The Annals of Statistics
1974, Vol. 2, No. 6, 1108–1126

PERIODIC SPLINES AND SPECTRAL ESTIMATION

BY ROBERT COGBURN AND HERBERT T. DAVIS

University of New Mexico

The theory of periodic smoothing splines is presented, with application to the estimation of periodic functions. Several theorems relating the order of the differential operator defining the spline to the saturation (order of bias) of the estimator are proven. The linear operator which maps a function to its periodic continuous smoothing spline approximation is represented as a convolution operator with a given convolution kernel. This operator is shown to be the limit of a sequence of operators which map a function into the periodic version of the usual lattice smoothing spline. The convolution kernel above appears as the kernel in a kernel type estimate of the spectral density. Thus, it is shown that, a smoothing spline spectral density estimate, is also asymptotically a kernel type spectral density estimate. Some numerical results are presented.

1. Introduction. While the use of interpolatory splines has recently become quite common, still little is known about the use of smoothing splines with statistical data. In this paper the problem of estimating a periodic $[2\pi]$ function $f(x)$, where $h(x) = f(x) + \varepsilon(x)$ ($E\varepsilon(x) \equiv 0$) is observed either on a lattice of points, or continuously in x, using periodic smoothing splines, is discussed.

There exists a variety of types of smoothing splines, so in this paper a "smoothing spline" is taken to be an "L spline" which is not constrained to go through the data points, but is, however, penalized for lack of fidelity to the data as well as roughness. The periodic version is defined in Section 2 of this paper, while a definition of non-periodic smoothing splines can be found in Kimeldorf and Wahba [3] and [4]. Spline functions are found to be a very flexible technique in that given an "order of smoothness" (i.e. order of saturation which is defined in Section 4), the appropriate spline gives the "smoothest" possible fit for that order. This is contrasted with the classical estimators where the order of smoothness of the estimate is fixed by the estimator. Conditions for matching the order of saturation of the spline to the smoothness of the function are also discussed.

Throughout the paper, the norm $||h||$ denotes the usual supremum norm

$$||h|| = \sup_{-\pi < x \leq \pi} |h(x)| .$$

2. Lattice smoothing splines. We begin by recalling a few standard results from Fourier analysis and introducing some notation as we do so.

Let $\mathscr{P}$ be the class of all real-valued functions in $(-\infty, \infty)$ of period 2π, and let $\mathscr{P}_0$ be the set of all Borel measurable g in $\mathscr{P}$ such that $\int_{-\pi}^{\pi} (g(x))^2\, dx < \infty$.

Received August 1972; revised October 1973.

[1] Research partially sponsored by NSF Grant GP-22825.

AMS 1970 *subject classifications*. Primary 62M15, 41A15.

Key words and phrases. Splines, smoothing splines, time series, spectral density estimation, saturation.

For g in $\mathscr{P}_0$, let

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} g(x)\, dx .$$

Then g coincides in $L_2(-\pi, \pi)$ with the series

$$\sum_{n=-\infty}^{\infty} a_n e^{inx} . \tag{1}$$

Moreover, the series (1) is real if, and only if, $a_{-n} = \bar{a}_n$ for each n (where the "bar" denotes complex conjugate), and this series is in $\mathscr{P}_0$ if, and only if, it is real and $\sum_{n=-\infty}^{\infty} |a_n|^2 < \infty$.

Now, for $k = 1, 2, \cdots$, let

$$\mathscr{P}_k = \{g(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx} \in \mathscr{P}_0 \colon \sum_{n=-\infty}^{\infty} |a_n|^2 n^{2k} < \infty\} .$$

Clearly $\mathscr{P}_0 \supset \mathscr{P}_1 \supset \mathscr{P}_2 \supset \cdots$. It follows from Hölder's inequality that, for $g(x) = \sum a_n e^{inx}$ in $\mathscr{P}_k$, $\sum |a_n| n^{k-1} < \infty$. In fact, g is in $\mathscr{P}_k$ if, and only if, g is in $\mathscr{P}$, has $k - 1$ continuous derivatives, and the kth derivative of g, $g^{(k)}$, is in $\mathscr{P}_0$.

Letting $C^{(k)}$ denote the set of all g in $\mathscr{P}$ such that g has k continuous derivatives, we have

$$\mathscr{P}_k = \{g \in \mathscr{P} \colon g \in C^{(k-1)} \text{ and } g^{(k)} \in \mathscr{P}_0\} .$$

The space $\mathscr{P}_k$ is a periodic Sobolov space.

Also, note that, for any g in $\mathscr{P}_1$, the series (1) converges pointwise and uniformly to g.

Let L be a linear differential operator

$$L = \frac{d^\nu}{dx^\nu} + \gamma_1 \frac{d^{\nu-1}}{dx^{\nu-1}} + \cdots + \gamma_\nu . \tag{2}$$

Throughout this paper ν is used to denote the order of L and $\gamma_1, \cdots, \gamma_\nu$ to denote its coefficients. (The value $\nu = 0$ is not allowed, so L is at least first order.)

We consider approximating a function h in $\mathscr{P}$ by a "smooth" function $g \in \mathscr{P}_\nu$ and introduce the following measure of closeness, which balances the accuracy of the approximation with the smoothness of the approximating function:

$$\Delta_{n,\lambda,L}(g, h) = \frac{1}{n} \sum_{k=-n+1}^{n} \left(g\left(\frac{k\pi}{n}\right) - h\left(\frac{k\pi}{n}\right)\right)^2 + \frac{1}{\lambda^{2\nu}\pi} \int_{-\pi}^{\pi} (Lg(x))^2\, dx .$$

A function g in $\mathscr{P}_\nu$ minimizing $\Delta_{n,\lambda,L}(g, h)$ will be called a *periodic lattice smoothing spline* (LSS) *to h*. To specify a particular LSS the operator L and constants n, λ must be given (always $n \geqq 1$, $\lambda > 0$). These will be stated when necessary, and omitted otherwise, in order to simplify notation.

Of course, $\Delta_{n,\lambda,L}(g, h)$ depends on h only through its values on the lattice

$$\mathscr{L}_{\pi/n} = \left\{0, \pm\frac{\pi}{n}, \pm\frac{2\pi}{n}, \cdots\right\} .$$

The ensuing discussion, through Theorem 1, will show that a minimizing g always exists, that the solution is unique for n sufficiently large, and that it is a linear function of the values of h restricted to $\mathscr{L}_{\pi/n}$.

Any $h \in \mathscr{P}$ coincides on $\mathscr{L}_{\pi/n}$ with a function $\tilde{h}(x) = \sum_{k=-n+1}^{n} \tilde{a}_k e^{ikx}$. To see this, first note that

$$(3) \qquad \frac{1}{2n} \sum_{l=-n+1}^{n} e^{i(j-k)l\pi/n} = 1 \qquad \text{if } j-k \in \{0, \pm 2n, \pm 4n, \cdots\}$$
$$= 0 \qquad \text{otherwise.}$$

Let

$$(4) \qquad \tilde{a}_k = \frac{1}{2n} \sum_{l=-n+1}^{n} e^{-ikl\pi/n} h\left(\frac{l\pi}{n}\right)$$
$$\tilde{h}(x) = \sum_{k=-n+1}^{n} \tilde{a}_k e^{ikx} .$$

Then, for every j, using (3) and the fact that h is in $\mathscr{P}$,

$$\tilde{h}\left(\frac{i\pi}{n}\right) = \sum_{k=-n+1}^{n} \tilde{a}_k e^{ikj\pi/n} = \sum_{k=-n+1}^{n} \left(\frac{1}{2n} \sum_{l=-n+1}^{n} e^{-ikl\pi/n} h\left(\frac{l\pi}{n}\right)\right) e^{ikj\pi/n}$$
$$= \sum_{l=-n+1}^{n} h\left(\frac{l\pi}{n}\right)\left(\frac{1}{2n} \sum_{k=-n+1}^{n} e^{i(j-l)k\pi/n}\right) = h\left(\frac{j\pi}{n}\right).$$

If $h = \sum a_m e^{imx}$ and $\sum |a_m| < \infty$, then it follows from (3) and (4) that

$$(5) \qquad \tilde{a}_k = \sum_{l=-\infty}^{\infty} a_{k+2nl} .$$

Thus the $\tilde{a}_k$ are sometimes called the "folded back" coefficients of the Fourier expansion of h. *The superscript tilde added to any function $h \in \mathscr{P}$ will hereafter be used to denote the function $\tilde{h}$ obtained from h by* (4) (*and h may be replaced by f, g,* etc.). *Similarly, the tilde added to coefficients of a Fourier expansion denotes the coefficients defined by* (5).

Let

$$\mathscr{P}(x) = x^\nu + \gamma_1 x^{\nu-1} + \cdots + \gamma_\nu$$

be the characteristic polynomial of L and let

$$(6) \qquad Q(k) = \mathscr{P}(ik)\mathscr{P}(-ik) = |\mathscr{P}(ik)|^2$$

for integers k. When $L = d^\nu/dx^\nu$, we have $Q(k) = k^{2\nu}$. In any case, $Q(k)$ is a nonnegative polynomial in even powers of k with leading term $k^{2\nu}$, so $Q(k) \sim k^{2\nu}$ as $|k| \to \infty$. Now let

$$N_0 = \{k : Q(k) = 0\} ,$$

and, if N_0 is not empty, let n_0 be the largest integer in N_0. (For example, if $L = d^\nu/dx^\nu$, then $N_0 = \{0\}$ and $n_0 = 0$.) If N_0 is empty let $n_0 = 0$.

Now let $f(x) = \sum a_m e^{imx}$ and $g(x) = \sum b_m e^{imx}$ be two functions in $\mathscr{P}_0$. Then the usual inner product is

$$\langle f, g \rangle = \sum a_m \bar{b}_m .$$

If f and g are in $\mathscr{P}_\nu$, let

$$\langle f, g \rangle_L = \sum a_m \bar{b}_m Q(m) .$$

This quantity is of interest in our problem since, for any g in $\mathscr{P}_\nu$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}(Lg(x))^2\,dx = \langle g, g\rangle_L .$$

But $\langle \cdot, \cdot\rangle_L$ is not quite an inner product on $\mathscr{P}_\nu$ since $\langle f, f\rangle_L = 0$ does not imply $f = 0$ when $N_0 \neq \emptyset$.

Let

$$\mathscr{P}_\nu^{(0)} = \{f \in \mathscr{P}_\nu : f(x) = \textstyle\sum_{m\in N_0} a_m e^{imx}\}$$
$$\mathscr{P}_\nu^{(1)} = \{f \in \mathscr{P}_\nu : f(x) = \textstyle\sum_{m\notin N_0} a_m e^{imx}\} .$$

Then $\mathscr{P}_\nu^{(0)}$ and $\mathscr{P}_\nu^{(1)}$ provide an orthogonal decomposition of $\mathscr{P}_\nu$ under $\langle \cdot, \cdot\rangle$, and $\langle \cdot, \cdot\rangle_L$ is an inner product on $\mathscr{P}_\nu^{(1)}$. Moreover, $(\mathscr{P}^{(1)}, \langle\ ,\ \rangle_L)$ is a Hilbert space, and this space has the reproducing kernel

$$\xi(x, y) = \sum_{m\notin N_0}\frac{1}{Q(m)}\,e^{im(y-x)} ,$$

since $\xi(x, \cdot)$ is in $\mathscr{P}_\nu^{(1)}$ for each x, and for any $g \in \mathscr{P}^{(1)}$, $\langle g, \xi(x, \cdot)\rangle_L = g(x)$.

Let

$$\xi_k(y) = \xi\left(\frac{k\pi}{n}, y\right)$$

and let Ξ be the space spanned by the $2n$ functions $\{\xi_{-n+1}, \cdots, \xi_n\}$.

LEMMA 1. *The functions* $\{\xi_{-n+1}, \cdots, \xi_n\}$ *are linearly independent. Moreover, for every* $h \in \mathscr{P}$ *there is a unique* $\xi \in \Xi$ *such that* $\xi = h$ *on* $\mathscr{L}_{\pi/n}$.

PROOF. Let $\delta_k(x) = 1$ if $x = k\pi/n \pmod{2\pi}$ and 0 otherwise. Then there exists an $f_k \in \mathscr{P}_\nu^{(1)}$ such that $\tilde{f}_k = \tilde{\delta}_k$. To see this, let

$$f_k(x) = \textstyle\sum_{m\in N_0} a_{k,m}e^{imx}$$
$$\tilde{f}_k(x) = \textstyle\sum_{m=-n+1}^{n} \tilde{a}_{k,m}e^{imx} .$$

The $\tilde{a}_{k,m}$ must be chosen to match the expansion (4) of $\tilde{\delta}_k$. Clearly it is then possible to choose the $a_{k,m}$ to satisfy (5) and so that $a_{k,m} = 0$ if $|m| \leqq n_0$ or $|m| > n_0 + 2n$, and so $a_{k,-m} = \bar{a}_{k,m}$. But then f_k is in $\mathscr{P}_\nu^{(1)}$ as required.

Now let $\sum_{k=-n+1}^{n} a_k\xi_k = 0$. Then

$$0 = \langle f_j, \textstyle\sum_{k=-n+1}^{n} a_k\xi_k\rangle_L = \sum_{k=-n+1}^{n}\bar{a}_k\langle f_j, \xi_k\rangle_L = \bar{a}_j$$

for each j, and it follows that $\{\xi_{-n+1}, \cdots, \xi_n\}$ are linearly independent. Let

$$M = [\langle\xi_j, \xi_k\rangle_L;\ j, k = -n+1, \cdots, n]$$

be the matrix of inner products of the ξ_k's. Since the ξ_k's are linearly independent, M is nonsingular. Now, given $h \in \mathscr{P}$, let $\xi = \sum_{j=-n+1}^{n} a_j\xi_j$ satisfy

$$\xi\left(\frac{k\pi}{n}\right) = \sum_{j=-n+1}^{n} a_j\xi_j\left(\frac{k\pi}{n}\right) = h\left(\frac{k\pi}{n}\right)$$

in every k. This is equivalent to

$$\langle \xi, \xi_k \rangle_L = \sum_{j=-n+1}^{n} a_j \langle \xi_j, \xi_k \rangle_L = h\left(\frac{k\pi}{n}\right) \tag{7}$$

in $k = -n+1, \cdots, n$. Letting $\alpha = (a_{-n+1}, \cdots, a_n)$ and η be the vector whose kth component is $h(k\pi/n)$ for $k = -n+1, \cdots, n$, (7) becomes

$$\alpha M = \eta$$

and has the unique solution $\alpha = \eta M^{-1}$.

The last part of this lemma corresponds to Lemma 3.1 of Kimeldorf and Wahba [3]. From that paper we also obtain the next result (corresponding to their Lemma 3.2).

LEMMA 2. *Let M be the matrix $[\langle \xi_j, \xi_k \rangle_L]$ of inner products of the ξ_j for $i = -n+1, \cdots, n$. Then there is a unique element ξ_0 of Ξ which minimizes*

$$\frac{1}{n} \sum_{k=-n+1}^{n} \left(\xi\left(\frac{k\pi}{n}\right) - h\left(\frac{k\pi}{n}\right)\right)^2 + \frac{2}{\lambda^{2\nu}} \langle \xi, \xi \rangle_L$$

for all ξ in Ξ, and

$$\xi_0 = (\eta_{-n+1}, \cdots, \eta_n)\left(M + \frac{2n}{\lambda^{2\nu}} I\right)^{-1} (\xi_{-n+1}, \cdots, \xi_n), \tag{8}$$

where $\eta_k = h(k\pi/n)$ for $k = -n+1, \cdots, n$.

THEOREM 1. *The solution ξ_0 given by* (8) *is a* LSS *to h. For $n > n_0$ this solution is unique.*

PROOF. Let $g \in \mathscr{P}_\nu$ and let $\xi \in \Xi$ be such that $g = \xi$ on $\mathscr{L}_{\pi/n}$. Let $f = g - \xi$. Then

$$\langle f, \xi_k \rangle_L = g\left(\frac{k\pi}{n}\right) - \xi\left(\frac{k\pi}{n}\right) = 0$$

for every k, so $\langle f, \xi \rangle_L = 0$. Then

$$\Delta_{n,\lambda,L}(g, h) = \frac{1}{n} \sum_{k=-n+1}^{n} \left(\xi\left(\frac{k\pi}{n}\right) - h\left(\frac{k\pi}{n}\right)\right)^2 + \langle \xi, \xi \rangle_L + \langle f, f \rangle_L .$$

For g to minimize $\Delta_{n,\lambda,L}$, necessarily $\xi = \xi_0$ and $\langle f, f \rangle_L = 0$. In particular ξ_0 is a minimizing solution. For the uniqueness, note that f on $\mathscr{L}_{\pi/n}$ equals 0, hence $\tilde{f} = 0$. But $f \in \mathscr{P}_\nu^{(0)}$, and it follows by (5) that $f = 0$ if $n > n_0$.

While (8) gives an explicit solution for the LSS it can be obtained in a more convenient form by noting that the problem of finding a LSS g to h is invariant under shifts of the function g and h by $k\pi/n$ for an integer k. Thus the problem can be reduced to finding the spline which best fits the impulse function.

$$\delta(x) = 1 \quad \text{if} \quad x = 0$$
$$\quad = 0 \quad \text{if} \quad x \neq 0 .$$

Let $s_{n,\lambda}$ denote the LSS to δ for given values of n, λ and call it the *lattice impulse spline* (LIS).

Then, for any h in $\mathscr{P}$, the LSS is given by $h \circledast^n s_{n,\lambda}$ where $\circledast^n$ is the discrete convolution defined by

$$f_1 \circledast^n f_2(x) = \sum_{k=-n+1}^{n} f_1\left(\frac{k\pi}{n}\right) f_2\left(x - \frac{k\pi}{n}\right).$$

Note that this operation has the properties of a convolution on the lattice $\mathscr{L}_{\pi/n}$ but is not commutative for other values of x.

Now let

$$s_{n,\lambda}(x) = \sum_{m=-\infty}^{\infty} a_{n,\lambda,m} e^{imx}$$
$$\tilde{s}_{n,\lambda}(x) = \sum_{m=-n+1}^{n} \tilde{a}_{n,\lambda,m} e^{imx}$$

be the expressions (1) and (4) for $s_{n,\lambda}$. Actually $s_{n,\lambda}$ is an even function and can be expressed as a cosine series, but it is easier to handle in exponential form. Using (8), it is possible to obtain an explicit solution for $s_{n,\lambda}$, but it is easier to proceed directly. (We use (8) and Theorem 1 only to establish that the LSS is linear and unique for $n > n_0$.)

Substituting the expressions for $s_{n,\lambda}$ in the expression for $\Delta_{n,\lambda,L}$,

$$\begin{aligned} \Delta(s_{n,\lambda}, \delta) &= \frac{1}{n} \sum_{k=-n+1}^{n} \left(s_{n,\lambda}\left(\frac{k\pi}{n}\right)\right)^2 - \frac{2}{n} s_{n,\lambda}(0) + \frac{1}{n} \\ &\quad + \frac{2}{\lambda^{2\nu}} \langle s_{n,\lambda}, s_{n,\lambda} \rangle_L \\ &= 2 \sum_{k=-n+1}^{n} \tilde{a}^2_{n,\lambda,k^2} - \frac{2}{n} \sum_{k=-\infty}^{\infty} a_{n,\lambda,k} + \frac{1}{n} \\ &\quad + \frac{2}{\lambda^{2\nu}} \sum_{k=-\infty}^{\infty} a_{n,\lambda,k} Q(k) . \end{aligned} \tag{9}$$

The $a_{n,\lambda,k}$ must minimize this expression. Setting partial derivatives, for each $a_{n,\lambda,k}$, equal to 0 and solving, it is found that there is a unique solution for $n > n_0$. To describe this solution it is convenient to introduce the quantities

$$q_{n,j} = 1 + Q(j)\left(\sum_{k=-\infty}^{-1} 1/Q(j + 2nk) + \sum_{k=1}^{\infty} 1/Q(j + 2nk)\right) \tag{10}$$

defined for $n > n_0$ and $|j| \leqq n$. Since $Q(k)$ is a polynomial of order 2ν with leading term $k^{2\nu}$, it follows that there exist constants $0 < C_0 \leqq C_1 < \infty$ such that

$$\begin{aligned} Q(k) &\leqq C_1 k^{2\nu} \quad \text{for all } |k| \geqq 1 \\ Q(k) &\geqq C_0 k^{2\nu} \quad \text{for all } |k| > n_0 . \end{aligned} \tag{11}$$

Then, for $j \neq 0$,

$$1 \leqq q_{n,j} \leqq 1 + \frac{C_1 j^{2\nu}}{C_0 n^{2\nu}}\left(1 + \frac{1}{2^{2\nu}} + \frac{1}{3^{2\nu}} + \cdots\right) = 1 + \beta \frac{j^{2\nu}}{n^{2\nu}} \tag{12}$$

for some finite β. Similarly, for some finite β',

(13) $$1 \leqq q_{n,0} \leqq 1 + \frac{\beta'}{n^{2\nu}}\,.$$

Returning to the solution of (9) for the minimizing $a_{n,\lambda,k}$, straightforward calculations yield

$$\tilde{a}_{n,\lambda,j} = \frac{1}{2n}\frac{\lambda^{2\nu}q_{n,j}}{Q(j) + \lambda^{2\nu}q_{n,j}} \qquad \text{for} \quad -n+1 \leqq j \leqq n$$

$$a_{n,\lambda,j} = \frac{1}{2n}\frac{\lambda^{2\nu}}{Q(j) + \lambda^{2\nu}q_{n,j}} \qquad \text{for} \quad |j| \leqq n$$

$$a_{n,\lambda,k} = \frac{Q(j)}{Q(k)}a_{n,\lambda,j} \qquad \text{for} \quad k \in \{j \pm 2n, j \pm 4n, \cdots\} \quad \text{and} \quad |j| \leqq n\,.$$

The polynomial Q is even and $q_{n,j} = q_{n,-j}$, so the $a_{n,\lambda,k}$ are symmetric around 0 and $s_{n,\lambda}$ is even, as remarked above. Moreover, $a_{n,\lambda,k} = O(k^{-2\nu})$ as $|k| \to \infty$, hence $s_{n,\lambda} \in \mathscr{P}_{2\nu-1} \subset \mathscr{P}_\nu$ (recall $\nu \geqq 1$) and $s_{n,\lambda} \in C^{(2\nu-2)}$. In fact, $s_{n,\lambda}$ is the "interpolatory L spline" to the values of δ on $L_{\pi/n}$, since it minimizes the second term of $\Delta_{n,\lambda,L}$ among all functions in $\mathscr{P}_\nu$ coinciding with $s_{n,\lambda}$ on $\mathscr{L}_{\pi/n}$. It follows from the theory of linear differential equations that, when $L = d^\nu/dx^\nu$, $s_{n,\lambda}$ is a piecewise polynomial of degree $2\nu - 1$ between the successive lattice points.

As $n \to \infty$, the shape of the $s_{n,\lambda}$, suitably normalized, converges to a limit s_λ. Let

$$a_{\lambda,k} = \lim_n na_{n,\lambda,k} = \frac{\lambda^{2\nu}}{2(Q(k) + \lambda^{2\nu})}$$

$$s_\lambda(x) = \frac{1}{\pi}\sum_{k=-\infty}^{\infty} a_{\lambda,k}e^{ikx}\,.$$

The order of approximation of $ns_{n,\lambda}$ and its derivatives to πs_λ will be useful in the theory to follow.

First note that, since $q_{n,j} \geqq 1$ and $Q(k)q_{nj}/Q(j) \geqq 1$ for $k = j + 2nl$ and $l = \pm 1, \pm 2, \cdots$ and $|j| \leqq n$ by (10), we have

(14) $$na_{n,\lambda,k} \leqq a_{\lambda,k}$$

for all k, while, for $|k| \leqq n$

(15) $$a_{\lambda,k} \leqq n\tilde{a}_{n,\lambda,k}\,.$$

LEMMA 3. *For* $0 \leqq m \leqq 2\nu - 2$,

$$||\pi s_\lambda{}^{(m)} - ns_{n,\lambda}^{(m)}|| = O\left(\frac{\lambda^{2\nu}}{n^{2\nu-m-1}}\right)$$

as $n \to \infty$, *uniformly in* λ.

Moreover, if $\lambda = O(n^{1/4\nu})$, *then* $ns_{n,\lambda}^{(2\nu-1)}$ *converges to* $\pi s_\lambda{}^{(2\nu-1)}$ *in* $L_2(-\pi, \pi)$ *and* $L_1(-\pi, \pi)$.

PROOF. From (14) and (15) we have

$$\sum_{k=-\infty}^{\infty} a_{\lambda,k}|k|^m \geqq \sum_{k=-\infty}^{\infty} na_{n,\lambda,k}|k|^m \geqq \sum_{-n+1}^{n} na_{n,\lambda,k}|k|^m \geqq \sum_{-n+1}^{n} a_{\lambda,k}|k|^m$$

when $m = 0$ interpret $O^0 = 1$). But then

$$\begin{aligned} \|\pi s_\lambda^{(m)} - n s_{n,\lambda}^{(m)}\| &\leqq \sum_{k=-\infty}^{\infty} (a_{\lambda,k} - n a_{n,\lambda,k})|k|^m \\ &\leqq a_{\lambda,n} n^m + \sum_{|k|>n} a_{\lambda,k}|k|^m \sim \int_n^\infty \frac{\lambda^{2\nu} x^m}{x^{2\nu} + \lambda^{2\nu}}\, ds \\ &= \lambda^{m+1} \int_{n/\lambda}^\infty \frac{y^m}{y^{2\nu} + 1}\, dy = O\left(\frac{\lambda^{2\nu}}{n^{2\nu-m-1}}\right). \end{aligned}$$

To prove the second assertion, observe that, by (12), for $1 \leqq |k| \leqq n$ and $n \geqq n_0$,

$$\begin{aligned} a_{\lambda,k} - n a_{n,\lambda,k} &\leqq \frac{1}{2}\left\{\frac{\lambda^{2\nu}}{Q(k) + \lambda^{2\nu}} - \frac{\lambda^{2\nu}}{Q(k) + \lambda^{2\nu}(1 + \beta k^{2\nu}/n^{2\nu})}\right\} \\ &\leqq \frac{\lambda^{4\nu}\beta k^{2\nu}}{2(Q(k) + \lambda^{2\nu})^2 n^{2\nu}} \leqq \frac{\beta}{2C_0} \frac{\lambda^{2\nu}}{n^{2\nu}} \end{aligned}$$

so

$$\sum_{k=-n}^{n} (a_{\lambda,k} - n a_{n,\lambda,k})^2 k^{4\nu-2} \leqq \left(\frac{\beta}{2C_0}\right) \frac{\lambda^{4\nu}}{n^{4\nu}} \sum_{|k|\leqq n} k^{4\nu-2} = O\left(\frac{\lambda^{4\nu}}{n}\right).$$

The convergence in $L_2(-\pi, \pi)$ follows from this and

$$\begin{aligned} \sum_{|k|>n}^{\infty} (n a_{n,\lambda,k} k^{2\nu-1})^2 &\leqq \sum_{|k|>n}^{\infty} (a_{\lambda,k} k^{2\nu-1})^2 \sim \int_n^\infty \left(\frac{\gamma^{2\nu} x^{2\nu-1}}{x^{2\nu} + \lambda^{2\nu}}\right) dx \\ &= \lambda^{4\nu-1} \int_{n/\lambda}^\infty \left(\frac{y^{2\nu-1}}{y^{2\nu} + 1}\right) dy = O\left(\frac{\lambda^{4\nu}}{n}\right), \end{aligned}$$

and the convergence in $L_1(-\pi, \pi)$ is an immediate consequence of this.

This lemma shows that πs_λ may be used as an approximation to $n s_{n,\lambda}$ when $\lambda \ll n$; however, its most important implication is the result in the next section.

3. Continuous smoothing splines. If h is known for all values of x, it is natural to define a smoothing spline to h as a function g minimizing

$$\Delta_{\lambda,L}(g, h) = \frac{1}{\pi} \int_{-\pi}^{\pi} (g(x) - h(x))^2\, dx + \frac{1}{\lambda^{2\nu}\pi} \int_{-\pi}^{\pi} (Lg(x))^2\, dx .$$

We call such a $g \in \mathscr{P}_\nu$ a *periodic continuous smoothing spline* (CSS) to h. It is completely specified by stating the value of λ and the operator L.

Provided $h \in L_2(-\pi, \pi)$, $\Delta_{\lambda,L}(g, h)$ is finite for all $g \in \mathscr{P}_\nu$ and is strictly convex in g, so at most one minimizing solution exists. The solution is again a convolution.

THEOREM 2. *Let $h \in L_2(-\pi, \pi)$, then the convolution defined by*

$$h \circledast s_\lambda(x) = \int_{-\pi}^{\pi} h(y) s_\lambda(x - y)\, dy$$

minimizes $\Delta_{\lambda,L}(\cdot, h)$ and is the CSS *to h.*

PROOF. First assume h is continuous on $[-\pi, \pi]$, hence bounded. The assertion of the theorem then results from the following relations valid for any $g \in \mathscr{P}_\nu$:

$$\begin{aligned}\Delta_{\lambda,L}(g, h) &= \lim_n \Delta_{n,\lambda,L}(g, h) \geqq \lim_n \Delta_{n,\lambda,L}(h \circledast^n s_{n,\lambda}, h) \\ &= \lim_n \Delta_{n,\lambda,L}\left(\frac{\pi}{n} h \circledast^n s_\lambda, h\right) = \lim_n \Delta_{n,\lambda,L}(h \circledast s_{n,\lambda}, h) \\ &= \Delta_{\lambda,L}(h \circledast s_{n,\lambda}, h) .\end{aligned}$$

The first and last equalities result from the definition of the Riemann integral since g, h, $h \circledast s_\lambda$ are continuous. The inequality follows from the definition of the LSS. The next equality follows from Lemma 3 since, using the triangle equality in the definition of $\circledast^n$,

$$\left\| h \circledast^n s_{n,\lambda} - \frac{\pi}{n} h \circledast^n s_\lambda \right\| \leqq 2\|h\| \cdot \|ns_{n,\lambda} - \pi s_\lambda\| = O\left(\frac{1}{n^{2\nu-1}}\right)$$

as $n \to \infty$ (note λ is fixed), and, for $\nu > 1$,

$$\left\| L(h \circledast^n s_{n,\lambda}) - L\left(\frac{\pi}{n} h \circledast s_\lambda\right) \right\| \leqq 2\|h\| \cdot \|nLs_{n,\lambda} - \pi LS_\lambda\| = O\left(\frac{1}{n^{\nu-1}}\right),$$

while for $\nu = 1$ we have $L(h \circledast^n s_{n,\lambda}) \to L(\pi h \circledast^n s_\lambda/n)$ in $L_2(-\pi, \pi)$. To establish the penultimate equality observe first that $h(y)s_\lambda(x - y)$ is uniformly continuous in x and y, so, as $n \to \infty$,

$$\left\| \frac{\pi}{n} h \circledast^n s_\lambda - h \circledast s_\lambda \right\| \to 0 .$$

Similarly, for $\nu > 1$,

$$\left\| L\left(\frac{\pi}{n} h \circledast^n s_\lambda\right) - L(h \circledast s_\lambda) \right\| = \left\| \frac{\pi}{n} h \circledast^n Ls_\lambda - h \circledast Ls_\lambda \right\| \to 0 .$$

For $\nu = 1$, it follows from Theorems 1, 3, 4 of Chapter 4, Section 3, of Tolstov [7] that Ls_λ is continuous except at $0, \pm 2\pi, \cdots$ and is in $L_1(-\pi, \pi)$, and it follows that the convergence of $L(\pi h \circledast s_\lambda/n)$ to $L(h \circledast s_\lambda)$ holds pointwise. Since s_λ is also in $L_2(-\pi, \pi)$ and

$$\left\| \frac{\pi}{n} h \circledast^n s_\lambda \right\|_{L_2} \leqq 2\pi\|h\| \cdot \|s_\lambda\|_{L_2} ,$$

where $\|\cdot\|_{L_2}$ denotes the L_2 norm, we have that $L(\pi h \circledast^n s_\lambda n) \to L(h \circledast s_\lambda)$ in $L_2(-\pi, \pi)$.

This proves the theorem for h continuous. For any $h \in L_2(-\pi, \pi)$ there exist continuous h_n corresponding to h in L_2. Then $h_n \circledast s_\lambda \to h \circledast s_\lambda$ and

$$L(h_n \circledast s_\lambda) = h_n \circledast Ls_\lambda \to h \circledast Ls_\lambda = L(h \circledast s_\lambda)$$

in L_2 since s_λ and Ls_λ are in L_1, and it follows that

$$\Delta_\lambda(g, h) \geqq \lim_n \Delta_\lambda(g, h_n) \geqq \lim_n \Delta_\lambda(h_n \circledast s_\lambda, h_n) = \Delta_\lambda(h \circledast s_\lambda, h)$$

for all $g \in \mathscr{P}_\nu$, and the proof is completed.

On the basis of this result we call s_λ a *continuous impulse spline* (CIS).

It is interesting to note that, as $\lambda \to \infty$, the shape of s_λ will depend on L only through its order ν. When $L = d^\nu/dx^\nu$,

$$s_\lambda(x) = \frac{1}{2\pi} \sum \frac{\lambda^{2\nu}}{k^{2\nu} + \lambda^{2\nu}} e^{ikx} .$$

We will denote this spline τ_λ in the following discussion. The difference between τ_λ and the continuous impulse spline for any L of order ν is uniformly bounded by

$$\frac{1}{2\pi} \sum \left| \frac{\lambda^{2\nu}}{k^{2\nu} + \lambda^{2\nu}} - \frac{\lambda^{2\nu}}{Q(k) + \lambda^{2\nu}} \right| = \frac{1}{2\pi} \sum \frac{\lambda^{2\nu} |k^{2\nu} - Q(k)|}{(k^{2\nu} + \lambda^{2\nu})(Q(k) + \lambda^{2\nu})} .$$

But $k^{2\nu} - Q(k)$ is a polynomial of order $2\nu - 2$ or lower, so the above difference is of order

$$\int \frac{\lambda^{2\nu} x^{2\nu-2}}{(x^{2\nu} + \lambda^{2\nu})^2} dx = O\left(\frac{1}{\lambda}\right) .$$

The result of Kimeldorf and Wahba [3] suggests that the general form of L may be of interest in some applications, but this requires quite specific information in order to choose L. Lacking this, the choice $L = d^\nu/dx^\nu$ is natural, and the shape of τ_λ will yield further insight into the nature of our splines.

Let

$$t(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{1 + y^{2\nu}} e^{iyx} dy ,$$

and

$$t_\lambda(x) = \lambda t(\lambda x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\lambda^{2\nu}}{\lambda^{2\nu} + y^{2\nu}} e^{iyx} dy$$

$$\hat{t}_\lambda(x) = \sum_{k=-\infty}^{\infty} t_\lambda(x + 2k\pi) .$$

The "folded back" function $\hat{t}_\lambda$ has period 2π, and for any function h of period 2π and in $L_1(-\pi, \pi)$, clearly

$$\int_{-\pi}^{\pi} h(x)\hat{t}_\lambda(x)\, dx = \int_{-\infty}^{\infty} h(x) t_\lambda(x)\, dx .$$

Taking $h(x) = e^{ikx}/2\pi$, we see that the Fourier coefficients of the expansion of $\hat{t}_\lambda$ are piecewise $\lambda^{2\nu}/(2\pi)(\lambda^{2\nu} + k^{2\nu})$. Thus $\hat{t}_\lambda = \tau_\lambda$. Moreover, $h \circledast \tau_\lambda$ becomes the convolution on the real line, $h * t_\lambda$. That is,

$$h \circledast \tau_\lambda(x) = h \circledast \hat{t}_\lambda(x) = \int_{-\infty}^{\infty} h(y) t_\lambda(x - y)\, dy .$$

Thus the CSS τ_ν of order ν is the real line convolution of h with a kernel t_λ, which is obtained from the kernel t by a change of scale $x \to \lambda x$.

For $\nu = 1$, $t(x) = e^{-|x|}/2$, while for $\nu > 1$

$$t(x) = \frac{1}{\nu} \sum_{\substack{1 \leq k < \nu - 1 \\ k \text{ odd}}} \exp\left[-|x| \sin \frac{k\pi}{2\nu} \right] \left\{ \sin \frac{k\pi}{2\nu} \cos\left(x \cos \frac{k\pi}{2\nu} \right) + \cos \frac{k\pi}{2\nu} \sin\left(x \cos \frac{k\pi}{2\nu} \right) \right\} + \chi_{\text{odd}}(\nu) \frac{1}{2\nu} e^{-|x|}$$

where $\chi_{\text{odd}}(\nu) = 1$ or 0 according as ν is odd or even. As $\nu \to \infty$, $(1 + y^{2\nu})^{-1}$ converges to 1 for $|y| < 1$ and 0 for $|y| > 1$ and the t kernels converge to the Dirichlet kernel $\sin x/\pi x$. For each fixed ν $t(x) = O[\exp(-|x| \sin(\pi/2\nu))]$ as $x \to \infty$, and it follows that

$$||s_\lambda - t_\lambda|| = ||\hat{t}_\lambda - t_\lambda|| = O[\lambda \exp(-\lambda\pi \sin(\pi/2\nu))] \quad \text{as} \quad \lambda \to \infty .$$

4. Saturation. Let g be a LSS or CSS to a function h. The question is, as $\lambda \to \infty$, how well does g approximate h? In the application to spectral estimation (Section 5) the difference between the LSS or CSS to h and h itself measures the bias of the estimating spline.

Saturation is a measure of the order of approximation of the spline g to the function h, expressed as a power of $1/\lambda$, as $\lambda \to \infty$ for a given value of ν. (Recall that ν is the order of the linear differential operator L used to measure smoothness of the spline.)

We begin with the most clear-cut result. The notion of saturation of a kernel t is given in Shapiro [6] and may be defined as follows: let $\int t(x)\,dx = 1$ and $\int |t(x)x^m|\,dx < \infty$. Let $t_\lambda(x) = \lambda t(\lambda x)$. If for every $h \in C_\infty^{(m)}$ (that is, h having a bounded, continuous mth derivative on the entire real line), $h * t_\lambda = h + O(\lambda^{-m})$, then t has *saturation* m. The following is Shapiro's Theorem 2:

THEOREM 3. *Let* $\int t(x)\,dx = 1$ *and* $\int |t(x)x^m|\,dx < \infty$. *Then* t *has saturation* m iff $\int t(x)x^k\,dx = 0$ *for* $k = 1, \cdots, m-1$ *and* $\int t(x)x^m\,dx = c \neq 0$. *Moreover, for any* $h \in C_\infty^{(m)}$

$$h * t_\lambda(x) = h(x) + \frac{c}{\lambda^m}\frac{h^{(m)}(x)}{m!} + o\left(\frac{1}{\lambda^m}\right).$$

We have seen that, when $L = d^\nu/dx^\nu$, $h \circledast \tau_\lambda = h * t_\lambda$ where t_λ is defined as in Section 3. Moreover, t has moments of all orders and by a standard theorem of Fourier Transforms,

$$\int t(x)x^k\,dx = (-i)^k \frac{d^k}{dy^k}\frac{1}{1 + y^{2\nu}}\bigg|_{y=0}.$$

Elementary calculations show that $\int t(x)x^k\,dx$ equals 0 except for $k = 0, 2\nu, 4\nu, \cdots$, and that $\int t(x)x^{2\nu}\,dx = (-1)^{\nu+1}(2\nu)!$ From this, Shapiro's theorem and an argument like that given by Shapiro, the following theorem follows by a straightforward argument.

THEOREM 4. *Let* $L = d^\nu/dx^\nu$. *Then for* $h \in C^{(2\nu)}$

$$(h \circledast s_\lambda)(x) = h(x) + (-1)^{\nu+1}\frac{h^{(2\nu)}}{\lambda^{2\nu}}(x) + O\left(\frac{1}{\lambda^{4\nu}}\right).$$

For other choices of L and for the LSS's similar results hold, based on Fourier analysis, but they are not quite as sharp. We will detail these results now.

First let $h_k(x) = e^{ikx}$. We then have that

$$h_j \circledast^n s_{n,\lambda} = \textstyle\sum_{k=-\infty}^{\infty} 2na_{n,\lambda,j+2nk} h_{j+2nk} \tag{16}$$

$$h_j \circledast s_\lambda = 2a_{\lambda,j} h_j . \tag{17}$$

In the next theorem we establish an approximation of order $\lambda^{-2\nu}$ when $h = \sum b_j h_j$ where $\sum |b_j| j^{2\nu} < \infty$. This condition implies that h has 2ν continuous derivatives and is periodic 2π, of course. Conversely, if h has a 2νth derivative satisfying any one of the following three conditions,

(1) $h^{(2\nu)}$ is of bounded variation and Lipschitz (α) for any $\alpha > 0$,

(2) $h^{(2\nu)}$ is Lipschitz (α) for any $\alpha > \frac{1}{2}$,

(3) $h^{(2\nu)}$ is absolutely continuous and $h^{(2\nu+1)} \in L_2$,

then $\sum |b_j| j^{2\nu} < \infty$ (see page 32 in [2]).

THEOREM 5. *Let $h(x) = \sum b_k e^{ikx}$. Then there exists a universal constant C such that*

$$||h - h \circledast^n s_{n,\lambda}|| \leqq C(|b_0| + \textstyle\sum_{|k|\geqq 1} |b_k| k^{2\nu}) \left(\dfrac{1}{\lambda^{2\nu}}\right).$$

Moreover,

$$||h - h \circledast s_\lambda|| \leqq (Q(0)|b_0| + C_1 \textstyle\sum_{|k|\geqq 1} |b_k| k^{2\nu}) \dfrac{1}{\lambda^{2\nu}}$$

where C_1 is the constant in (11).

PROOF. The second inequality follows from (17) since

$$||h - h \circledast s_\lambda|| \leqq (Q(0)|b_0| + C_1 \textstyle\sum_{|k|\geqq 1} |b_k| k^{2\nu}) \dfrac{1}{\lambda^{2\nu}}$$

and $Q(k) \leqq C_1 k^{2\nu}$ for $k \neq 0$.

To establish the first equality observe first that, for $1 \leqq |j| \leqq n$,

$$\begin{aligned} ||h_j - h_j \circledast^n s_{n,\lambda}|| &\leqq 1 - 2na_{n,\lambda,j} + \textstyle\sum_{|k|\geqq 1} 2na_{n,\lambda,j+2nk} \\ &\leqq C_2 j^{2\nu} \left(\frac{1}{\lambda^{2\nu}} + \frac{1}{n^{2\nu}}\right) \end{aligned}$$

for some finite C_2 since

$$\begin{aligned} 1 - na_{n,\lambda,j} &= \frac{Q(j) + \lambda^{2\nu}(q_{n,j} - 1)}{Q(j) + \lambda^{2\nu} q_{n,j}} \leqq \frac{C_1 j^{2\nu} + \lambda^{2\nu} \beta j^{2\nu} n^{-2\nu}}{\lambda^{2\nu}} \\ &\leqq (C_1 + \beta) j^{2\nu} \left(\frac{1}{\lambda^{2\nu}} + \frac{1}{n^{2\nu}}\right) \end{aligned}$$

and for some finite C_3

$$\begin{aligned} \textstyle\sum_{|k|\geqq 1} 2na_{n,\lambda,j+2nk} &\leqq \textstyle\sum_{|k|\geqq 1} \dfrac{Q(j)}{Q(j + 2nk)} \leqq \dfrac{C_1}{C_0} \dfrac{j^{2\nu}}{n^{2\nu}} \\ &\times \left(1 + \frac{1}{2^{2\nu}} + \frac{1}{3^{2\nu}} + \cdots\right) = C_3 \frac{j^{2\nu}}{n^{2\nu}}. \end{aligned}$$

Similarly,

$$\begin{aligned} ||h_0 - h_0 \circledast^n s_{n,\lambda}|| &\leqq 1 - 2na_{n,\lambda,0} + \textstyle\sum_{|k|\geqq 1} 2na_{n,\lambda,2nk} \\ &\leqq \frac{Q(0)}{\lambda^{2\nu}} + \frac{\beta'}{n^{2\nu}} + \frac{C_1 Q(0)}{n^{2\nu}} \left(1 + \frac{1}{2^{2\nu}} + \frac{1}{3^{3\nu}} + \cdots\right) \\ &\leqq C_4 \left(\frac{1}{\lambda^{2\nu}} + \frac{1}{n^{2\nu}}\right). \end{aligned}$$

Finally, for $|j| > n$,

$$||h_j - h_j \circledast^n s_{n,\lambda}|| \leqq 1 + \sum_k 2na_{n,\lambda,j+2nk} = C_5 .$$

Combining these estimates and using the inequality

$$||h - h \circledast^n s_{n,\lambda}|| \leqq \sum |b_j| \, ||h_j - h_j \circledast^n s_{n,\lambda}||$$

completes the proof.

COROLLARY 1. *If* $h(x) = \sum_{j=-\infty}^{\infty} b_j e^{ijx}$ *and* $\sum |j^r b_j| < \infty$, *where* $0 < r \leqq 2\nu$, *then, as* $\lambda \to \infty$

(18) $$||h \circledast s_\lambda - h|| = O\left(\frac{1}{\lambda^r}\right)$$

while, for $\lambda = O(n)$ *as* $n, \lambda \to \infty$,

(19) $$||h \circledast s_{n,\lambda} - h|| = O\left(\frac{1}{\lambda^r}\right) .$$

PROOF. We have

$$||h - \sum_{|j|<\lambda} b_j h_j|| \leqq \sum_{|j|\geqq\lambda} |b_j| \leqq \frac{1}{\lambda^r} \sum_{|j|\geqq\lambda} |j^r b_j| = O\left(\frac{1}{\lambda^r}\right)$$

$$||h \circledast s_\lambda - (\sum_{|j|<\lambda} b_j h_j) \circledast s_\lambda|| \leqq \sum_{|j|\geqq\lambda} |b_j| 2a_{\lambda,j} = O\left(\frac{1}{\lambda^r}\right)$$

$$||h \circledast^n s_{n,\lambda} - (\sum_{|j|<\lambda} b_j h_j) \circledast^n s_{n,\lambda}|| \leqq 2 \sum_{|j|\geqq\lambda} |b_j| = O\left(\frac{1}{\lambda^r}\right) .$$

The corollary then follows from the estimates in the theorem applied to $\sum_{|j|<\lambda} b_j h_j$ since

$$\frac{1}{\lambda^{2\nu}} \sum_{|j|<\lambda} |b_j| j^{2\nu} \leqq \frac{\lambda^{2\nu-r}}{\lambda^{2\nu}} \sum_{|j|<\lambda} |j^r b_j| = O\left(\frac{1}{\lambda^r}\right) .$$

The results of this section show that, for sufficiently smooth functions h, the order of approximation will depend on the choice of ν. Furthermore, the larger ν the better the order. This suggests choosing ν large, but then the impulse splines will resemble the Dirichlet kernel which is known to have bad properties as a smoothing kernel. The situation is paradoxical. A better understanding will be gained by observing that the error of the approximation at x will be asymptotically $h^{(2\nu)}(x)/\lambda^{2\nu}$, but even this value may apply only for extremely large λ. An upper bound on the error that does not depend on x is $||h^{(2\nu)}||/\lambda^{2\nu}$. Typically $||h^{(2\nu)}||$ will increase very rapidly (if it remains finite at all) with ν. This suggests that, even for large λ, ν should not be too large if h has rugged features. In the statistical problem h is unknown and may be relatively smooth or rugged. In this situation the appropriate procedure seems to be to fit several splines to h with different ν and λ values. This will be discussed further in Section 5 for the problem of spectral estimation.

5. Estimation of the spectral density of a second order stationary random sequence. Let $X_1, X_2, \cdots$ be a second order stationary random sequence with $EX_k = 0$, $EX_j X_{j+k} = \rho_k$. Then the ρ_k's are Fourier coefficients of a symmetric distribution function on $[-\pi, \pi]$:

$$\rho_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\omega}\, dF(\omega) = \frac{1}{\pi} \int_0^{\pi} \cos k\omega\, dF(\omega)\,.$$

(See [1].) When F is absolutely continuous, it is completely determined by the *spectral density*, $f(\omega) = dF(\omega)/d\omega$.

$$f(\omega) = \sum_{-\infty}^{\infty} \rho_k e^{ik\omega}\,.$$

The statistical problem is to estimate $f(\omega)$ on the basis of observations $X_1, \cdots, X_n$. Let $\hat{f}(\omega)$ denote the periodogram:

$$\hat{f}(\omega) = \sum_{k=-n+1}^{n-1} \hat{\rho}_k e^{ik\omega} = \hat{\rho}_0 + 2 \sum_{k=1}^{n-1} \hat{\rho}_k \cos k\omega$$

where

$$\hat{\rho}_k = \hat{\rho}_{-k} = \frac{1}{n} \sum_{j=1}^{n-k} X_j X_{j+k}\,, \qquad k = 0, \cdots, n-1\,.$$

Under very general conditions (see Walker [8]),

$$\hat{f}(\omega) = f(\omega) I_\varepsilon(\omega) + \eta_n(\omega)$$

where $\|\eta_n\| \to 0$ in probability as $n \to \infty$ and $I_\varepsilon(j\pi/n)$, $j = 0, \cdots, (n-1)/2$, are uncorrelated and have a mean and variance of 1. Also, if the process is Gaussian, then $I_\varepsilon(j\pi/n)$ will be independent and have exponential distributions (see Grenander and Rosenblatt [1].) Since the periodogram is an inconsistent estimator of f, some modification is required. "Smoothed" (consistent) estimators of $f(\omega)$ are obtained either by smoothing the periodogram

$$f^*(\omega) = \int_{-\pi}^{\pi} \hat{f}(\lambda) K(\omega - \lambda)\, d\lambda$$

or by the equivalent method of weighting the covariances by a "lag window" $k_M(r)$, $r = 0, 1, \cdots, M$,

$$f^*(\omega) = \frac{1}{2\pi} \sum_{r=-M}^{M} k_M(r) e^{-ir\omega} \hat{\rho}_r$$

where

$$K(\omega) = \frac{1}{2\pi} \sum_{r=-M}^{M} e^{-ir\omega} k_M(r)\,.$$

While the advent of the Fast Fourier Transform (FFT) has made the former method more popular, the advantages of the latter method, also using the FFT, have recently been noted (Parzen [5]) and will be discussed further in the next section.

We may consider estimating f by a LSS or CSS to $\hat{f}$: $f \circledast^n s_{n,\lambda}$ or $\hat{f} \circledast \tau_\lambda = \hat{f} * t_\lambda$. We will consider only $L = d^\nu/dx^\nu$ for various ν.

Based on the properties of the periodogram stated above, it follows that as

$n, \lambda \to \infty$ with $\lambda = o(n)$, the bias of either estimator is asymptotically $f^{(2\nu)}/\lambda^{2\nu}$. Similarly, the variance is asymptotically $f^2\lambda\sigma_\nu^2/n$, where

$$\sigma_\nu^2 = \int_{-\infty}^{\infty} \frac{1}{(1+y^{2\nu})^2}\,dy\,,$$

for either estimator. (See [1].) This gives asymptotic mean squared error (MSE) of

$$f^2 \frac{\lambda}{n} \sigma_\nu^2 + \frac{(f^{(2\nu)})^2}{\lambda^{4\nu}}\,.$$

For a fixed ν, the standard deviation and bias are of the same order, when $\lambda = O(n^{1/(4\nu+1)})$, and the MSE is then $O(n^{-4\nu/(4\nu+1)})$.

This asymptotic expression may be valid only for extremely large values of n, but the corresponding estimate of integrated MSE

$$\int_{-\pi}^{\pi} \text{MSE}\,(\hat{f}(x))\,dx \cong \frac{\lambda}{n}\sigma_\nu^2 \int_{-\pi}^{\pi} f^2(x)\,dx + \frac{1}{\lambda^{4\nu}} \int_{-\pi}^{\pi} (f^{(2\nu)}(x))^2\,dx$$

should apply for much smaller n. The value of λ minimizing the right-hand side is

$$\lambda_0 = \left(\frac{4n\nu}{\sigma_\nu^2} \cdot \int_{-\pi}^{\pi} (f^{(2\nu)}(x)^2\,dx \Big/ \int_{-\pi}^{\pi} f^2(x)\,dx\right)^{1/(4\nu+1)}$$

and the resulting MSE is

$$\frac{\lambda_0}{n}\sigma_\nu^2\left(\frac{4\nu+1}{4\nu}\right)\int_{-\pi}^{\pi} f^2(x)\,dx\,.$$

While this result may be of some interest in comparing estimators, it can be very misleading. The values of λ and ν minimizing this expression for integrated MSE, even if they were known or could be estimated, might lead to unsatisfactory estimators of f. Thus if f has one or several sharp peaks but is predominantly smooth then an estimator minimizing integrated MSE may smooth the periodogram so much that these peaks fail to appear in the estimate. From the standpoint of spectral analysis this would be a serious failing.

In practice it is desirable to compute several estimators from the data, corresponding to different λ and ν values. One advantage of the spline function approach is that it provides a highly flexible family of estimators that can be built into a single program with the coefficients λ and ν to be specified.

Empirical studies suggest that the best resolution of extremely sharp peaks (on the order of 5 to 10 times π/n wide) is obtained when $\nu = 2$ and λ is about $n/20$. Such an estimator will contain much spurious oscillation. For the estimation of less rugged features of f, values of $\nu = 3$ and $\lambda = n/100$ or $\nu = 4$ and $\lambda = n/1000$ may be tried. An important consideration is that the width of the smoothing function s_λ (roughly $9/\lambda$ between the first minima—see Table 1) should be smaller than the width of the sharpest peak of f; otherwise the familiar side lobe phenomena will appear in the estimate.

TABLE 1
Zeros, minima and maxima of $t(x)$ and values of σ_ν^2

			1st minimum			1st maxima			2nd minimum		
ν	$\pi t(0)$	1st zero	x	$\pi t(x)$	2nd zero	x	$\pi t(x)$	3rd zero	x	$\pi t(x)$	σ_ν^2
2	1.111	3.332	4.443	−.048	7.775	8.886	.002	—	—	—	1.666
3	1.047	3.143	4.300	−.102	6.628	7.848	.018	10.282	11.489	−.003	1.746
4	1.026	3.116	4.318	−.138	6.392	7.660	.038	9.771	11.049	−.011	1.796
5	1.017	3.114	4.352	−.161	6.314	7.618	.056	9.583	10.903	−.020	1.830
10	1.004	3.130	4.444	−.201	6.269	7.656	.102	9.420	10.833	−.059	1.908
20	1.001	3.138	4.480	−.213	6.278	7.703	.121	9.418	10.876	−.081	1.942
∞	1.000	3.1416	4.493	−.217	6.283	7.725	.128	9.425	10.904	−.091	2.000

6. Numerical algorithms. Two alternative methods of spectral estimation were discussed in the last section: smoothing the periodogram and weighting the covariances. Both approaches, using periodic splines, are discussed in this section.

The first method of smoothing the periodogram is given by

$$f^*(\omega) = \int_{-\pi}^{\pi} \hat{f}(x) K(\omega - x)\, dx\,,$$

for a given smoothing kernel $K(\cdot)$. Since the periodogram is typically computed on a lattice $\{\hat{f}(r\pi/n),\ -n < r \leqq n\}$, the actual spectral estimator used is the Riemann approximation

$$f_1^*\left(\frac{r\pi}{n}\right) = \sum_{k=-n+1}^{n} \hat{f}\left(\frac{k\pi}{n}\right) K\left(\frac{r-k}{n}\pi\right) = \hat{f}_1 \circledast^n K\left(\frac{r\pi}{n}\right)$$

where $r = 0, 1, \cdots, n-1$ and f_1^* is periodic (2π) and symmetric about the origin.

Three asymptotically equivalent smoothing kernels have been presented in this paper: The Lattice Impulse Spline (LIS) $s_{n,\lambda}(\cdot)$, the Continuous Impulse Spline (CIS) $s_\lambda(\cdot)$, and the function $t(\cdot)$. The first two may be evaluated on the desired lattice $\{r\pi/n,\ -n < r \leqq n\}$ by using the Fast Fourier Transform (FFT) on either $a_{n,\lambda,k}$ (for $s_{n,\lambda}(\cdot)$) or $a_{\lambda,k}$ (for $s_\lambda(\cdot)$). Since both kernels are symmetric, a real cosine transformation is needed, and can be programmed quite efficiently.

The above two kernels require order $n \log n$ operations, while the third window, $t(\cdot)$, requires only order n operations in practice, and hence is asymptotically superior from the point of view of computation time. This estimator is given by

$$f_\lambda^*\left(\frac{r\pi}{n}\right) = \lambda \sum_{k=-n+1}^{n} \hat{f}\left(\frac{k\pi}{n}\right) t\left(\lambda\pi\frac{r-k}{n}\right).$$

In practice, it has been found that $t(\cdot)$ needs to be computed from the origin to the second zero (see Table 1) for $\nu = 2$ and 3, and to the third zero for $\nu = 4$ and 5. If $\nu = 2$, for example, then $t(\lambda r\pi/n)$ is needed for $r = 0, 1, \cdots, v$ where $v = [7.775n/\pi]$, hence an order of n operations.

The alternative approach of covariance weighting is given by

$$f_2^*(\omega) = \frac{1}{2\pi} \sum_{r=-n}^{n} k_\lambda(r) e^{-ir\omega} \hat{\rho}_r \, .$$

The principal difference between spline smoothing and the classical smoothing methods is that rather than smooth by truncation, the smoothing is done by shaping the "lag window". If one uses the Fast Fourier Transform, then it is necessary even with "truncated" windows to augment the array with zeros; hence there is no longer a real computational advantage to truncated windows.

The weighting coefficients $k_\lambda(\cdot)$ are given for the Lattice Smoothing Spline and the Continuous Smoothing Spline as the Fourier Coefficients $2a_{n,\lambda,r}$ of the LIS and $2a_{\lambda,r}$ of the CIS respectively. These coefficients are given in Sections 2 and 3. The two splines are asymptotically equivalent, and for samples of $n =$ 500 were found to give highly similar results. So in numerical studies, the computationally simpler

$$a_{\lambda,k} = \frac{\lambda^{2\nu}}{2(Q(k) + \lambda^{2\nu})}$$

was used. Also, since $Q(k)$ is asymptotically dominated by the leading term (which is equivalent to $L = D^\nu$),

$$k_\lambda(r) = 2a_{\lambda,k} = 1/(1 + (r/\lambda)^{2\nu})$$

was used.

The advent of the FFT has made smoothing the periodogram quite popular. However, recent results (e.g., see Parzen [5]) are noting that by using the FFT in covariance weighting, there is no clearcut advantage to smoothing the periodogram. For small n, the method of smoothing the periodogram is preferable but asymptotically requires n^2 operations as opposed to $n \log n$ for covariance weighting. The deciding factor in the authors' numerical studies was the ease with which a general program could be written using the covariance weighting technique.

The data for the numerical study was generated as a first order autoregression ($\alpha = .5$) with a Gaussian innovation process and with two moderate peaks superimposed in the frequency near $.4\pi$ and $.8\pi$ radians. A sample of size 500 was used.

While several spectral estimators were studied, we present the results of only two—the Parzen spectral estimator, and the CIS estimator with $\nu = 3$. These two estimators were selected as they provide the greatest contrast, the Parzen window with saturation of order 2 and the spline with saturation of order 6 ($\nu = 3$). From this comparison, a better understanding of saturation can be obtained.

The Parzen spectral estimate was made with $M = 35$ lags to have a variance of .377 and a bandwidth of .23. For comparison, the CIS estimate used $\lambda =$ 10.75 ($\nu = 3$) to have a variance of .376 and a bandwidth of .28.

Even though the two estimators have comparable statistical properties (i.e. the same variance and similar bandwidth), the spline estimate is seen to be much smoother (a higher correlation between adjacent frequencies). The spline gives a better picture of the autoregressive part where infinite saturation is perhaps appropriate. However, the higher saturation tends to lose resolution about the peaks at $.4\pi$ and $.8\pi$.

From this the intuitive meaning of saturation and the smoothness properties of splines can be seen. Also the role of splines as a step intermediate to the classical estimators with fixed saturation of order 2 or 4, and the revived autoregressive spectral estimators with infinite saturation (see Parzen [5]) is better understood. The greatest advantage of spline estimators seems to be the flexibility in selecting order of saturation (from $\nu = 1$ with saturation of order 2 on up) as a tool for better understanding the spectrum being estimated.

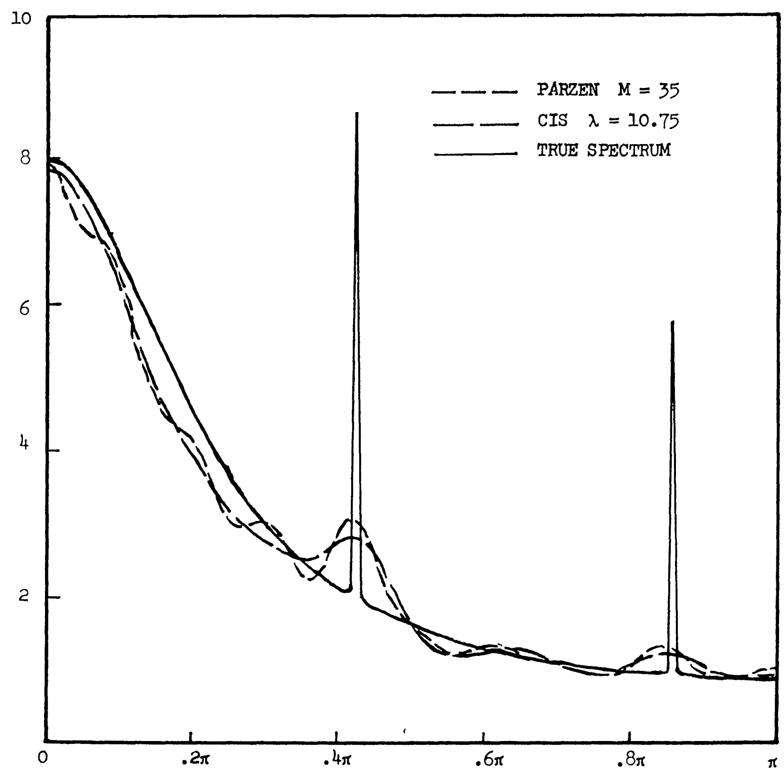

FIG. 1.

REFERENCES

[1] GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

[2] KATZNELSON (1968). *An Introduction to Harmonic Analysis*. Wiley, New York.

[3] KIMELDORF, G. and WAHBA, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495-502.

[4] KIMELDORF, G. and WAHBA, G. (1970). Spline functions and stochastic processes. *Sankhyā* **132** 173-180.

[5] PARZEN, E. (1972). Some recent advances in time series analysis. *Statistical Models and Turbulence*, (Murray Rosenblatt, ed.). Springer-Verlag, New York.

[6] SHAPIRO, H. S. (1969). *Smoothing and Approximation of Functions*. Van Nostrand Reinhold, New York.

[7] TOLSTOV, G. P. (1962). *Fourier Series*. Prentice Hall, Englewood Cliffs, New Jersey.

[8] WALKER, A. M. (1965). Some asymptotic results for the periodogram of a stationary time series. *J. Austral. Math. Soc.* **5** 107–128.

DEPARTMENT OF MATH. AND STAT.
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NEW MEXICO 87131