DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT NO. 381

July 1974

PERIODIC SPLINES FOR SPECTRAL DENSITY ESTIMATION: THE USE OF CROSS VALIDATION FOR DETERMINING THE DEGREE OF SMOOTHING

by

Grace Wahba* and Svante Wold**
University of Wisconsin
Madison, Wisconsin

Typist: Mary Ann Clark

---

*Research was supported by the Air Force Office of Scientific Research under Grant No. AFOSR 72-2363B.

**Statistician-in-Residence 1973-74.

Permanent address: Research Group for Chemometrics, Institute of Institute of Chemistry,
Umeå University, S 901 87, Umeå, Sweden.

PERIODIC SPLINES FOR SPECTRAL DENSITY ESTIMATION: THE USE OF CROSS VALIDATION FOR DETERMINING THE DEGREE OF SMOOTHING

G. Wahba

Department of Statistics, University of Wisconsin, Madison

S. Wold

Research Group for Chemometrics
Institute of Chemistry, Umea University, Umea, Sweden

ABSTRACT

The cross validation mean square error technique is shown to be appropriate for choosing the smoothing or "bandwidth" parameter, in estimating the log spectral density with periodic splines.

1. INTRODUCTION

Let X(t), t = . . . - 1, 0, 1, . . . be a zero-mean stationary Gaussian stochastic process with spectral density f(x), $-\frac{1}{2} < x < \frac{1}{2}$, given by

$$f(x) = \sum_{\tau=-\infty}^{\infty} R(\tau)\, e^{2\pi i x \tau} ,$$

$$R(\tau) = EX(t)X(t+\tau), \qquad \tau = \ldots, -1, 0, 1, \ldots$$

We suppose that $g(x) \equiv \log f(x)$ possess a power series expansion of the form

$$g(x) = \sum_{\nu=-\infty}^{\infty}{}' \, g_\nu e^{2\pi i \nu x}, \qquad x \in [-\tfrac{1}{2}, \tfrac{1}{2}],$$

and furthermore, for some (given) $m \geq 2$,

$$\sum_{\nu=-\infty}^{\infty}{}' |g_\nu|^2 (2\pi\nu)^{2m} < \infty .$$

Our goal is to estimate $g(x)$, $x \in [-\frac{1}{2}, \frac{1}{2}]$ from values of the periodogram $I(x)$,

$$I(x) = \frac{1}{2n} \left| \sum_{\tau=1}^{2n}{}' X(\tau) e^{2\pi i x \tau} \right|^2$$

at the points $x = x_j = j/2n$, $j = -(n-1), \ldots, n$. We will make no attempt to discuss the extensive literature on spectral density estimation here, other than to point out that every method in common use requires that the user choose a bandwidth or smoothing parameter. The purpose of this note is to present an argument that the cross validation mean square error technique (CVMSE) of Wahba and Wold [13] is ideally suited to make this choice. The CVMSE is here used in conjunction with a method introduced by Cogburn and Davis [3] for estimating f. We feel that this is an important idea, since the problem of selecting the degree of smoothing in <u>any</u> spectral density estimation procedure seems to be incompletely solved. Thus, we present the application of the CVMSE idea without direct experimental evidence, relying on the similarity of the formulation of the model here to that in Wahba and Wold (13), and the experimental results there, to persuade the reader of its reasonableness. We present a sketch of a proof that the smoothing parameter chosen by CVMSE converges to the smoothing parameter which minimizes the mean square error.

2. THE PERIODIC SMOOTHING SPLINE FOR LOG SPECTRAL DENSITY ESTIMATION

Let

$$I_j = I(j/2n), \qquad j = -(n-1), \ldots, -1, 0, 1, \ldots n.$$

Then

$$I_j = I_{-j}, \qquad j = 1, 2, \ldots n-1$$

and, to a good approximation

$$I_j = f(j/2n)\, U_j$$

where U_j, $j = 1, 2, \ldots n-1$, are independently distributed as one half times a χ^2 random variable with 2 degrees of freedom, and U_o and U_n are distributed independently of each other and U_j, $j = 1, 2, \ldots, n-1$, and as χ^2 with one degree of freedom. See Walker [12]. Let

$$Y_j = \log I_j + C_j$$

where $C_j = C$, the Euler-Mascheroni constant, for $j = \pm 1, 2, \ldots, n-1$, $C = .57721\ldots$ and let $C_o = C_n = \frac{1}{\pi}(\ln 2 + C)$.
Then

$$Y_j = g(j/2n) + \epsilon_j, \qquad j = -(n-1), \ldots, n, \qquad (2.1a)$$

where $\epsilon_j = \log U_j + C_j$. Using Bateman [2] Vol. I, Section 4.6, and the density of a χ^2 random variable, it is seen that

$$E\epsilon_j = 0,\ j = -(n-1), \ldots, n,\ E\epsilon_j^2 = \pi^2/6 \equiv \sigma^2,\ j = \pm 1, \pm 2, \ldots, \pm n-1. \quad (2.1b)$$

Now, let $\mathcal{H}^{(m)}$ be the linear space of <u>periodic</u> Hermitian functions on $[-\frac{1}{2}, \frac{1}{2}]$ of the form

$$h(x) = \sum_{\nu=-\infty}^{\infty}{}' h_\nu e^{2\pi i \nu x},$$

with $h_\nu = h_{-\nu}^*$, and

$$\sum_{\nu=-\infty}^{\infty}' |h_\nu|^2 (2\pi\nu)^{2m} < \infty .$$

$\mathcal{H}^{(m)}$ is a Hilbert space with the norm defined by

$$\|h\|_m^2 = \sum_{\nu=-\infty}^{\infty} (2\pi\nu)^{2m} |h_\nu|^2 + \frac{|h_o|^2}{\lambda_o}$$

$$\equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} |h^{(m)}(x)|^2 dx + 1/\lambda_o \left[\int_{-\frac{1}{2}}^{\frac{1}{2}} h(x)\, dx \right]^2 ,$$

where λ_o is a fixed positive constant whose choice will be discussed later.

The estimate $g_{n,\lambda}$ of g will be taken as the (unique) solution to the problem: Find $g \in \mathcal{H}^{(m)}$ to

$$\min \frac{1}{2n} \sum_{j=-(n-1)}^{n} (g(j/2n) - Y_j)^2 +$$

$$\lambda \left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} |g^{(m)}(x)|^2 dx + \frac{1}{\lambda_o} \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} g(x) dx \right)^2 \right\}. \qquad (2.2)$$

Cogburn and Davis [3] use the solution to the problem: Find $g \in \mathcal{H}^{(m)}$ to

$$\min \frac{1}{2n} \sum_{j=-(n-1)}^{n}' (g(j/2n) - Y_j)^2 + \lambda \int_{-\frac{1}{2}}^{\frac{1}{2}} |g^{(m)}(x)|^2 dx,$$

which can be shown to be the limit as $\lambda_o \to \infty$ of the solution of (2.2). We discuss (2.2) because it simplifies the technical details to follow.

It can be verified that $\mathcal{H}^{(m)}$ possesses the reproducing kernel $Q(x, y)$ given by

$$Q(x, y) = \sum_{\nu=-\infty}^{\infty}' \lambda_\nu \Phi_\nu(x) \Phi_\nu^*(y) .$$

where

$$\Phi_\nu(x) = e^{2\pi i\nu x}, \qquad \nu = \ldots -1, 0, 1, \ldots$$

$$\lambda_\nu = \lambda_{-\nu} = 1/(2\pi\nu)^{2m}, \quad \nu = 1, 2, \ldots$$

and λ_o is the same as in the definition of $\|h\|_m^2$.

It is well known from e.g. the properties of reproducing kernels, (see Kimeldorf and Wahba, [8]) that the solutionto the problem of (2.2) is

$$g_{n,\lambda}(x) = (Q_{x_1}(x), Q_{x_2}(x), \ldots Q_{x_{2n}}(x))(Q_n+2n\lambda I)^{-1} \begin{pmatrix} Y_{-(n-1)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} \tag{2.3}$$

where $Q_{x_i}(x) = Q(x_i, x)$, $x_i = (i-n)/2n$, $i = 1, 2, \ldots, 2n$, and Q_n is the $2n \times 2n$ matrix with i, jth entry $Q(x_i, x_j)$.

Since

$$Q(x,y) = 2\sum_{\nu=1}^{\infty} \frac{\cos 2\pi\nu(x-y)}{(2\pi\nu)^{2m}} + \lambda_o ,$$

the series expression for $Q(x,y)$ can be summed (see Golomb [6], also Jolley [7]). We have

$$2\sum_{\nu=1}^{\infty} \frac{\cos 2\pi\nu t}{(2\pi\nu)^{2m}} = \frac{(-1)^{r-1}B_{2r}(t)}{(2r)!} \qquad -1 \leq t \leq 1$$

where

$$B_{2r}(t) = \sum_{p=0}^{2r} \binom{2r}{p} B_p |t|^{2r-p},$$

$B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_3 = B_5 = \ldots = 0$, and

$$B_{2p} = (-1)^{p-1}\, 2(2p)! \sum_{\nu=1}^{\infty} \frac{1}{(2\pi\nu)^{2m}}, \qquad p = 1, 2, \ldots$$

The B_p are Bernoulli numbers. Thus

$$B_2(x) = x^2 - |x| + 1/6$$

$$B_4(x) = x^4 - 2|x|^3 + x^2 - 1/30$$

$$B_6(x) = x^6 - 3|x|^5 + 5/2\, x^4 - 1/2 x^2 + 1/42$$

$g_{n,\lambda}(x)$ is thus the sum of a polynomial spline of degree $2m-1$ and continuity class C^{m-2} with knots at x_i, and a multiple of x^{2m}. As $\lambda_o \to \infty$, the coefficient of $x^{2m} \to 0$ so that $g_{n,\lambda}(x)$ is a spline. See Kimeldorf and Wahba [8, Lemma 5.1] for the $\lambda_o = \infty$ solution. See Cogburn and Davis [3] for a simplified representation, as well as a discussion of computational considerations and experimental results with varying λ. Golomb [6] has discussed interpolating periodic splines ($\lambda = 0$). Note that $Q_n + 2n\lambda I$ is a circulant matrix so that its inverse can be written exactly.

We next give an approximate large n formula for $g_{n,\lambda}$ for the purpose of studying the CVMSE.

Let

$$Q^n(x,y) = \sum_{\nu=-(n-1)}^{n}{}' \lambda_\nu \Phi_\nu(x) \Phi_\nu^*(y)$$

and let Q_n^n be the $2n \times 2n$ matrix with $i\,j$th entry $Q^n(x_i, y_j)$. Then

$$Q_n^n = W D W^*$$

where W is the $2n \times 2n$ dimensional unitary matrix with $j\,k$th entry

$$(1/\sqrt{2n})\, \Phi_{-n+j}(x_k) = (1/\sqrt{2n})\, e^{2\pi i (j-n)(k-n)/2n},$$

and D is the diagonal matrix with $j\,j$th entry $2n\,\lambda_{j-n}$, $j = 1, 2, \ldots, 2n$. Note that the diagonal entries of $Q_n - Q_n^n$ are bounded in absolute value by

$$2 \sum_{\nu=n}^{\infty}{}' \frac{1}{(2\pi\nu)^{2m}} \le \frac{1}{(2\pi)^{2m}} \quad \frac{1}{(2m-1)n^{2m-1}},$$

which can be used to show rigorously that all of the following approximations are "good" for large n.

Then, approximating Q by Q^n in (2.3) gives

$$g_{n,\lambda}(x_j) \simeq g^n_{n,\lambda}(x_j) \equiv \frac{1}{\sqrt{2n}} \sum_{\nu=(n-1)}^{n}{}' e^{2\pi i(j-n)\nu/2n} \frac{\lambda_\nu}{(\lambda_\nu + \lambda)} \tilde{Y}_\nu$$

where

$$\tilde{Y}_\nu = \frac{1}{\sqrt{2n}} \sum_{k=-(n-1)}^{n}{}' e^{-2\pi i \nu k/2n} Y_k ,$$

equivalently

$$g^n_{n,\lambda}\left(\frac{j}{2n}\right) = \frac{1}{2n} \sum_{k=-(n-1)}^{n}{}' Y_k a_\lambda\left(\frac{j-k}{2n}\right) \qquad (2.5a)$$

where

$$a_\lambda\left(\frac{\tau}{2n}\right) = \sum_{\nu=-(n-1)}^{n}{}' e^{2\pi i(\tau-n)\nu/2n} \frac{\lambda_\nu}{\lambda_\nu + \lambda} . \qquad (2.5b)$$

Equation (2.5) with $\frac{\lambda_o}{\lambda_o+\lambda} = 1$ is the approximation used by Cogburn and Davis to compute their spectral density estimate. (Note that our λ is Cogburn and Davis' $1/\lambda^{2m}$).

3. THE OPTIMUM THEORETICAL CHOICE OF λ.

The criteria we adopt for choosing λ is to minimize the average mean square error

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

The optimum $\lambda = \lambda^*$ is given by Cogburn and Davis [3] who gave (3.4) - (3.5) below for the estimation of f, assuming $f^{(2m)}$ continuous. To be able to make precise our claims concerning the CVMSE, we give the formula for λ^*.

We sketch the proof under the slightly weaker condition $g \in \mathcal{H}^{(2m)}$.

Let

$$\bar{g}_n = (g(x_1),\ g(x_2),\ \ldots\ ,\ g(x_{2n}))'$$

$$\bar{\epsilon}_n = (\epsilon_1,\ \epsilon_2,\ \ldots\ \epsilon_{2n})'$$

Then

$$E\ \frac{1}{2n} \sum_{j=1}^{2n}{}' (g_{n,\lambda}(x_j) - g(x_j))^2 = \frac{1}{2n}\ E\{(2n\lambda)^2 \bar{g}_n'(Q_n + 2n\lambda I)^{-2}\bar{g}_n$$

$$- 4n\lambda\bar{\epsilon}_n'(Q_n + 2n\lambda I)^{-1} Q_n (Q_n + 2n\lambda I)^{-1}\bar{g}_n'$$

$$+ \bar{\epsilon}_n'(Q_n + 2n\lambda I)^{-1} Q_n^2 (Q_n + 2n\lambda I)^{-1}\bar{\epsilon}_n\} \qquad (3.1)$$

Let

$$g^n(x) = \sum_{\nu=-(n-1)}^{n}{}' g_\nu \Phi_\nu(x) \qquad \text{and}$$

$$\bar{g}_n^n = (g^n(x_1), g^n(x_2), \ldots, g^n(x_{2n}))'$$

$$\equiv \sqrt{2n}\ W(g_{-(n-1)}, g_{-(n-2)}, \ldots, g_n).$$

Approximating $\bar{g}_n$ by $\bar{g}_n^n$ and Q_n by Q_n^n, it can be shown that the right hand side of Equation (3.1) is $(1 + o(1))$ times

$$\sum_{\nu=-(n-1)}^{n}{}' \frac{(2n\lambda)^2 g_\nu^2}{(2n\lambda_\nu + 2n\lambda)^2} + E\ \frac{1}{2n} \sum_{\nu=-(n-1)}^{n}{}' |\tilde{\epsilon}_\nu|^2 \frac{(2n\lambda_\nu)^2}{(2n\lambda_\nu + 2n\lambda)^2}$$

where

$$\tilde{\epsilon}_\nu = \frac{1}{\sqrt{2n}} \sum_{j=-(n-1)}^{n}{}' e^{2\pi i\nu j/2n}\ \epsilon_j\ .$$

Now, making the approximation $E\epsilon_j^2 = \sigma^2 \equiv \pi^2/6$, all j, gives

$$E|\tilde{\epsilon}_\nu|^2 = \sigma^2$$

and

$$E\,\frac{1}{2n}\sum_{j=1}^{2n}{}' (g_{n,\lambda}(x_j) - g(x_j))^2$$

$$= \lambda^2\left\{\sum_{\nu=-(n-1)}^{n}{}' \frac{g_\nu^2}{(\lambda_\nu+\lambda)^2} + \frac{\sigma^2}{2n}\sum_{\nu=-(n-1)}^{n}{}' \frac{\lambda_\nu^2}{(\lambda_\nu+\lambda)^2}\right\}(1+o(1)) \tag{3.2}$$

$$\leq \left(\lambda^2 \langle g, g\rangle_{2m} + \frac{k_m\sigma^2}{2n\,\lambda^{1/2m}}\right)(1+o(1)) \tag{3.3}$$

where

$$\langle g,g\rangle_{2m} = \int_{-\frac{1}{2}}^{\frac{1}{2}} |g^{(2m)}(x)|^2\,dx + 1/\lambda_0^2\left[\int_{-\frac{1}{2}}^{\frac{1}{2}} g(x)dx\right]^2 \equiv \sum_{\nu=-\infty}^{\infty}{}' \frac{g_\nu^2}{\lambda_\nu^2}\,,$$

and

$$k_m = \frac{1}{\pi}\int_0^\infty \frac{dy}{(1+y^{2m})^2}\,.$$

Expression (3.3) is minimized for $\lambda = \lambda^*(1+o(1))$,

$$\lambda^* = \left\{\left(\frac{k_m}{4m}\right)\cdot\frac{\sigma^2}{\langle g,g\rangle_{2m}}\quad\frac{1}{2n}\right\}^{2m/(4m+1)}. \tag{3.4}$$

It is shown in Wahba [8] that, as $n \to \infty$, λ^* also tends to the minimizer of (3.2). Then

$$\frac{1}{2n}\,E\sum_{\nu=-(n-1)}^{n}{}' (g_{n,\lambda^*}(x_i) - g(x_i))^2$$

$$= \left(\frac{k_m}{4m}\right)^{4m/(4m+1)}(4m+1)\left[\frac{\langle g,g\rangle_{2m}}{\sigma^2}\right]^{1/(4m+1)}\frac{1}{n^{4m/(4m+1)}}\,. \tag{3.5}$$

Now this result does not in practice tell us how to choose λ since $\langle g, g \rangle_{2m}$ is unknown. However, we make the following claims

1) The rate $O(\frac{1}{n}^{4m/(4m+1)})$ is the best achievable for average mean-square error convergence, that is uniformly good over all $g \in \mathcal{H}^{(2m)}$ with $\langle g, g \rangle_{2m} \leq$ some constant. See Wahba [11] for proof of a related theorem for density estimates.
2) The CVMSE technique in Wahba and Wold [13] and described in the next section <u>estimates λ^*</u>.

4. THE CVMSE PROCEDURE FOR FINDING A GOOD λ.

The CVMSE procedure in its "purest" form goes as follows:

1) Guess a λ

2) Delete the kth data point. Here the kth data point means Y_k and Y_{-k} for $k \neq 0$ or n. (In Wahba and Wold [13] we have deleted groups of points at a time to decrease the amount of computation).
3) Compute $g^k_{n,\lambda}$, the solution to the minimization problem of (2. 2), with the remaining data. (That is, in (2. 2) the term(s) in the sum corresponding to the missing data is (are) deleted.) $g^k_{n,\lambda}(x)$ is given by (2. 3) upon deleting rows and columns corresponding to the missing data.

4) Let $CV(\lambda) = \frac{1}{2n} \sum_{k=-(n-1)}^{n} (g^k_{n,\lambda}(x_k) - Y_k)^2$

5) Compute $CV(\lambda)$ for various values of λ until a minimum $\tilde{\lambda}$ is found.

5. COMPUTATIONAL CONSIDERATIONS

Let

$$\underset{\sim}{Y} = (Y_{-(n-1)}, Y_{-(n-2)}, \cdots Y_n)$$

Observe that $S(\lambda)$ defined by

$$S(\lambda) \equiv \frac{1}{2n} \sum_{j=-(n-1)}^{n}{}' (g_{n,\lambda}(x_j) - Y_j)^2$$

satisfies

$$S(\lambda) = \lambda^2 (2n) \overline{Y}' (Q_n + 2n\lambda I)^{-2} \overline{Y} \quad .$$

$$\simeq \lambda^2 (2n) \overline{Y}' (Q_n^n + 2n\lambda I)^{-2} \overline{Y}$$

$$= \lambda^2 \frac{1}{2n} \sum_{\nu=-(n-1)}^{n}{}' \frac{|\tilde{Y}_\nu|^2}{(\lambda_\nu + \lambda)^2}$$

where

$$\tilde{Y}_\nu = \frac{1}{\sqrt{2n}} \sum_{j=-(n-1)}^{n}{}' e^{2\pi i \nu j/2n} Y_j \quad .$$

It is known from the experiments of Gasser [3], Wahba and Wold [13] and, Wold [14], and theoretical work of Wahba [10] that $S(\lambda^*) \approx k\sigma^2$, where k is generally less than 1. See Wahba and Wold [13] for experimental verification of the dependence of k on the sample size. Thus, to obtain a starting guess for this problem, one might make a plot of $S(\lambda)$ and find λ for which $S(\lambda) \simeq k\sigma^2$, with $k \approx .8$ say, to pick a number from nowhere.

Now, we find an approximate explicit expression for $CV(\lambda)$. Let M be the $2n \times 2n$ dimensional matrix

$$M = Q_n + 2n\lambda I$$

and let M_k be the $2n-2 \times 2n-2$ dimensional matrix formed from M by deleting the n-kth and the n + kth row and column, that is, the rows and columns corresponding to Y_k, $k = 1, 2, \ldots, n-1$. Let q_{n-k} and q_{n+k} be the $2n-2$ dimensional column vectors formed from the n-k th and n + k th columns of Q_n respectively by deleting the n-k th and the n + k th entries in each, $k = 1, 2, \ldots, n-1$. Similarly define M_o, g_n and M_n, g_{2n} for $k = 0, n$, where only entries corresponding to Y_o and then Y_n are deleted. (M_o is of dimension 2n-1). Let $\overline{Y}^k$ be formed from $\overline{Y}$ by deleting Y_k.

Then

$$CV(\lambda) = \frac{1}{2n} \sum_{k=-(n-1)}^{n}{}' (g^k_{n,\lambda}(x_k) - Y_k)^2$$

$$= \frac{1}{2n} \sum_{k=1}^{n-1}{}' (q'_{n-k} M_k^{-1} \overline{Y}_k - Y_{-k})^2 + (q'_{n+k} M_k^{-1} \overline{Y}_k - Y_k)^2$$

$$+ (q'_n M_o^{-1} \overline{Y}_o - Y_o)^2 + (q'_{2n} M_k^{-1} \overline{Y}_n - Y_n)^2 \quad ..$$

It will next be important to note that if we had formed the q's from columns of Q_n instead of from M, the result would be the same, since these two matrices differ only on the diagonal. We can then use the following: For any $(2n) \times (2n)$ symmetric positive definite matrix partitioned as follows:

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

we have

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

with

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

It follows, upon rearranging so that $M'_{12} = (g_{n-k} : g_{n+k})$, g_o or g_n, that

$$\begin{pmatrix}(M^{-1}\overline{Y})_{n-k}\\ (M^{-1}\overline{Y})_{n+k}\end{pmatrix}=\begin{pmatrix}m^{n-k,\,n-k} & m^{n-k,\,n+k}\\ m^{n-k,\,n+k} & m^{n+k,\,n+k}\end{pmatrix}\begin{pmatrix}Y_{-k}-g'_{n-k}M_k^{-1}\overline{Y}\\ Y_k-g'_{n+k}M_k^{-1}\overline{Y}\end{pmatrix},\ k=1,2,\ldots,n-1,$$

$$(M^{-1}\overline{Y})_n = m^{n,n}(Y_o-g'_n\ M_o^{-1}\overline{Y})$$

$$(M^{-1}\overline{Y})_{2n}= m^{2n,2n}(Y_n-g'_{2n}\ M_n^{-1}\overline{Y}),$$

where $(M^{-1}\overline{Y})_k$ is the k th entry of $M^{-1}\overline{Y}$, and m^{ij} is the ij th entry of M^{-1}.

Now, it is clear by symmetry that $Y_{-k}-g'_{n-k}M_k^{-1}\overline{Y} = Y_k-g'_{n+k}M_k^{-1}\overline{Y}$. Also, since M is a circulant matrix, so is M^{-1}, which entails that $m^{n-k,\,n-k} = m^{n+k,\,n+k} = m^{oo}$, all k, $m^{n-k,\,n+k} = m^{o,\,2k}$, and hence

$$Y_k-g'_{n-k}\ M_k^{-1}\overline{Y} = \frac{1}{m^{oo}+m^{o,\,2k}}(M^{-1}\overline{Y})_{n-k}\,.$$

Let $\epsilon_k = m^{o,\,2k}(2m^{oo} + m^{o,\,2k})/(m^{oo} + m^{o,\,2k})^2$, $k = \pm 1,\ldots,\pm n-1$, $\epsilon_o = \epsilon_n = 0$. Then

$$CV(\lambda) = \frac{1}{2n}\sum_{k=-(n-1)}^{n}{}' (M^{-1}Y)_k^2/(m^{oo})^2 - \frac{1}{2n}\sum_{k=-(n-1)}^{n}{}' \epsilon_k(M^{-1}Y)_k^2/(m^{oo})^2. \tag{5.1}$$

We have $|\epsilon_k| \le 2\,|m^{o,\,2k}|/m^{oo}$. Making the approximation

$$M \simeq W(D + 2n\lambda I)\,W^* \equiv Q_n^n + 2n\lambda I \tag{5.2}$$

gives

$$m^{oo} = \frac{1}{2n}\sum_{j=-(n-1)}^{n}{}' \frac{1}{(2n\lambda_j + 2n\lambda)}$$

$$m^{o,\,2k} = \frac{1}{2n}\sum_{j=-(n-1)}^{n}{}' \frac{e^{2\pi i(2kj)/2n}}{(2n\lambda_j + 2n\lambda)}\,.$$

The next step in our argument is to claim, for λ in the range we will be interested in, and n large, that $m^{o,2k}$ is small compared to m^{oo} for all but a few values of k near 0, and the second term in (5.1) can be neglected compared to the first. (Note that $1/(2n\lambda_j + 2n\lambda) \simeq 1/2n\lambda$ = constant for all but a few λ_j, i.e. whenever $1/(2\pi j)^{2m} << \lambda$. We omit a rigorous argument). Neglecting the second term in (5.1) and using (5.2) gives

$$CV(\lambda) \simeq \frac{1}{2n} \sum_{\nu=-(n-1)}^{n}{}' \frac{|\tilde{Y}_\nu|^2}{(2n\lambda_\nu + 2n\lambda)^2} \Bigg/ \left[\frac{1}{2n} \sum_{\nu=-(n-1)}^{n}{}' \frac{1}{(2n\lambda_\nu + 2n\lambda)} \right]^2$$

$$= \frac{1}{2n} \sum_{\nu=-(n-1)}^{n}{}' \frac{|\tilde{Y}_\nu|^2}{(\lambda_\nu + \lambda)^2} \Bigg/ \left[\frac{1}{2n} \sum_{\nu=-(n-1)}^{n}{}' \frac{1}{(\lambda_\nu + \lambda)} \right]^2 . \quad (5.3)$$

Thus, one can directly seek to find the minimum of (5.3).

6. WHY DOES $\min_\lambda$ CV(λ) MINIMIZE THE EXPECTED MEAN SQUARE ERROR?

From (3.2) the expected mean square error when λ is used, to be denoted by E TR(λ), is

$$E\, TR(\lambda) \equiv E \frac{1}{2n} \sum_{j=1}^{2n}{}' (g_{n,\lambda}(x_j) - g(x_j))^2$$

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

where "$\simeq$" means $(1 + o(1))$ as $n \to \infty$, $\lambda \to 0$ in such a way that $n\lambda \to \infty$. (TR = true)

Using (5.1), the expected value of $CV(\lambda)$ is

$$E\,CV(\lambda) \simeq E\,\frac{1}{2n}\sum_{\nu=-(n-1)}^{n}{}' \frac{|\tilde{Y}_\nu|^2}{(\lambda_\nu+\lambda)^2} \Bigg/ \left[\frac{1}{2n}\sum_{\nu=-(n-1)}^{n}{}' \frac{1}{(\lambda_\nu+\lambda)}\right]^2$$

$$\simeq \frac{1}{2n}\sum_{\nu=-(n-1)}^{n}{}' \frac{(2n)g_\nu^2 + E|\tilde{\epsilon}_\nu|^2}{(\lambda_\nu+\lambda)^2} \Bigg/ \left[\frac{1}{2n}\sum_{\nu=-(n-1)}^{n}{}' \frac{1}{(\lambda_\nu+\lambda)}\right]^2 .$$

Now, suppose $E|\tilde{\epsilon}_\nu|^2 = \sigma^2$, all ν, and let

$$\Psi_g(\lambda) = \lambda^2 \sum_{\nu=-(n-1)}^{n}{}' \frac{g_\nu^2}{(\lambda_\nu+\lambda)^2}, \quad c_\nu = \frac{\lambda_\nu}{\lambda_\nu+\lambda} .$$

Then

$$E\,TR(\lambda) \simeq \Psi_g(\lambda) + \frac{\sigma^2}{2n}\sum_{\nu=-(n-1)}^{n}{}' c_\nu^2$$

$$E\,CV(\lambda) \simeq \left[\Psi_g(\lambda) + \frac{\sigma^2}{2n}\sum_{\nu=-(n-1)}^{n}{}' (1-c_\nu)^2\right] \Bigg/ \left[1-\frac{1}{2n}\sum_{\nu=-(n-1)}^{n}{}' c_\nu\right]^2 .$$

Now

$$\frac{1}{2}\sum_{\nu=-(n-1)}^{n}{}' c_\nu^2 \simeq \frac{k_m}{n\lambda^{1/2m}}, \quad \frac{1}{2}\sum_{\nu=-(n-1)}^{n}{}' c_\nu \simeq \frac{\tilde{k}_m}{n\lambda^{1/2m}}$$

where

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

thus

$$E\,TR(\lambda) \simeq \Psi_g(\lambda) + \frac{\sigma^2 k_m}{n\lambda^{1/2m}} \tag{6.1}$$

$$E\,CV(\lambda) \simeq \left\{\Psi_g(\lambda) + \sigma^2\left(1 - \frac{2\tilde{k}_m}{n\lambda^{1/2m}} + \frac{k_m}{n\lambda^{1/2m}}\right)\right\} \Bigg/ \left[1-\frac{\tilde{k}_m}{n\lambda^{1/2m}}\right]^2 . \tag{6.2}$$

Now, the minimum λ^* of (6.1) occurs for the (smallest) solution of

$$\Psi_g'(\lambda) - \frac{\sigma^2 k_m}{2mn\lambda^{(2m+1)/2m}} = 0 ,$$

and is

$$\lambda^* \simeq \left(\frac{k_m}{4m} \frac{\sigma^2}{\|g^{(2m)}\|_2^2} \right)^{2m/(4m+1)} \frac{1}{n^{2m/(4m+1)}} .$$

See [8] for details, it is shown there that $\Psi_g(\lambda)$ can be replaced by $\lambda^2 \|g^{(2m)}\|^2$, with vanishingly small error.

Differentiating the right hand side of (6.2) with respect to λ gives

$$\frac{d}{d\lambda} CV(\lambda) = \frac{1}{V^{3/2}} \left\{ \Psi_g'(\lambda) - \frac{\sigma^2 k_m}{2m\, n\lambda^{(2m+1)/2m}} \left[1 - \frac{2\tilde{k}_m}{\sigma^2 k_m} \left(\sigma^2 - \frac{U}{V^{\frac{1}{2}}} \right)^2 \right] \right\}$$

where

$$U = \Psi_g(\lambda) + \sigma^2 \left(1 - \frac{1}{n\lambda^{1/2m}} (2\tilde{k}_m - k_m) \right)$$

$$V = \left(1 - \frac{\tilde{k}_m}{n\lambda^{1/2m}} \right)^2 .$$

Now $\Psi_g(\lambda) \le \lambda^2 \|g^{(2m)}\|_2^2$, thus, in the neighborhood of

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

$$U = \sigma^2 + O(n^{-4m/(4m+1)})$$

$$V = 1 + O(n^{-4m/(4m+1)})$$

$$\sigma^2 - \frac{U}{V^{\frac{1}{2}}} = O(n^{-4m/(4m+1)}) .$$

and

$$\frac{d}{d\lambda} CV(\lambda)$$

$$= \left(1+O(n^{-4m/4m+1})\right) \left\{ \Psi_g'(\lambda) - \frac{\sigma^2 k_m}{2mn\lambda^{(2m+1)/2m}} \left(1+O(n^{-4m/4m+1})\right) \right\} .$$

Thus, $\frac{d}{d\lambda} CV(\lambda) = 0$ for $\lambda = \lambda^{**} = \lambda^* \left(1+ O(n^{-4m/4m+1})\right)$.

Note that this argument holdsfor fixed $\sigma^2, \|g^{(2m)}\|_2^2$, but does not necessarily hold for the general case of curve estimation by periodic splines if $\sigma^2 \to 0$.

We would like to hear of any computational results.

7. REMARKS

Remark 1: The reader who is familiar with the ridge regression approach to regression of Marquardt [9] will note the similarity of the problem considered here to that of choosing the ridge parameter k in a ridge regression.

Remark 2: The CVMSE technique for estimating λ has a certain intimate relationship to a method studied by Anderssen and Bloomfield [1, Sections 4 and 5], related to work of Cullum [4], for recovering smooth curves and their derivatives from noisy data. Anderssen and Bloomfield's technique can be thought of as based on the model

$$Y_j = g(j/2n) + \epsilon_j, \qquad j = -(n-1), \ldots, n,$$

where $g(j/2n)$ is not a value of a fixed function, but a realization of a random variable G_j, where $\{G_j\}_{j=-\infty}^{\infty}$ is a zero mean stationary Gaussian process with spectral density $f_G(\omega)$ given by

$$f_G(\omega) = \frac{\sigma^2}{\lambda\left[\left(\frac{2\pi\omega}{\Delta}\right)^2 + \left(\frac{2\pi\omega}{\Delta}\right)^4\right]} \qquad -\tfrac{1}{2} \le \omega \le \tfrac{1}{2} \tag{7.1}$$

or

$$f_G(\omega) = \frac{\sigma^2}{\lambda[(\frac{2\pi\omega}{\Delta})^2 + (\frac{2\pi\omega}{\Delta})^4 + (\frac{2\pi\omega}{\Delta})^6]}, \tag{7.2}$$

and $\{\epsilon_i\}_{i=-\infty}^{\infty}$ is a Gaussian white noise process with variance σ^2. Δ is fixed and for comparison we set $\Delta = 1/n$. Their estimate of $g'(x)$ corresponds to the derivative of $\tilde{g}_{n,\lambda}$ given by

$$\tilde{g}_{n,\lambda}(x) = \frac{1}{2n} \sum_{k=-(n-1)}^{n}{}' Y_k b_\lambda(x - \frac{k}{2n})$$

where, if (7.1) holds

$$b_\lambda(x) = \sum_{\nu=-(n-1)}^{n}{}' e^{2\pi i(x-\frac{1}{2})\nu} \frac{1}{1 + \lambda\left[\left(\frac{2\pi x_{\nu+n}}{\Delta}\right)^2 + \left(\frac{2\pi x_{\nu+n}}{\Delta}\right)^4\right]}$$

$$= \sum_{\nu=-(n-1)}^{n}{}' e^{2\pi i(x-\frac{1}{2})\nu} \frac{\xi_\nu}{(\xi_\nu + \lambda)}$$

where $\xi_\nu = 1/[(\pi\nu)^2 + (\pi\nu)^4]$. If (7.2) holds, $\xi_\nu = 1/[(\pi\nu)^2 + (\pi\nu)^4 + (\pi\nu)^6]$. They find ([1], eq. (15), upon taking the anti log), that an approximate maximum likelihood estimate for λ with this model is obtained by minimizing

$$\sum_{\substack{\nu=-(n-1)\\ \nu\neq 0}}^{n}{}' \frac{|\tilde{Y}_\nu|^2}{1 + \left[\frac{1}{\lambda[(\pi\nu)^2 + (\pi\nu)^4]}\right]} \Bigg/ \left\{\prod_{\substack{\nu=-(n-1)\\ \nu\neq 0}}^{n} \left[\frac{1}{1 + \frac{1}{\lambda[(\pi\nu)^2 + (\pi\nu)^4]}}\right]\right\}^{1/(2n-1)}$$

$$\equiv \lambda \sum_{\substack{\nu=-(n-1)\\ \nu\neq 0}}^{n}{}' \frac{|\tilde{Y}_\nu|^2}{(\xi_\nu + \lambda)} \Bigg/ \left\{\prod_{\substack{\nu=-(n-1)\\ \nu\neq 0}}^{n} \left(1 - \frac{\xi_\nu}{(\xi_\nu + \lambda)}\right)\right\}^{1/(2n-1)}.$$

Compare with

$$CV(\lambda) = \frac{\lambda^2}{2n} \sum_{\nu=-(n-1)}^{n}{}' \frac{|\tilde{Y}_\nu|^2}{(\lambda_\nu+\lambda)^2} \Bigg/ \left[1 - \frac{1}{2n} \sum_{\nu=-n-1}^{n}{}' \frac{\lambda_\nu}{\lambda_\nu + \lambda} \right]^2 .$$

8. ACKNOWLEDGEMENT

This work was supported by the Air Force Office of Scientific Research under Grant No. AFOSR 72-2363B and the Swedish Natural Science Research Foundation and The Institute of Applied Mathematics, Stockholm. This work was done while Professor Wold was Statistician-in-Residence at the University of Wisconsin, Madison.

BIBLIOGRAPHY

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

[2] Bateman, H. (1954). Tables of Integral Transforms, McGraw Hill, New York.

[3] Cogburn, R., and Davis, H. (1973). Periodic splines and spectral estimation. University of New Mexico-Albuquerque, Mathematics and Statistics Department TR #253 (Rev.).

[4] Cullum, J. (1971). Numerical differentation and regularization, SIAM J. Numer. Anal. 8, 254-265.

[5] Gasser, Theo, (no date given). Smooth estimation of spectra with splines, preprint, Eidgenössische Techn. Hochschule, Fachgruppe fur Statistik, Zürich.

[6] Golomb, M. (1968). Approximation by periodic spline interpolants on uniform meshes, J. Approx. Thy. 1, 26-65.

[7] Jolley, L. B. W. (1961). Summation of Series, Second Revised Edition. Dover Publications, New York.

[8] Kimeldorf, G., and Wahba, G. (1971). Some results on Tchebycheffian spline functions, J. Math Anal. Appl. 33, 82-95.

[9] Marquardt, D. (1970). Generalized inverses, ridge regression, biassed linear estimation, and non-linear estimation. Technometrics 12, 3, 591-611.

[10] Wahba, G. (1973). Smoothing noisy data by spline functions II. University of Wisconsin-Madison, Statistics Department TR #380.

[11] Wahba, G. (1974). Optimal convergence properties of orthgonal series, kernel, and variable knot methods for density estimation. To appear, Ann. Statist.

[12] Walker, A. M. (1965). Some asymptotic results for the periodogram of a stationary time series, J. Australian Math. Soc., 5, 107-128.

[13] Wahba, G., and Wold, S. (1974). A completely automatic French curve: Fitting spline functions by cross-validation. University of Wisconsin-Madison, Statistics Department TR #379.

[14] Wold, S. (1974). Spline functions in data analysis, Technometrics, 16, 1, 1-11.