

additional  $N/2$  complex multiplications necessary for algorithm (A) cause only a slight increase in the overall computation time. In applications where the fast Fourier transform is used repeatedly with the same table of complex exponentials, the increase in computation time becomes insignificant.

The main advantage of algorithm (A) is that it is less susceptible to roundoff errors than schemes such as (B). If the cumulative roundoff error due to finite precision arithmetic is proportional to the number of operations, then the maximum error for algorithm (A) is proportional to  $\log_2 N$ . The maximum error for (B) is proportional to  $N$ . This is because the largest number of complex multiplications required to produce any exponential factor with algorithm (A) is  $2(\log_2 N - 2)$ , while the number of complex multiplications per exponential factor varies from 1 to  $N/2$  for algorithm (B).

As a final note, algorithm (A) is easily implemented in a high-level computer language such as Fortran.

#### SUMMARY

An algorithm for calculating the table of complex exponentials for a fast Fourier transform has been developed. The proposed algorithm requires  $N - 4$  complex multiplications while a standard algorithm uses  $N/2$  operations, accompanied by a bit-reversal scheme. This increase in computational effort is offset by a considerable gain in improved numerical stability. The algorithm developed in this paper accumulates roundoff errors at the rate of  $\log_2 N$ , while the standard algorithm has an error growth rate of order  $N$ . In applications where the same table of complex exponentials are used repetitively, the additional operations required by the proposed algorithm amount to only a slight increase in the overall computation time.

#### REFERENCES

- [1] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297-301, Apr. 1965.
- [2] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975, pp. 290-320.

## Optimal Estimation of Time Series Functions

EDWARD J. WEGMAN

**Abstract**—The principal methods of spectral estimation include kernel smoothing, ME-AR methods, and the Prony algorithm. In this paper we consider not only the estimation of spectral density matrices, but also related time series functions such as transfer functions, gain, phase, cepstrum, and so on. Computational algorithms are developed for use of splines as estimators. Smoothing splines are shown to be special cases of kernel smoothers so that properties of kernel smoothers carry over to spline estimators. Optimality of these estimators is discussed.

### I. INTRODUCTION

In this paper, we are concerned with the optimal estimation of spectral density matrices and related time series functions such as transfer functions, gain, phase, cepstrum, and so on.

Manuscript received July 17, 1979; revised May 9, 1980.

The author is with the Department of the Navy, Office of Naval Research, Arlington, VA 22217.

Much literature exists on the estimation of spectral densities. Three principal approaches include the smoothed FFT (Blackman-Tukey) approach, the Burg-Parzen MEM-AR algorithm, and the so-called Prony method. See [7] for a competitive evaluation of these approaches. Each method has its adherents and its advantages. The Prony method is successful in identifying line spectra in a relatively noise-free environment, but is less successful in noisy environments. Similarly, the MEM-AR approaches of Burg [3] and Parzen [6] are successful at identifying lines in a low-noise environment, although they may exhibit line splitting. Because of the intimate connection with the autoregressive model, the MEM-AR technique can also estimate well the smooth spectral densities provided the spectra are autoregressive like. The smoothed FFT approach does not adapt well to spectra with strong line components since the smoothing tends to eliminate the lines as well. Much data are needed to resolve lines in such a case and, of course, knowledge of smoothness cannot compensate for lack of data. The approaches of Burg, Parzen, and Prony are more appropriate on these cases. For smooth spectra and related functions with noisy data, the smoothing approach appears to be a viable alternative as we shall see later in this paper.

Cogburn and Davis [4] consider estimating a periodic function (of period  $2\pi$ )  $f(\omega)$  using periodic smoothing splines which they apply to univariate spectral density estimation. As we shall show later, the smoothing spline approach is equivalent to the smoothing window approach with the smoothing window determined in some optimal sense. We extend this approach to the multidimensional case and apply results not only to spectral density matrices, but also to nonparametric estimation of such time series functions as gain, phase, transfer function, cepstrum, and the like. It is worthwhile to point out that our analysis applies to physical systems without strong resonances where smoothness of an appropriate filter-related function may be assumed.

In Section II, we define the filter-related functions with which we shall be concerned. Section III contains a description of the computational algorithms for the smoothing splines while Section IV discusses the parallel between the use of smoothing splines and smoothing kernels. Section V deals with multiplicative models while Section VI concludes the paper with a discussion of optimality papers.

### II. SOME PERIODIC TIME SERIES FUNCTIONS

Following the notation of Brillinger [2], we consider an  $(r + s)$  vector-valued stationary series

$$\begin{bmatrix} X(t)^\dagger \\ Y(t)^\dagger \end{bmatrix}$$

where  $t = 0, \pm 1, \pm 2, \dots$  with  $X(t)$   $r$ -vector valued and  $Y(t)$   $s$ -vector valued. All vectors in this paper are row vectors unless otherwise specified.  $\dagger$  is used to indicate transpose. We let  $\mu_X = EX(t)$ ,  $\mu_Y = EY(t)$  where  $E(\cdot)$  is the statistical expectation and with covariances given by

$$E([X(t+u) - \mu_X]^\dagger [X(t) - \mu_X]) = \gamma_{XX}(u)$$

$$E([X(t+u) - \mu_X]^\dagger [Y(t) - \mu_Y]) = \gamma_{XY}(u) = \gamma_{YX}(-u)$$

$$E([Y(t+u) - \mu_Y]^\dagger [Y(t) - \mu_Y]) = \gamma_{YY}(u) \quad u = 0, \pm 1, \dots$$

We may define the spectral densities by

$$f_{XX}(\omega) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma_{XX}(u) e^{-i\omega u}$$

$$f_{XY}(\omega) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma_{XY}(u) e^{-i\omega u}$$

$$f_{YY}(\omega) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma_{YY}(u) e^{-i\omega u} \quad -\pi < \omega < \pi.$$

Suppose we believe that  $Y(t)$  is a filtered version of  $X(t)$ , i.e.,

$$Y(t) = \mu + \sum_{u=-\infty}^{\infty} X(u) a(t-u)^{\dagger} + \epsilon(t). \quad (2.1)$$

The  $\mu$  and  $a(t)$  that minimize

$$E(\epsilon(t) \epsilon^{\dagger}(t))$$

are given by

$$\mu^{\dagger} = \mu_Y^{\dagger} - A(0) \mu_X^{\dagger}$$

and

$$a(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega) e^{i\omega u} d\omega \quad (2.2)$$

where

$$A(\omega) = f_{YX}(\omega) f_{XX}(\omega)^{-1}, \quad -\pi < \omega < \pi.$$

$A(\omega)$  is the *matrix transfer function* of the filter. The  $s$ -vector-valued series

$$\epsilon(t) = Y(t) - \mu - \sum_u X(u) a(t-u)^{\dagger} \quad t = 0, \pm 1, \pm 2, \dots$$

is the *error series* and has spectral density matrix

$$f_{\epsilon\epsilon}(\omega) = f_{YY}(\omega) - f_{YX}(\omega) f_{XX}(\omega)^{-1} f_{XY}(\omega), \quad -\pi < \omega < \pi. \quad (2.3)$$

In the case  $s = 1$ , we may write

$$|R_{YX}(\omega)|^2 = \frac{f_{YX}(\omega) f_{XX}(\omega)^{-1} f_{XY}(\omega)}{f_{YY}(\omega)}, \quad -\pi < \omega < \pi \quad (2.4)$$

which is the *multiple coherence* of  $Y(t)$  with  $X(t)$  at frequency  $\omega$ . In the case where  $r, s$  both are 1,

$$R_{YX}(\omega) = f_{YX}(\omega) / [f_{XX}(\omega) f_{YY}(\omega)]^{1/2}, \quad -\pi < \omega < \pi \quad (2.5)$$

is the *coherency*. Again, if  $r, s = 1$ , we may define

$$G(\omega) = |A(\omega)| = |f_{YX}(\omega) / f_{XX}(\omega)|, \quad -\pi < \omega < \pi$$

and

$$\phi(\omega) = \arg A(\omega), \quad -\pi < \omega < \pi \quad (2.6)$$

which are, respectively, the *gain* of  $Y(t)$  over  $X(t)$  at  $\omega$  and the *phase between*  $Y(t)$  and  $X(t)$  at  $\omega$ .  $G$  is symmetric about 0 and periodic with period  $2\pi$ . Additionally, if  $\epsilon(t) \equiv 0$ ,

$$|G(\omega)|^2 = f_{YY}(\omega) / f_{XX}(\omega), \quad -\pi < \omega < \pi. \quad (2.7)$$

$\phi$  is also periodic with period  $2\pi$  and satisfies  $\phi(-\omega) = -\phi(\omega)$ . Also, we may write

$$\phi(\omega) = \arg f_{YX}(\omega), \quad -\pi < \omega < \pi. \quad (2.8)$$

### III. VECTOR SPLINES

Clearly, a wide variety of time series related functions are periodic vector- (or matrix-) valued functions with period  $2\pi$ . Provided we are willing to assume some smoothness conditions

on these functions, we may use a periodic smoothing spline approach to their estimation. In the discussion which follows we let  $f$  be a generic representation for any of these time series functions. That is, we let  $f(\omega) = (f_1(\omega), \dots, f_k(\omega))$  be a  $k$ -dimensional vector of real-valued functions of period  $2\pi$ . We are assuming the multivariate model

$$h(\omega) = f(\omega) + \epsilon(\omega)$$

where  $E\epsilon(\omega) = 0$  and  $E\epsilon(\omega)^{\dagger} \epsilon(\omega) = \sigma^2 I$ . The function  $h(\omega)$  is known either at lattice frequency-sampling points or continuously and it is desired to estimate  $f(\omega)$ . Let  $\mathcal{P}$  be the class of all such functions  $f$  and let  $\mathcal{P}_0$  be the set of Borel measurable  $g$  in  $\mathcal{P}$  such that  $\int_{-\pi}^{\pi} g(\omega) g^{\dagger}(\omega) d\omega < \infty$ . We let  $C^{(m)}$  denote the set of all vectors  $g$  in  $\mathcal{P}$  such that  $g$  has  $m$  continuous derivative (componentwise) and define

$$\mathcal{P}_m = \{g \in \mathcal{P} : g \in C^{(m-1)} \text{ and } g^{(m)} \in \mathcal{P}_0\}.$$

Let  $L$  be a linear differential operator

$$L = \frac{d^v}{d\omega^v} + \gamma_1 \frac{d^{v-1}}{d\omega^{v-1}} + \dots + \gamma_v, \quad 0 < v \leq m.$$

Of course,  $Lg(\omega) = (Lg_1(\omega), \dots, Lg_k(\omega))$ . We want to approximate  $h$  in  $P$  by a smooth function  $g \in \mathcal{P}_v$ . The following measure of closeness is the analog of the one-dimensional Cogburn-Davis measure which attempts to balance accuracy (first term) with smoothness (second term):

$$\Delta_{n,\lambda,L}(g, h) = \frac{1}{n} \sum_{j=-n+1}^n \left( g\left(\frac{j\pi}{n}\right) - h\left(\frac{j\pi}{n}\right) \right) \cdot \left( g\left(\frac{j\pi}{n}\right) - h\left(\frac{j\pi}{n}\right) \right)^{\dagger} + \frac{1}{\lambda^{2v}\pi} \int_{-\pi}^{\pi} (Lg(\omega)) (Lg(\omega))^{\dagger} d\omega. \quad (3.1)$$

A function  $g$  in  $\mathcal{P}_v$  minimizing  $\Delta_{n,\lambda,L}(g, h)$  will be called a *periodic lattice smoothing vector spline* (LSVS). The parameter  $\lambda$  balances the degree of smoothing with the degree of accuracy. The choice of  $\lambda$  may be made by the method of (generalized) cross validation which is studied extensively by Wahba and Wold [10], [11] and Wahba [9].

It is also worth pointing out that (3.1) could be extended by replacing the first term on the right-hand side with

$$\frac{1}{n} \sum_{j=-n+1}^n \left( g\left(\frac{j\pi}{n}\right) - h\left(\frac{j\pi}{n}\right) \right) \wedge \left( g\left(\frac{j\pi}{n}\right) - h\left(\frac{j\pi}{n}\right) \right)^{\dagger}$$

to get cross coupling. This generalization complicates computational algorithms, but, in principle, could be dealt with by using quadratic programming algorithms. We will not pursue this generalization here.

Similarly, if  $h$  is known for all values of  $\omega$ , it is natural to define a smoothing spline to  $h$  as a function  $g$  minimizing

$$\Delta_{\lambda,L}(g, h) = \frac{1}{\pi} \int_{-\pi}^{\pi} (g(\omega) - h(\omega)) (g(\omega) - h(\omega))^{\dagger} d\omega + \frac{1}{\lambda^{2\lambda}\pi} \int_{-\pi}^{\pi} (Lg(\omega)) (Lg(\omega))^{\dagger} d\omega. \quad (3.2)$$

We call such a  $g \in P_v$  a *periodic continuous smoothing vector spline* (CSVS) to  $h$ .

It is clear that (3.1) may be written as

$$\begin{aligned} \Delta_{n,\lambda,L}(g,h) &= \frac{1}{n} \sum_{j=-n+1}^n \sum_{i=1}^k \left( g_i \left( \frac{j\pi}{n} \right) - h_i \left( \frac{j\pi}{n} \right) \right)^2 \\ &\quad + \frac{1}{\lambda^{2v}\pi} \int_{-\pi}^{\pi} \sum_{i=1}^k (Lg_i(\omega))^2 d\omega \\ &= \sum_{i=1}^k \left[ \frac{1}{n} \sum_{j=-n+1}^n \left( g_i \left( \frac{j\pi}{n} \right) - h_i \left( \frac{j\pi}{n} \right) \right)^2 \right. \\ &\quad \left. + \frac{1}{\lambda^{2v}\pi} \int_{-\pi}^{\pi} (Lg_i(\omega))^2 d\omega \right] \\ &= \sum_{i=1}^k \Delta_{n,\lambda,L}(g_i, h_i) \end{aligned} \tag{3.3}$$

where, of course,  $\Delta_{n,\lambda,L}(g_i, h_i)$  is the one-dimensional measure applied to the  $i$ th component of  $g$  and  $h$ . Similarly,

$$\Delta_{\lambda,L}(g,h) = \sum_{i=1}^k \Delta_{\lambda,L}(g_i, h_i). \tag{3.4}$$

Equations (3.3) and (3.4) imply that we can fit the lattice smoothing vector spline  $g$  and  $h$  by fitting the lattice smoothing spline  $g_i$  to  $h_i$ ,  $i = 1, 2, \dots, k$ . For if not, suppose  $g^*$  is the function which minimizes  $\Delta_{n,\lambda,L}(g,h)$  but  $g_i^*$  is not the spline fit to  $h_i$ . Then there exists a spline  $g_i$  such that

$$\Delta_{n,\lambda,L}(g_i, h_i) < \Delta_{n,\lambda,L}(g_i^*, h_i).$$

Define  $g^{**} = (g_1, \dots, g_{i-1}^*, g_i, g_{i+1}^*, \dots, g_k^*)$  so the clearly

$$\Delta_{n,\lambda,L}(g^{**}, h) < \Delta_{n,\lambda,L}(g^*, h)$$

which contradicts the assumption that  $g^*$  was the LSVS to  $h$ . Thus, the LSVS is the vector of univariate LSS. Similarly, the CSVS to  $h$  is just the vector of CSS to  $h_i$ ,  $i = 1, 2, \dots, k$ .

Cogburn and Davis discuss algorithms for fitting the LSS and the CSS. In particular, let  $P(s)$  be the characteristic polynomial of  $L$  so that

$$P(s) = s^v + \gamma_1 s^{v-1} + \dots + \gamma_v$$

and let

$$Q(k) = |P(ik)|^2, \quad i = \sqrt{-1}.$$

Let

$$q_{n,j} = 1 + Q(j) \left[ \sum_{l=-\infty}^{-1} 1/Q(j+2nl) + \sum_{l=1}^{\infty} 1/Q(j+2nl) \right]$$

and let

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

and

$$\begin{aligned} a_{n,\lambda,l} &= \frac{Q(j)}{Q(l)} a_{n,\lambda,j} \quad \text{for } l \in \{j \pm 2n, j \pm 4n, \dots\} \\ &\quad \text{with } |j| \leq n. \end{aligned}$$

Then the LSS to  $h_i$  is given by

$$h_i \otimes^n s_{n,\lambda}(\omega) = \sum_{l=-n+1}^n h_i \left( \frac{l\pi}{n} \right) s_{n,\lambda} \left( \omega - \frac{l\pi}{n} \right) \tag{3.5}$$

where

$$s_{n,\lambda}(\omega) = \sum_{j=-\infty}^{\infty} a_{n,\lambda,j} e^{ij\omega}.$$

Letting

$$a_{\lambda,l} = \lim_{n \rightarrow \infty} n a_{n,\lambda,l} = \frac{\lambda^{2v}}{2(Q(l) + \lambda^{2v})}$$

and

$$s_{\lambda}(\omega) = \frac{1}{\pi} \sum_{l=-\infty}^{\infty} a_{\lambda,l} e^{il\omega},$$

we can closely approximate  $n s_{n,\lambda}(\omega)$  by  $\pi s_{\lambda}(\omega)$ .  $s_{\lambda}(\omega)$  has importance as an approximation to  $s_{n,\lambda}(\omega)$  but also because the CSS to  $h_i$  is given by

$$h_i \otimes^n s_{\lambda}(\omega) = \int_{-\pi}^{\pi} h_i(y) s_{\lambda}(\omega - y) dy. \tag{3.6}$$

If it is known that the function  $f$  to be estimated has derivatives of order  $v$ , but if no specific operator  $L$  is known, a natural choice is  $L = d^v/d\omega^v$ , in which case  $s_{\lambda}(\omega)$  becomes

$$s_{\lambda}(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \frac{\lambda^{2v}}{l^{2v} + \lambda^{2v}} e^{il\omega} = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \frac{\lambda^{2v}}{l^{2v} + \lambda^{2v}} e^{-il\omega}. \tag{3.7}$$

Letting

$$t(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{1 + y^{2v}} e^{iy\omega} dy,$$

we may define

$$t_{\lambda}(\omega) = \lambda t(\lambda\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\lambda^{2v}}{\lambda^{2v} + y^{2v}} e^{iy\omega} dy$$

and

$$\hat{t}_{\lambda}(\omega) = \sum_{k=-\infty}^{\infty} t_{\lambda}(\omega + 2k\pi). \tag{3.8}$$

It follows (see Cogburn and Davis [4]) that

$$h \otimes^n s_{\lambda}(\omega) = h \otimes^n \hat{t}_{\lambda}(\omega) = \int_{-\infty}^{\infty} h(y) t_{\lambda}(\omega - y) dy. \tag{3.9}$$

Explicit functional forms are given by Cogburn and Davis for  $t(\omega)$ . Notice that we may write the estimate of  $f$  as a convolution with  $t_{\lambda}$ . Hence, provided that  $t_{\lambda}$  satisfies appropriate properties, consistency of the spectral density estimators follows automatically by classical methods. We discuss this more fully in Section IV.

We close this section by noting again that the choice of  $\lambda$  is of critical importance. For  $\lambda$  small, much weight is given to the smoothness term so that the resultant estimator may lack fidelity to the data, increasing bias. Conversely, if  $\lambda$  is large, more weight is given to the accuracy term, increasing variance. A balance must be struck between these two extremes, but regrettably the optimal choice of  $\lambda$  depends on the function  $f$  we are trying to estimate. Wahba and Wold [10], [11] and Wahba [9] suggest the method of cross validation as a solution. We refer to those papers for specific details of the method.

## IV. WINDOW ESTIMATORS, SPLINES, AND CONSISTENCY

Equations (3.5) and (3.9) establish the connection of spline estimators with the classic smoothing kernel estimators. In principle, then, provided the splines are chosen with some care, the classic results appropriate to kernel smoothers carry over directly to spline smoothers. We detail this connection below in order to establish some consistency results for spline smoothers.

Several approaches to estimating the time series functions may be pursued. Note first that if a suitable raw estimate  $h$  of the appropriate time series function (e.g., spectral density, coherence, transfer function, etc.) can be identified, then a spline smoother may be applied directly. The suitable raw  $h$  for the spectral density is, of course, the periodogram. In the other cases, the raw estimate is not so obvious. Alternatively, the filter functions may be estimated by the corresponding functionals of the spectral density estimate. That this is a sensible procedure is demonstrated by the consistency results below. To see this, let  $X(1), X(2), \dots$  be a stationary, discrete, vector-valued stochastic process. We may define the *finite Fourier transform* (of sample size  $n$ ) by

$$d_X^{(n)}(\omega) = \sum_{j=1}^n e^{-i\omega j} X(j)$$

and in turn we may define the *periodogram* by

$$I_{XX}^{(n)}(\omega) = \frac{1}{2\pi n} d_X^{(n)}(-\omega)^\dagger d_X^{(n)}(\omega).$$

The *cross periodogram* is defined in a natural way as

$$I_{XY}^{(n)}(\omega) = \frac{1}{2\pi n} d_X^{(n)}(-\omega)^\dagger d_Y^{(n)}(\omega).$$

It is easy to show that

$$I_{XX}^{(n)}(\omega) = \frac{1}{2\pi} \sum_{u=-n+1}^{n-1} \gamma_{XX}^{(n)}(u) e^{-i\omega u}$$

where the covariance is estimated by

$$\gamma_{XX}^{(n)}(u) = \frac{1}{n} \sum_{j=1}^{n-u} X(j+u)^\dagger X(j).$$

The periodogram is well known to be inconsistent (having too large a variance) and a usual method of establishing consistency is to smooth by means of a spectral window,

$$f_{XX}^{(n)}(\omega) = \int_{-\pi}^{\pi} I_{XX}^{(n)}(y) K(\omega - y) dy = I_{XX}^{(n)} \otimes K(\omega) \quad (4.1)$$

or equivalently by weighting the covariances by a lag window

$$f_{XX}^{(n)}(\omega) = \frac{1}{2\pi} \sum_{u=-M}^M \gamma_{XX}^{(n)}(u) k_M(u) e^{-i\omega u}. \quad (4.2)$$

The resemblance of formula (4.1) to (3.6) is obvious so that, in fact, the continuous smoothing spline estimate of the spectral density matrix is just the classic spectral window estimate with the spectral window determined in a way which minimizes (3.2).

Letting  $L = d^v/d\omega^v$ , we may take

$$f_{XX}^{(n)}(\omega) = I_{XX}^{(n)} \otimes \hat{t}_\lambda(\omega) \quad (4.3)$$

where  $\hat{t}_\lambda(\omega)$  is determined by (3.8). In this case,  $1/\lambda$  becomes the usual "bandwidth" parameter. The mean-square error is minimized (asymptotically) if  $\lambda$  is chosen  $O(n^{1/4v+1})$ . In this

case the mean-square error is  $O(n^{-4v/4v+1})$ , hence  $f_{XX}^{(n)}$  is weakly consistent.

While (4.3) is certainly a feasible computational formula, the construction of  $\hat{t}_\lambda$  is rather messy. We may take advantage of the alternative form (4.2). In this case,

$$f_{XX}^{(n)}(\omega) = \frac{1}{2\pi} \sum_{u=-n}^n \gamma_{XX}^{(n)}(u) k_\lambda(u) e^{-i\omega u}$$

where  $k_\lambda(u)$  is related to  $\hat{t}_\lambda$  by

$$\hat{t}_\lambda(\omega) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} k_\lambda(u) e^{-i\omega u}.$$

Comparing this expression with (3.7), we may write

$$k_\lambda(u) = \frac{\lambda^{2v}}{u^{2v} + \lambda^{2v}} = \frac{1}{1 + \left(\frac{u}{\lambda}\right)^{2v}}$$

so that the CSVS becomes

$$f_{XX}^{(n)}(\omega) = \frac{1}{2\pi} \sum_{u=-n}^n \frac{\gamma_{XX}^{(n)}(u)}{1 + \left(\frac{u}{\lambda}\right)^{2v}} e^{-i\omega u}. \quad (4.4)$$

In practice, the periodogram is often computed on a lattice  $\{j\pi/n, -n < j \leq n\}$  and the CSVS may then be approximated by

$$\begin{aligned} \hat{f}_{XX}^{(n)}\left(\frac{r\pi}{n}\right) &= \sum_{j=-n+1}^n I_{XX}^{(n)}\left(\frac{j\pi}{n}\right) \hat{t}_\lambda\left(\frac{r-j}{n}\pi\right) \\ &= I_{XX}^{(n)} \otimes^n \hat{t}_\lambda\left(\frac{r\pi}{n}\right) \end{aligned} \quad (4.5)$$

which, of course, corresponds to the LSVS. Again, consistency follows from classical results for kernel estimates (see Anderson [1]).

Given the appropriate spline estimates of the spectral densities  $f_{XX}^{(n)}(\omega)$ ,  $f_{YY}^{(n)}(\omega)$ ,  $f_{XY}^{(n)}(\omega)$ ,  $f_{YX}^{(n)}(\omega)$ , we may construct estimates of other periodic functions, namely, the transfer function

$$A^{(n)}(\omega) = f_{YX}^{(n)}(\omega) f_{XX}^{(n)}(\omega)^{-1} \quad (4.6)$$

the spectral density of the error series

$$f_{ee}^{(n)}(\omega) = f_{YY}^{(n)}(\omega) - f_{YX}^{(n)}(\omega) f_{XX}^{(n)}(\omega)^{-1} f_{XY}^{(n)}(\omega), \quad (4.7)$$

the multiple coherence

$$|R_{YX}^{(n)}(\omega)|^2 = \frac{f_{YY}^{(n)}(\omega) f_{XX}^{(n)}(\omega)^{-1} f_{XY}^{(n)}(\omega)}{f_{YY}^{(n)}(\omega)}, \quad (4.8)$$

the coherency

$$R_{YX}^{(n)}(\omega) = f_{YX}^{(n)}(\omega) / [f_{XX}^{(n)}(\omega) f_{YY}^{(n)}(\omega)]^{1/2}, \quad (4.9)$$

the gain

$$G^{(n)}(\omega) = |A^{(n)}(\omega)| = |f_{YX}^{(n)}(\omega) / f_{XX}^{(n)}(\omega)|, \quad (4.10)$$

and finally, the phase

$$\phi^{(n)}(\omega) = \arg f_{YX}^{(n)}(\omega). \quad (4.11)$$

Brillinger [2] provides strong consistency results for all of these functions when the density matrix is estimated by kernel smoothed estimates. His theorem [2, Theorem 8.11.1] requires that  $\lambda$  increases to  $\infty$  as  $n$  increases, say through the sequence,  $\lambda_n$  such that  $\sum_n \lambda_n^{-m} < \infty$  for some  $m > 0$ ,  $(\lambda_n^{-1} n)^{1/2} \lambda_n^{-\epsilon} \rightarrow \infty$  as  $n \rightarrow \infty$  for some  $\epsilon > 0$ . For such a choice,

all the estimators (4.6)-(4.11) are strongly consistent estimators.

The filter weights  $a(u)$  are related to the transfer function via

$$a(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega) e^{iu\omega} d\omega$$

where, of course,  $A(\omega) = f_{YX}(\omega) f_{XX}(\omega)^{-1}$ . Since Brillinger's consistency results apply to lattice estimators, we have  $A$  available on a lattice. Thus,  $a$  may be approximated by a sum. In addition to results of Brillinger [2], Hannan [5], and Wahba [8] consider estimates of the form

$$\hat{a}(u) = \frac{1}{2m} \sum_{j=-m+1}^{m-1} f_{YX}^{(n)}(\omega_j) [f_{XX}^{(n)}(\omega_j)]^{-1} e^{i\omega_j u}$$

where  $f^{(n)}$  are spectral density estimates based on kernel smoothed periodograms. Our suggested estimator for  $\hat{a}$  is essentially this with the  $f^{(n)}$  being spline estimators. The results of Hannan and Wahba carry over directly.

### V. MULTIPLICATIVE MODELS

As we have mentioned earlier, the use of spline estimates for spectral densities, even matrix spectral densities, is not totally satisfactory in the presence of lines in the spectra. There is another sense, however, in which the discussion of the previous section is unsatisfactory. It is well known that  $2I_{XX}^{(n)}(\omega)/f_{XX}(\omega)$  is asymptotically chi-squared with two degrees of freedom,  $\omega \in (-\pi, \pi)$ ,  $\omega \neq 0$ . It is also known for  $\omega_1, \dots, \omega_j$ , the random variables  $I_{XX}^{(n)}(\omega_1), \dots, I_{XX}^{(n)}(\omega_j)$  are asymptotically independent. Thus, a multiplicative model of the form

$$\frac{h(\omega)}{f(\omega)} = \epsilon(\omega)$$

holds rather than the additive form  $h(\omega) = f(\omega) + \epsilon(\omega)$ , postulated in the beginning of Section III. Taking logarithms yields a model of the form

$$\log I_{XX}^{(n)}(\omega) = \log f_{XX}(\omega) + \epsilon(\omega) \tag{5.1}$$

which, strictly speaking, fits the spline model much better. Thus, with a multiplicative model of this type, we are estimating not the spectral density, but the log spectral density. This is the cepstrum. We are thus using our smoother not on the periodogram but on the log periodogram. The main disadvantage of this model is the loss of the extremely convenient computation of the spline given by (4.4). The logged model (5.1), however, fits almost exactly the theoretical model required of the spline. The fact that  $I_{XX}^{(n)}(\omega_1), \dots, I_{XX}^{(n)}(\omega_j)$  are asymptotically independent guarantees that, asymptotically, the covariance matrix of  $\epsilon(\omega_1), \dots, \epsilon(\omega_j)$  is diagonal, fulfilling the conditions of the model in the beginning of Section III. It should be pointed out that were these errors correlated, expressions (3.1) and (3.2) would have to be quadratic forms involving the inverse covariance matrix. This would eliminate the result that the vector spline is the vector of univariate splines, thus making the whole problem considerably more difficult computationally. Of course, because of the kernel interpretation, the additive model retains a theoretical validity, even though strictly speaking, the multiplicative model is most appropriate.

The logged model has an interesting practical advantage, however. In particular (assuming  $r, s = 1$ ), we may write the transfer function using the following:

$$\log \frac{I_{YX}^{(n)}(\omega)}{I_{XX}^{(n)}(\omega)} = \log A(\omega) + \epsilon_1(\omega)$$

and coherency

$$\log \frac{I_{YX}^{(n)}(\omega)}{[I_{XX}^{(n)}(\omega) I_{YY}^{(n)}(\omega)]^{1/2}} = \log R_{YX}(\omega) + \epsilon_2(\omega).$$

That is to say, each of these functions fits the spline model directly, so that each may be estimated directly with a spline rather than as products and ratios of spline estimates. Since as we pointed out in Section I, the spline estimates of spectral densities may miss sharp peaks, this latter procedure seems more satisfactory when appropriate.

### VI. OPTIMALITY

The statistical quality of an estimator is usually measured against some type of optimality criterion. We have used the phrase "optimal estimation" in our title. Our estimators are optimal in the sense that they minimize the mixed norm criterion given in (3.1) and (3.2). These error criterion represent least squares augmented by a penalty for lack of smoothness. Because of the kernel-smoother interpretation, we can view these error criterion as determining, in the sense of (3.1) or (3.2), the optimal kernel smoother and regard the fact that they are also splines as incidental. Of course, the maximum entropy method is also optimal in a different sense. The point is that kernel smoothers are not to be regarded as simply ad hoc estimators, but as estimators which are optimized against a penalized least squares criterion.

There is another sense in which our estimators can be contrasted with maximum entropy estimators. Maximum entropy estimators are parametric estimators. They depend on a Gaussian assumption to carry out the maximum entropy computation and are in fact parameterized as an autoregressive model. By contrast, the spline-based estimates are truly nonparametric, making no assumption about the underlying functional form or the finite dimensional distributions of the stochastic process. Wegman [12] discusses the nonparametric character of spline estimators at some length and discusses the role of optimality in a nonparametric setting.

### REFERENCES

- [1] T. W. Anderson, *The Statistical Analysis of Time Series*. New York: Wiley, 1971.
- [2] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston, 1975.
- [3] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Dep. Geophys., Stanford Univ., Stanford, CA, 1975.
- [4] R. Cogburn and H. T. Davis, "Periodic splines and spectral estimation," *Ann. Statist.*, vol. 2, pp. 1108-1126, 1974.
- [5] E. J. Hannan, "The estimation of a lagged regression relation," *Biometrika*, vol. 54, pp. 409-418, 1967.
- [6] E. Parzen, "Multivariate time series modeling," in *Multivariate Analysis*, vol. II, P. R. Krishnaiah, Ed. New York: Academic, 1970.
- [7] *Proc. RADC Spectrum Estimation Workshop*, Rome Air Development Center, Griffiss Air Force Base, NY, 1978.
- [8] G. Wahba, "Estimation of lag coefficients in a multidimensional distributed lag model," *Econometrica*, vol. 37, pp. 398-407, 1969.
- [9] —, "A survey of some smoothing problems and the method of generalized cross validation for solving them," in *Applications of Statistics*, P. R. Krishnaiah, Ed. New York: North-Holland, 1977.
- [10] G. Wahba and S. Wold, "A completely automatic French curve: Fitting spline functions by cross-validation," *Commun. Statist.*, vol. 4, pp. 1-17, 1975.
- [11] —, "Periodic splines for spectral density estimation: The use of cross validation for determining the degree of smoothing," *Commun. Statist.*, vol. 4, pp. 125-141, 1975.
- [12] E. J. Wegman, "Two approaches to nonparametric regression: Splines and isotonic inference," in *Recent Development in Statistical Inference and Data Analysis*, K. Matusita, Ed. New York: North-Holland, 1980.