^{(m)}$, $t \in [0, 1]$.

(For further properties of reproducing kernels, see Aronszajn, [2], Kimeldorf and Wahba [7].) Part i) is obvious upon noting that Q_t is a monospline of degree $2m$ and hence has $2m-2$ continuous derivatives. To verify ii), we calculate

$$L_r Q_t = k_r(t), \qquad r = 0, 1, \ldots, m-1$$

$$\frac{\partial^m}{\partial u^m} Q_t(u) = k_m(t) + (-1)^{2m-1} k_m(t, u)$$

and so

$$\begin{aligned}\langle Q_t, f\rangle &= \sum_{r=0}^{m-1} k_r(t)(L_r f) + \int_0^1 (k_m(t) - k_m(t, u))\, f^{(m)}(u)\, du \\ &= \sum_{r=0}^{m} k_r(t)(L_r f) - \int_0^1 k_m(t, u)\, f^{(m)}(u)\, du \qquad \text{(A2.1)} \\ &= h(t), \text{ say.}\end{aligned}$$

We wish to show that $h(t) \equiv f(t)$. We are allowed to differentiate (A2.1) $m-1$ times under the integral sign, giving

$$h^{(m)}(t) = (L_m f) - \frac{\partial}{\partial t} \int_0^1 k_1(t, u)\, f^{(m)}(u)\, du.$$

Since $B_1(t) = t - 1/2$, $k_1(t, u) = (t-u) - 1/2$, $u < t$, $= (t-u) + 1/2$, $u > t$, and hence, if t is a point of continuity of $f^{(m)}$,

$$\begin{aligned}h^{(m)}(t) &= (L_m f) - \Bigg[\int_0^1 \frac{\partial}{\partial t} k_1(t, u)\, f^{(m)}(u)\, du + k_1(t, t_-)\, f^{(m)}(t) \\ &\quad + \int_0^1 \frac{\partial}{\partial t} k_1(t, u)\, f^{(m)}(u)\, du - k_1(t, t_+)\, f^{(m)}(t)\Bigg] \\ &= (L_m f) - \int_0^1 f^{(m)}(u)\, du + f^{(m)}(t) \\ &= f^{(m)}(t).\end{aligned}$$

It is easy to see that $L_r(f-h) = 0$ for $r = 0, 1, \ldots, m-1$ so that $f = h$. (This lemma corrects an error in [13], p. 391, line 2.)

Lemma 4.1. For any $g \in W_2^{(m)}$,

$$b^2(\lambda) \equiv \frac{1}{n} \|(I - A(\lambda))\, g\|^2 \leqq \lambda \int_0^1 (g^{(m)}(u))^2\, du.$$

Proof. Note that $A(\lambda)\, g$ is the vector of values of the function, call it $g^*_{n,\lambda}$, which is the solution to the problem; Find $f \in W_2^{(m)}$ to minimize

$$\frac{1}{n} \sum_{j=1}^{n} (g(t_j) - f(t_j))^2 + \lambda \int_0^1 (f^{(m)}(u))^2\, du.$$

Therefore,

$$\begin{aligned} &\frac{1}{n} \sum_{j=1}^{n} (g(t_j) - g^*_{n,\lambda}(t_j))^2 + \lambda \int_0^1 (g^{*(m)}_{n,\lambda}(u))^2\, du \\ &\quad \equiv \frac{1}{n} \|(I - A(\lambda))\, g\|^2 + \lambda \int_0^1 (g^{*(m)}_{n,\lambda}(u))^2\, du \\ &\quad \leqq \frac{1}{n} \sum_{j=1}^{n} (g(t_j) - g(t_j))^2 + \lambda \int_0^1 (g^{(m)}(u))^2\, du = \lambda \int_0^1 (g^{(m)}(u))^2\, du. \end{aligned}$$

Lemma 4.2. Let $\{t_{in}\}_{i=1}^{n}$ satisfy

$$\int_0^{t_{in}} w(u)\, du = i/n, \qquad i = 1, 2, \ldots, n, \;\; n = 1, 2, \ldots$$

where $w(u)$ is a strictly positive continuous weight function. Then if $g(\cdot) \notin \pi_{m-1}$ (and not identically 0), and λ is bounded away from 0 as $n \to \infty$ then $\frac{1}{n} g'(I - A(\lambda))^2 g$ is also bounded away from 0 as $n \to \infty$.

Proof. Let $g^*_{n,\lambda}$ be as in the proof of Lemma 4.1. Then $g^*_{n,\lambda}$ converges in $W_2^{(m)}$, as $n \to \infty$, to the minimizer, call it g^*_λ, of

$$J_{\infty,g}(f) = \int_0^1 \frac{(g(u) - f(u))^2}{\omega(u)}\, du + \lambda \int_0^1 (f^{(m)}(u))^2\, du.$$

Now if $g \in \pi_{m-1}$, it is easy to see that $g^*_\lambda = g$, since, in that case $J_{\infty,g}(g) = 0$. However, if $g \notin \pi_{m-1}$ then

$$J_{\infty,g}(\theta g) < J_{\infty,g}(g)$$

for

$$\theta = \frac{\displaystyle\int_0^1 \frac{g^2(u)}{\omega(u)}\, du}{\displaystyle\int_0^1 \frac{g^2(u)}{\omega(u)}\, du + \lambda \int_0^1 g^{(m)}(u)\, du}$$

so that g^*_λ is not equal to g. Furthermore

$$\frac{1}{n} g'(I-A(\lambda))^2 g = \frac{1}{n} \sum_{j=1}^{n} (g(t_i) - g^*_{n,\lambda}(t_i))^2 \to \int_0^1 \frac{(g(u)-g^*_\lambda(u))^2}{\omega(u)} > 0 \quad \text{for } \lambda > 0$$

Lemma 4.3. Let $\{t_{in}\}_{i=1}^n$ satisfy the hypothesis of Lemma 4.2 with $0<\alpha \leqq w(t) \leqq \beta < \infty$. Then

$$\frac{k_m}{\beta^{1/2m}} + O(\lambda) + O(1/n\lambda^{1/2m}) \leqq n\lambda^{1/2m}\left(\frac{1}{n}\operatorname{Tr} A(\lambda)\right) \equiv n\lambda^{1/2m}\mu_1(\lambda)$$

$$\leqq \frac{k_m}{\alpha^{1/2m}} + O(\lambda) + O(1/n\lambda^{1/2m})$$

$$\frac{l_m}{\beta^{1/2m}} + O(\lambda) + O(1/n\lambda^{1/2m}) \leqq n\lambda^{1/2m}\left(\frac{1}{n}\operatorname{Tr} A^2(\lambda)\right) \equiv n\lambda^{1/2m}\mu_2(\lambda)$$

$$\leqq \frac{l_m}{\alpha^{1/2m}} + O(\lambda) + O(1/n\lambda^{1/2m})$$

as $\lambda \to 0$, $n\lambda^{1/2m} \to \infty$

where

$$k_m = \frac{1}{\pi}\int_0^\infty \frac{dx}{(1+x^{2m})}, \qquad l_m = \frac{1}{\pi}\int_0^\infty \frac{dx}{(1+x^{2m})^2}.$$

Conversely, if $n\lambda^{1/2m}$ is bounded away from 0 as $n \to \infty$, then so are $\frac{1}{n}\operatorname{Tr} A(\lambda)$ and $\frac{1}{n}\operatorname{Tr} A^2(\lambda)$.

Proof.

$$A(\lambda) = (n\lambda P + K) M^{-1},$$

where

$$M = K + n\lambda I$$

and

$$P = M^{-1} T (T' M^{-1} T + \Delta)^{-1} T'.$$

Let

$$n\lambda P M^{-1} = E$$
$$K M^{-1} = A_0.$$

Now, since $0 \leqslant A_0 \leqslant A \equiv A_0 + E \leqslant I$, (where $B \leqslant C$ means $C - B$ is non-negative definite) and $0 \leqslant E \leqslant I$, with E of rank $m+1$, we have

$$\operatorname{Tr} A_0 \leqq \operatorname{Tr} A \leqq \operatorname{Tr} A_0 + (m+1)$$

$$\operatorname{Tr} A_0^2 \leqq \operatorname{Tr} A^2 = \operatorname{Tr} A_0^2 + 2\operatorname{Tr} A_0 E + \operatorname{Tr} E^2$$

$$\leqq \operatorname{Tr} A_0^2 + 3\operatorname{Tr} E$$

$$\leqq \operatorname{Tr} A_0^2 + 3(m+1)$$

and so

$$\lambda^{1/2} \sum_{v=1}^{n} \left(\frac{\lambda_{vn}}{\lambda_{vn}+n\lambda}\right) \leqq n\lambda^{1/2m} \left[\frac{1}{n}\operatorname{Tr} A(\lambda)\right] \leqq \lambda^{1/2m} \sum_{v=1}^{n} \left(\frac{\lambda_{vn}}{\lambda_{vn}+n\lambda}\right) + \lambda^{1/2m}(m+1),$$

where λ_{vn}, $v=1,2,\ldots,n$, are the eigenvalues of K. We continue the proof under the assumption that the eigenvalues λ_{vn} satisfy

$$\alpha \frac{(\pi v)^{2m}}{n} \leqq \lambda_{vn}^{-1} \leqq \beta \frac{(\pi v)^{2m}}{n} \tag{A 4.3.1}$$

for some α, β, $0<\alpha\leqq\beta<\infty$. Then we give an outline of an argument to show that the hypothesis of the theorem on $\{t_{in}\}$ guarantees that (A 4.3.1) holds with α, β given by

$$\alpha = \min_t w(t)(1+o(1))$$
$$\beta = \max_t w(t)(1+o(1)) \tag{A 4.3.2}$$

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

Using (A 4.3.1) gives

$$\lambda^{1/2m} \sum_{v=1}^{n} \frac{1}{(1+\lambda\beta(\pi v)^{2m})} \leqq \lambda^{1/2m} \sum_{v=1}^{n} \left(\frac{1}{1+n\lambda\lambda_{vn}^{-1}}\right)^2 \leqq \lambda^{1/2m} \sum_{v=1}^{n} \frac{1}{(1+\lambda\alpha(\pi v)^{2m})}$$

Since, for any fixed $\gamma>0$ we have

$$\int_{(\gamma\lambda)^{1/2m}\pi}^{(n-1)(\gamma\lambda)^{1/2m}\pi} \frac{dx}{(1+x^{2m})^2} \leqq (\gamma\lambda)^{1/2m}\pi \sum_{v=1}^{n} \frac{1}{(1+\lambda\gamma(\pi)^{2m}v^{2m})^2} \leqq \int_0^\infty \frac{dx}{(1+x^{2m})^2},$$

we obtain

$$\frac{1}{\beta^{1/2m}\pi} \int_{(\beta\lambda)^{1/2m}\pi}^{(n-1)(\beta\lambda)^{1/2m}\pi} \frac{dx}{(1+x^{2m})} \leqq n\lambda^{1/2m}\left[\frac{1}{n}\operatorname{Tr} A(\lambda)\right]$$
$$\leqq \frac{1}{\alpha^{1/2m}\pi} \int_0^\infty \frac{dx}{(1+x^{2m})^2} + \lambda^{1/2m}(m+1),$$

and so

$$\frac{k_m}{\beta^{1/2m}} + O(\lambda) + O(1/n\lambda^{1/2m}) \leqq n\lambda^{1/2m}\left[\frac{1}{n}\operatorname{Tr} A(\lambda)\right]$$
$$\leqq \frac{k_m}{\alpha^{1/2m}} + O(\lambda) + O(1/n\lambda^{1/2m}).$$

A similar argument gives the inequality involving $\operatorname{Tr} A^2(\lambda)$. We now give a heuristic argument to show that A.4.3.1 holds with α, β given by A.4.3.2. The jkth entry K_{jk} of K is given (n even) by

$$K_{jk}=(-1)^{m-1}\,k_{2m}(t_j,t_k)=\sum_{\substack{\nu=-\infty\\ \nu\neq 0}}^{\infty}\frac{1}{(2\pi\nu)^{2m}}\,e^{2\pi i\nu(t_j-t_k)}$$
$$\simeq\sum_{\substack{\nu=-n/2\\ \nu\neq 0}}^{n/2}\frac{1}{(2\pi\nu)^{2m}}\,e^{2\pi i\nu(t_j-t_k)},$$

and so

$$K\approx\Phi D\Phi^*$$

where Φ is the $n\times n$ matrix with $j\nu$th entry $\frac{1}{\sqrt{n}}e^{2\pi i\nu t_{jn}}$, and D is a diagonal matrix with $\nu\nu^{\text{th}}$ entry $D_{\nu\nu}\approx\frac{n}{(2\pi\nu)^{2m}}$, $\nu=-n/2,\dots,n/2$, $\nu\neq 0$, (n even). Since $t_{j+1,n}-t_{jn}=\frac{1}{nw(t_*)}$, for some $t_*\in[t_{jn},t_{j+1,n}]$ we have

$$\frac{1}{n}\sum_{j=1}^{n}e^{2\pi i\nu t_{jn}}e^{-2\pi i\mu t_{jn}}\frac{1}{w(t_{jn})}\approx\int_0^1 e^{2\pi i(\nu-\mu)s}\,ds=1,\quad \mu=\nu$$
$$=0\quad\text{otherwise}$$

and so, letting D_w be the diagonal matrix with jjth entry $\frac{1}{w(t_{jn})}$, we have

$$\Phi^*D_w\Phi\approx 1.$$

Letting $U=D_w^{1/2}\Phi$, we have that U is (approximately) unitary

$$K\approx D_w^{-1/2}\,UDU^*\,D_w^{-1/2}. \qquad \text{(A 4.4.3)}$$

If equality were to hold in (A 4.4.3) and U were unitary, then we would have that the eigenvalues $\lambda_{\nu n}$ of K satisfy

$$\min_t\left(\frac{1}{w(t)}\right)D_{\nu\nu}\leqq\lambda_{\nu n}\leqq\max_t\left(\frac{1}{w(t)}\right)D_{\nu\nu}.$$

or

$$\min_t w(t)\,D_{\nu\nu}\leqq\lambda_{\nu n}^{-1}\leqq\max_t w(t)\,D_{\nu\nu}.$$

Since the $2\nu^{\text{th}}$ and the $2\nu-1$st largest $D_{\nu\nu}$ are $\frac{n}{(2\pi\nu)^{2m}}$, we then would have (A 4.3.1) with α, β as in (A 4.3.2).

It remains to show that if $1/(n\lambda^{1/2m})$ is bounded below away from 0 as $n\to\infty$, then so is $\frac{1}{n}\operatorname{Tr}A^2(\lambda)$. We have

$$\frac{1}{n}\sum_{\nu=1}^{n}\frac{1}{(1+\beta\pi^{2m}\lambda\nu^{2m})^2}\leqq\frac{1}{n}\operatorname{Tr}A^2(\lambda).$$

Let $\lambda=\lambda(n)$ satisfy $n\lambda^{1/2m}=c^{1/2m}$, equivalently $\lambda=c/n^{2m}$. Then

$$\frac{1}{(1+\beta\pi^{2m}c)^2}\leqq\frac{1}{n}\sum_{\nu=1}^{n}\frac{1}{(1+\beta\pi^{2m}c(\nu^{2m}/n^{2m}))^2}\leqq\frac{1}{n}\operatorname{Tr}A^2(\lambda).$$

Acknowledgements. This work originated in the Common Room at the Oxford University Mathematical Institute, and the ministrations of the tea lady are remembered fondly by the American author (GW). She wishes to express her sincere appreciation for the hospitality of the Oxford University Mathematical Institute and St. Cross College, Oxford, and the support of the Science Research Council of Great Britain. We are grateful to Professor Gene Golub for several helpful discussions which led to the computational approach of Sect. 5.

References

1. Abramowitz, M., Stegun, I.: Handbook of mathematical functions with formulas, graphs, and mathematical tables. U.S. Department of Commerce, National Bureau of Standards Applied Mathematics Series, No. **55**, pp. 803–819, 1964
2. Aronszajn, N.: Theory of reproducing kernels. Trans. Amer. Math. Soc. **68**, 337–404 (1950)
3. Golomb, M.: Approximation by periodic spline interpolants on uniform meshes. J. Approximation Theory **1**, 26–65 (1968)
4. Golub, G., Heath, M., Wahba, G.: Generalized cross validation as a method for choosing a good ridge parameter, to appear, Technometrics
5. Golub, G., Reinsch, C.: Singular value decomposition and least squares solutions. Numer. Math. **14**, 403–420 (1970)
6. Hudson, H.M.: Empirical Bayes estimation. Technical Report #58, Stanford University, Department of Statistics, Stanford, Cal., 1974
7. Kimeldorf, G., Wahba, G.: A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Inst. Statist. Math. **41**, 495–502 (1970)
8. Mallows, C.L.: Some comments on C_p. Technometrics **15**, 661–675 (1973)
9. Reinsch, C.M.: Smoothing by spline functions. Numer. Math. **10**, 177–183 (1967)
10. Reinsch, C.M.: Smoothing by spline functions, II. Numer. Math. **16**, 451–454 (1971)
11. Schoenberg, I.J.: Spline functions and the problem of graduation. Proc. Nat. Acad. Sci. (USA) **52**, 947–950 (1964)
12. Wahba, G.: Convergence rates for certain approximate solutions to first kind integral equations. J. Approximation Theory **7**, 167–185 (1973)
13. Wahba, G.: Smoothing noisy data with spline functions. Numer. Math. **24**, 383–393 (1975)
14. Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy. SIAM J. Numer. Anal. **14**, 651-667 (1977)
15. Wahba, G., Wold, S.: A completely automatic French curve: Fitting spline functions by cross-validation. Comm. Statist. **4**, 1–17 (1975)
16. Wahba, G., Wold, S.: Periodic splines for spectral density estimation: The use of cross-validation for determining the degree of smoothing. Comm. Statist. **4**, 125–141 (1975)
17. Wahba, G.: A survey of some smoothing problems and the method of generalized cross validation for solving them. University of Wisconsin-Madison, Statistics Dept., Technical Report #457. In: Proceedings of the Conference on Applications of Statistics, Dayton, Ohio (P.R. Krishnaiah, ed.) June 14–18, 1976
18. Wahba, G.: Improper priors, spline smoothing and the problem of guarding against model errors in regression. J. Roy. Statist. Soc., Ser. B. To appear

Received January 10, 1978