SIAM J. APPL. MATH.
Vol. 44, No. 2, April 1984

012

SPECTRAL ESTIMATORS THAT EXTEND THE MAXIMUM ENTROPY AND MAXIMUM LIKELIHOOD METHODS*

CHARLES L. BYRNE† AND RAYMOND M. FITZGERALD‡

Abstract. The theory of best linear approximation in weighted L^2 spaces is used to obtain a general procedure, the PDFT, for linearly reconstructing the Fourier transform from sampled data. The PDFT can be used either directly to reduce sidelobe structure and to extrapolate the data or indirectly to obtain high resolution spectral estimators. The direct and indirect PDFT include as special cases many of the commonly used spectral techniques, including Burg's maximum entropy method, Capon's maximum likelihood method, the spectral estimators based on bandlimited extrapolation, the eigenvalue/eigenvector methods for detecting sinusoids in noise (Pisarenko method, Schmidt's MUSIC, eigenvector power beamforming), and the best linear unbiased estimator (BLUE) for regression coefficients. By exploiting their relationship to the linear PDFT, these nonlinear techniques can be analyzed in terms of linear approximation theory. In addition to providing a unifying formulation for many different spectral estimators, the PDFT approach provides new techniques which expand the class of available high resolution spectral estimators.

1. Introduction. Using the theory of best linear approximation in Hilbert space we develop a general class of high resolution spectral estimators that contains, as special cases, several of the well-known linear and nonlinear methods, including Burg's maximum entropy (MEM) estimator [1], [4], [5], [25], the maximum likelihood method (MLM) of Capon [11], [25], the bandlimited extrapolation procedures of Cadzow [9], Papoulis [28], Kolba and Parks [24], and others, and the eigenvalue/eigenvector techniques of Pisarenko [30], Schmidt [33] and Bienvenu [2]. The recent tutorial article by Kay and Marple [23] summarized concisely most of the modern spectral estimators. In this paper we provide a common mathematical framework within which many of these techniques can be analyzed and extended.

In [23] it was noted that most modern spectral estimators are based on models that involve a small number of parameters. As we shall see, these same estimators can also be based on linear approximation in weighted L^2 spaces. The weights reflect the user's prior information about the spectrum to be estimated and his objectives in estimation. Because the sampled data are always insufficient to specify the unique, correct spectrum and because no estimator will reconstruct equally well all the characteristics of the spectrum, it is essential that prior knowledge and the aims of the user play a role in the design of the estimators employed. The approximation theoretic approach we present in this paper provides the user with the opportunity to participate in the design of the estimator, so that it will have the properties appropriate for his or her application.

The basic problem to be solved is the following. We observe the function $x(t)$ at the "times" $t_1, \cdots, t_N$. From the data, $x(t_1) \cdots x(t_N)$, we must either reconstruct (estimate, approximate) the Fourier transform (FT) of $x(t)$, which is defined by

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{i\omega t}\, dt, \tag{1}$$

or, equivalently, we must extrapolate and interpolate the unobserved values of $x(t)$. Clearly the data fails to determine $X(\omega)$ and $x(t)$ uniquely, and our success in reconstructing from the data will depend in large part on the effective use of prior

* Received by the editors August 31, 1982, and in revised form April 28, 1983.

† U.S. Naval Research Laboratory, Washington, DC 20375 and Department of Mathematics, The Catholic University of America, Washington, DC 20064.

‡ U.S. Naval Research Laboratory, Washington, DC 20375.

information in the design of the estimator. For example, suppose we know that $X(\omega)=0$ for $|\omega|>\omega_0$ and our data are $x(n\Delta)$, $n=1,\cdots,N$, for some sampling interval $0<\Delta\leqq \pi/\omega_0$. Because we can write $X(\omega)$ as

$$X(\omega)=\begin{cases}\Delta \sum_{n=-\infty}^{\infty} x(n\Delta)\, e^{in\Delta\omega}, & |\omega|\leqq \pi/\Delta,\\ 0, & |\omega|>\pi/\Delta,\end{cases} \tag{2}$$

it is reasonable to consider the discrete Fourier transform (DFT) estimator of $X(\omega)$, which is defined by

$$X_{DFT}(\omega)=\begin{cases}\Delta \sum_{n=1}^{N} x(n\Delta)\, e^{in\Delta\omega}, & |\omega|\leqq \pi/\Delta,\\ 0, & |\omega|>\pi/\Delta.\end{cases} \tag{3}$$

If it should happen that $\Delta<\pi/\omega_0$, then the DFT does not vanish, as $X(\omega)$ does, for $\omega_0<|\omega|\leqq \pi/\Delta$. Can we find a better estimator for $X(\omega)$? In particular, can we find one that does vanish off of $[-\omega_0, \omega_0]$?

Consider as an estimator of $X(\omega)$ the function of the form

$$X_{\mathrm{BL}}(\omega)=\begin{cases}\sum_{n=1}^{N} a_n\, e^{in\Delta\omega}, & |\omega|\leqq \omega_0,\\ 0, & |\omega|>\omega_0,\end{cases} \tag{4}$$

which is consistent with the data. That is, choose $a_1,\cdots,a_N$ so that

$$x(n\Delta)=\int_{-\omega_0}^{\omega_0} X_{\mathrm{BL}}(\omega)\, e^{-in\Delta\omega}\frac{d\omega}{2\pi}, \qquad n=1,\cdots,N. \tag{5}$$

This bandlimited (BL) estimator has the important approximation theoretic property that the coefficients that minimize the mean square error on $[-\omega_0, \omega_0]$, given by

$$\text{error}=\int_{-\omega_0}^{\omega_0}\left|X(\omega)-\sum_{n=1}^{N} a_n\, e^{in\Delta\omega}\right|^2 d\omega, \tag{6}$$

are precisely the ones that satisfy (5). This BL estimator of $X(\omega)$ is essentially the Fourier transform estimator determined by the various bandlimited extrapolation procedures of Cadzow [9], Papoulis [28], Kolba and Parks [24] and others [7], [16]. When the sampling rate is Nyquist (so that $\Delta=\pi/\omega_0$), then $X_{\mathrm{BL}}=X_{\mathrm{DFT}}$. When the sampling rate is above Nyquist (so that $\Delta<\pi/\omega_0$), then the DFT estimator wastes degrees of freedom approximating zero on $\omega_0\leqq|\omega|\leqq\pi/\Delta$. By using the band $[-\omega_0, \omega_0]$ in the definition of the approximation error, we release those degrees of freedom and make it possible to better reconstruct $X(\omega)$ where it is known to be supported.

In practice the assumption that $X(\omega)=0$ for $|\omega|>\omega_0$ is usually too strict. Roundoff error and other perturbations quite easily introduce out-of-band components to which the BL reconstruction is very sensitive. In addition, the equations that must be solved for the parameters $a_1,\cdots,a_N$,

$$x(m\Delta)=\sum_{n=1}^{N} a_n \frac{\sin(m\Delta-n\Delta)\omega_0}{\pi(m\Delta-n\Delta)}, \qquad m=1,\cdots,N, \tag{7}$$

involve a matrix that becomes ill-conditioned as $\pi/\Delta-\omega_0$ increases. In an attempt to stabilize this BL method we are led to consider a general technique for the linear estimation of the FT. This technique, which we call the PDFT, provides a framework for a unified treatment of linear and nonlinear spectral estimators.

Most high resolution spectral estimators, such as MEM and MLM, are based on a knowledge of autocorrelation, rather than time series, values. That is, the data, $r(m)$, $m=0,\cdots,M$, are samples of the autocorrelation function, defined by

$$r(t)=\int_{-\infty}^{\infty}|X(\omega)|^2\,e^{-it\omega}\,d\omega/2\pi, \tag{8}$$

and the problem is to reconstruct the power spectrum, $R(\omega)=|X(\omega)|^2$, from these data.

Frequently the problem is posed in a stochastic context. The autocorrelation function, $r(t)$, is defined to be the expected value of the product, $x(s+t)\overline{x(s)}$, and $x(t)$ is viewed as a realization of a zero mean, weakly stationary random process [26] (overbars are used to indicate complex conjugates). For the purposes of this discussion, however, the deterministic interpretation will suffice.

We rarely observe $r(m)$ directly. In practice $r(m)$ is usually estimated from measurements of $x(t)$. An alternative to dealing with estimates of the autocorrelation lags is to replace $r(0),\cdots,r(M)$ with an equivalent set of parameters—reflection coefficients in the Burg algorithm, for example [6]—and to estimate these parameters from the $x(t)$ data. Because it provides an extrapolation of the data, the PDFT can be used to improve the autocorrelation estimates which are then used to estimate the spectrum.

The PDFT is a linear Fourier transform estimator and, equivalently, a linear extrapolator of the data. When used directly it provides essentially a "numerator-type" estimator, i.e. one whose parameters appear in the numerator. Such estimators are useful for suppressing sidelobe structure and enhancing the less prominent features and for improving autocorrelation estimates. As we shall see, it is also possible to use the PDFT indirectly to obtain a large class of "denominator-type" high resolution spectral estimators. These indirect PDFT estimators (IPDFT) include as special cases Burg's MEM and the various eigenvalue/eigenvector techniques for detecting sinusoids in noise. The maximum likelihood (MLM) estimator will be shown to be a special case of calibrated or normalized PDFT spectral estimators and the best linear unbiased estimate of regression coefficients (BLUE) will be derived as a limiting case of PDFT amplitude estimation. We discuss first the direct PDFT.

For completeness we include a brief sketch of both the maximum entropy method and the maximum likelihood method in Appendix A. These topics are treated rather completely in [12], [19]. Appendix B contains an iterative algorithm for inverting the matrices that arise in the PDFT development.

2. The direct PDFT. In discussing the BL estimator of the Fourier transform we showed how the prior knowledge that $X(\omega)$ was supported on a subinterval of $[-\pi/\Delta,\pi/\Delta]$ could be employed to design an optimal estimator. To form the BL estimator we may employ our prior knowledge about the general shape of $|X(\omega)|$ by writing (4) in the form

$$X_{\mathrm{BL}}(\omega)=P(\omega)\sum_{n=1}^{N}a_n\,e^{in\Delta\omega}, \tag{9}$$

where $P(\omega)=1$ for $|\omega|\leqq\omega_0$ and $P(\omega)=0$ for $|\omega|>\omega_0$. That is, the prior knowledge that $X(\omega)$ vanishes off of $[-\omega_0,\omega_0]$ is here incorporated in the function $P(\omega)$. This technique is readily extended by allowing $P(\omega)$ to be a more general estimate of $|X(\omega)|$ in accordance with what is known a priori. In particular, by making $P(\omega)=1$ for $|\omega|\leqq\omega_0$ and $P(\omega)\geqq 0$ small, but not zero, for $\omega_0\leqq|\omega|\leqq\pi/\Delta$, we obtain an estimator which has the desirable properties of X_{BL} but which is considerably more stable than X_{BL}. For simplicity, let us assume from now on that $\Delta=1$, so that the data are $x(1), \cdots, x(N)$ and we have that $X(\omega)=0$ for $|\omega|>\pi$.

Having chosen $P(\omega)\geqq 0$ to share with $|X(\omega)|$ whatever broad features are known a priori, we take as our FT estimator the data-consistent function of the form

(10) $$X_{PDFT}(\omega)=P(\omega)\sum_{n=1}^{N} a_n e^{in\omega}, \qquad |\omega|\leqq\pi.$$

With $p(t)$ defined by

(11) $$p(t)=\int_{-\pi}^{\pi} P(\omega)\, e^{-it\omega}\frac{d\omega}{2\pi},$$

the requirement of data consistency leads to the set of equations

(12) $$x(m)=\sum_{n=1}^{N} a_n p(m-n), \qquad m=1,\cdots,N,$$

which determine the parameters a_n, $n=1,\cdots,N$. As with X_{BL}, this PDFT Fourier transform estimator is optimal in an approximation theoretic sense. The coefficients $a_1,\cdots,a_N$ that minimize the weighted error,

(13) $$\text{error}=\int_{-\pi}^{\pi}\left|X(\omega)-P(\omega)\sum_{n=1}^{N} a_n e^{in\omega}\right|^2 P^{-1}(\omega)\,d\omega,$$

are precisely those satisfying (12). The appearance in (13) of $P^{-1}(\omega)$ is significant. As $P(\omega)$ takes on the prominent features of $|X(\omega)|$, the weighting by $P^{-1}(\omega)$ directs the sensitivity to error toward the less prominent components of $X(\omega)$. This sensitivity to the less obvious features suggests a role for the PDFT in computerized tomography as discussed in [8].

Taking the inverse FT of $X_{PDFT}(\omega)$ gives the data extrapolation as

(14) $$x_{PDFT}(t)=\sum_{n=1}^{N} a_n p(t-n),$$

which can be used to improve estimates of the autocorrelation function.

The biased lag-product estimate of $r(m)$, for $|m|\leqq N-1$, given by

(15) $$\hat{r}(m)=\sum_{n=1}^{N-|m|} x(n+m)\overline{x(n)},$$

is closely related to the DFT estimator of the Fourier transform. From (3) we have that

(16) $$|X_{DFT}(\omega)|^2=\sum_{m=-N+1}^{N-1}\left(\sum_{n=1}^{N-|m|} x(n+m)\overline{x(n)}\right)e^{im\omega},$$

which leads to the use of $\hat{r}(m)$ as an estimator for $r(m)$. It is well known that $\hat{r}(m)$ is not always a good estimator for $r(m)$ and that the use of $\hat{r}(m)$ can degrade the performance of high resolution techniques such as MEM. For this reason several authors have published MEM algorithms that use reflection coefficient estimates instead of autocorrelation estimates [6], [27], [35]. Instead of trying to bypass the use of autocorrelation estimates, we can try to improve them. One possibility is to use (14) to increase the number of time series values used in the biased lag-product estimates (15).

The PDFT estimator in (10) is the Fourier transform of the extrapolated time function in (14). The use of extrapolation procedures is not new and an extensive study of this subject can be found in [20]. The PDFT estimator (10) is closely related to Jain's mean-square extrapolation filter (MSEF), although the use of a prior weighting function to derive the autocorrelation function is not considered in [20].

The autocorrelation function, $p(t)$, in (11) can be used to define a kernel, $K(s,t) = p(s-t)$, for a reproducing kernel Hilbert space (RKHS). The analysis of such spaces and their uses in estimation and detection have been discussed by Parzen [29] and Kailath [22].

The use of windowing techniques to analyze the MEM and MLM estimators has recently been reported by Durrani and Arslanian [15]. A "windowing" interpretation of the direct and indirect PDFT methods appears feasible, but we do not consider it here.

We conclude our discussion of the direct PDFT with an example taken from the Rome Air Development Center (RADC) 1979 Workshop on Spectrum Estimation. The data consist of 32 complex, equi-spaced samples of a time series $x(t)$ corresponding to a signal from a Doppler radar corrupted by low-Doppler clutter. The clutter spectrum is known to be supported on $[-\pi/3, \pi/3]$ and the targets (sinusoids) that are present are known to be at frequencies outside the clutter band, but within $[-\pi, \pi]$. Figure 1 gives the magnitude of the DFT estimator for the 32 data values and indicates the clutter region between the dotted lines. The three actual targets (indicated by arrows) are located at $\omega = -1.84$, -1.15 and $+2.46$ and have amplitudes of 0.0873, 0.2184, and 0.1203 respectively. A prior $P(\omega)$ was chosen (indicated in Fig. 2) and the PDFT spectrum was calculated (Fig. 3). The effect of the PDFT on sidelobe background is the most obvious feature.

Because the direct PDFT is essentially a "numerator-type" estimator, $X_{\text{PDFT}}(\omega)$ does not exhibit the highly resolved peaked structure familiar to users of MEM. In order to obtain such high resolution and yet retain the flexibility associated with the choice of the weighting function, $P(\omega)$, we consider next an indirect application of the PDFT method.

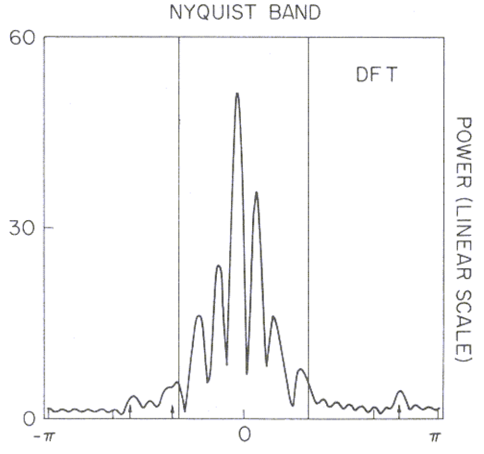

FIG. 1. *Magnitude of* DFT; 32 *complex samples*, $|\omega| \leq \pi$ (*linear scale*). *Clutter region between lines*; 3 *targets indicated by* ↑.

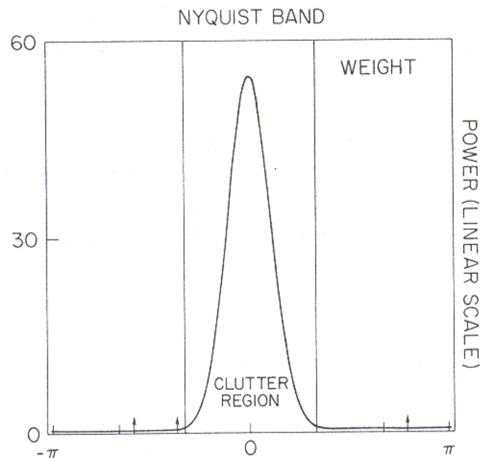

FIG. 2. *Prior weight function*, $P(\omega)=0.01+$ *Gaussian* (*mean* 0.0, *standard deviation* 0.3), $|\omega|\leqq\pi$ (linear scale).

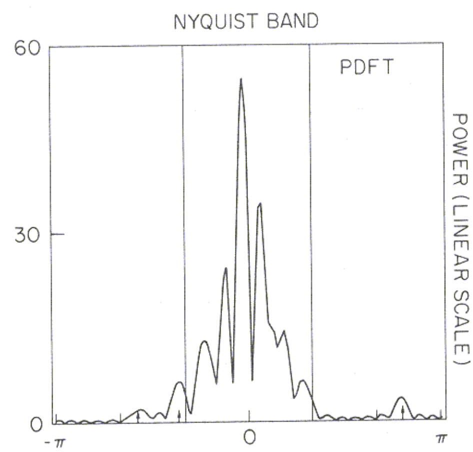

FIG. 3. *Magnitude of* PDFT, 32 *complex samples*, $|\omega|\leqq\pi$, *prior as in Fig.* 2 (*linear scale*).

3. The indirect PDFT high resolution estimators. We assume now that the data, $r(m), |m|\leqq M$, are exact or estimated autocorrelation values and that the problem is to estimate $R(\omega)$, the FT *of* $r(t)$. We could select a function $P(\omega)\geqq 0$ as a prior estimate of $R(\omega)$ and use the PDFT directly to obtain the estimate

$$R_{\mathrm{PDFT}}(\omega)=P(\omega)\sum_{n=-M}^{M}a_n e^{in\omega}, \tag{17}$$

with $a_{-M},\cdots,a_M$ chosen to make $R_{\mathrm{PDFT}}(\omega)$ consistent with the conjugate symmetric autocorrelation data. This estimator is necessarily real, but it may be negative for some values of ω. Unless $P(\omega)$ itself is highly structured, $R_{\mathrm{PDFT}}(\omega)$ will not provide the sort of resolution desired in many applications. To obtain such resolution the polynomial in (17) must somehow be put into the denominator.

Let us switch the roles of $R(\omega)$ and $P(\omega)$ and imagine that $R(\omega)$ is a prior weighting function which is being used to reconstruct $P(\omega)$ from samples of $p(t)$. If the "data" consist of $M+1$ contiguous values of $p(t)$, then the matrix involved in this

PDFT is known. For example, if $M=2K$ and the "data" are $p(-K),\cdots,p(K)$, then the system of equations to be solved for a_n, $n=-K,\cdots,K$ is given by

$$p(m)=\sum_{m=-K}^{K} a_n r(m-n), \qquad m=-K,\cdots,K. \tag{18}$$

The resulting PDFT estimate of $P(\omega)$ will be

$$P_{\mathrm{PDFT}}(\omega)=R(\omega)\sum_{n=-K}^{K} a_n\, e^{in\omega}. \tag{19}$$

Since it is $R(\omega)$ that we really want, we solve (19) for $R(\omega)$ and replace $P_{\mathrm{PDFT}}(\omega)$ by the known $P(\omega)$ to get as the estimator of $R(\omega)$

$$\hat{R}(\omega)=P(\omega)\Big/\sum_{n=-K}^{K} a_n\, e^{in\omega}. \tag{20}$$

This estimator is necessarily real, although the polynomial has been observed to take on negative values (in a high signal-to-noise environment the polynomial sometimes has a zero at or near frequencies of sinusoidal components.) This is not necessarily bad, however, for it indicates that the polynomial can be used to locate the frequencies of sinusoids in noise. We shall take up this topic again when we discuss the relation between the PDFT and the eigenvalue/eigenvector techniques.

The "data" set used above to estimate $R(\omega)$ is redundant in that it is conjugate symmetric. Suppose we take the "data" to be just the values $p(0), p(1),\cdots,p(M)$. Solving the system

$$p(m)=\sum_{n=0}^{M} a_n r(m-n), \qquad m=0,1,\cdots,M, \tag{21}$$

for a_n, $n=0,1,\cdots,M$, we obtain the PDFT estimator

$$P_{\mathrm{PDFT}}(\omega)=R(\omega)\sum_{n=0}^{M} a_n\, e^{in\omega}. \tag{22}$$

This function approximates not only $P(\omega)$ itself but also every other function consistent with the data $p(0),\cdots,p(M)$. In fact, as $M\to\infty$, $P_{\mathrm{PDFT}}(\omega)$ converges, not to $P(\omega)$, but to a function having the same causal part as $P(\omega)$. (The causal part of $P(\omega)$ is the function $P^+(\omega)=\sum_{n=0}^{\infty} p(n)\, e^{in\omega}, |\omega|\leqq\pi$.) Consider therefore the causal part of $P_{\mathrm{PDFT}}(\omega)$ given by

$$P^+_{\mathrm{PDFT}}(\omega)=R^+(\omega)\sum_{n=0}^{M} a_n\, e^{in\omega}+G^+(\omega), \tag{23}$$

where $G^+(\omega)$ is defined by

$$\begin{aligned}G^+(\omega)&=a_1\overline{r(1)}+a_2(\overline{r(2)}+\overline{r(1)}\, e^{i\omega})\\&\quad+\cdots+a_M(\overline{r(M)}+\overline{r(M-1)}\, e^{i\omega}+\cdots+\overline{r(1)}\, e^{i(M-1)\omega}).\end{aligned} \tag{24}$$

This $G^+(\omega)$ can be computed from the data and (21). Solving for $R^+(\omega)$, the causal part of $R(\omega)$, we obtain

$$R^+(\omega) = (P^+_{\mathrm{PDFT}}(\omega) - G^+(\omega)) \Big/ \sum_{n=0}^{M} a_n e^{in\omega}. \tag{25}$$

Replacing $P^+_{\mathrm{PDFT}}(\omega)$ with the known function $P^+(\omega)$ we obtain the indirect PDFT(IPDFT) estimate of $R^+(\omega)$ as

$$R^+_{\mathrm{IPDFT}}(\omega) = (P^+(\omega) - G^+(\omega)) \Big/ \sum_{n=0}^{N} a_n e^{in\omega}. \tag{26}$$

The IPDFT estimate for $R(\omega)$ itself is then

$$R_{\mathrm{IPDFT}}(\omega) = 2\,\mathrm{Real}\,(R^+_{\mathrm{IPDFT}}(\omega)) - r(0). \tag{27}$$

If the polynomial $\sum_{n=0}^{M} a_n z^n$ has all its zeros outside the unit circle (minimum phase), then $R^+_{\mathrm{IPDFT}}(\omega)$ is causal (is equal to its causal part) and $R_{\mathrm{IPDFT}}(\omega)$ is consistent with the autocorrelation data, $r(m)$, $|m| \leqq M$. The minimum phase condition holds whenever the real part of $\sum_{n=0}^{M} a_n e^{in\omega}$ does not change sign on $[-\pi, \pi]$([34, p. 123]). The IPDFT estimator is real, although not necessarily nonnegative. In particular applications $R_{\mathrm{IPDFT}}(\omega)$ is frequently both nonnegative and consistent with the data. For discussion of a closely related minimum phase problem see [32].

If we select as our prior weight function $P(\omega) = 1$ for $|\omega| \leqq \pi$ and $P(\omega) = 0$ otherwise, then we can show that

$$R_{\mathrm{IPDFT}}(\omega) = a_0 \Big/ \left| \sum_{n=0}^{M} a_n e^{in\omega} \right|^2. \tag{28}$$

The equations (21) are then the same ones that occur in the derivation of the MEM estimator [25] and indeed for this particular weighting function we have that $R_{\mathrm{IPDFT}}(\omega) = R_{\mathrm{MEM}}(\omega)$. Table 1 indicates the mean square error terms that are minimized in DFT, PDFT, MEM and IPDFT spectrum estimation.

TABLE 1
Mean square error terms.

$\int \left| R(k) - \sum_{n=-N}^{N} a_n e^{ikn} \right|^2 dk$ (DFT)	$\int \left| \frac{R(k)}{P(k)} - \sum_{n=-N}^{N} a_n e^{ikn} \right|^2 P(k)\, dk$ (PDFT)
$\int \left| \frac{1}{R(k)} - \sum_{n=0}^{N} a_n e^{ikn} \right|^2 R(k) dk$ (MEM)	$\int \left| \frac{P(k)}{R(k)} - \sum_{n=0}^{N} a_n e^{ikn} \right|^2 R(k) dk$ (IPDFT)

To illustrate the use of the IPDFT we consider once again the RADC problem discussed earlier. From the 32 complex samples we estimate $r(m)$, $|m| \leqq M$, using the biased lag-products estimates (15), for $M = 11$. Figure 4 is the DFT of $r(m)$, $|m| \leqq 11$, in dB scale (0 dB is 2500.0 throughout). Figure 5 is the prior weighting. Figures 6 and 7 show MEM and IPDFT. The IPDFT involves both numerator and denominator.

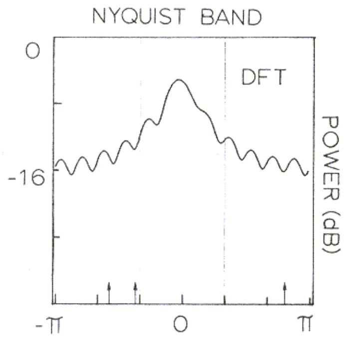

FIG. 4. DFT *of biased lag-product estimates,* $m=0,\cdots,11$, $|\omega|\leqq\pi$ (dB *scale,* 0 dB = 2500).

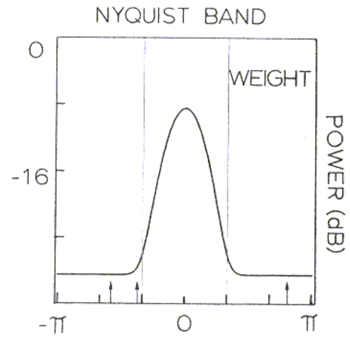

FIG. 5. *Prior weight function of Fig.* 2 (dB *scale*).

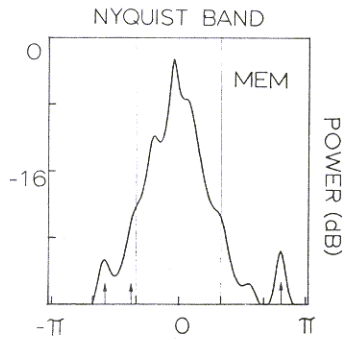

FIG. 6. MEM *estimate, same data as Fig.* 4 (dB *scale*).

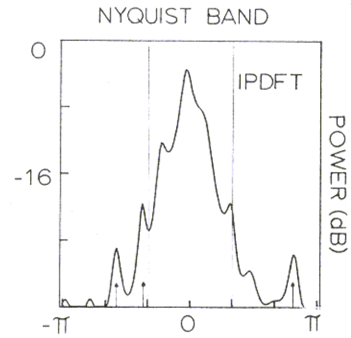

FIG. 7. IPDFT *estimate, same data as Fig.* 4, *prior as in Fig.* 2 (dB *scale*).

The improvement over MEM is reminiscent of the advantage obtained through the use of ARMA models in [10], [13] and the IPDFT can be viewed as a computationally simpler alternative to ARMA.

If our goal is to obtain a reconstruction of $R(\omega)$ that presents a good overall picture of $R(\omega)$ then it helps to take $P(\omega)$ as close to $R(\omega)$ as possible. In the extreme case, in which $P(\omega)$ equals $R(\omega)$, the polynomial reduces to unity and $R_{\text{IPDFT}}(\omega)=R(\omega)$. Judging from the reported success of MEM in reconstructing spectra, it appears that even a noncommittal choice of $P(\omega)$ can lead to a useful estimate. At times, however, a good overall picture of $R(\omega)$ is not our primary concern. Consider for example the problem of detecting and localizing sinusoids in noise and of estimating their amplitudes. In the detection problem resolution is important. The choice of $P(\omega)$ can be made to involve not only our prior knowledge of $R(\omega)$ but also our goals in estimation. A full discussion of the use of the IPDFT for high resolution estimation requires a brief sketch of the eigenvalue/eigenvector approaches for the detection and localization of sinusoids in noise. This is the topic of the next section.

4. IPDFT and eigenvector analysis of sinusoids in noise. Because most modern spectral estimation methods are based on models that involve a small set of parameters, errors arise either because the parameters are poorly estimated or because the model itself is inappropriate. Not surprisingly, the methods that are best able to resolve two closely spaced sinusoids are those that are based on a model of sinusoids in noise. Such methods have been presented under a variety of names. Generally they are eigenvalue/eigenvector methods, and are usually associated with the names Pisarenko [23], [30], Schmidt [33] and Bienvenu [2] (see also [3], [21]).

The function $x(t)$ being measured is taken to be

$$x(t)=\sum_{j=1}^{J} A_j e^{i\theta_j} e^{it\omega_j}+\eta(t), \tag{29}$$

where $B_j=A_j e^{i\theta_j}$ is the complex amplitude, A_j is the amplitude, θ_j is the phase, ω_j is the angular frequency $(-\pi \leqq \omega_j \leqq \pi)$, and $\eta(t)$ is one realization of a zero mean, weakly stationary noise process with autocorrelation $r_{\eta\eta}(t)$. We assume here that the θ_j are independent random variables uniformly distributed on $[-\pi, \pi]$, although more involved models can be and sometimes are employed. The autocorrelation function for the process $x(t)$ is then given by

$$r_{xx}(t)=E\{x(t+s)\overline{x(s)}\}=\sum_{j=1}^{J} |A_j|^2 e^{it\omega_j}+r_{\eta\eta}(t). \tag{30}$$

With the matrix notation $R_{xx}=[r_{xx}(n-m)]$, $R_{\eta\eta}=[r_{\eta\eta}(n-m)]$, $S=[S_{m,j}]=[e^{im\omega_j}]$, and $A=\operatorname{diag}\{A_1,\cdots,A_J\}$, where $m, n=0,\cdots,M$ and $j=1,\cdots,J$, and with * denoting conjugate transpose, we have

$$R_{xx}=SAA^*S^*+R_{\eta\eta}. \tag{31}$$

Typically $J \ll M+1$ so that SAA^*S^*, with rank J, is singular. Consequently there are $M+1-J$ linearly independent solutions y to the matrix equation $yR_{xx}=yR_{\eta\eta}$, or equivalently, the matrix $Q=R_{\eta\eta}R_{xx}^{-1}$ has $M+1-J$ mutually orthogonal eigenvectors corresponding to the eigenvalue $\lambda=1$, which repeats $M+1-J$ times. Because SAA^*S^* is nonnegative definite we can show that $\lambda=1$ is the largest eigenvalue of Q. For this reason the nonzero eigenvectors satisfying $yR_{xx}=yR_{\eta\eta}$ are called the "largest" eigenvectors of Q. It follows that $ySAA^*S^*=0$ or $ySAA^*Sy^*=0$, so that $yS=0$ for each largest eigenvector y. With $y=(y_0, y_1,\cdots, y_M)$ we then have

$$0=\sum_{m=0}^{M} y_m e^{im\omega_j}, \tag{32}$$

for each $j=1,\cdots,J$. This means that the z-transform (32) of each largest eigenvector y has zeros at precisely the sinusoidal frequencies, ω_j, $j=1,\cdots,J$.

In practice R_{xx} must be estimated from data and $R_{\eta\eta}$ must be approximated by the user. Frequently $R_{\eta\eta}=\sigma^2 I$ is chosen, even though the success of the method in the presence of colored noise depends heavily on the choice of a good estimate for $R_{\eta\eta}$. Eigenvalues of $Q=R_{\eta\eta}R_{xx}^{-1}$ are then calculated. Because of the approximations involved, the largest eigenvalue may not repeat. One does count on the largest $M+1-J$ eigenvalues clustering visibly, however, in order to make it possible to estimate J. Once J is determined, the $M+1-J$ largest eigenvectors are then calculated from Q. Theoretically the zeros of their z-transforms indicate the sinusoidal frequencies. In practice the magnitudes squared of the $M+1-J$ z-transforms are averaged to add stability to the procedure. In array processing this is called "eigenvector beamforming" [17].

The largest eigenvalue of $Q=R_{\eta\eta}R_{xx}^{-1}$ can be computed using any of several variants of the power method. Starting with any vector v one calculates vQ, vQ^2, etc. Ignoring certain technicalities ([14, p. 192]) we can say that this sequence of vectors converges to a largest eigenvector y of Q, corresponding to largest eigenvalue λ. Because $yQ=\lambda y$ we get $\lambda=yQy^*/yy^*$. If our initial choice of v was good, that is, if v itself is close to y, then vQ or vQ^2 should estimate y quite well, or at least, the z-transforms should begin to indicate sinusoidal frequency locations. We relate this technique to the IPDFT.

In computing the R_{IPDFT} spectral estimator we solve the system of equations (21), which, using $p=(p(0),\cdots,p(M))$, $a=(a_0,\cdots,a_M)$ and $R=[R_{n,m}]=[r(m-n)]$, $m, n=0,\cdots,M$, we can write as $p=aR$. The vector a whose z-transform appears in the denominator of the IPDFT is then $a=pR^{-1}$. If $x(t)$ has the form (29) then $R=R_{xx}$. For the z-transform of a to provide a good indication of the locations of the ω_j, it is necessary that a be close to y, a largest eigenvector of $Q=R_{\eta\eta}R_{xx}^{-1}$. If $a=vQ$ for some v then $p=vR_{\eta\eta}$. The task then is to select p of the form $p=vR_{\eta\eta}$ corresponding to a v that is close to y. If we are only interested in detecting sinusoids within a particular frequency band, it is necessary that v "mimic" y only to the extent of having its z-transform relatively small near the frequencies of interest. One way to do this is to select $V(\omega)\geqq 0$ to be relatively small in the frequency band of interest, larger elsewhere. Let $v=(v(0), v(1),\cdots, v(M))$ be samples from $v(t)$, the inverse FT of $V(\omega)$. Choose $R_{\eta\eta}$ to estimate, as best we can, the noise correlation matrix and take $p=vR_{\eta\eta}$. If $a=pR^{-1}$ then a represents one further step along the iteration sequence converging to y. If a does not provide the desired resolution, we can iterate again, setting $v^{(1)}=a$, $p^{(1)}=v^{(1)}R_{\eta\eta}$, $a^{(1)}=pR^{-1}$; generally we can have $v^{(k)}=a^{(k-1)}$, $p^{(k)}=v^{(k)}R_{\eta\eta}$, and $a^{(k)}=p^{(k)}R^{-1}$. The z-transform of $a^{(k)}$ should begin to resolve the sinusoids after a small number of iterations, if the initial v is chosen as described above.

To illustrate this method, exact autocorrelation values corresponding to unit amplitude sinusoids at $\omega=1.26$, 1.33 were considered, for $m=0,\cdots,14$. The value of $r(0)$ was increased by 10% to increase the stability of the matrix inversion. The reciprocal magnitude of the z-transform of $a^{(k)}$ was calculated for $k=1, 2, 3, 4$. Figure 8 displays the results. The prior noise matrix was taken as $0.1I$ and v was selected initially to correspond to $0\leqq V(\omega)=1.1-(1.5)\mathcal{N}(1.3, 0.25)$, an inverted Gaussian centered at 1.3 with standard deviation of 0.25. The two frequencies are resolved at $k=2$ and improvement is rapid as k increases. The MEM estimator failed to resolve the two sinusoids.

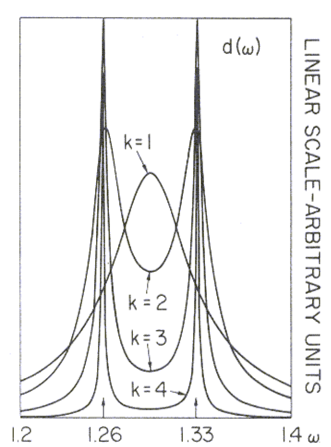

FIG. 8. *Iterative* IPDFT, $d(\omega)=1/|\sum_{n=0}^{14} a_n e^{in\omega}|$, $a=pR^{-k}$, $k=1,2,3,4$, $1.2\leqq\omega\leqq 1.4$ (linear scale).

As with any high resolution method, there is a price to pay. The above analysis depends heavily on our knowing $R_{\eta\eta}$ well. If our guess of $R_{\eta\eta}$ is bad the eigenvalue/eigenvector method can produce completely erroneous answers, whereas the IPDFT, because it is less committed to the specific model of (24), can still produce reasonable estimates.

If, in the eigenvalue/eigenvector method described above, we take $R_{\eta\eta}=I$ and we use all $M+1$ orthogonal eigenvectors of R^{-1}, instead of the "largest" $M+1-J$, weighting each z-transform by its corresponding eigenvalue, we obtain Capon's MLM estimator. In the next section we rederive the MLM as a calibrated or normalized PDFT spectral estimator.

5. The PDFT, minimum weighted leakage and the MLM. In this section we develop a Fourier transform estimator we call the minimum weighted leakage (MWL) estimator. We then relate it to the PDFT method and to Capon's maximum likelihood estimator of the power spectrum ([11], [25]).

First let us return to the problem of estimating $X(\omega)$ from $x(1),\cdots,x(N)$. Suppose we decide that for each ω the estimate $\hat{X}(\omega)$ of $X(\omega)$ should be a linear combination of the data values. That is, we seek an estimate of the form

$$\hat{X}(\omega)=\sum_{n=1}^{N} b_n(\omega)x(n). \tag{33}$$

The PDFT is such an estimator. Now let us hold ω fixed and try to determine what set of coefficients $b_1(\omega),\cdots,b_N(\omega)$ is best to use. If we are mainly concerned with sinusoids and, in particular, with estimating the power in each sinusoidal component, and if we want to eliminate as much "leakage" as possible from the presence of power at one frequency into the estimate of the power at another frequency, then the MWL is a reasonable approach.

For each frequency $-\pi\leqq\alpha\leqq\pi$ write

$$\hat{X}_\alpha(\omega)=\sum_{n=1}^{N} b_n(\omega)\, e^{-in\alpha}. \tag{34}$$

Because of the way we define the FT, a delta-function at frequency α corresponds to the sinusoid $e^{-it\alpha}$. Consequently $\hat{X}_\alpha(\omega)$ is the value of the FT estimate at frequency ω when the data comes from the sinusoid $e^{-it\alpha}$. As the conditions which determine the optimal set of coefficients, $b(\omega)=(b_1(\omega),\cdots,b_N(\omega))$, select

(35)
1) $$\hat{X}_\omega(\omega)=1,$$
2) $$\int_{-\pi}^{\pi}|\hat{X}_\alpha(\omega)|^2 W(\alpha)\,d\alpha/2\pi \quad \text{is minimized,}$$

where $W(\alpha)$ is an arbitrary nonnegative function of our choosing. The quantity $|\hat{X}_\alpha(\omega)|$ we term "leakage" in the estimator. Ideally, if $\alpha\neq\omega$, then $\hat{X}_\alpha(\omega)$ would be zero. A nonzero value indicates the effect of a sinusoidal component at α on our estimate of $X(\omega)$. The aim in 2) is to weight the "leakage" from each frequency α according to its severity (at least as far as we can determine it a priori) and then to minimize over all α. We rewrite (35) as

(36)
1′) $$\sum_{n=1}^{N} b_n(\omega)\, e^{-in\omega}=1,$$
2′) $$\sum_{n=1}^{N}\sum_{m=1}^{N} b_n(\omega)\overline{b_m(\omega)}\int_{-\pi}^{\pi} W(\alpha)\, e^{-i(n-m)\alpha}\,d\alpha/2\pi \quad \text{is minimized.}$$

With

$$w(t)=\int_{-\pi}^{\pi} W(\alpha)\, e^{-it\alpha}\,d\alpha/2\pi, \tag{37}$$

and with the matrix notation

$$W=[W_{n,m}]=[w(n-m)], \qquad m, n=1,\cdots,N,$$

and

$$S_\omega=(e^{i\omega}, e^{i2\omega},\cdots, e^{iN\omega}),$$

(36) becomes

$$(38)\qquad \begin{array}{ll} 1') & b(\omega)\cdot S_\omega^*=1, \\ 2') & b(\omega)\cdot W\cdot b(\omega)^* \quad \text{is minimized.} \end{array}$$

The solution is given by

$$(39)\qquad b(\omega)=\lambda S_\omega W^{-1}, \qquad 1=b(\omega)S_\omega^*=S_\omega W^{-1}S_\omega^*,$$

so that

$$(40)\qquad b(\omega)=S_\omega W^{-1}/S_\omega W^{-1}S_\omega^*.$$

The optimal estimate of $X(\omega)$ is then

$$(41)\qquad X_{\text{MWL}}(\omega)=\hat{X}(\omega)=b(\omega)x^T=xb(\omega)^T,$$

where $x=(x(1),\cdots,x(N))$. This estimator we call the minimum weighted leakage estimator (MWL).

Writing $X_{\text{MWL}}(\omega)$ as

$$(42)\qquad X_{\text{MWL}}(\omega)=\frac{S_\omega W^{-1}x^T}{S_\omega W^{-1}S_\omega^*}=\frac{W(\omega)S_\omega W^{-1}x^T}{W(\omega)S_\omega W^{-1}S_\omega^*}$$

we see that $X_{\text{MWL}}(\omega)$ is related to the PDFT. The numerator, $W(\omega)S_\omega W^{-1}x^T$, is the PDFT estimator corresponding to the data vector x and prior weighting function $P(\omega)=W(\omega)$. The denominator is the same, except that x^T is replaced by S_ω^*. This means that the denominator is a PDFT corresponding to different data at each frequency ω. The MWL estimator is simply a normalized PDFT estimator, in which at each frequency ω the PDFT estimate of $X(\omega)$ is divided by what the PDFT estimate of $X(\omega)$ would have been, if the data had come from the sinusoid $e^{-it\omega}$ instead of from $x(t)$.

Next we apply the MWL to the problem of estimating $R(\omega)$ from $r(m)$, $|m|\leqq M$. Take the data vector to be $x=(r(0), r(1),\cdots, r(M))$, take $S_\omega=(1, e^{i\omega},\cdots, e^{iM\omega})$, and choose a suitable weight function, $W(\omega)$. As seen in (38) not all of $W(\omega)$ actually enters the calculation of the MWL. Only the values $w(0), w(1),\cdots, w(M)$ are involved in the present problem. In that case, let us take as $W(\omega)$ the actual spectrum $R(\omega)$. Surely the actual spectrum will serve as a reasonable weighting of leakage. The MWL estimator that results is

$$(43)\qquad R_{\text{MWL}}(\omega)=1/S_\omega(R^T)^{-1}S_\omega^*,$$

which is the maximum likelihood spectral estimator of Capon (MLM), [11], [25]. The MLM is, in general, not consistent with the data, is smoother than MEM, and is an estimator of the power in spectral peaks [1], [25]. We see that the MLM, as a special case of MWL, is a normalized PDFT estimator, based on the choice of the true spectrum as the weight function. The MLM does not have the resolving capability of MEM or the eigenvector methods, because the normalization causes a smoothing that reduces resolution. The MLM is related to the eigenvector/eigenvalue method in yet another way, as noted above. If, instead of taking only those eigenvectors corresponding to the largest eigenvalue of Q we take $M+1$ orthogonal eigenvectors and $R_{\eta\eta}=I$ and we average the magnitude squared of the z-transforms, weighted according to the square of the corresponding eigenvalue, we get the reciprocal of (43). This is nothing more than diagonalizing $(R^T)^{-1}$. As we mentioned earlier, MLM is designed to estimate well the power in sinusoidal components. The PDFT can also be used for this purpose, and in that regard, is related to the best linear unbiased estimator (BLUE) of regression coefficients. We discuss this briefly in the next section.

6. The PDFT and unbiased estimation in regression. Once the number and locations of the sinusoidal components have been determined we want to estimate the amplitudes of those components, i.e. the $B_j=A_j e^{i\theta_j}$ of (29). For this purpose let us choose as our prior weighting function

$$P(\omega)=2\pi \sum_{j=1}^{J} \delta(\omega+\omega_j)+\varepsilon N(\omega), \tag{44}$$

where $2\pi\delta(\omega+\omega_j)$ corresponds to the sinusoid $e^{-it\omega_j}$, $N(\omega)$ is our prior estimate of the noise spectrum, and $\varepsilon>0$ is an arbitrary scale factor. If $\rho(t)$ is the inverse FT of $N(\omega)$, then the inverse FT of $P(\omega)$ is given by

$$p(t)=\sum_{j=1}^{J} e^{it\omega_j}+\varepsilon\rho(t), \tag{45}$$

and the matrix $P=[p(n-m)]$ can be written as

$$P=FF^*+\varepsilon N, \tag{46}$$

where $N=[\rho(n-m)]$ is the noise autocorrelation matrix, and F is the N by J matrix whose n, j entry is $e^{-in\omega_j}$. The PDFT estimate of $X(\omega)$ now becomes

$$X_{\mathrm{PDFT}}(\omega)=P(\omega) \sum_{n=1}^{N} a_n e^{in\omega}, \tag{47}$$

where

$$a=xP^{-1}=x(FF^*+\varepsilon N)^{-1}.$$

Thus

$$\begin{aligned} X_{\mathrm{PDFT}}(\omega)&=\left(2\pi \sum_{j=1}^{J} \delta(\omega+\omega_j)+\varepsilon N(\omega)\right)\left(\sum_{n=1}^{N} a_n e^{in\omega}\right) \\ &=2\pi \sum_{j=1}^{J}\left(\sum_{n=1}^{N} a_n e^{-in\omega_j}\right)\delta(\omega+\omega_j)+\varepsilon\left(\sum_{n=1}^{N} a_n e^{in\omega}\right)N(\omega). \end{aligned} \tag{48}$$

A reasonable choice for the complex amplitude estimate is then

$$B_j=A_j e^{i\theta_j}\approx 2\pi \sum_{n=1}^{N} a_n e^{-in\omega_j}. \tag{49}$$

The vector A of complex amplitude estimates is then $A = 2\pi a \cdot F$ or

$$A = 2\pi x(FF^* + \varepsilon N)^{-1}F = 2\pi x N^{-1}(FF^*N^{-1} + \varepsilon I)^{-1}F. \tag{50}$$

An easy calculation shows that

$$(FF^*N^{-1} + \varepsilon I)^{-1}F = F(\varepsilon I + F^*N^{-1}F)^{-1}, \tag{51}$$

so that

$$A = 2\pi x N^{-1}F(F^*N^{-1}F + \varepsilon I)^{-1}. \tag{52}$$

As $\varepsilon \downarrow 0$, A approaches the vector

$$A_0 = 2\pi x N^{-1}F(F^*N^{-1}F)^{-1},$$

which is the best (minimum variance) linear unbiased estimate (BLUE) of the amplitudes [31, p. 181]. We may not want to take the unbiased estimator, corresponding as it does to $\varepsilon \downarrow 0$. The ε represents our estimate of the relative level of the noise, and if we have a reasonable estimate that is not 0 we can use it in (52) to obtain a biased estimate of the amplitudes. In choosing a biased estimate we deliberately acknowledge the presence of noise and attempt to estimate its level. The BLUE is designed to provide the exact answer when noise disappears entirely. There is noise, however, and if we can estimate the SNR, the biased estimate will have smaller variance.

7. Conclusion. We have used the theory of best linear approximation in weighted L^2 spaces to obtain the PDFT method for linearly estimating the Fourier transform from finite samples. Special cases of this method are the DFT and the bandlimited extrapolation techniques of Cadzow, Papoulis, Kolba and Parks and others. Using the PDFT indirectly, on autocorrelation data, we rederive and considerably generalize the MEM technique of Burg and the eigenvalue/eigenvector approaches to sinusoids in noise. Using the PDFT directly on autocorrelation data and calibrating the result gives the MWL estimator, a special case of which is the maximum likelihood method of Capon. When used to estimate the amplitudes of known sinusoids, the PDFT provides, in the limit as $\varepsilon \downarrow 0$, the BLUE estimate from the theory of statistical regression, and for larger ε, approximating the true reciprocal SNR, a biased amplitude estimate.

By relating well-known nonlinear spectral techniques to the PDFT we facilitate the analysis of these techniques in terms of approximation theory and provide a much broader class of estimators from which to choose. Because the high resolution spectral estimators are nonlinear functions of the autocorrelation data, much of what is known concerning the behavior of these estimators has come from the study of mathematically tractable special cases (one or two sinusoids in white noise) and from numerical simulations. By basing the choice of spectral estimators on the user's prior knowledge and primary goals, rather than on the selection of models, the task of the would-be user is simplified and the design of an estimator (or estimators) appropriate for the problem at hand is made easier.

Appendix A. MEM and MLM. The MEM spectral estimator of $R(\omega)$, given data $r(0), \cdots, r(M)$, is that nonnegative function $R_{\text{MEM}}(\omega)$ for which the entropy integral,

$$\text{entropy} = \int_{-\pi}^{\pi} \log R_{\text{MEM}}(\omega)\, d\omega,$$

is maximized, subject to the constraints

$$r(m)=\int_{-\pi}^{\pi} R_{\mathrm{MEM}}(\omega)\, e^{-im\omega}\, d\omega/2\pi, \qquad m=-M,\cdots,M.$$

From the calculus of variations it follows that $R_{\mathrm{MEM}}(\omega)$ equals the $R_{\mathrm{IPDFT}}(\omega)$ of (28).

The MLM estimator of $R(\omega)$ is derived in a different manner. For each fixed value of ω we seek a linear filter $b(\omega)=(b_0, b_1,\cdots, b_m)$ that will pass undistorted the component with frequency ω, while optimally suppressing all other components. Viewing $R(\omega)$ as the spectrum of the process to be filtered, we then minimize total output power, $b(\omega)R^Tb(\omega)^*$, subject to $b(\omega)$. $S_\omega^*=1$. The resulting optimal filter is $b(\omega)=S_\omega(R^T)^{-1}/S_\omega(R^T)^{-1}S_\omega^*$ and the minimum output power is then the estimate of $R(\omega)$,

$$R_{\mathrm{MLM}}(\omega)=1/S_\omega(R^T)^{-1}S_\omega^*.$$

Appendix B. Recursive computation of PDFT. Having chosen the weight function $P(\omega)$, we must determine the coefficients $a_1,\cdots,a_N$ in (10) by solving system (12). The special properties of the matrix involved can be exploited to greatly reduce the computational burden. The algorithm we use is similar to others in the literature and is based on the recursive formulae associated with orthogonal polynomials, as discussed in [8].

Suppose we find complex polynomials,

$$f_n(z)=b_0^n+b_1^nz+\cdots+b_n^nz^n,$$

with $b_n=b_n^n>0$ for $n=0,1,\cdots$, and satisfying the orthogonality conditions,

$$\int_{-\pi}^{\pi} f_n(e^{i\omega})\overline{f_m(e^{i\omega})}P(\omega)\, d\omega/2\pi=\begin{cases}1, & m=n,\\ 0, & m\neq n.\end{cases}$$

Then using

$$\sum_{n=1}^{N} a_n\, e^{in\omega}=e^{i\omega}\sum_{n=0}^{N-1} a_{n+1}\, e^{in\omega}$$

and

$$\sum_{n=0}^{N-1} a_{n+1}\, e^{in\omega}=\sum_{n=0}^{N-1} c_nf_n(e^{i\omega})$$

we have that for some choice of coefficients, $c_0,\cdots,c_{N-1}$, the PDFT estimator (10) can be written

$$X_{\mathrm{PDFT}}(\omega)=e^{i\omega}\sum_{n=0}^{N-1} c_nf_n(e^{i\omega})P(\omega), \qquad |\omega|\leqq\pi.$$

The coefficients c_n do not depend on the number N, and the coefficients b_k^n do not depend on the data $x(1),\cdots,x(N)$ and can be computed recursively.

If we denote by $f_n^*(z)$ the function $z^n\overline{f_n(1/z)}$, $f_n^*(0)=b_n$, we have [18]

$$b_n f_{n+1}(z)=b_{n+1}zf_n(z)+f_{n+1}(0)f_n^*(z),$$

and

$$b_n f_{n+1}^*(z)=b_{n+1}f_n^*(z)+\overline{f_{n+1}(0)}zf_n(z).$$

For

$$\lambda_n=\int_{-\pi}^{\pi} e^{i\omega}f_n(e^{i\omega})P(\omega)\,d\omega/2\pi$$

we have, using (11),

$$\lambda_n=\sum_{k=0}^{n} b_k^n\overline{p(k+1)}, \qquad n=0,1,\cdots.$$

It follows that

$$b_{n+1}^2=b_n^2/(1-b_n^2|\lambda_n|^2), \qquad n=0,1,\cdots,$$

where $b_n^2|\lambda_n|^2<1$ if and only if the $(n+1)$ by $(n+1)$ matrix with entries $p(m-k)$, m, $k=0,1,\cdots,n$, is invertible.

REFERENCES

Articles followed by (*) are reproduced in [12].

[1] A. B. BAGGEROER, *Confidence intervals for regression* (MEM) *spectral estimates*, IEEE Trans. Inform. Theory, Vol. IT-22 (Sept. 1976), pp. 534–545. (*)

[2] E. BIENVENU AND L. KOPP, *Adaptive high resolution spatial descrimination of passive sources*, Underwater Acoustics and Signal Processing, D. Reidel, Boston, 1981, pp. 509–515.

[3] T. P. BRONEZ AND J. A. CADZOW, *An algebraic approach to data-adaptive array processing*, Proc. First ASSP Workshop on Spectral Estimation, McMaster University, Hamilton, Ontario, Aug. 1981, pp. 5.2.1.–5.2.8.

[4] J. P. BURG, *Maximum entropy spectrum analysis*, Ph.D. dissertation, Stanford Univ., Palo Alto, Calif., 1975.

[5] ———, *Maximum entropy spectral analysis*, in Proc. 37th Meeting Soc. of Exploration Geophysicists, Oklahoma City, OK, Oct. 1967. (*)

[6] ———, *A new analysis technique for time series data*, NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, the Netherlands, Aug. 1968. (*)

[7] C. L. BYRNE AND R. M. FITZGERALD, *A unifying model for spectral estimation*, in Proc. RADC Spectrum Estimation Workshop, (Oct. 1979), pp. 157–162.

[8] ———, *Reconstruction from partial information, with applications to tomography*, SIAM J. Appl. Math., 42 (Aug. 1982), pp. 933–940.

[9] J. A. CADZOW, *An extrapolation procedure for bandlimited signals*, IEEE Trans. Acoust., Speech, Signal Process., ASSP-27 (Feb. 1979), pp. 4–12.

[10] ———, ARMA *spectral estimation: an iterative procedure*, in Proc. RADC Spectrum Estimation Workshop, (Oct. 1979), pp. 67–80.

[11] J. CAPON, *High resolution frequency-wavenumber spectrum analysis*, Proc. IEEE, 57 (Aug. 1969), pp. 1408–1418. (*)

[12] D. G. CHILDERS (editor), *Modern Spectrum Analysis*, IEEE Press, New York, 1978.

[13] D. G. Childers, J. J. Yea and A. Krishnamurthy, *Spectral analysis:* AR, MA, ARMA, in Proc. of First ASSP Workshop on Spectral Estimation, McMaster University, Hamilton, Ontario, Canada, (Aug. 1981), pp. 2.2.1–2.2.10.

[14] S. D. Conte and C. de Boor, *Elementary Numerical Analysis—An Algorithmic Approach*, (3rd edition), McGraw-Hill, New York, 1980.

[15] T. S. Durrani and A. S. Arslanian, *Windows associated with high resolution spectral estimators*, pre-print.

[16] R. Fitzgerald and C. Byrne, *Extrapolation of bandlimited signals: a tutorial* in Signal Processing: Theories and Applications, M. Kunt and F. de Coulon, eds., North-Holland, Amsterdam, 1980.

[17] W. F. Gabriel, *Spectral analysis and adaptive array superresolution techniques*, Proc. IEEE, Vol. 68 (June 1970), pp. 654–666.

[18] L. Ya. Geronimus, *Orthogonal polynomials* (translated from the Russian), Consultants Bureau, New York (1961).

[19] S. Haykin (editor), *Nonlinear methods of spectral analysis*, Springer-Verlag, New York (1979).

[20] A. K. Jain and S. Ranganath, *Extrapolation algorithms for discrete signals, with application in spectral estimation*, IEEE Trans. Acoust. Speech, Signal Process., Vol. ASSP-29, No. 4 (Aug. 1981), pp. 830–845.

[21] D. Johnson, T. Parks and S. DeGraaf, *Improving the resolution of bearing in passive sonar arrays*, in Proc. First ASSP Workshop on Spectral Estimation, McMaster University, Hamilton, Ontario, Aug. 1981, pp. 5.6.1–5.6.9.

[22] T. Kailath, RKHS *approach to detection and estimation problems—Part* I: *Deterministic signals in Gaussian noise*, IEEE Trans. Inform. Theory, IT-17 (1971), pp. 530–549.

[23] S. M. Kay and S. L. Marple, *Spectrum analysis—a modern perspective*, Proc. IEEE, 69 (1981), pp. 1380–1419.

[24] D. P. Kolba and T. W. Parks, *Extrapolation and spectral estimation for bandlimited signals*, Conference Record IEEE International Conference on Acoustics, Speech and Signal Processing, April 1978, pp. 372–374.

[25] R. T. Lacoss, *Data adaptive spectral analysis methods*, Geophys., 36 (1971), pp. 661–675. (*)

[26] J. Lamperti, *Stochastic Processes: Survey of the Mathematical Theory*, Springer-Verlag, New York, 1977.

[27] S. L. Marple, *A new autoregressive spectrum analysis algorithm*, IEEE Trans. Acoust., Speech, Signal Process, ASSP-28 (1980), pp. 441–454.

[28] A. Papoulis, *A new algorithm in spectral analysis and bandlimited extrapolation*, IEEE Trans. Circuits and Systems, CAS-22, 4 (1975), pp. 735–742.

[29] E. Parzen, *Extraction and detection problems and reproducing kernel Hilbert spaces*, SIAM J. Control, 1 (1962), pp. 35–62.

[30] V. F. Pisarenko, *The retrieval of harmonics from a covariance function*, Geophys. J. Royal Astron. Soc., 33 (1973), pp. 347–366.

[31] C. R. Rao, *Linear Statistical Inference and its Applications*, John Wiley, New York, 1965.

[32] E. A. Robinson, *Iterative and* CDE ARMA *algorithms*, in Proc. First ASSP Workshop on Spectral Estimation, Aug. 1981, pp. 2.3.1–2.3.4.

[33] R. O. Schmidt, *Multiple emitter location and signal parameter estimation*, in Proc. RADC Spectrum Estimation Workshop, Oct. 1979, pp. 243–258.

[34] E. C. Titchmarsh, *The Theory of Functions*, 2nd ed., Oxford Univ. Press, Cambridge, 1939.

[35] T. J. Ulrych and R. W. Clayton, *Time series modelling and maximum entropy*, Physics of Earth, Planetary Interiors, 12 (1976), pp. 188–200.