Spectrum Analysis—A Modern Perspective

STEVEN M. KAY, MEMBER, IEEE, AND STANLEY LAWRENCE MARPLE, JR., MEMBER, IEEE

Abstract—A summary of many of the new techniques developed in the last two decades for spectrum analysis of discrete time series is presented in this tutorial. An examination of the underlying time series model assumed by each technique serves as the common basis for understanding the differences among the various spectrum analysis approaches. Techniques discussed include the classical periodogram, classical Blackman-Tukey, autoregressive (maximum entropy), moving average, autoregressive-moving average, maximum likelihood, Prony, and Pisarenko methods. A summary table in the text provides a concise overview for all methods, including key references and appropriate equations for computation of each spectral estimate.

I. INTRODUCTION

ESTIMATION of the power spectral density (PSD), or simply the spectrum, of discretely sampled deterministic and stochastic processes is usually based on procedures employing the fast Fourier transform (FFT). This approach to spectrum analysis is computationally efficient and produces reasonable results for a large class of signal processes. In spite of these advantages, there are several inherent performance limitations of the FFT approach. The most prominent limitation is that of frequency resolution, i.e., the ability to distinguish the spectral responses of two or more signals. The frequency resolution in hertz is roughly the reciprocal of the time interval in seconds over which sampled data is available. A second limitation is due to the implicit windowing of the data that occurs when processing with the FFT. Windowing manifests itself as "leakage" in the spectral domain, i.e., energy in the main lobe of a spectral response "leaks" into the sidelobes, obscuring and distorting other spectral responses that are present. In fact, weak signal spectral responses can be masked by higher sidelobes from stronger spectral responses. Skillful selection of tapered data windows can reduce the sidelobe leakage, but always at the expense of reduced resolution.

These two performance limitations of the FFT approach are particularly troublesome when analyzing short data records. Short data records occur frequently in practice because many measured processes are brief in duration or have slowly time-varying spectra that may be considered constant only for short record lengths. In radar, for example, only a few data samples are available from each received radar pulse. In sonar, the motion of targets results in a time-varying spectral response due to Doppler effects.

In an attempt to alleviate the inherent limitations of the FFT approach, many alternative spectral estimation procedures have been proposed within the last decade. A comparison of the spectral estimates shown in Fig. 1 illustrate the improvement that may be obtained with nontraditional approaches. The three spectra illustrated were computed using the first nine autocorrelation lags[1] of a process consisting of two equi-amplitude sinusoids at 3 and 4 Hz in additive white noise. The conventional spectral estimate based on the nine known lags $R_{xx}(0), \cdots, R_{xx}(8)$ is shown in Fig. 1(a). The spectrum is a plot of 512 values obtained by application of a 512-point FFT to the nine lags, zero-padded with 503 zeros. This spectrum, often termed the Blackman–Tukey (BT) estimate of the PSD, is characterized by sidelobes, some of which produce negative values for the PSD, and by an inability to distinguish the two sinusoidal responses.

Fig. 1(b) shows the spectral response of the autoregressive (AR) method based on the same nine lags. The improvement in resolution over that shown in Fig. 1(a) has contributed to the popularity of this alternative spectral estimate. Although the AR spectral estimate was originally developed for geophysical data processing, where it was termed the maximum entropy method (MEM) [16], [37]–[39], [50], [84], [136], [138], [158], [221], [231], [246], [247], it has been used for applications in radar [75], [92], [99], [116], [125], [126], [216], sonar [122], [198], imaging [98], radio astronomy [162], [264], [265], biomedicine [71], [74], oceanography [96], ecological systems [88], and direction finding [70], [128], [233]. The AR approach to spectrum analysis is closely related to linear prediction coding (LPC) techniques used in speech processing [80], [130], [143], [145]. The AR PSD estimator fits an AR model to the data. The origin of AR models may be found in economic time series forecasting [31], [276] and statistical estimation [189]–[191]. The MEM approach makes different assumptions about the lags, but for practical purposes, the MEM and AR spectral estimators are identical for one-dimensional analysis of wide sense stationary, Gaussian processes.

The ultimate resolution of the two sinusoidal signals into two delta function responses in a uniform spectral floor, representing the white noise PSD level, is achieved with the Pisarenko harmonic decomposition (PHD) method shown in Fig. 1(c). This technique yields the most accurate estimate of the spectrum of sinusoids in noise, at least when the autocorrelation lags are known.

As evidenced by the spectrum examples of Fig. 1, the development of alternative spectral estimates in widely different application areas has led to a confusion of conflicting terminology and different algorithm development viewpoints. Thus

Manuscript received September 8, 1980; revised June 2, 1981. The submission of this paper was encouraged after the review of an advanced proposal.

S. M. Kay is with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI 02881.

S. L. Marple, Jr. is with the Analytic Sciences Corporation, McLean Operation, 8301 Greensboro Drive, Suite 1200, McLean, VA 22102.

[1] The autocorrelation function $R_{xx}(k)$ of a stochastic wide sense stationary discrete process x_n at lag k is defined in this paper as the expectation of the product $x_{n+k}x_n^*$, or $R_{xx}(k) = E[x_{n+k}x_n^*]$, where x_n is assumed to have zero mean. The $*$ denotes complex conjugate, since complex processes are assumed in general, and $E(\ )$ denotes the expectation operator.

0018-9219/81/1100-1380$00.75

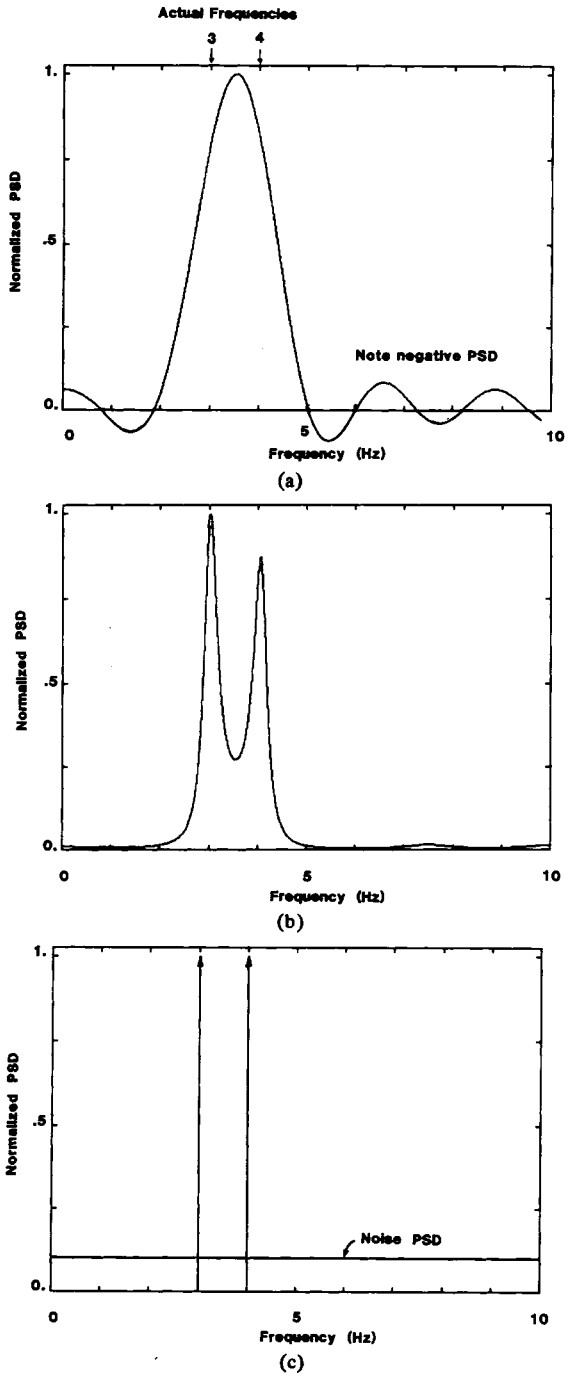

Fig. 1. Examples of three spectral estimates based on nine known autocorrelation lags of a process consisting of two equi-amplitude sinusoids in additive white noise (the variance of the noise is 10 percent of the sinusoid power). (a) BT PSD. (b) Autoregressive PSD. (c) Pisarenko harmonic decomposition PSD.

two purposes of this review are 1) to establish a common framework of terminology and symbols and 2) to unify the various approaches and algorithm developments that have evolved in various disciplines.

Claims have been made concerning the degree of improvement obtained in the spectral resolution and the signal detectability when AR and Pisarenko techniques are applied to sampled data [36], [206], [250], [251]. These performance advantages, though, strongly depend upon the signal-to-noise ratio (SNR), as might be expected. In fact, for low enough SNR's the modern spectral estimates are often no better than those obtained with conventional FFT processing [122], [150]. Even in those cases where improved spectral fidelity is achieved by use of an alternative spectral estimation procedure, the computational requirements of that alternative method may be significantly higher than FFT processing. This may make some modern spectral estimators unattractive for real-time implementation. Thus a third objective of this paper is to present tradeoffs among the various techniques. In particular, the performance advantages and disadvantages will be highlighted for each method, the computational complexity will be summarized, and criteria will be presented for determining if the selected spectral estimator is appropriate for the process being analyzed.

Some historical perspective is instructive for an appreciation of the basis for modern spectral estimation. The illustrious history of the Fourier transform can be traced back over 200 years [34], [223]. The advent of spectrum analysis based on Fourier analysis can be traced to Schuster, who was the first to coin the term "periodogram" [218], [219]. Schuster made a Fourier series fit to the variation in sun-spot numbers in an attempt to find "hidden periodicities" in the measured data. The next pioneering step was described in Norbert Wiener's classic paper on "generalized harmonic analysis" [269]. This work established the theoretical framework for the treatment of stochastic processes by using a Fourier transform approach. A major result was the introduction of the autocorrelation function of a random process and its Fourier transform relationship with the power spectral density. Khinchin [127] defined a similar relationship independently of Wiener.

Blackman and Tukey, in a classical publication in 1958 [25], provided a practical implementation of Wiener's autocorrelation approach to power spectrum estimation when using sampled data sequences. The method first estimates the autocorrelation lags from the measured data, windows (or tapers) the autocorrelation estimates in an appropriate manner, and then Fourier transforms the windowed lag estimates to obtain the PSD estimate. The BT approach was the most popular spectral estimation technique until the introduction of the FFT algorithm in 1965, generally credited to Cooley and Tukey [53]. This computationally efficient algorithm renewed an interest in the periodogram approach to PSD estimation. The periodogram spectral estimate is obtained as the squared magnitude of the output values from an FFT performed directly on the data set (data may be weighted). Currently, the periodogram is the most popular PSD estimator [17], [24], [32], [105]-[107], [109].

Conventional FFT spectral estimation is based on a Fourier series model of the data, that is, the process is assumed to be composed of a set of harmonically related sinusoids. Other time series models have been used in nonengineering fields for many years. Yule [276] and Walker [258] both used AR models to forecast trends in economic time series. Baron de Prony [202] devised a simple procedure for fitting exponential models to data obtained from an experiment in gas chemistry. Other models have arisen in the statistical and numerical analysis fields. The modern spectral estimators have their roots in these nonengineering fields of time series modeling.

The use of nontraditional spectral estimation techniques in a significant manner began in the 1960's. Parzen [189], in 1968, formally proposed AR spectral estimation. Independently in 1967, Burg [37] introduced the maximum entropy method, motivated by his work with linear prediction filtering

in geoseismological applications. The one-dimensional MEM was shown formally by Van den Bos [255] to be equivalent to the AR PSD estimator. Prony's method also bears some mathematical similarities to the AR estimation algorithms. An area of current research is that of autoregressive-moving average (ARMA) models. The ARMA model is a generalization of the AR model. It appears that methods based upon these may provide even better resolution and performance than AR methods. The PHD [194], [195] is one example of a spectral estimation technique based upon a special case ARMA model.

The unifying approach employed in this paper is to view each spectral estimation technique as being based on the fitting of measured data to an assumed model. The variations in performance among the various spectral estimates may often be attributed to how well the assumed model matches the process under analysis [173]. Different models may yield similar results, but one may require fewer model parameters and is therefore more efficient in its representation of the process. Spectral estimates of various techniques computed from samples of a process consisting of sinusoids in colored Gaussian noise are presented in Section III to illustrate these variations. The process has both narrow-band and broad-band components. This process helps to illustrate how some spectral estimates tend to better estimate the narrow-band components while other spectral estimates better estimate the broad-band components of the spectra. This example process emphasizes the need to understand the underlying model before passing judgement on a spectral estimation method.

This tutorial is divided into five sections. Section II is the largest section. It contains a tutorial review of all the methods considered in this paper. Section III provides a summary table and illustration that highlights and compares the various modern spectral estimation methods. Section IV briefly examines other application areas that utilize the spectral estimation methods discussed in this paper.

A table of contents of these three sections is included below to enable the reader to quickly locate topics of interest.

No discussion of band-limited extrapolation techniques for spectral estimation is presented here since a good tutorial is already available [103]. The conclusion, Section V, makes observations concerning trends in research and application of modern spectral estimation.

II. Review of Spectral Estimation Techniques

A. Spectral Density Definitions and Basics

Traditional spectrum estimation, as currently implemented using the FFT, is characterized by many tradeoffs in an effort to produce statistically reliable spectral estimates. There are tradeoffs in windowing, time-domain averaging, and frequency-domain averaging of sampled data obtained from random processes in order to balance the needs to reduce sidelobes, to perform effective ensemble averaging, and to ensure adequate spectral resolution. To summarize the basics of conventional spectrum analysis, consider first the case of a deterministic analog waveform $x(t)$, that is a continuous function of time. For generality, $x(t)$ will be considered complex-valued in this paper. If $x(t)$ is absolute integrable, i.e., the signal energy $\mathcal{E}$ is finite

$$\mathcal{E} = \int_{-\infty}^{\infty} |x(t)|^2 \, dt < \infty \tag{2.1}$$

then the continuous Fourier transform (CFT) $X(f)$ of $x(t)$ exists and is given by

$$X(f) = \int_{-\infty}^{\infty} x(t) \exp(-j2\pi ft) \, dt. \tag{2.2}$$

(Note that (2.1) is a sufficient, but not a necessary condition for the existence of a Fourier transform [33].) The squared modulus of the Fourier transform is often termed the spectrum, $\delta(f)$, of $x(t)$,

$$\delta(f) = |X(f)|^2. \tag{2.3}$$

Parseval's energy theorem, expressed as

$$\int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |X(f)|^2 \, df \tag{2.4}$$

is a statement of the conservation of energy; the energy of the time domain signal is equal to the energy of the frequency domain transform, $\int_{-\infty}^{\infty} \delta(f) \, df$. Thus $\delta(f)$ is an *energy spectral density* (ESD) in that it represents the distribution of energy as a function of frequency. If the signal $x(t)$ is sampled at equispaced intervals of Δt s to produce a discrete sequence $x_n = x(n\Delta t)$ for $-\infty < n < \infty$, then the sampled sequence can be represented as the product of the original time function $x(t)$ and an infinite set of equispaced Dirac delta functions $\delta(t)$. The Fourier transform of this product may be written, using distribution theory [33], as

$$X'(f) = \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} x(t)\delta(t - n\Delta t)\Delta t \right] \exp(-j2\pi ft) \, dt$$

$$= \Delta t \sum_{n=-\infty}^{\infty} x_n \exp(-j2\pi fn\Delta t). \tag{2.5}$$

Expression (2.5) corresponds to a rectangular integration approximation of (2.2); the factor Δt ensures conservation of

integrated area between (2.2) and (2.5) as $\Delta t \to 0$. Expression (2.5) will be identical in value to the transform $X(f)$ of (2.2) over the interval $-1/(2\Delta t) \leqslant f \leqslant 1/(2\Delta t)$ Hz, as long as $x(t)$ is band limited and all frequency components are in this interval. Thus the continuous energy spectral density

$$\mathcal{S}'(f) = |X'(f)|^2 \tag{2.6}$$

for data sampled from a band-limited process is identical to that of (2.3).

If a) the data sequence is available from only a finite time window over $n = 0$ to $n = N - 1$, and b) the transform is discretized also for N values by taking samples at the frequencies $f = m\Delta f$ for $m = 0, 1, \cdots, N - 1$ where $\Delta f = 1/N\Delta t$, then one can develop the familiar discrete Fourier transform (DFT) [33] from (2.5),[2]

$$X_m = \Delta t \sum_{n=0}^{N-1} x_n \exp(-j2\pi m\Delta f n\Delta t)$$

$$= \Delta t \sum_{n=0}^{N-1} x_n \exp(-j2\pi mn/N)$$

$$\text{for } m = 0, \cdots, N - 1. \tag{2.7}$$

Both (2.7) and its associated inverse transform are cyclic with period N. Thus by using (2.7), we have forced a periodic extension to both the discretized data and the discretized transform values, even though the original continuous data may not have been periodic. A discrete ESD may then be defined as

$$\mathcal{S}_m = |X_m|^2 \tag{2.8}$$

also for $0 \leqslant m \leqslant N - 1$. Both the discrete $\mathcal{S}_m$ and the continuous $\mathcal{S}'(f)$ have been termed *periodogram* spectral estimates. Note however that $\mathcal{S}_m$ and $\mathcal{S}'(f)$, when evaluated at $f = m/N\Delta t$ for $m = 0, \cdots, N - 1$, do not yield identical values. $\mathcal{S}_m$ is, in effect, a sampled version of a spectrum determined from the convolution of $X(f)$ with the transform of the rectangular window that contains the data samples. Thus the discrete spectrum $\mathcal{S}_m$ based on a finite data set is a distorted version of the continuous spectrum $\mathcal{S}'(f)$ based on an infinite data set.

A different viewpoint must be taken when the process $x(t)$ is a wide sense stationary, stochastic process rather than a deterministic, finite-energy waveform. The energy of such processes are usually infinite, so that the quantity of interest is the power (time average of energy) distribution with frequency. Also, integrals such as (2.2) normally do not exist for a stochastic process. For the case of stationary random processes, the autocorrelation function

$$R_{xx}(\tau) = E[x(t + \tau)x^*(t)] \tag{2.9}$$

provides the basis for spectrum analysis, rather than the random process $x(t)$ itself. The Wiener–Khinchin theorem relates $R_{xx}(\tau)$ via the Fourier transform to $\mathcal{P}(f)$, the PSD,

$$\mathcal{P}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) \exp(-j2\pi f\tau)\, d\tau. \tag{2.10}$$

[2] The inverse transform is given by $x_n = \Delta f \sum_{m=0}^{N-1} X_m \exp(+j2\pi mn/N)$ and the energy theorem is

$$\sum_{n=0}^{N-1} |x_n|^2 \Delta t = \sum_{n=0}^{N-1} |X_m|^2 \Delta f.$$

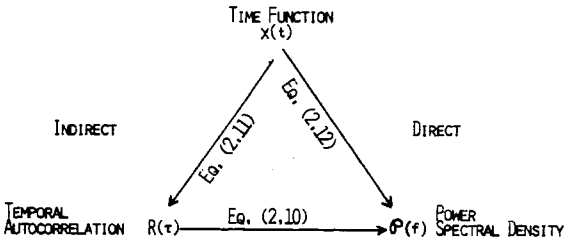

Fig. 2. Direct and indirect methods to obtain PSD (stationary and ergodic properties assumed).

As a practical matter, one does not usually know the statistical autocorrelation function. Thus an additional assumption often made is that the random process is ergodic in the first and second moments. This property permits the substitution of time averages for ensemble averages. For an ergodic process, then, the statistical autocorrelation function may be equated to

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t + \tau)x^*(t)\, dt. \tag{2.11}$$

It is possible to show [107], [132], [187], with the use of (2.11), that (2.10) may be equivalently expressed as

$$\mathcal{P}(f) = \lim_{T\to\infty} E\left\{\frac{1}{2T}\left|\int_{-T}^{T} x(t) \exp(-j2\pi ft)\, dt\right|^2\right\}. \tag{2.12}$$

The expectation operator is required since the ergodic property of $R_{xx}(\tau)$ does not couple through the Fourier transform; that is, the limit in (2.12) without the expected value does not converge in any statistical sense. Fig. 2 depicts the direct and indirect approaches to obtain the PSD from the signal $x(t)$, based on the formal relationships (2.10), (2.11), and (2.12).

Difficulties may arise if (2.12) is applied to finite data sets without regard to the expectation and limiting operations. Statistically inconsistent (unstable) estimates result if no statistical averaging is performed; i.e., the variance of the PSD estimate will not tend to zero as T increases without bound [183].

B. Traditional Methods

Two spectral estimation techniques based on Fourier transform operations have evolved. The PSD estimate based on the indirect approach via an autocorrelation estimate was popularized by Blackman and Tukey [14]. The other PSD estimate, based on the direct approach via an FFT operation on the data, is the one typically referred to as the periodogram.

With a finite data sequence, only a finite number of discrete autocorrelation function values, or lags, may by estimated. Blackman and Tukey proposed the spectral estimate

$$\hat{\mathcal{P}}_{BT}(f) = \Delta t \sum_{n=-M}^{M} \hat{R}_{xx}(m) \exp(-j2\pi fm\Delta t) \tag{2.13}$$

based on the available autocorrelation lag estimates $\hat{R}_{xx}(m)$, where $-1/(2\Delta t) \leqslant f \leqslant 1/(2\Delta t)$ and $\hat{}$ denotes an estimate. This spectral estimate is the discrete-time version of the Wiener–Khinchin expression (2.10). An obvious companion autocorrelation estimate, based on (2.11), is the unbiased estimator

$$\hat{R}_{xx}(m) = \frac{1}{N - m} \sum_{n=0}^{N-m-1} x_{n+m}x_n^* \tag{2.14}$$

for $m = 0, \cdots, M$, where $M \leqslant N-1$. The negative lag estimates are determined from the positive lag estimates as follows:

$$\hat{R}_{xx}(-m) = \hat{R}^*_{xx}(m) \tag{2.15}$$

in accordance with the conjugate symmetric property of the autocorrelation function of a stationary process. Instead of (2.14), both Jenkins-Watts [107] and Parzen [188], [189] provide arguments for the use of the autocorrelation estimate

$$\hat{R}'_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x_{n+m} x_n^* \tag{2.16}$$

defined for $m = 0, \cdots, M$, since it tends to have less mean-square error than (2.14) for many finite data sets. $\hat{R}'_{xx}(m)$ is a biased estimator since $E[\hat{R}'_{xx}(m)] = [(N-m)/N]\, R_{xx}(m)$. The mean value is a triangular window weighting (sometimes called a Bartlett weighting) of the true autocorrelation function.

The direct method of spectrum analysis is the modern version of Schuster's periodogram. A sampled data version of expression (2.12), for which measured data is available only for samples $x_0, \cdots, x_{N-1}$, is

$$\hat{\mathcal{P}}_{\mathrm{PER}}(f) = \frac{1}{N\Delta t} \left| \Delta t \sum_{n=0}^{N-1} x_n \exp(-j2\pi f n \Delta t) \right|^2 \tag{2.17}$$

also defined for the frequency interval $-1/(2\Delta t) \leqslant f \leqslant 1/(2\Delta t)$. Note that the expectation operation in (2.12) has been ignored for the moment. Use of the fast Fourier transform (FFT) will permit evaluation of (2.17) at the discrete set of N equally spaced frequencies $f_m = m\Delta f$ Hz, for $m = 0, 1, \cdots, N-1$ and $\Delta f = 1/N\Delta t$,

$$\hat{\mathcal{P}}_m = \hat{\mathcal{P}}_{\mathrm{PER}}(f_m) = \frac{1}{N\Delta t} |X_m|^2 \tag{2.18}$$

where X_m is the DFT of (2.7). $\hat{\mathcal{P}}_m$ is identical to the energy spectral density δ_m of (2.8) except for the division by the time interval of $N\Delta t$ seconds required to make $\hat{\mathcal{P}}_m$ a power spectral density. The total power in the process, which is assumed periodic due to the DFT property, is

$$\text{Power} = \sum_{m=0}^{N-1} \hat{\mathcal{P}}_m \Delta f \tag{2.19}$$

based on rectangular integration approximation of $\hat{\mathcal{P}}_{\mathrm{PER}}$. If the Δf factor is incorporated into $\hat{\mathcal{P}}_m$, then

$$\hat{\hat{\mathcal{P}}}_m = \hat{\mathcal{P}}_m \Delta f = \frac{1}{(N\Delta t)^2} |X_m|^2$$

$$= \left| \frac{1}{N} \sum_{n=0}^{N-1} x_n \exp(-j2\pi m n/N) \right|^2. \tag{2.20}$$

This is the quantity often computed as the periodogram, but it is not scaled appropriately as a PSD. Using (2.20), it is the *peak* in the PSD plot, rather than the *area* under the plot, that is equal to the power of the assumed periodic signal. The computational economy of the FFT algorithm has made this approach a popular one.

Often a periodogram of N data samples is computed using (2.18) when the measured process has deterministic components imbedded in random noise. As pointed out earlier, care must be taken since statistically inconsistent results can occur if (2.18) is used literally without regard to the expectation operation. This need for some sort of ensemble averaging, or smoothing of the sample spectrum, is illustrated by Oppenheim and Schafer [183, p. 546] and Otnes and Enochson [184, p. 328] with examples of a white-noise process in which the variance of the spectral estimate does not decrease, even though longer and longer data sequences are used. Bartlett [18] had recognized the statistical problems with (2.18) and suggested splitting the data into segments, computing $\hat{\mathcal{P}}_m$ for each segment, and averaging the periodograms of all segments. Welch [262], [263] suggested a special digital procedure with the FFT that involves averaging periodograms.

Other mechanisms for approximating an ensemble average make use of windows in the time or frequency domain, or both [183], [184]. Overlapped weighted segment averaging is advocated by Nuttall and Carter [43], [44], [175], [179], [181] to give stability and to minimize the impact of window sidelobes.

Other references for the FFT and its application for PSD estimation may be found in Bergland [19], Bertram [22], [23], Brigham and Morrow [32], [33], Cochran *et al.* [52], Cooley *et al.* [54], [55], Glisson *et al.* [77], Nuttall and Carter [43], [181], Richards [207], Rife and Vincent [208], Webb [259], and Yuen [274], [275].

In general, the spectral estimates $\hat{\mathcal{P}}_{\mathrm{BT}}(f)$ and $\hat{\mathcal{P}}_{\mathrm{PER}}(f)$ are not identical. However, if the biased autocorrelation estimate (2.16) is used and as many lags as data samples ($M = N-1$) are computed, then the BT estimate and the periodogram estimate yield identical numerical results [184]. Thus the periodogram can be viewed as a special case of the BT procedure. It is for this reason that the BT and periodogram estimates are occasionally termed taper and transform (TT) approaches [116].

Many of the problems of the periodogram PSD estimation technique can be traced to the assumptions made about the data outside the measurement interval. The finite data sequence may be viewed as being obtained by windowing an infinite length sample sequence with a boxcar function. The use of only this data implicitly assumes the unmeasured data to be zero, which is usually not the case. This multiplication of the actual time series by a window function means the overall transform is the convolution of the desired transform with the transform of the window function. If the true power of a signal is concentrated in a narrow bandwidth, this convolution operation will spread that power into adjacent frequency regions. This phenomena, termed leakage, is a consequence of the tacit windowing inherent in the computation of the periodogram.

In addition to the distorting effects of leakage on the spectral estimate, leakage has a detrimental impact on power estimation and detectability of sinusoidal components [201], [224], [234]. Sidelobes from adjacent frequency cells add in a constructive or destructive manner to the main lobe of a response in another frequency cell of the spectrum, affecting the estimate of power in that cell. In extreme cases, the sidelobes from strong frequency components can mask the main lobe of weak frequency components in adjacent cells, as illustrated in Fig. 3. Sidelobes characteristic of the $\sin \pi f/\pi f$ function (the transform of a rectangular time-domain window) are evident in this illustration.

Data windowing is also the fundamental factor that determines the frequency resolution of the periodogram. The convolution of the window transform with that of the actual signal transform means that the most narrow spectral response of the resultant transform is limited to that of the main-lobe width of

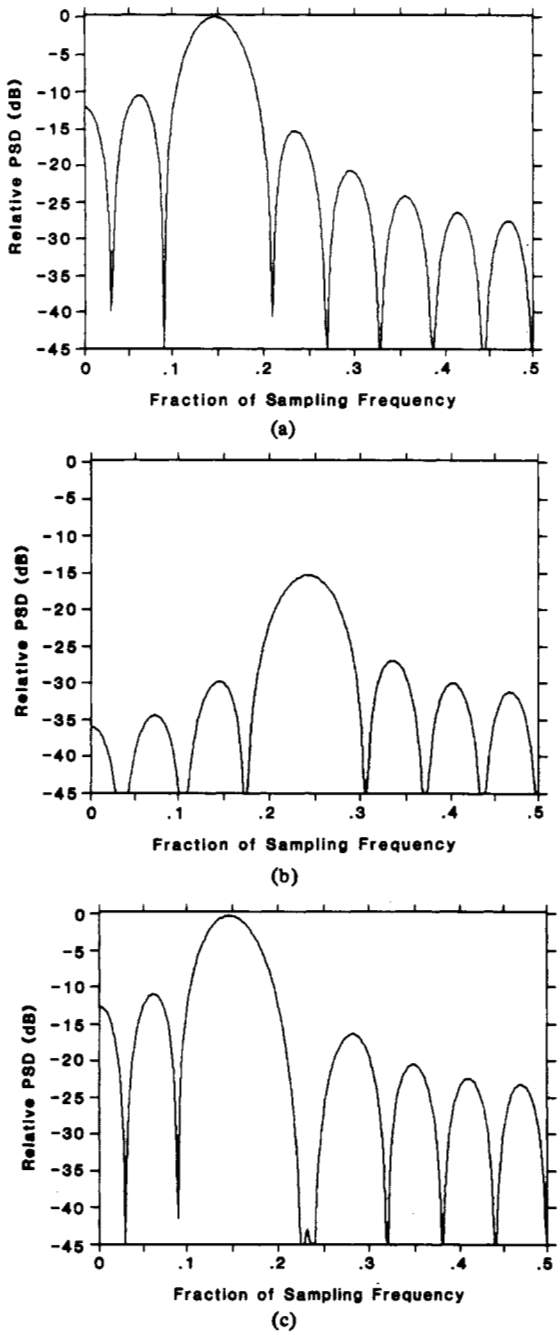

Fig. 3. Illustration of weak signal masking by adjacent strong signals when using periodogram spectrum analysis. Number of samples used in spectra (a)-(c) was 16. (a) PSD of a single sinusoid of amplitude unity, fractional sampling frequency 0.15, and initial phase 45°. (b) PSD (relative to previous PSD level) of a single sinusoid of amplitude 0.19, fractional sampling frequency 0.24, and initial phase 162°. (c) Combined signal PSD—note that there is little response at the weaker signal frequency location.

the window transform, independent of the data. For a rectangular window, the main-lobe width between 3-dB levels (and therefore, the resolution) of the resulting $(\sin \pi f)/\pi f$ transform is approximately the inverse of the observation time of $N\Delta t$ seconds. Other windows may be used, but the resolution will always be proportional to $1/N\Delta t$ Hz. Leakage effects due to data windowing can be reduced by the selection of windows with nonuniform weighting. Harris [90] has provided a good summary of the merits of various windows. Nuttall [182] provides a correction to the sidelobe behavior for some of the windows described by Harris. Other references are [12], [61], [161], [224], [260], [273]. No attempt is made here to summarize the relative merits of various window functions. The price paid for a reduction in the sidelobes is always a broadening in the main lobe of the window transform, which in turn means a decrease in the resolution of the spectral estimate.

There is a common misconception that zero-padding the data sequence before transforming will improve the resolution of the periodogram. Transforming a data set with zeros only serves to interpolate additional PSD values within the frequency interval $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$ Hz between those that would be obtained with a non-zero-padded transform. Fig. 4 shows periodograms of an N-point data set with no zero padding, data padded with N zeros, $7N$ zeros, and $31N$ zeros. In each case, the additional values of the periodogram, computed by an FFT applied to the zero-padded data set, fill in the shape of the continuous-frequency periodogram as defined by expression (2.17). In no case of zero padding, however, is there an improvement in the fundamental frequency resolution (reciprocal of the measurement interval). Zero padding is useful for 1) smoothing the appearance of the periodogram estimate via interpolation, 2) resolving potential ambiguities as illustrated in Fig. 4, and 3) reducing the "quantization" error in the accuracy of estimating the frequencies of spectral peaks. Mathematically, a $2N$-point DFT of a $2N$-point sequence $x_0, \cdots, x_{2N-1}$ is

$$X_m = \Delta t \sum_{n=0}^{2N-1} x_n \exp(-j2\pi mn/2N) \qquad (2.21)$$

for $m = 0, 1, \cdots, 2N-1$.. If the data set had been zero padded with N zeros, $x_n = 0$ for $n = N, \cdots, 2N-1$, then (2.21) becomes

$$X_m = \Delta t \sum_{n=0}^{N-1} x_n \exp\left(-j2\pi \left[\frac{m}{2}\right] n/N\right) \qquad (2.22)$$

which is the same as the N-point transform (2.7), but evaluated over the interval $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$ at twice as many frequencies as (2.7). By eliminating the operations on zeros introduced by zero padding, or pruning as it is called, a more efficient FFT algorithm is possible [148].

In summary, the conventional BT and periodogram approaches to spectral estimation have the following advantages: 1) computationally efficient if only a few lags are needed (BT) or if the FFT is used (periodogram), 2) PSD estimate directly proportional to the power for sinusoid processes, and 3) a good model for some applications (to be explained in more detail in the next section). The disadvantages of these techniques are: 1) suppression of weak signal main-lobe responses by strong signal sidelobes, 2) frequency resolution limited by the available data record duration, independent of the characteristics of the data or its SNR, 3) introduction of distortion in the spectrum due to sidelobe leakage, 4) need for some sort of pseudo ensemble averaging to obtain statistically consistent periodogram spectra, and 5) the appearance of negative PSD values with the BT approach when some autocorrelation sequence estimates are used.

C. Modeling and the Parameter Identification Approach

At this point, we shall depart from the traditional perspective of spectrum analysis as presented in the last section. The conventional approach used FFT operations on either windowed data or windowed lag estimates. Windowing of data or

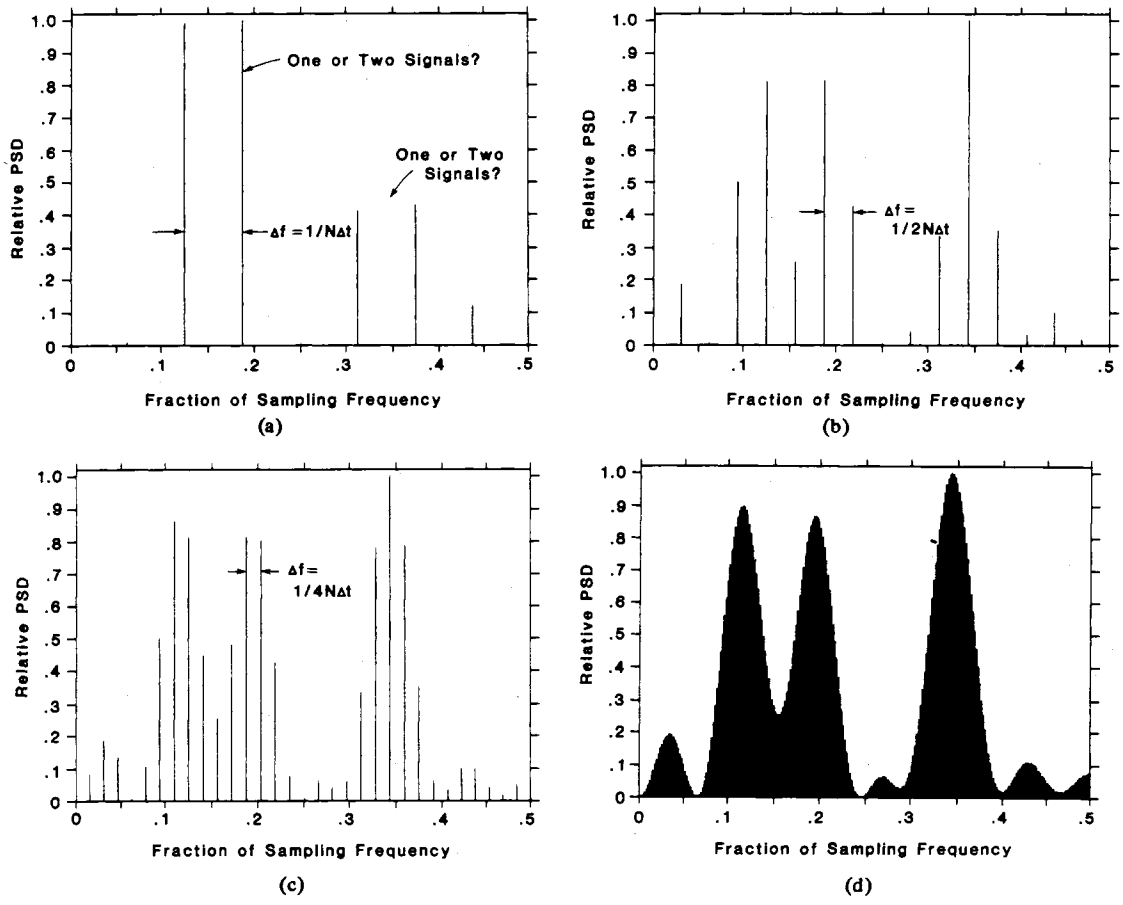

Fig. 4. Impact of zero padding the periodogram to interpolate the spectral shape and to resolve ambiguities. The spectra were estimated using the same 16 samples of a process consisting of three sinusoids of fractional sampling frequencies 0.1335, 0.1875, 0.3375 and initial phases 0°, 90°, 0°, respectively. (a) No zero padding; ambiguities are present in the spectrum. (b) Double padding; ambiguities resolved. (c) Quadruple padding; smoothest spectrum seen. (d) 32-times padding; envelope is approximation to continuous Fourier transform.

lags makes the implicit assumption that the unobserved data or lag values outside the window are zero, which is normally an unrealistic assumption. A smeared spectral estimate is a consequence of the windowing.

Often one has more knowledge about the process from which the data samples are taken, or at least is able to make a more reasonable assumption other than to assume the data is zero outside the window. Use of *a priori* information (or assumptions) may permit selection of an exact model for the process that generated the data samples, or at least a model that is a good approximation to the actual underlying process. It is then usually possible to obtain a better spectral estimate based on the model by determining the parameters of the model from the observations. Thus spectrum analysis, in the context of modeling, becomes a three step procedure. The first step is to select a time series model. The second step is to estimate the parameters of the assumed model using either the available data samples or autocorrelation lags (either known or estimated from the data). The third step is to obtain the spectral estimate by substituting the estimated model parameters into the theoretical PSD implied by the model. One major motivation for the current interest in the modeling approach to spectral estimation is the higher frequency resolution achievable with these modern techniques over that achievable with the traditional techniques previously discussed. The degree of improvement in resolution and spectral fidelity, if any, will be determined by the ability to fit an assumed model with a few parameters to the measured data. The selection of a model for the spectral estimate is intimately tied to estimation and identification techniques employed in linear system theory [271], [272].

To illustrate the modeling viewpoint of spectral estimation, the discrete periodogram PSD estimate (2.18) will be shown to be equivalent to a least squares fit of the data to a harmonic model, namely the discrete Fourier series. The least squares fit to a Fourier series is well known [28]. Only the essential ideas are presented here. If N samples $x_0, \cdots, x_{N-1}$ of a continuous-time process $x(t)$ are modeled by a discrete sequence $\hat{x}_n$ composed of N complex sinusoids of arbitrary frequencies $f_0, \cdots, f_{N-1}$, then

$$\hat{x}(n\Delta t) = \hat{x}_n = \sum_{m=0}^{N-1} a_m \exp(j2\pi f_m n\Delta t) \qquad (2.23)$$

for $n = 0, \cdots, N-1$. Thus the signal $x(t)$ over the interval $N\Delta t$ seconds is represented with periodic functions, whether or not $x(t)$ is itself cyclic. The weights a_m, to be determined, are assumed to be complex-valued for generality. The N terms of (2.23) can be expressed in matrix form as

$$\hat{X} = \Phi A \qquad (2.24)$$

where

$$\hat{X} = \begin{bmatrix} \hat{x}_0 \\ \vdots \\ \hat{x}_{N-1} \end{bmatrix}, \quad A = \begin{bmatrix} a_0 \\ \vdots \\ a_{N-1} \end{bmatrix}, \quad \Phi = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \exp(\lambda_0) & \exp(\lambda_1) & \cdots & \exp(\lambda_{N-1}) \\ \vdots & \vdots & & \vdots \\ \exp(\lambda_0[N-1]) & \exp(\lambda_1[N-1]) & \cdots & \exp(\lambda_{N-1}[N-1]) \end{bmatrix}$$

and $\lambda_i = j2\pi f_i \Delta t$. Given the N frequencies f_m, the amplitude vector A determined by minimizing the total squared estimation error,

$$\sum_{n=0}^{N-1} |x_n - \hat{x}_n|^2 \tag{2.25}$$

is provided by the well-known solution

$$A = (\Phi^H \Phi)^{-1} \Phi^H X \tag{2.26}$$

where X is the data vector

$$X = \begin{bmatrix} x_0 \\ \vdots \\ x_{N-1} \end{bmatrix}$$

and H denotes complex conjugate transpose.

The Fourier series fit selects the N sinusoidal frequencies to be the preassigned, harmonically related frequencies $f_m = m\Delta f$ Hz, where $\Delta f = 1/N\Delta t$. It is well known [28] that such a selection of harmonic frequencies makes each column (row) vector of Φ orthogonal to all other column (row) vectors so that

$$(\Phi^H \Phi)^{-1} = \frac{1}{N} I \tag{2.27}$$

where I is the identity matrix. The amplitude vector A is then given by

$$A = \frac{1}{N} \Phi^H X \tag{2.28}$$

or

$$a_m = \frac{1}{N} \sum_{n=0}^{N-1} x_n \exp(-j2\pi mn/N)$$

for $m = 0, \cdots, N-1$. The power of the sinusoidal component at the preassigned frequency f_m is

$$|a_m|^2 = \left| \frac{1}{N} \sum_{n=0}^{N-1} x_n \exp(-j2\pi mn/N) \right|^2 \tag{2.29}$$

which is identical to expression (2.20). Thus the discrete periodogram spectral estimate may be viewed as a least squares fit of a harmonic set of complex sinusoids to the data.

The case where frequencies are not preselected and are not necessarily harmonic is treated in Section II-J, which provides a discussion of Prony's method. The harmonic model preassigned the frequencies and number of sinusoids so that only estimation of the sinusoidal powers was necessary. The nonharmonic model of Prony's method will require estimation of not only the powers, but also the number of sinusoids present and their frequencies. Another aspect of the harmonic model that is noteworthy is the fact that noise is not accounted for in the model. Any noise present must also be modeled by the harmonic sinusoids. Thus, to decrease the fluctuations due to noise, one must average over a set of periodograms made from the data.

One key feature of the modeling approach to spectral estimation that differentiates it from the general identification problem is that only the output process of the model is available for analysis; the input driving process is not assumed available as it is for general system identification.

One of the promising aspects of the modeling approach to spectral estimation is that one can make more realistic assumptions concerning the nature of the measured process outside the measurement interval, other than to assume it is zero or cyclic. Thus the need for window functions can be eliminated, along with their distorting impact. As a result, the improvement over the conventional FFT spectral estimate can be quite dramatic, especially for short data records.

D. Rational Transfer Function Modeling Methods

Many deterministic and stochastic discrete-time processes encountered in practice are well approximated by a rational transfer function model. In this model, an input driving sequence $\{n_n\}$ and the output sequence $\{x_n\}$ that is to model the data are related by the linear difference equation,

$$x_n = \sum_{l=0}^{q} b_l n_{n-l} - \sum_{k=1}^{p} a_k x_{n-k}. \tag{2.30}$$

This most general linear model is termed an ARMA model. The interest in these models stems from their relationship to linear filters with rational transfer functions.

The system function $H(z)$ between the input n_n and output x_n for the ARMA process of (2.30) is the rational expression

$$H(z) = \frac{B(z)}{A(z)} \tag{2.31}$$

where

$$A(z) = z\text{ - transform of AR branch} = \sum_{m=0}^{p} a_m z^{-m}$$

$$B(z) = z\text{ - transform of MA branch} = \sum_{m=0}^{q} b_m z^{-m}.$$

It is well known that the power spectrum at the output of a linear filter, $P_x(z)$, is related to the power spectrum of the input stochastic process, $P_n(z)$, as follows:

$$P_x(z) = H(z) H^*(1/z^*) P_n(z) = \frac{B(z) B^*(1/z^*)}{A(z) A^*(1/z^*)} \cdot P_n(z). \tag{2.32}$$

Expression (2.32) is normally evaluated along the unit circle, $z = \exp(j2\pi f \Delta t)$ for $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$. Often the driving process is assumed to be a white-noise sequence of zero mean and variance σ^2. The PSD of the noise is then $\sigma^2 \Delta t$. (Note that we have included the Δt factor in the expression for

power spectral density of the noise so that $P_x(\exp[j2\pi f\Delta t])$, when integrated over $-1/2\Delta t \leq f \leq 1/2\Delta t$, yields the true power of an analog signal). The PSD of the ARMA output process is then

$$\mathcal{P}_{\mathrm{ARMA}}(f) = \mathcal{P}_x(f) = \sigma^2\Delta t|\mathcal{B}(f)/\mathcal{A}(f)|^2 \qquad (2.33)$$

where $\mathcal{A}(f) = A(\exp[j2\pi f\Delta t])$ and $\mathcal{B}(f) = B(\exp[j2\pi f\Delta t])$. Specification of the parameters $\{a_k\}$ (termed the autoregressive coefficients), the parameters $\{b_k\}$ (termed the moving-average coefficients), and σ^2 is equivalent to specifying the spectrum of the process $\{x_n\}$. Without loss of generality, one can assume $a_0 = 1$ and $b_0 = 1$ since any filter gain can be incorporated into σ^2.

If all the $\{a_k\}$ terms except $a_0 = 1$ vanish, then

$$x_n = \sum_{l=0}^{q} b_l n_{n-l} \qquad (2.34)$$

and the process is strictly a moving average of order q, and

$$\mathcal{P}_{\mathrm{MA}}(f) = \sigma^2\Delta t|\mathcal{B}(f)|^2. \qquad (2.35)$$

This model is sometimes termed an all-zero model [266].

If all the $\{b_i\}$, except $b_0 = 1$, are zero, then

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} + n_n \qquad (2.36)$$

and the process is strictly an autoregression of order p. The process is termed AR in that the sequence x_n is a linear regression on itself with n_n representing the error. With this model, the present value of the process is expressed as a weighted sum of past values plus a noise term. The PSD is

$$\mathcal{P}_{\mathrm{AR}}(f) = \frac{\sigma^2\Delta t}{|\mathcal{A}(f)|^2}. \qquad (2.37)$$

This model is sometimes termed an all-pole model.

The Wold decomposition theorem [266] relates the ARMA, MA, and AR models. Basically, the theorem asserts that any stationary ARMA or MA process of finite variance can be represented as a unique AR model of possibly infinite order; likewise, any ARMA or AR process can be represented as a MA process of possibly infinite order. This theorem is important because if we choose the wrong model among the three, we may still obtain a reasonable approximation by using a high order. Thus an ARMA model can be approximated by an AR model of higher order. Since the estimation of parameters for an AR model results in linear equations, as will be shown, it has a computational advantage over ARMA and MA parameter estimation techniques. The largest portion of research effort on rational transfer function modeling has therefore been concerned with the AR model.

E. Autoregressive PSD Estimation

Introduction: Since this section is detailed, reflecting the extensive research on this PSD estimation method, it is worthwhile to briefly outline the material to be presented. The Yule–Walker equations are first derived. They describe the linear relationship between the AR parameters and the autocorrelation function. The solution of these equations is provided by the computationally efficient Levinson–Durbin algorithm, the details of which reveal some fundamental properties of AR processes.

Next, the relationship between AR modeling, linear prediction theory, and maximum entropy spectral estimation (MESE) is examined. Selection of the AR model order and the associated AR parameters is then addressed. Included in the discussion are batch estimation techniques based upon linear prediction theory and sequential estimation methods based on recursive least squares and adaptive algorithms. Tradeoffs between the various approaches are noted.

Finally, some limitations of AR spectral estimation that reduce its applicability in practice are described. These involve the degrading effect of observation noise, spurious peaks, and some anomolous effects which occur when the data are dominated by sinusoidal components. Some techniques for reducing these effects are presented.

Yule–Walker Equations: If an autoregression is a reasonable model for the data, then the AR power spectral density estimate based on (2.37) may be rewritten as

$$\mathcal{P}_{\mathrm{AR}}(f) = |H(\exp[j2\pi f\Delta t])|^2\mathcal{P}_n(f) = \frac{\sigma^2\Delta t}{\left|1 + \sum_{k=1}^{p} a_k \exp(-j2\pi fk\Delta t)\right|^2}. \qquad (2.38)$$

Thus, to estimate the PSD one need only estimate $\{a_1, a_2, \cdots, a_p, \sigma^2\}$. To do this, a relationship between the AR parameters and the autocorrelation function (known or estimated) of x_n is now presented. This relationship is known as the Yule–Walker equations [31]. The derivation proceeds as follows:

$$R_{xx}(k) = E[x_{n+k}x_n^*] = E\left[x_n^*\left(-\sum_{l=1}^{p} a_l x_{n-l+k} + n_{n+k}\right)\right]$$

$$= -\sum_{l=1}^{p} a_l R_{xx}(k-l) + E[n_{n+k}x_n^*].$$

Since $H(z)$ is assumed to be a stable, causal filter, we have

$$E(n_{n+k}x_n^*) = E\left[n_{n+k}\sum_{l=0}^{\infty} h_l^* n_{n-l}^*\right]$$

$$= \sum_{l=0}^{\infty} h_l^*\sigma^2\delta_{k+l}$$

$$= \sigma^2 h_{-k}^*$$

$$= \begin{cases} 0, & \text{for } k > 0 \\ h_0^*\sigma^2, & \text{for } k = 0. \end{cases}$$

Note that δ_m is the discrete delta function, i.e., $\delta_m = 1$ if $m = 0$ or 0 if $m \neq 0$. But $h_0 = \lim_{z\to\infty} H(z) = 1$, and therefore,

$$R_{xx}(k) = \begin{cases} -\sum_{l=1}^{p} a_l R_{xx}(k-l), & \text{for } k > 0 \\ -\sum_{l=1}^{p} a_l R_{xx}(-l) + \sigma^2, & \text{for } k = 0. \end{cases} \qquad (2.39)$$

Expression (2.39) is the Yule–Walker equations. To determine the AR parameters, one need only choose p equations from (2.39) for $k > 0$, solve for $\{a_1, a_2, \cdots, a_p\}$, and then find σ^2 from (2.39) for $k = 0$. The set of equations which require the fewest lags of the autocorrelation function is the selection

$k = 1, 2, \cdots, p$. They can be expressed in matrix form as

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \cdots & R_{xx}(-(p-1)) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(-(p-2)) \\ \vdots & \vdots & & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ \vdots \\ R_{xx}(p) \end{bmatrix}. \tag{2.40}$$

Note that the above autocorrelation matrix, $\boldsymbol{R}_{xx}$, is Hermitian ($\boldsymbol{R}_{xx}^H = \boldsymbol{R}_{xx}$) and it is Toeplitz since the elements along any diagonal are identical. Also, the matrix is positive definite (assuming x_n is not purely harmonic) which follows from the positive definite property of the autocorrelation function [41], [81], [92].

It should be noted that (2.40) can also be augmented to incorporate the σ^2 equation, yielding

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \cdots & R_{xx}(-p) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(-(p-1)) \\ & & & \vdots \\ R_{xx}(p) & R_{xx}(p-1) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.41}$$

which follows from (2.39). This form will be useful later. Thus, to determine the AR parameters and σ^2, one must solve (2.41) with the $p+1$ estimated autocorrelation lags $R_{xx}(0), \cdots, R_{xx}(p)$ and use $R_{xx}(-m) = R_{xx}^*(m)$.

Levinson–Durbin Algorithm [56], [60], [142], [270]: The Levinson–Durbin algorithm provides an efficient solution for (2.41). The algorithm requires only order p^2 operations, denoted $o(p^2)$, as opposed to $o(p^3)$ for Gaussian elimination. Although appearing at first to be just an efficient algorithm, it reveals fundamental properties of AR processes. The algorithm proceeds recursively to compute the parameter sets $\{a_{11}, \sigma_1^2\}, \{a_{21}, a_{22}, \sigma_2^2\}, \cdots, \{a_{p1}, a_{p2}, \cdots, a_{pp}, \sigma_p^2\}$. Note that an additional subscript has been added to the AR coefficients to denote the order. The final set at order p is the desired solution. In particular, the recursive algorithm is initialized by

$$a_{11} = -R_{xx}(1)/R_{xx}(0) \tag{2.42}$$

$$\sigma_1^2 = (1 - |a_{11}|^2) R_{xx}(0) \tag{2.43}$$

with the recursion for $k = 2, 3, \cdots, p$ given by

$$a_{kk} = -\left[R_{xx}(k) + \sum_{l=1}^{k-1} a_{k-1,l} R_{xx}(k-l) \right] \Big/ \sigma_{k-1}^2 \tag{2.44}$$

$$a_{ki} = a_{k-1,i} + a_{kk} a_{k-1,k-i}^* \tag{2.45}$$

$$\sigma_k^2 = (1 - |a_{kk}|^2) \sigma_{k-1}^2. \tag{2.46}$$

It is important to note that $\{a_{k1}, a_{k2}, \cdots, a_{kk}, \sigma_k^2\}$, as obtained above, is the same as would be obtained by using (2.41) for $p = k$. Thus the Levinson–Durbin algorithm also provides the AR parameters for all the lower order AR model fits to the data. This is a useful property when one does not know *a priori* the correct model order, since one can use (2.42)–(2.46) to generate successively higher order models until the modeling error σ_k^2 is reduced to a desired value. In particular, if a process is actually an AR(p) process (an AR process of order p), then $a_{p+1,k} = a_{pk}$ for $k = 1, 2, \cdots, p$ and hence $a_{p+1,p+1} = 0$. In general for an AR(p) process, $a_{kk} = 0$ and $\sigma_k^2 = \sigma_p^2$ for $k > p$. Hence, the variance of the excitation noise is a constant for a model order equal to or greater than the correct order. Thus the point at which σ_k^2 does not change would appear to be a good indicator of the correct model order. It can be shown that $|a_{kk}| \leq 1$, so that $\sigma_{k+1}^2 \leq \sigma_k^2$ [9], [41]. This means that σ_k^2 first reaches its minimum at the correct model order. This point is discussed further under the topic of model order determination.

TABLE I
SUMMARY OF AR PROCESS PROPERTIES (EXCLUDING PURELY HARMONIC PROCESSES)

- Autocorrelation matrix is positive definite: $\mathbf{X}^H \mathbf{R}_{xx} \mathbf{X} > 0$ for all $\mathbf{X}$ vectors
- Reflection coefficient sequence satisfies $|K_i| < 1$ for $i = 1, 2, \ldots, p$
- Zeros of $A(z)$ lie within unit circle: $|z_i| < 1$ for $i = 1, 2, \ldots, p$
- Prediction error powers monotically decrease: $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_p^2 \geq 0$

The parameters $\{a_{11}, a_{22}, \cdots, a_{pp}\}$ are often called the reflection coefficients and are designated as $\{K_1, K_2, \cdots, K_p\}$. They have the property that for $\{R_{xx}(0), R_{xx}(1), \cdots, R_{xx}(p)\}$ to be a valid autocorrelation sequence, i.e., the autocorrelation matrix is positive semidefinite, then it is necessary and sufficient that $|a_{kk}| = |K_k| \leq 1$ for $k = 1, 2, \cdots, p$ [41]. Furthermore, a necessary and sufficient condition for the poles of $A(z)$ to be on or within the unit circle of the z plane is $|K_k| \leq 1$ for $k = 1, 2, \cdots, p$ [41], [140]. It should be noted that if $|K_k| = 1$ for some k, then the recursion (2.42)–(2.45) must terminate since $\sigma_k^2 = 0$. The process in this case is purely harmonic (consists only of sinusoids). These properties are summarized in Table I.

The problem of AR parameter identification is closely related to the theory of linear prediction. Assume x_n is an AR(p) process. If one wishes to predict x_n on the basis of the previous p samples [92],

$$\hat{x}_n = -\sum_{k=1}^{p} \alpha_k x_{n-k} \tag{2.47}$$

then $\{\alpha_1, \alpha_2, \cdots, \alpha_p\}$ can be chosen to minimize the prediction error power Q_p where

$$Q_p = E[|x_n - \hat{x}_n|^2]. \tag{2.48}$$

By the orthogonality principle [187]

$$E[(x_n - \hat{x}_n) x_k^*] = 0, \qquad \text{for } k = n-1, \cdots, n-p$$

or

$$R_{xx}(k) = -\sum_{l=1}^{p} \alpha_l R_{xx}(k-l), \quad \text{for } k = 1, 2, \cdots, p. \tag{2.49}$$

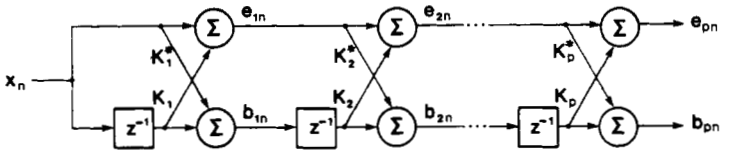

Fig. 5. Lattice formulation of prediction error (whitening, or inverse) filter.

The minimum prediction error power is

$$Q_{p_{\min}} = E[(x_n - \hat{x}_n)x_n^*] = R_{xx}(0) + \sum_{k=1}^{p} \alpha_k R_{xx}(-k). \quad (2.50)$$

These equations are identical to (2.39). Thus it must be true that $\alpha_k = a_{pk}$ for $k = 1, \cdots, p$ and $Q_{p_{\min}} = \sigma_p^2$, so that the best linear predictor is just $\hat{x}_n = -\sum_{k=1}^{p} a_{pk} x_{n-k}$. The error sequence, although uncorrelated with the linear estimate, is not necessarily a white process (it will be if x_n is a AR(p) process). In the limit as $p \to \infty$, the error sequence becomes white.

Note that $\{a_{k1}, a_{k2}, \cdots, a_{kk}\}$ and σ_k^2 constitute the parameters for the optimum kth-order linear predictor and the corresponding minimum prediction error power, respectively. Therefore, AR parameter identification and linear prediction of an AR(p) process yield identical results and the theory of one is applicable to the other.

The theory of linear prediction lends an important interpretation to the Levinson–Durbin algorithm. Denote the prediction error for a pth-order linear predictor as e_{pn}. Then, using (2.45)

$$e_{pn} = x_n + \sum_{k=1}^{p} a_{pk} x_{n-k} \quad (2.51)$$

$$= x_n + \sum_{k=1}^{p-1} (a_{p-1,k} + K_p a_{p-1,p-k}^*) x_{n-k} + K_p x_{n-p} \quad (2.52)$$

$$= e_{p-1,n} + K_p b_{p-1,n-1}$$

where

$$b_{pn} = x_{n-p} + \sum_{k=1}^{p} a_{pk}^* x_{n-p+k}. \quad (2.53)$$

The term b_{pn} is the backward prediction error, i.e., the error when one attempts to "predict" x_{n-p} on the basis of samples $x_{n-p+1}, \cdots, x_n$. It is seen that the predictor coefficients for the backwards predictor are complex conjugates of those of the forward predictor, which is a consequence of the stationary autocorrelation function assumption [41]. Similarly, it can be shown

$$b_{pn} = b_{p-1,n-1} + K_p^* e_{p-1,n}. \quad (2.54)$$

The relationships of (2.52) and (2.54) give rise to the so called lattice filter structure shown in Fig. 5 [146]. Note that the transfer function of the entire filter is just

$$A(z) = 1 + \sum_{k=1}^{p} a_{pk} z^{-k}$$

which is the inverse of $H(z) = 1/A(z)$. This follows from (2.51). This filter is termed either the "inverse" filter or "prediction error" filter. The lattice filter interpretation of the Levinson–Durbin algorithm leads to an important time recursive formulation for estimating the AR parameters. By updating the reflection coefficients using (2.52) and (2.54), the AR parameters are obtained via the order recursion (2.45). More details are presented under the topic of sequential estimation of AR parameters.

If one minimizes $E(|e_{pn}|^2)$, as given by (2.52), with respect to K_p, one obtains

$$K_p = \frac{-E(e_{p-1,n} b_{p-1,n-1}^*)}{E(|b_{p-1,n-1}|^2)}. \quad (2.55)$$

As stated previously, both e_{pn} and b_{pn} have the same statistical properties, so that (2.55) can be written as

$$K_p = \frac{-E(e_{p-1,n} b_{p-1,n-1}^*)}{\sqrt{E(|e_{p-1,n}|^2) E(|b_{p-1,n-1}|^2)}}. \quad (2.56)$$

Thus K_p is the negative of the normalized correlation coefficient between $e_{p-1,n}$ and $b_{p-1,n-1}$, so we must have $|K_p| \leq 1$. In the statistical literature, $-K_p$ is known as a partial correlation coefficient since it is the normalized correlation between x_n and x_{n-p} with the correlation of $x_{n-1}, x_{n-2}, \cdots, x_{n-p+1}$ removed [31].

Maximum Entropy Spectral Estimation [41], [62], [78], [171], [172], [205], [215], [248], [249]: MESE is based upon an extrapolation of a segment of a known autocorrelation function for lags which are not known. In this way the characteristic smearing of the estimated PSD due to the truncation of the autocorrelation function can be removed. If we assume $\{R_{xx}(0), R_{xx}(1), \cdots, R_{xx}(p)\}$ are known, the question arises as to how $\{R_{xx}(p+1), R_{xx}(p+2), \cdots\}$ should be specified in order to guarantee that the entire autocorrelation sequence is positive semi-definite. In general, there are an infinite number of possible extrapolations, all of which yield valid autocorrelation functions. Burg [37], [41] argued that the extrapolation should be made so that the time series characterized by the extrapolated autocorrelation function has maximum entropy. The time series will then be the most random one which has the known autocorrelation lags for its first $p+1$ lags. Alternately, the power spectral density is the one with the flattest (whitest) spectrum of all spectra for which $\{R_{xx}(0), R_{xx}(1), \cdots, R_{xx}(p)\}$ is equal to the known lags. The resultant spectral estimate is termed the maximum entropy spectral estimate. The rationale for the choice of the maximum entropy criterion is that it imposes the fewest constraints on the unknown time series by maximizing its randomness, thereby producing a minimum bias solution.

In particular, if one assumes a Gaussian random process, then the entropy per sample is proportional to

$$\int_{-1/2\Delta t}^{1/2\Delta t} \ln \mathcal{P}_x(f)\, df \quad (2.57)$$

where $\mathcal{P}_x(f)$ is the PSD of x_n. $\mathcal{P}_x(f)$ is found by maximizing (2.57) subject to the constraints that the $(p+1)$ known lags

satisfy the Wiener–Khinchin relationship [62],

$$\int_{-1/2\Delta t}^{1/2\Delta t} \mathcal{P}_x(f) \exp(-j2\pi fn\Delta t)\, df = R_{xx}(n), \quad \text{for } n = 0, 1, \cdots, p. \quad (2.58)$$

The solution is found by the Lagrange multiplier technique and is

$$\mathcal{P}_x(f) = \frac{\sigma_p^2 \Delta t}{\left|1 + \sum_{k=1}^{p} a_{pk} \exp(-j2\pi fk\Delta t)\right|^2} \quad (2.59)$$

where $\{a_{p1}, \cdots, a_{pp}\}$ and σ_p^2 are just the pth-order predictor parameters and prediction error power, respectively, With knowledge of $\{R_{xx}(0), R_{xx}(1), \cdots, R_{xx}(p)\}$, the MESE will be equivalent to an AR PSD, with $\mathcal{P}_x(f)$ given by (2.59) and based on parameters found by solving (2.41). The maximum entropy relationship to AR PSD analysis is only valid for Gaussian random processes and known autocorrelation lags.

AR Autocorrelation Extension: An alternative representation for (2.38) is [59]

$$\mathcal{P}_{AR}(f) = \frac{\sigma_p^2 \Delta t}{\left|1 + \sum_{k=1}^{p} a_{pk} \exp(-j2\pi fk\Delta t)\right|^2} = \Delta t \sum_{n=-\infty}^{\infty} r_{xx}(n) \exp(-j2\pi fn\Delta t) \quad (2.60)$$

where

$$r_{xx}(n) = \begin{cases} R_{xx}(n), & \text{for } |n| \leq p \\ -\sum_{k=1}^{p} a_{pk}\, r_{xx}(n-k), & \text{for } |n| > p. \end{cases} \quad (2.61)$$

From this, it is easy to see that the AR PSD preserves the known lags and recursively extends the lags beyond the window of known lags. The AR PSD function (2.60) summation is identical to the BT PSD function (2.13) up to lag p, but continues with an infinite extrapolation of the autocorrelation function rather than windowing it to zero. Thus AR spectra do not exhibit sidelobes due to windowing. Also, it is the implied extrapolation given by (2.61) that is responsible for the high resolution property of the AR spectral estimator.

AR Parameter Estimation: In most practical situations, one has data samples rather than known autocorrelation lags available for the spectral estimation procedure. To obtain reliable estimates of the AR parameters, standard statistical estimation theory can be used. The usual estimator for a nonrandom set of parameters is the maximum likelihood estimator (MLE). However, the exact solution of the MLE for the parameters of an AR(p) process is difficult to obtain [31]. If $N \gg p$, an approximate MLE can be found which amounts to nothing more than solving the Yule–Walker equation with the autocorrelation function replaced by a suitable estimate. For long data records, this AR parameter estimator produces good spectral estimates. For short data records, which are more commonly encountered in practice (and for which the periodogram spectral estimate has the poorest resolution), the use of the Yule–Walker approach produces poor resolution spectral estimates.

To improve upon the approximate MLE approach for short data sets, a variety of batch estimation techniques based on least squares techniques have been proposed that operate on a block of data samples. For longer data sets, a variety of sequential estimation techniques are available for updating the AR estimates as new data are received. These techniques are especially useful for tracking processes that slowly vary with time. The next two topic areas cover batch and sequential AR estimation methods.

Many additional references on the statistical properties of AR spectral estimates may be found in [13], [100], [116], [186], [212]. Some references dealing with frequency estimation accuracy of AR spectral estimates are [124], [212], [239]. It has been empirically shown by Sakai [212] that the frequency variance of an AR spectral estimate is inversely proportional to both the data length and the square of the SNR; Keeler [124] has empirical evidence that the variance is inverse to both the length and the SNR (rather than the square of SNR).

Batch Estimation of the AR Parameters: Although the Yule–Walker equations could be solved using lag estimates, several least squares estimation procedures are available that operate directly on the data to yield better AR parameter estimates. These techniques often produce better AR spectra than that achievable with the Yule–Walker approach. Two types of least squares estimators will be considered. The first type utilizes forward linear prediction for the estimate, while the second type employs a combination of forward and backward linear prediction.

Assume the data sequence $x_0, \cdots, x_{N-1}$ is used to find the pth-order AR parameter estimates. The foward linear predictor will have the usual form

$$\hat{x}_n = -\sum_{k=1}^{p} a_{pk} x_{n-k}. \quad (2.62)$$

The prediction is forward in the sense that the prediction for the current data sample is a weighted sum of p previous samples. The forward linear prediction error is

$$e_{pn} = x_n - \hat{x}_n = \sum_{k=0}^{p} a_{pk} x_{n-k} \quad (2.63)$$

where

$$a_{p0} = 1.$$

We may compute e_{pn} for $n = 0$ to $n = N + p - 1$ if one assumes the terms outside the measurements are zero, i.e., $x_n = 0$ for $n < 0$ and $n > N - 1$. There is an implied windowing of the data sequence in order to extend the index range for e_{pn} from 0 to $N + p - 1$. Using a matrix formulation for (2.63),

$$\boldsymbol{E} = \begin{bmatrix} e_0 \\ \vdots \\ e_p \\ \vdots \\ e_{N-1} \\ \vdots \\ e_{N+p-1} \end{bmatrix} = \boldsymbol{X}_2 \boldsymbol{X}_1 = \begin{bmatrix} x_0 & & \\ \vdots & \ddots & \\ x_p & \cdots & x_0 \\ \vdots & & \vdots \\ x_{N-1} & \cdots & x_{N-p-1} \\ & \ddots & \vdots \\ & & x_{N-1} \end{bmatrix} \boldsymbol{X}_3 \boldsymbol{X}_4 \begin{bmatrix} 1 \\ a_{p1} \\ \vdots \\ a_{pp} \end{bmatrix} = \boldsymbol{X}\boldsymbol{A} \quad (2.64)$$

or

$$\boldsymbol{E} = \boldsymbol{X}\boldsymbol{A}. \tag{2.65}$$

The prediction error energy is simply

$$\mathcal{E}_p = \sum_n |e_{pn}|^2 = \sum_n \left| \sum_{k=0}^{p} a_{pk} x_{n-k} \right|^2. \tag{2.66}$$

The summation range for $\mathcal{E}_p$ is purposely not specified for the moment. To minimize $\mathcal{E}_p$, the derivatives of $\mathcal{E}_p$ with respect to the $\{a_{pk}\}$ are set to zero and the resultant equations solved for the AR parameters. The result is

$$\sum_{k=0}^{p} a_{pk} \left(\sum_n x_{n-k} x_{n-i}^* \right) = 0, \qquad \text{for } 1 \leq i \leq p \tag{2.67}$$

with minimum error energy,

$$\mathcal{E}_p = \sum_{k=0}^{p} a_{pk} \left(\sum_n x_{n-k} x_n^* \right). \tag{2.68}$$

Expressions (2.67) and (2.68) can be reformulated in normal equation form

$$(\boldsymbol{X}_l^H \boldsymbol{X}_l)\boldsymbol{A} = \begin{pmatrix} \mathcal{E}_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.69}$$

for which four special indexing ranges $l = 1, 2, 3, 4$ are selected, as indicated in (2.64). Note that (2.69) has the same structure as (2.41); however, the data matrix product $(\boldsymbol{X}_l^H \boldsymbol{X}_l)$ is not necessarily Toeplitz as are the Yule-Walker equations.

If the data matrix $\boldsymbol{X}_1$ is selected, the normal (2.69) are termed the covariance equations, often encountered in LPC of speech (see Makhoul [144]). If the data matrix $\boldsymbol{X}_2$ is selected, the resulting normal equations are called autocorrelation equations since the product matrix $(\boldsymbol{X}_2^H \boldsymbol{X}_2)/N$ reduces exactly to the Yule-Walker equations, for which the biased autocorrelation estimator (2.16) has been used in lieu of the known autocorrelation function. Note that a data window has been assumed for this case. This data window reduces the resolution of AR spectra estimated with data matrix $\boldsymbol{X}_2$, as will be illustrated in Section III. If the data matrix $\boldsymbol{X}_3$ is selected, the normal equations are termed the prewindowed normal equations due to the zero value assumptions made for the missing data prior to x_0. If the data matrix $\boldsymbol{X}_4$ is selected, the normal equations are termed the postwindowed normal equations since a zero data assumption is made for the data beyond x_{N-1}.

It would appear that only the data matrix $\boldsymbol{X}_2$ will yield normal equations with Toeplitz structure to permit an efficient recursive solution (namely, the Levinson recursion); as outlined in Fig. 6. However, even though the product matrix $(\boldsymbol{X}_i^H \boldsymbol{X}_i)$ may not be Toeplitz, each of the four data matrices $\boldsymbol{X}_i$ have Toeplitz structure. This property allows one to develop recursive algorithms with $o(p^2)$ operations in each of the four cases. Morf *et al.* have provided the details of the recursive algorithms for the covariance case [163], [164], [167], [169] and for the prewindowed case [66], [168]. The interested reader may consult these references for details of these algorithms. They are not discussed in detail here because the forward and backward prediction approaches, to be discussed next, yield better spectral estimates in most cases.

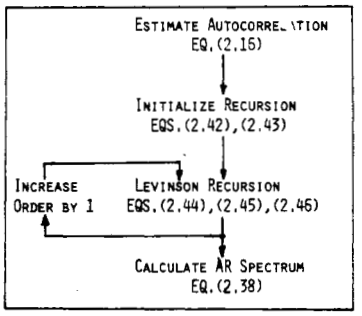

Fig. 6. Summary of Levinson recursion algorithm for AR spectral estimation.

Many problems with the forward only prediction approach to AR spectral estimation exist. The autocorrelation matrix $(\boldsymbol{X}_2^H \boldsymbol{X}_2)$ solution yields AR spectra with the least resolution among these four AR least squares estimates when the data sequence is short. The decrease in resolution is due to the inherent windowing in the data matrix $\boldsymbol{X}_2$. The covariance matrix $(\boldsymbol{X}_1^H \boldsymbol{X}_1)$ solution produces AR parameters whose resulting spectra have been observed by several authors [227], [228], [251] to have more false peaks and greater perturbations of spectral peaks from their correct frequency locations than other AR estimation approaches. The covariance normal equations, which are also used in the Prony method (discussed in Section II-J), lead to AR parameter estimates with greater sensitivity to noise. Spectral line splitting, the placement of two or more closely spaced peaks in the spectrum where only one should be present, has been observed in all four forward prediction cases considered here. The reasons for line splitting have been documented by Kay and Marple [120].

If the process is wide sense stationary, the coefficients of the optimum backward prediction error filter are identical to the coefficients of the optimum forward prediction error filter, but conjugated and reversed in time [41]. The use of the backward prediction errors when estimating the AR parameters was introduced by Burg [37], [41].

The most popular approach for AR parameter estimation with N data samples was introduced by Burg in 1967. The Burg algorithm, separate and distinct from the maximum entropy viewpoint discussed earlier, may be viewed as a constrained least squares minimization. Assuming a wide sense stationary process, the forward linear prediction error of (2.63) is defined for $p \leq n \leq N-1$ (p is the predictor order) and the backward linear prediction error is given by

$$b_{pn} = \sum_{k=0}^{p} a_{pk}^* x_{n-p+k} \tag{2.70}$$

also for $p \leq n \leq N-1$. Recall that a_{po} is defined as unity, the a_{pk} are the predictor parameters at order p, the x_n are the data samples, and the nonwindowed prediction error range has been assumed.

To obtain estimates of the predictor (or AR) parameters, Burg minimized the sum of the forward and backward prediction error energies,

$$\mathcal{E}_p = \sum_{n=p}^{N-1} |e_{pn}|^2 + \sum_{n=p}^{N-1} |b_{pn}|^2 \tag{2.71}$$

subject to the constraint that the AR parameters satisfy the

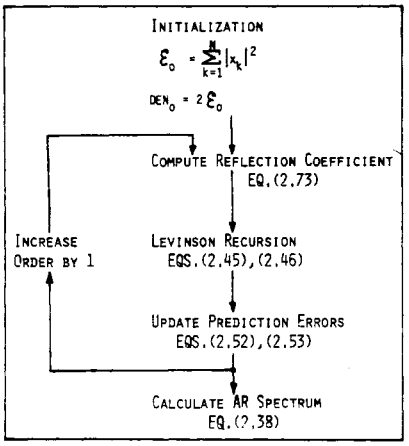

Fig. 7. Summary of Burg algorithm for AR spectral estimation.

Levinson recursion

$$a_{pk} = a_{p-1,k} + a_{pp}a^*_{p-1,p-k} \tag{2.72}$$

for all orders from 1 to p. This constraint was motivated by Burg's desire to ensure a stable AR filter (poles within the unit circle). By substituting the lattice recursion expressions (2.52) and (2.54) into (2.71), $\mathcal{E}_p$ becomes a function of only the unknown reflection coefficient a_{pp} and the prediction errors at order $p-1$, which are assumed known. Thus one need only estimate a_{ii} for $i = 1, 2, \cdots, p$ if (2.72) is used. Setting the derivative of $\mathcal{E}_i$ with respect to a_{ii} to zero then yields

$$a_{ii} = \hat{K}_i = \frac{-2\sum_{k=i}^{N-1} b^*_{i-1,k-1}e_{i-1,k}}{\sum_{k=i}^{N-1}(|b_{i-1,k-1}|^2 + |e_{i-1,k}|^2)}. \tag{2.73}$$

Note that $|a_{ii}| \leq 1$, which may be easily shown using (2.73). Thus (2.72) and (2.73) together will guarantee a stable all-pole filter. A recursive relationship for the denominator of (2.73) was found by Anderson [10]

$$\begin{aligned}\mathrm{DEN}(i) &= \sum_{k=i}^{N-1}(|b_{i-1,k-1}|^2 + |e_{i-1,k}|^2)\\ &= \mathrm{DEN}(i-1)[1-|a_{i-1,i-1}|^2]\\ &\quad - |b_{i-1,N-i}|^2 - |e_{i-1,i}|^2.\end{aligned} \tag{2.74}$$

Fig. 7 is an outline of the Burg procedure for AR spectral estimation. A computational complexity analysis indicates that $3Np - p^2 - 2N - p$ complex adds, $3Np - p^2 - N + 3p$ complex multiplications, and p real divisions are required. Storage of $3N + p + 2$ complex values is also required [9], [57], [91], [92]. Multichannel versions of Burg's algorithm may be found in [166], [177], [178], [180], [210], [225].

The Burg algorithm has several problems associated with it, including spectral line splitting and biases in the frequency estimate. If one minimizes $\mathcal{E}_p$ in (2.71) with respect to all the a_{pk} for $k = 1, \cdots, p$, then these problems can be mitigated. Ulrych and Clayton [251] and Nuttall [176] independently suggested this least squares procedure for forward and backward prediction in which the Levinson recursion constraint imposed by Burg is removed.

To obtain the p normal equations for the LS (also called forward-backward) algorithm, determine the minimum of $\mathcal{E}_p$ by setting the derivatives of $\mathcal{E}_p$ with respect to all the AR parameters a_{p1} through a_{pp} to zero. This yields

$$\frac{\partial \mathcal{E}_p}{\partial a_{pi}} = 2\sum_{j=0}^{p} a_{pj}r_p(i,j) = 0 \tag{2.75}$$

for $i = 1, \cdots, p$, where $a_{po} = 1$ by definition and

$$r_p(i,j) = \sum_{k=0}^{N-p-1}(x_{k+p-j}x^*_{k+p-i} + x_{k+i}x^*_{k+j}) \tag{2.76}$$

for $0 \leq i, j \leq p$. The minimum prediction error energy may be determined to be

$$\mathcal{E}_p = \sum_{j=0}^{p} a_{pj}r_p(0,j). \tag{2.77}$$

Expressions (2.75) and (2.77) can be combined into a single $(p+1)$ by $(p+1)$ matrix expression

$$\boldsymbol{R}_p\boldsymbol{A}_p = \boldsymbol{E}_p \tag{2.78}$$

where

$$\boldsymbol{A}_p = \begin{bmatrix}1\\ a_{p1}\\ \vdots\\ a_{pp}\end{bmatrix}, \boldsymbol{E}_p = \begin{bmatrix}\mathcal{E}_p\\ 0\\ \vdots\\ 0\end{bmatrix}, \boldsymbol{R}_p = \begin{bmatrix}r_p(0,0) & \cdots & r_p(0,p)\\ \vdots & & \vdots\\ r_p(p,0) & \cdots & r_p(p,p)\end{bmatrix}. \tag{2.79}$$

Direct solution of (2.78) by Gaussian elimination requires a computational complexity proportional to $o(p^3)$. Barrodale and Erickson [16] discuss such a solution and its numerical stability. However, expression (2.79) has a structure that can be exploited to generate an algorithm of $o(p^2)$ operations. Basically, $\boldsymbol{R}_p$ can be expressed as sums and products of Toeplitz and Hankel matrices. This structure enabled a recursive algorithm of $o(p^2)$ to be developed (see Marple [154]). A flowchart of the algorithm may be found in the same reference and is not provided here due to its complexity. The LS algorithm is almost as computationally efficient as the Burg algorithm, requiring typically about 20 percent more computations. The improvement obtained from the LS approach over the Burg algorithm is well worth the slight additional computation. These improvements include less bias in the frequency estimate of spectral components, reduced variance in frequency estimation, and absence of observed spectral line splitting [141], [154], [227].

Sequential Estimation of AR Parameters: Three major approaches to time updating the AR parameter estimates on a data sample by data sample basis are available. These are the recursive least squares method, the gradient adaptive approach, and sequential identification using a lattice filter.

The recursive least squares method [14], [20], [211], [220] in appearance resembles a Kalman filtering procedure. By eliminating the $\mathcal{E}_p$ term from the normal equations (2.69), the least squares solution for the AR parameter estimate vector $\hat{\boldsymbol{A}}_m$ for order p is

$$\hat{\boldsymbol{A}}_m = [\mathfrak{X}_m^H\mathfrak{X}_m]^{-1}\mathfrak{X}_m^H\boldsymbol{Y}_m \tag{2.80}$$

where an m subscript has been added to all the vectors and matrices to indicate that the time samples up to index m are in-

cluded. The vector Y_m is composed of data samples,

$$Y_m = \begin{bmatrix} x_0 \\ \vdots \\ x_m \end{bmatrix}$$

and $\mathfrak{X}_m$ is a modified version of data matrices X_1 or X_3,

$$\mathfrak{X}_m = \begin{bmatrix} x_0 & & \\ \vdots & \ddots & \\ x_{m-1} & \cdots & x_{m-p} \end{bmatrix} \text{ or } \begin{bmatrix} x_{p-1} & \cdots & x_0 \\ \vdots & & \vdots \\ x_{m-1} & \cdots & x_{m-p} \end{bmatrix}$$

corresponding to the covariance or prewindowed cases discussed under batch processing methods.

The addition of a new time sample x_{m+1} can be accounted for by partitioning Y_{m+1} and $\mathfrak{X}_{m+1}$ as follows:

$$Y_{m+1} = \begin{bmatrix} Y_m \\ \hline x_{m+1} \end{bmatrix}, \quad m+1 = \begin{bmatrix} \mathfrak{X}_m \\ \hline H_{m+1} \end{bmatrix}$$

where $H_{m+1} = [x_m \cdots x_{m-p+1}]$. If we define $P_m = [\mathfrak{X}_m^H \mathfrak{X}_m]^{-1}$ and substitute the partitioned Y_{m+1} and $\mathfrak{X}_{m+1}$, then

$$\begin{aligned} \hat{A}_{m+1} &= P_{m+1} \mathfrak{X}_{m+1}^H Y_{m+1} \\ &= P_{m+1} [\mathfrak{X}_m^H Y_m + H_{m+1}^H x_{m+1}] \\ &= P_{m+1} [P_m^{-1} P_m \mathfrak{X}_m^H Y_m + H_{m+1}^H x_{m+1}]. \end{aligned} \tag{2.81}$$

Noting that $P_m \mathfrak{X}_m^H Y_m = \hat{A}_m$ and $P_m^{-1} = P_{m+1}^{-1} - H_{m+1}^H H_{m+1}$, then

$$\begin{aligned} \hat{A}_{m+1} &= P_{m+1} [(P_{m+1}^{-1} - H_{m+1}^H H_{m+1}) \hat{A}_m + H_{m+1}^H x_{m+1}] \\ &= \hat{A}_m + P_{m+1} H_{m+1}^H (x_{m+1} - H_{m+1} \hat{A}_m). \end{aligned} \tag{2.82}$$

Using the matrix inversion lemma, $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$, then an alternative recursive formulation for P_{m+1} is

$$\begin{aligned} P_{m+1} &= (P_m^{-1} + H_{m+1}^H H_{m+1})^{-1} \\ &= P_m - P_m H_{m+1}^H (1 + H_{m+1} P_m H_{m+1}^H)^{-1} H_{m+1} P_m \\ &= (I - K_{m+1} H_{m+1}) P_m \end{aligned} \tag{2.83}$$

where by definition,

$$K_{m+1} = P_m H_{m+1}^H (1 + H_{m+1} P_m H_{m+1}^H)^{-1}. \tag{2.84}$$

The sequential recursion (2.82) then reduces to

$$\begin{aligned} \hat{A}_{m+1} &= \hat{A}_m + K_{m+1} (x_{m+1} - H_{m+1} \hat{A}_m) \\ &= \hat{A}_m + K_{m+1} e_{p,m+1}. \end{aligned} \tag{2.85}$$

Equations (2.13)–(2.85) are similar in structure to a Kalman filter in which the data vector H_m and data matrix P_m are similar to the covariance vector and matrix assumed available in the Kalman formulation. In fact, the least squares formulation presented here could be modified to incorporate *a priori* knowledge of any statistics available concerning the linear prediction error noise statistics.

The recursive least squares technique can be applied to other parameter estimation problems other than this AR application, of on-line spectral estimation, as long as the model is linear in the parameters [170]. In order to start the recursion, $\hat{A}_0$ and P_0 must be specified. If P_0^{-1} is other than an all zero matrix, then the selection of $\hat{A}_0$ must be made carefully since the $\hat{A}_m$ estimate will be biased toward $\hat{A}_0$. This bias can be removed by setting P_0^{-1} to an all zero matrix, but this requires an alternative formulation of (2.84) and (2.85) in terms of P_0^{-1} rather than P_0.

The update relationships in the least squares recursion require $o(p^3)$ operations with each new data point, which may be formidable. An alternative approach that requires $o(p)$ operations for each sequential update is the adaptive linear prediction filter approach [8], [83], [85], [243]. The adaptive approach recursively estimates the AR parameter vector using a gradient technique

$$\hat{A}_{m+1} = \hat{A}_m - \mu \nabla E(|e_{pm}|^2) \tag{2.86}$$

where μ is the step size and ∇ denotes the gradient. By substituting

$$e_{pm} = x_m + \sum_{k=1}^{p} a_{pk} x_{m-k}$$

and taking expectations, then

$$E(|e_{pm}|^2) = R_{xx}(0) + \hat{A}_m^H R_{xx} \hat{A}_m + 2 \operatorname{Re}(\hat{A}_m^H r_{xx})$$

where R_{xx} is the autocorrelation matrix and r_{xx} is the autocorrelation vector $r_{xx} = (R_{xx}(1) \cdots R_{xx}(p))^T$. The gradient of $E(|e_{pm}|^2)$ is

$$\nabla E(|e_{pm}|^2) = 2 r_{xx} + 2 R_{xx} \hat{A}_m. \tag{2.87}$$

When the gradient technique converges, $\nabla E(|e_{pm}|^2) = 0$ and

$$r_{xx} = -R_{xx} A$$

which is the same as the Yule-Walker equations (2.40). In practice, r_{xx} and R_{xx} are unavailable, so instantaneous estimates are substituted. This yields

$$r_{xx} \to x_m X_{m-1}^* \quad \text{where} \quad X_{m-1}^* = \begin{bmatrix} x_{m-1}^* \\ \vdots \\ x_{m-p}^* \end{bmatrix}$$

$$R_{xx} \to X_{m-1}^* X_{m-1}^T.$$

Noting that $e_{pm} = x_m + X_{m-1}^T \hat{A}_m$ and using (2.87), then (2.86) becomes

$$\hat{A}_{m+1} = \hat{A}_m - 2\mu e_{pm} X_{m-1}^*. \tag{2.88}$$

The above gradient formulation is called the least mean square (LMS) algorithm. It has a similar structure to (2.85), except μ is fixed whereas the gain K_{m+1} is variable.

Convergence is guaranteed as long as the constant μ is selected between 0 and $1/\lambda_{\max}$, where $\lambda_{\max}$ is the maximum eigenvalue of R_{xx}. The choice of μ involves a tradeoff between rate of convergence of $E[\hat{A}_m]$ to A_m and the amount of steady-state variance (sometimes termed misadjustment) to be tolerated once convergence is achieved. Thus the price paid for a much reduced computational burden over the recursive least squares approach is a slow convergence requiring the need for a longer data record to achieve reliable AR parameter estimates.

In an attempt to achieve the accuracy of the recursive least squares technique and the computational savings of the adaptive gradient techniques, a third method has been advocated based on the lattice filter. However, the lattice recursive relationships update only the reflection coefficients with $o(p)$ operations. If the AR parameter estimates are to be updated

with each new data sample, then the Levinson recursion (2.45) will need to be used, which requires $o(p^2)$ operations. Thus computational savings with the lattice technique are achievable only if the AR parameter estimates are updated infrequently rather than with each new sample.

A sequential algorithm based on a lattice structure may be developed for each of the least squares techniques presented under the batch estimation topic [164], [168]. An illustration of one such algorithm based on Burg's algorithm is given here. The reflection coefficient K_{mn} for order m and time index n was determined in the Burg algorithm by computing

$$K_{mn} = \frac{-2\sum_{i=m}^{n} e_{m-1,i}\, b^*_{m-1,i-1}}{\sum_{i=m}^{n} (|b_{m-1,i-1}|^2 + |e_{m-1,i}|^2)}. \tag{2.89}$$

A time update recursive formulation for (2.89) is given by [222]

$$K_{m+1,n+1} = K_{m+1,n} - \frac{[K_{m+1,n}(|e_{mn}|^2 + |b_{m,n-1}|^2) + 2e_{mn}b^*_{m,n-1}]}{\sum_{i=m}^{n} [|e_{mi}|^2 + |b_{m,i-1}|^2]}. \tag{2.90}$$

Thus, (2.90) in combination with (2.52) and (2.54) for $m = 1, \cdots, p$ and with initial conditions $e_{0n} = b_{0n} = x_n$, form a sequential time-update algorithm for the reflection coefficients.

AR Spectral Power Estimation: It has been shown [136] that, unlike conventional Fourier spectral estimates, the peak amplitudes in AR spectral estimates are not linearly proportional to the power when the input process consists of sinusoids in noise. Lacoss [136] has shown that for high SNR, the peak is proportional to the square of the power, although the area under the peak is proportional to power. One method for obtaining an estimate of the actual power of real sinusoids from an AR spectrum was suggested by Andersen and Johnsen [108]. The method works best for high SNR components in the process. The AR PSD in z-transform notation is

$$P_{AR}(z) = \frac{\sigma^2 \Delta t}{A(z)A^*(1/z^*)} \tag{2.91}$$

where

$$A(z) = 1 + \sum_{k=1}^{p} a_{pk} z^{-k}.$$

If a peak is at f_i in the AR spectral estimate, i.e., $z_i = \exp(j2\pi f_i \Delta t)$, then the estimated power is approximately

$$\text{Power}\,(f_i) = 2 \times \text{Real}\left\{\text{Residue of } \frac{P_{AR}(z)}{z} \text{ at } z_i\right\} \tag{2.92}$$

where

$$\text{Residue } \frac{P_{AR}(z)}{z} = (z - z_i)\left|\frac{P_{AR}(z_i)}{z}\right|_{z=z_i}$$

$$z_i = \text{Root of } A(z) = \exp(\alpha_i + j2\pi f_i)\Delta t.$$

Note that it is assumed that the peak occurs at the angular location of the pole z_i. Negative power estimates can occur with this technique if peaks are very close together. This approach is closely related to the power estimation procedure utilized in the Prony method presented in Section II-J.

Model Order Selection [21], [100], [110], [114], [133], [235], [237], [238]: Since the best choice of filter order p is not generally known *a priori*, it is usually necessary in practice to postulate several model orders. Based on these, one then computes some error criterion that indicates which model order to choose. Too low a guess for model order results in a highly smoothed spectral estimate. Too high an order introduces spurious detail into the spectrum. One intuitive approach would be to construct AR models of increasing order until the computed prediction error power reaches a minimum. However, all the least squares estimation procedures discussed in this paper have prediction error powers that decrease monotonically with increasing order p. For example, the Burg algorithm and Yule-Walker equations involve the relationship

$$\mathcal{E}_i = \mathcal{E}_{i-1}[1 - |a_{ii}|^2]. \tag{2.93}$$

As long as $|a_{ii}|^2$ is nonzero (it must be ≤ 1), the prediction error power decreases. Thus the prediction error power alone is not sufficient to indicate when to terminate the search.

Several criteria have been introduced as objective bases for selection of the AR model order. Akaike [1]–[7] has provided two criteria. His first criterion is the final prediction error (FPE). This criterion selects the order of the AR process so that the average error for a one step prediction is minimized. He considers the error to be the sum of the power in the unpredictable (or innovation) part of the process and a quantity representing the inaccuracies in estimating the AR parameters. The FPE for an AR process is defined as

$$\text{FPE}_p = \mathcal{E}_p\left(\frac{N + p + 1}{N - p - 1}\right) \tag{2.94}$$

where N is the number of data samples. Note that (2.94) assumes one has subtracted the sample mean from the data. The term in parentheses increases the FPE as p approaches N, reflecting the increase in the uncertainty of the estimate $\mathcal{E}_p$ of the prediction error power. The order p selected is the one for which the FPE is minimum. The FPE has been studied for application by Gersch and Sharpe [73], Jones [111], Fryer *et al.* [69], and Ulrych and Bishop [250]. For AR processes, the FPE works fairly well. However, when processing actual geophysical data, both Jones [111] and Berryman [21] found the order selected tended to be too low.

Akaike suggested a second-order selection criterion using a maximum likelihood approach to derive a criterion termed the Akaike information criterion (AIC). The AIC determines the model order by minimizing an information theoretic function. Assuming the process has Gaussian statistics, the AIC is

$$\text{AIC}_p = \ln(\mathcal{E}_p) + 2(p + 1)/N. \tag{2.95}$$

The term $(p + 1)$ in (2.95) is sometimes replaced by p, since $2/N$ is only an additive constant which accounts for the subtraction of the sample mean. The second term in (2.95) represents the penalty for the use of extra AR coefficients that do not result in a substantial reduction in the prediction error power. Again, the order p selected is the one that minimizes the AIC. As $N \to \infty$, the AIC and FPE are equivalent. Kashyap [115] claims the AIC is statistically inconsistent in that the probability of error in choosing the correct order does not tend to zero as $N \to \infty$.

A third method was proposed by Parzen [192] and is termed the criterion autoregressive transfer (CAT) function. The order p is selected to be the one in which the estimate of the difference of the mean-square errors between the true prediction error filter (which may be of infinite length) and the estimated filter is a minimum. Parzen showed that this difference can be calculated, without explicitly knowing the true prediction error filter, by

$$\mathrm{CAT}_p = \left(\frac{1}{N}\sum_{j=1}^{p}\frac{1}{\hat{\mathcal{E}}_j}\right) - \frac{1}{\hat{\mathcal{E}}_p} \tag{2.96}$$

where $\hat{\mathcal{E}}_j = (N/(N-j))\mathcal{E}_j$. Again p is chosen to minimize CAT_p.

The results of spectra using the FPE, AIC, and CAT have been mixed, particularly against actual data rather than simulated AR processes. Ulrych and Clayton [251] have found that for short data segments, none of the criteria work well. For harmonic processes in noise, the FPE and AIC also tend to underestimate the order if the SNR is high [92], [138]. Ulrych and Ooe [92] suggest in the case of short data segments that an order selection between $N/3$ to $N/2$ often produces satisfactory results. In the final analysis, more subjective judgment is still required in the selection of order for data from actual processes than that required for controlled simulated computer processes.

Anomalies of and Patches for the AR Spectral Estimator: Several anomalies of the AR spectral estimator have been observed by researchers. When the model order is chosen to be too large relative to the number of data points, the AR spectral estimate exhibits spurious peaks [212], [250]. Ideally, if the autocorrelation lags, or equivalently, the reflection coefficents, were estimated without error, then the estimated AR parameters for an AR(p) model would be

$$\hat{a}_{pi} = \begin{cases} a_{pi}, & \text{for } i = 1, 2, \cdots, p \\ 0, & \text{for } i = p+1, \cdots, n \end{cases} \tag{2.97}$$

where a_{pi} are the AR(p) parameters. However, when estimation errors are present, then $a_{pi} \neq 0$ in general for $i > p$. Correspondingly, there will be n-p "extra" poles. When the estimated extra poles occur near the unit circle, spurious spectral peaks result. It is this possibility of spurious peaks that is the basis of the recommendation that the maximum model order should be no greater than $N/2$, where N is the data record length [251].

It has been observed for a process consisting of a sinusoid in noise that the peak location in the AR spectral estimate depends critically on the phase of the sinusoid [45], [229]. Also, it has been observed that the spectral estimate sometimes exhibits two closely spaced peaks, falsely indicating a second sinusoid. The latter phenomenon is known as spectral line splitting (SLS) [64].

The phase dependence of the AR spectral estimate decreases as the data record length increases. The amount of phase dependence varies for the dfiferent AR estimation procedures. For the Burg algorithm, the shift in peak location can be as much as 16 percent [230]. The forward-backward prediction error approach is least dependent on phase [154], [251]. Two techniques have been proposed to reduce this effect. In the first approach, the phase dependence is attributed to the interaction between the positive and negative frequency components of the real sinusoid, much in the same way as the peak of the periodogram depends upon phase [119], [231], [234], Based on this premise, the solution is then to replace the real valued signal by the analytic signal. The analytic signal process is then down sampled by two and the AR spectral estimate for complex data used. The model order for complex data need only be half as large as for real data since the complex conjugate pole pairs in the real case are not required in the complex approach. Using this approach, the phase dependence of the Burg spectral estimate can be decreased [119].

The other alternative procedure (for the Burg spectral estimate) is to employ the estimator

$$\hat{K}_i = -2\sum_{n=i}^{N-1} v_n e_{i-1,n} b^*_{i-1,n-1} \Big/ \sum_{n=i}^{N-1} v_n \left[|e_{i-1,n}|^2 + |b_{i-1,n-1}|^2\right] \tag{2.98}$$

which weights the reflection coefficient terms with the real sequence $\{v_n\}$. This windowing of the residual time series, suggested by Swingler [228], has the effect of reducing the end effects of the short data record. Simulations indicate the phase dependence is also reduced by this method.

The problem of spectral line splitting in AR spectra produced by the Burg algorithm was first documented by Fougere *et al.* [64]. They noted that spectral line splitting was most likely to occur when 1) the SNR is high, 2) the initial phase of sinusoidal components is some odd multiple of 45°, 3) the time duration of the data sequence is such that sinusoidal components have an odd number of quarter cycles, and 4) the number of AR parameters estimated is a large percentage of the number of data values used for the estimation. Many spurious spectral peaks often accompany spectra that exhibit line splitting. Again the phenomenon is associated with short data records since it tends to disappear as the data record increases. SLS has been observed in the Burg and Yule–Walker spectral estimates for multiple sinusoids in white noise [120]. A solution to the problem has been proposed by Fougere [65]. He attributes the splitting to the fact that the prediction error power is not truly minimized using Burg's estimate for the reflection coefficients. His technique minimizes the prediction error power by varying all reflection coefficients simultaneously. From his simulation, the technique appears to eliminate SLS for at least one sinusoid. Another method of eliminating SLS for one sinusoid is to use the analytic signal approach [94], [120]. The analytic signal approach yields the true reflection coefficients for the Burg spectral estimate when noise is negligible, the condition where SLS is most likely to be observed. For the Yule–Walker spectral estimate, if the analytic signal approach and the unbiased autocorrelation estimate are used, the true autocorrelation function is obtained. Thus, in either case, SLS is eliminated. For multiple sinusoids, the performance of Fougere's algorithm is undocumented and the use of complex data in conjunction with Burg's reflection coefficient estimate can still exhibit SLS [94]. Using the forward–backward LS approach in conjunction with the recursive algorithm of Marple, SLS has not been observed [154].

Besides phase, signals with large dc levels or a linear trend have also been found to corrupt AR spectra [113], particularly the low-frequency end of the spectral estimate. These components should be removed before applying AR spectral analysis techniques.

A very important problem with the AR spectral estimator, which limits its utility, is its sensitivity to the addition of observation noise to the time series [186]. An example is given in Fig. 8. It is seen that the spectral peaks are broadened and dis-

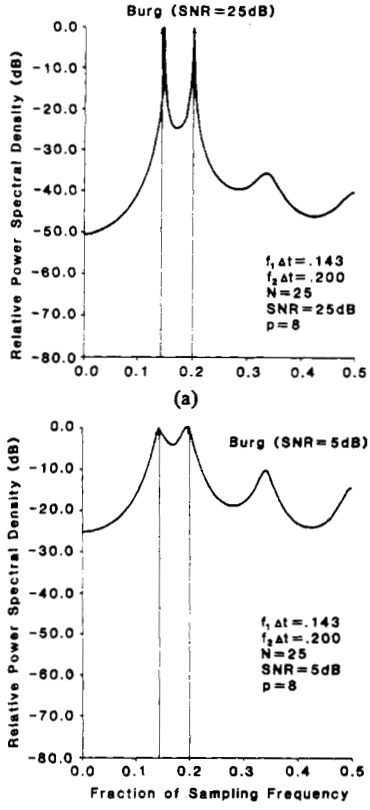

Fig. 8. Spectral estimates for two sinusoids in white Gaussian noise. (a) High SNR. (b) Low SNR.

placed from their true positions (indicated by arrows). In particular, it has been shown that the resolution of the AR spectral estimate for two equiamplitude sinusoids in white noise decreases as the SNR decreases [149], [152]. For low SNR, the resolution is no better than that of the periodogram. The reason for the degradation is that the all-pole model assumed in AR spectral analysis is no longer valid when observation noise is present. To see this, assume y_n denotes the noise corrupted AR process, x_n. Thus

$$y_n = x_n + w_n \tag{2.99}$$

where w_n is the observation noise. If w_n is white noise with variance σ_w^2 and is uncorrelated with x_n,

$$P_y(z) = \frac{\sigma^2 \Delta t}{A(z)A^*(1/z^*)} + \sigma_w^2 \Delta t = \frac{[\sigma^2 + \sigma_w^2 A(z)A^*(1/z^*)]\,\Delta t}{A(z)A^*(1/z^*)}. \tag{2.100}$$

Thus the PSD of y_n is characterized by poles and zeros, i.e., y_n is an ARMA (p, p) process. The inconsistency of the AR model for a noise corrupted AR process leads to the degradation observed in Fig. 8 [121]. The phenomenon is explained as follows. The effect of noise is to reduce the dynamic range of the PSD of x_n. Since the prediction error filter $\hat{A}(z)$ attempts to whiten the PSD, it is not surprising that for low SNR, the zeros of $\hat{A}(z)$ are located near the origin of the z plane, i.e., $\hat{A}(z) \approx 1$. This is because the PSD of y_n is already relatively flat due to noise so that subsequent filtering operations, i.e., the use of a prediction error filter, will not significantly whiten the PSD further.

To reduce the degradation of the AR spectral estimate in the presence of noise, four general approaches have been proposed. One can

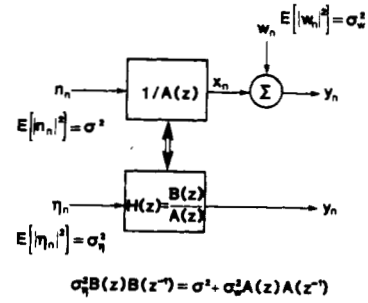

Fig. 9. ARMA model for AR process in white noise.

1) use the ARMA spectral estimate;
2) filter the data to reduce the noise;
3) use a large order AR model;
4) compensate either the autocorrelation function estimates or the reflection coefficient estimates for the noise effects.

The ARMA approach assumes that the noise corrupted AR process is a general ARMA (p, p) process even though the AR and MA parameters are related by (2.100). The most common approach has been the use of the modified Yule–Walker equations as described in Section II-F. [58], [72]. For an ARMA (p, p) process, this means solving the set of equations

$$R_{yy}(k) = -\sum_{l=1}^{p} a_l R_{yy}(k-l) \tag{2.101}$$

for $k = p+1, p+2, \cdots, 2p$ in order to obtain the AR parameters. Although simple to implement, this approach has met with only moderate success. Reasonable results are obtained for long data records and/or high SNR's. A more suitable solution to the noise problem is to use the maximum likelihood ARMA estimate. However, this procedure leads to a set of highly nonlinear equations [31]. In the case for which the maximum likelihood equations are determined specifically for an AR process in white noise, a suboptimal solution to these equations leads to an iterative filtering scheme as described in [143]. Other ARMA filtering schemes can be found in references [118], [151], [159], [172]. All the methods rely on a boot-strapping approach to design the filter since the power spectral density of x_n, which is what we are attempting to estimate, is unknown.

Another technique to combat noise is to employ an AR model with a model order larger than the true AR model. This is because an ARMA(p, p) process is equivalent to an AR(∞) model, as guaranteed by the Wold decomposition. Using (2.100), let

$$\sigma^2 + \sigma_w^2 A(z)A^*(1/z^*) = \sigma_\eta^2 B(z)B^*(1/z^*)$$

where

$$B(z) = 1 + \sum_{k=1}^{p} b_{pk} z^{-k}$$

so that y_n can be represented as the output of a pole-zero filter, $H(z)$, driven by white noise (with variance σ_η^2) as shown in Fig. 9. If we divide $A(z)$ by $B(z)$, we have

$$H(z) = 1/(A(z)/B(z)) = 1/C(z)$$

where

$$C(z) = \frac{A(z)}{B(z)} = 1 + \sum_{k=1}^{\infty} c_k z^{-k}.$$

Thus, y_n can be modeled by an AR(∞) process with parameters $\{c_k\}$. Clearly, as the assumed AR model order increases, the estimated AR PSD will approach the true PSD of y_n. This property is also evident from the maximum entropy formulation since it is shown there that

$$\hat{R}_{yy}(k) = R_{yy}(k), \quad \text{for} \quad |k| \leqslant p$$

where $\hat{R}_{yy}(k)$ is the autocorrelation function corresponding to the AR(p) model and $R_{yy}(k)$ is the true autocorrelation function of y_n. Thus, as $p \to \infty$, the autocorrelation function of the model matches the true autocorrelation function. Hence, the spectra must also match each other.

It would seem that a model order as large as possible should be used. However, due to the spurious peak problem, one should limit the maximum model order to no more than one half the number of data points, as discussed previously.

In practice a larger order model will be needed when the zeros of $B(z)$ are near the unit circle of the z plane. In this case, the c_k sequence will die out slowly. Since the zeros $B(z)$ move outward as the SNR decreases [121], increasing the model order will be necessary as the SNR decreases. To quantify the effect of model order on the AR spectral estimate for an AR process in white noise, consider two equiamplitude sinusoids in white noise. It has been shown [149] for this case that the resolution δf in hertz of the AR spectral estimate, assuming a known autocorrelation function, is approximately

$$\delta f \approx \frac{[p(p+1)]^{0.31}}{6.471 \times 2\pi p \Delta t} \tag{2.102}$$

where ρ is the SNR. As expected, the resolution increases with increasing model order. An example of this behavior is shown in Fig. 10. Note that the extra poles are approximately uniformly spaced within the unit circle, producing an equiripple approximation to the flat noise spectrum.

Many noise cancellation schemes that compensate the autocorrelation lags for the noise are available [123], [151], [159], [214], [240], [268]. Details may be found in these references. The PHD is a special case of these schemes. In general, these noise cancellations schemes can reduce the bias, but will increase the variance of the spectral estimate. A serious deficiency is that, in general, one does not know how much noise power to remove. Thus, if σ_w^2 is too large, the estimated AR spectrum will exhibit sharper peaks than the true spectrum. Thus one must be careful in applying these techniques.

F. Moving Average PSD Estimation

As presented in the introduction to Section II, a MA process is a stochastic process obtained from the output of a filter whose transfer function contains only zeros, and whose input is a white noise process, i.e.,

$$x_n = \sum_{m=0}^{q} b_m n_{n-m} \tag{2.103}$$

with

$$E[n_n] = 0 \qquad E[n_{n+m} n_n^*] = \sigma^2 \delta_m$$

where δ_m is 1 for $m = 0$, and 0 otherwise. Based on (2.103),

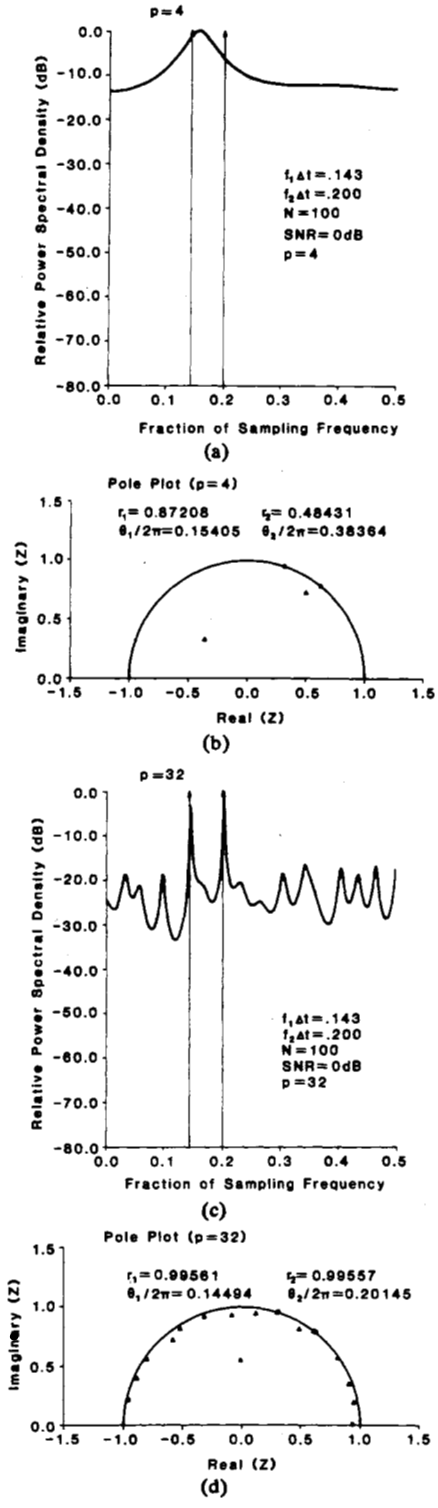

Fig. 10. Burg spectral estimate for two sinusoids in white noise. Effect of model order. (a) $p = 4$. (b) Pole plot for $p = 4$. (c) $p = 32$. (d) Pole plot for $p = 32$.

the autocorrelation function of a MA process of order q is

$$R_{xx}(k) = \begin{cases} \sigma^2 \sum_{i=0}^{q-k} b_i^* b_{i+k}, & \text{for} \quad k = 0, 1, \cdots, q. \\ 0, & \text{for} \quad k > q \end{cases} \tag{2.104}$$

Thus, if $(q+1)$ lags of the autocorrelation function are known, the parameters of a qth-order MA process are determined by solving the nonlinear set of equations (2.104), often called the method of moments [31]. However, if only a spectral estimate is desired, then there is no need to solve for the MA parameters, but only to determine the autocorrelation function, since

$$\mathcal{P}_{\mathrm{MA}}(f) = \sum_{m=-q}^{q} R_{xx}(m) \exp(-j2\pi f m \Delta t)$$

which is identical to a BT spectral estimate. The method of moments is then not applicable to the spectral estimation problem. If one uses a MLE of the MA parameters, this also corresponds to a MLE of the autocorrelation function, since a one-to-one transformation is given by (2.104) (assuming the zeros are within the unit circle of the z plane). In this case, it is appropriate to determine the MA parameters as an intermediate step to estimating the spectrum. This approach has not been employed since the MLE for the MA parameters is highly nonlinear [31]. Furthermore, for MA modeling, too many coefficients are necessary to represent narrow-band spectra, leading to poor spectral estimates for these situations.

One must determine the order of the MA model when only data samples are available. One intuitive method suggested by Chow [48] would be to use the unbiased autocorrelation lag estimator (2.14) and check that the lag estimates approach zero rapidly after a small number of terms, since from (2.104) we know that $R_{xx}(m) = 0$ for lags greater than the order of the MA process. If not, an AR or ARMA model may be more appropriate. Chow suggested a hypothesis test on successive lags to determine if lag $R_{xx}(q)$ is sufficiently close to zero relative to the variance of the lags indexed less than q. If so, then the order of the MA process is considered to be q. The lag estimates are used in (2.104) to find the MA parameters. Further refinements of the MA parameter estimation can be made [236], once the order q has been determined, by enforcing the constraint on the lag estimates that $R_{xx}(m) = 0$ for $m > q$.

G. ARMA PSD Estimation [11], [31], [35], [72], [79], [86], [87], [89], [117], [165], [242], [266], [267]

Yule–Walker Equations: Recall that the ARMA model assumes that a time series x_n can be modeled as the output of a p pole and q zero filter excited by white noise, i.e.,

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} + \sum_{k=0}^{q} b_k n_{n-k} \tag{2.105}$$

where $R_{nn}(k) = \sigma^2 \delta_k$ and $b_0 \equiv 1$. The poles of the filter are assumed to be within the unit circle of the z-plane. The zeros of the filter may lie anywhere in the z-plane.

Once the parameters of the ARMA (p, q) model are identified, the spectral estimate is obtained as

$$\mathcal{P}_x(f) = |H(\exp[j2\pi f \Delta t]|^2 \, \mathcal{P}_n(f)$$

$$= \frac{\sigma^2 \Delta t \left| 1 + \sum_{k=1}^{q} b_k \exp(-j2\pi f k \Delta t) \right|^2}{\left| 1 + \sum_{k=1}^{p} a_k \exp(-j2\pi f k \Delta t) \right|^2}. \tag{2.106}$$

The relationship of the ARMA parameters to the autocorrelation function is easily found as follows. Multiply (2.105) by x_{n-l}^* and take the expectation to yield

$$R_{xx}(l) = -\sum_{k=1}^{p} a_k R_{xx}(l-k) + \sum_{k=0}^{q} b_k R_{nx}(l-k) \tag{2.107}$$

where

$$R_{nx}(k) = E(n_n x_{n-k}^*).$$

But $R_{nx}(k) = 0$ for $k > 0$ since a future input to a causal, stable filter cannot affect the present output and n_n is white noise. Therefore,

$$R_{xx}(l) = \begin{cases} -\sum_{k=1}^{p} a_k R_{xx}(l-k) + \sum_{k=0}^{q} b_k R_{nx}(l-k), & \text{for } l = 0, \cdots, q \\ -\sum_{k=1}^{p} a_k R_{xx}(l-k), & \text{for } l = q+1, q+2, \cdots. \end{cases}$$

From the derivation of the Yule–Walker equations, it was shown

$$R_{nx}(k) = \sigma^2 h_{-k}^*$$

and therefore

$$R_{xx}(l) = \begin{cases} -\sum_{k=1}^{p} a_k R_{xx}(l-k) + \sigma^2 \sum_{k=l}^{q} b_k h_{k-l}^*, & \text{for } l = 0, 1, \cdots, q \\ -\sum_{k=1}^{p} a_k R_{xx}(l-k), & \text{for } l = q+1, q+2, \cdots. \end{cases} \tag{2.108}$$

These normal equations for an ARMA process are analogous to the Yule–Walker equations for an AR process.

Estimation of ARMA Parameters: Many ARMA parameter estimation techniques have been formulated theoretically, which usually involve many matrix computations and/or iterative optimization techniques. These approaches are normally not practical for real-time processing. Suboptimum techniques have therefore been developed to make the computational load more manageable. These techniques are usually based on a least squares error criterion and require solutions of linear equations. These methods generally estimate the AR and MA parameters separately rather than jointly as required for optimal parameter estimation. The AR parameters can be estimated independently of the MA parameters first if one uses the Yule–Walker equations as given by (2.108). A final point in favor of the suboptimal linear approaches is that iterative optimization techniques are not guaranteed to converge or may converge to the wrong solution. The nonlinearity of the equations encountered is typified by (2.108). Since the impulse response is a function of $a_1, \cdots, a_p, b_1, \cdots, b_q$, the equations given by (2.108) are nonlinear in the ARMA parameters. As an

example, consider an ARMA (1, 1) process. In this case,

$$h_k = (-a_1)^k u(k) + b_1(-a_1)^{k-1} u(k-1)$$
$$R_{xx}(0) = -a_1 R_{xx}(-1) + \sigma^2(1 + |b_1|^2 - a_1^* b_1)$$
$$R_{xx}(1) = -a_1 R_{xx}(0) + \sigma^2 b_1$$
$$R_{xx}(l) = -a_1 R_{xx}(l-1), \quad \text{for } l \geq 2 \tag{2.109}$$

where $u(k)$ is the unit step function. Although numerous researchers have proposed means of solving these equations, there appear to be few successful applications of these approaches.

A more popular approach to this problem is to use (2.108) for $l > q$ to find $(a_1, a_2, \cdots, a_p)$ and then to apply some appropriate technique to find $(b_1, b_2, \cdots, b_q)$ or an equivalent parameter set. For example, to find the AR parameters, using (2.108) and $l = q+1, q+2, \cdots, q+p$, we solve the following matrix expression [72]:

$$\underbrace{\begin{bmatrix} R_{xx}(q) & R_{xx}(q-1) & \cdots & R_{xx}(q-p+1) \\ R_{xx}(q+1) & R_{xx}(q) & \cdots & R_{xx}(q-p+2) \\ \vdots & \vdots & & \vdots \\ R_{xx}(q+p-1) & R_{xx}(q+p-2) & \cdots & R_{xx}(q) \end{bmatrix}}_{\boldsymbol{R}'_{xx}} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = -\begin{bmatrix} R_{xx}(q+1) \\ R_{xx}(q+2) \\ \vdots \\ R_{xx}(q+p) \end{bmatrix}. \tag{2.110}$$

These equations have been called the extended, or modified, Yule–Walker equations. The matrix is Toeplitz, although not symmetric, and is therefore not guaranteed to be either positive–definite or nonsingular. An algorithm requiring $o(p^2)$ operations has been developed by Zohar [278] for solving (2.110).

In order to choose an appropriate model order p for the AR portion of the ARMA model, the property [48]

$$|\boldsymbol{R}'_{xx}| = 0$$

for dimension of $\boldsymbol{R}'_{xx}$ greater than the AR order p can be used. Here $|\boldsymbol{R}'_{xx}|$ denotes the determinant of the matrix $\boldsymbol{R}'_{xx}$. This means that one need only monitor the determinant, $|\boldsymbol{R}'_{xx}|$, for $i = 1, 2, \cdots$ until it becomes sufficiently small. Once the AR parameter estimates $\{\hat{a}_k\}$ have been found, the MA parameters may be found by filtering the data with the all-zero filter $\hat{A}(z)$, where

$$\hat{A}(z) = 1 + \sum_{k=1}^{p} \hat{a}_k z^{-k}$$

to yield a purely MA process. Having performed this operation, the techniques of Section II-F for MA processes can be applied. A spectral factorization is required to determine the MA parameters. To avoid the spectral factorization, note that for spectral estimation one is only concerned with finding $A(z)A^*(1/z^*)$ and $B(z)B^*(1/z^*)$, since the spectral estimate is [117], [129]

$$\mathcal{P}_x(f) = \left. \frac{\sigma^2 \Delta t B(z) B^*(1/z^*)}{A(z) A^*(1/z^*)} \right|_{z = \exp(j2\pi f \Delta t)}.$$

If

$$B_k = Z^{-1}[\sigma^2 \Delta t B(z) B^*(1/z^*)]$$

where Z and Z^{-1} denote the z-transform and inverse z-transform, respectively, then the spectral estimate is

$$\mathcal{P}_x(f) = \frac{\sum_{k=-q}^{q} B_k \exp(-j2\pi f k \Delta t)}{|A(\exp[j2\pi f \Delta t])|^2}$$

where $B_{-k} = B_k^*$. To obtain B_k, observe that

$$B_k = Z^{-1}[A(z) A^*(1/z^*) P_x(z)].$$

Letting $A_k = Z^{-1}[A(z)A^*(1/z^*)]$, which is known, then

$$B_k = \sum_{n=-p}^{P} A_n R_{xx}(k-n), \quad \text{for } k = 0, 1, \cdots, q.$$

To determine B_k requires knowledge of $\{R_{xx}(0), R_{xx}(1), \cdots, R_{xx}(p+q)\}$. To insure a nonnegative spectral estimate, B_k must be a positive–semidefinite sequence.

The performance of the modified Yule–Walker approach as applied to ARMA modeling varies greatly. For some processes, the estimates of the ARMA parameters obtained will be quite accurate. However, for some processes, this will not be the case. As an example, consider the asymptotic variance (as $N \to \infty$) of the AR parameter estimate for a real ARMA (1, 1) process. The estimate as given by (2.110)

$$\hat{a}_1 = -\hat{R}_{xx}(2)/\hat{R}_{xx}(1)$$

can be shown to have a variance [72]

$$\mathrm{Var}(\hat{a}_1) = \frac{\sigma^2}{N} \frac{(1 + b_1^2) R_{xx}(0) + 2b_1 R_{xx}(1)}{R_{xx}^2(1)}. \tag{2.111}$$

But

$$R_{xx}(0) = \sigma^2 \left[\frac{1 + b_1^2 - 2a_1 b_1}{1 - a_1^2}\right] \text{ and } R_{xx}(1) = -a_1 R_{xx}(0) + \sigma^2 b_1$$

so (2.111) may be rewritten as

$$\mathrm{Var}(\hat{a}_1) = \frac{1}{N} \frac{\left[1 + \frac{(a_1 - b_1)^2}{(1 - a_1^2)}\right]^2 + \frac{2b_1^2}{(1 - a_1^2)}}{\left(\frac{1}{1 - a_1^2}\right)\left[b_1 - a_1\left(1 + \frac{(a_1 - b_1)^2}{(1 - a_1^2)}\right)\right]^2}.$$

For a reasonably accurate estimate of a_1, one might require the rms error to be no greater than 0.1,

$$\sqrt{\mathrm{Var}(\hat{a}_1)} \leq 0.1.$$

To meet this requirement for $b_1 = 0.5$, the minimum number of samples versus a_1 is shown in Fig. 11. It may be seen that the statistical fluctuation will vary greatly for a given N, depending on the spectral shape. As $N \to \infty$, the ARMA(1, 1) model is seen to be inappropriate as $a_1 \to 0.5$, since the pole and zero cancel resulting in a white noise process. This example illustrates that care must be taken when using the modified Yule–Walker approach, especially when the model order p is unknown. This is because the modified Yule–Walker matrix $\boldsymbol{R}'_{xx}$ formed from exactly known lags will be singular for a dimension in excess of the true model order.

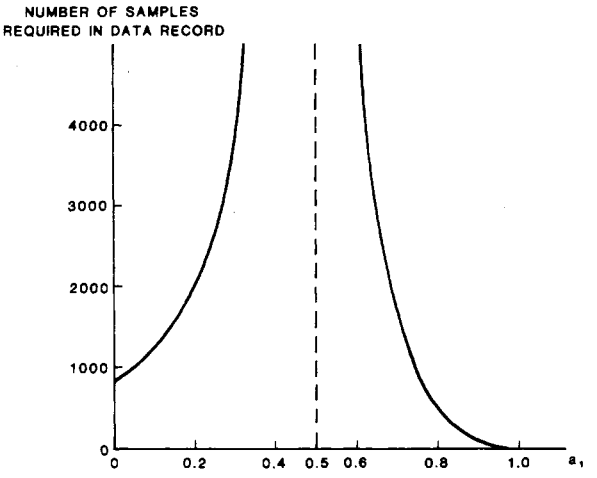

Fig. 11. Required number of samples in data record for accurate estimation of AR parameters for ARMA (1, 1) process using modified Yule-Walker equations.

A second technique for estimating the ARMA parameters utilizes the identify

$$\frac{B(z)}{A(z)} = \frac{1}{C(z)}$$

where

$$C(z) = 1 + \sum_{k=1}^{\infty} c_k z^{-k}$$

to equate an ARMA model to an infinite order AR model. The $\{c_k\}$ may be estimated using AR techniques only and then related to the ARMA parameters. Specifically, let $\hat{C}(z) = 1 + \sum_{k=1}^{M} \hat{c}_k z^{-k}$ be the estimated AR parameters, where $M \geq p + q$. Assuming $p > q$, then [79]

$$\frac{\hat{B}(z)}{\hat{A}(z)} = \frac{1}{\hat{C}(z)} \quad \text{or}$$

$$\sum_{k=0}^{q} \hat{b}_k \hat{c}_{n-k} = \hat{a}_n, \quad \text{for } n = 1, 2, \cdots.$$

where $\hat{b}_0 = 1$. Since a_n should be equal zero for $n > p$, set

$$\sum_{k=0}^{q} \hat{b}_k \hat{c}_{n-k} = 0, \quad \text{for } n = p+1, p+2, \cdots, p+q.$$

This expression may be written in matrix form,

$$\begin{bmatrix} \hat{c}_p & \hat{c}_{p-1} & \cdots & \hat{c}_{p+1-q} \\ \hat{c}_{p+1} & \hat{c}_p & \cdots & \hat{c}_{p+2-q} \\ \vdots & \vdots & & \vdots \\ \hat{c}_{p+q-1} & \hat{c}_{p+q-2} & \cdots & \hat{c}_p \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_q \end{bmatrix} = - \begin{bmatrix} \hat{c}_{p+1} \\ \hat{c}_{p+2} \\ \vdots \\ \hat{c}_{p+q} \end{bmatrix}. \tag{2.112}$$

Once $\{\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_q\}$ are found, then $\{\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p\}$ may be found by solving

$$\sum_{k=0}^{q} \hat{b}_k \hat{c}_{n-k} = \hat{a}_n, \quad \text{for } n = 1, 2, \cdots, p \tag{2.113}$$

where

$$\hat{c}_0 \equiv 1.$$

In matrix form, this is

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} \hat{c}_1 & 1 & 0 & \cdots & 0 \\ \hat{c}_2 & \hat{c}_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{c}_p & \hat{c}_{p-1} & \hat{c}_{p-2} & \cdots & \hat{c}_{p-q} \end{bmatrix} \begin{bmatrix} 1 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_q \end{bmatrix}. \tag{2.114}$$

Since $C(z) = A(z)/B(z)$, a very large-order AR model must be used when the zeros of $B(z)$ are near the unit circle. In this case, the c_k sequence will not die out rapidly. This will usually be the case of interest, for if the zeros of $B(z)$ are near the origin, they will have negligible effect upon the PSD. In this case, an AR model would suffice. Nevertheless, some promising results have been obtained with this method [79], [197].

A third technique based upon least squares input–output identification has also been proposed [131], [155]. From (2.107) it may be seen that the nonlinear character of the normal equations is due to the unknown cross correlation between the input and output. If n_n is unobservable, then $R_{nx}(k)$ cannot be estimated. If, however, n_n were known, so that $R_{nx}(k)$ could be estimated, then the ARMA parameters could be found as the solutions of a set of linear equations. In practice, n_n is estimated from x_n in a boot-strap approach to be discussed later. To set up the linear equations, rewrite x_n from (2.105) as

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} + \sum_{k=1}^{q} b_k n_{n-k} + n_n,$$

$$\text{for } n = 0, \cdots, N-1. \tag{2.115}$$

From (2.115), one can observe that

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{v} \tag{2.116}$$

where

$$\boldsymbol{z} = [x_0 \;\; x_1 \;\; \cdots \;\; x_{N-1}]^T$$

$$\boldsymbol{v} = [n_0 \;\; n_1 \;\; \cdots n_{N-1}]^T$$

$$\boldsymbol{\theta} = [-a_1 \;\; -a_2 \;\; \cdots \;\; -a_p \;\; b_1 \;\; b_2 \;\; \cdots \;\; b_q]^T$$

and

$$\boldsymbol{H} = \begin{bmatrix} x_{-1} & x_{-2} & \cdots & x_{-p} & n_{-1} & n_{-2} & \cdots & n_{-q} \\ x_0 & x_{-1} & \cdots & x_{-p+1} & n_0 & n_{-1} & \cdots & n_{-q+1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ x_{N-1} & x_{N-3} & \cdots & x_{N-p-1} & n_{N-2} & n_{N-3} & \cdots & n_{N-q-1} \end{bmatrix} = \text{Input/Output Data Matrix.}$$

Note that $\boldsymbol{H}$ has dimensions $N \times (p+q)$. Equation (2.115) is the standard form for a linear least squares problem, for which the solution is

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^H\boldsymbol{H})^{-1}\boldsymbol{H}^H\boldsymbol{z}. \tag{2.117}$$

This approach is similar to the least squares approach for AR parameter estimation of pure AR processes. The correlation matrix $\boldsymbol{H}^H\boldsymbol{H}$ of the ARMA process involves the estimate of the cross-correlation function $R_{nx}(k)$. The initial conditions $\{x_{-p}, x_{-p+1}, \cdots, x_{-1}, n_{-q}, n_{-q+1}, \cdots n_{-1}\}$ need to be specified, or assumed equal to zero (in a similar fashion to least squares estimation of AR parameters). To estimate n_n [155], one models x_n by a large AR model and sets n_n equal to the prediction error time series. The ARMA parameter estimates are then further improved by an iterative procedure. The technique works well only if the zeros of $B(z)$ are well within the unit circle. It should be observed that (2.117) is the least squares solution corresponding to (2.107), in which the exactly known statistical correlations have been replaced by their estimates. The least squares approach does not utilize the additional information that $R_{nx}(k) = 0$ for $k > 0$.

ARMA parameter estimation continues to be an active area of research as no one method seems to stand out over another method in terms of its performance and/or lower computational complexity.

H. Pisarenko Harmonic Decomposition

If a stochastic process consists solely of sinusoids in additive white noise, then it is possible to model it as a special case ARMA process. Unlike the model for the periodogram, this model assumes the sinusoids are, in general, nonharmonically related. The mathematical properties of this special ARMA process leads to an eigenanalysis for the estimation of its parameters. Hence, a separate treatment from the general ARMA model discussion is given to this process.

Sinusoids in additive noise is a frequently used test process for evaluating spectrum analysis techniques. To motivate the selection of an ARMA process as the appropriate model for sinusoids in white noise, consider the following trigonometric identity:

$$\sin(\Omega n) = 2\cos\Omega \sin(\Omega[n-1]) - \sin(\Omega[n-2]) \tag{2.118}$$

for $-\pi < \Omega \leq \pi$. By letting $\Omega = 2\pi f \Delta t$, where $-1/2\Delta t < f \leq 1/2\Delta t$, $\sin \Omega n$ represents a sinusoid sampled at increments of Δt s. By setting $x_n = \sin(\Omega n)$, (2.118) may be rewritten as a second-order difference equation

$$x_n = (2\cos\Omega)x_{n-1} - x_{n-2} \tag{2.119}$$

permitting the current sinusoid value to be recursively computed from the two previous values x_{n-1} and x_{n-2}. If the z transform of (2.119) is taken, then

$$X(z)[1 - 2\cos\Omega z^{-1} + z^{-2}] = D(z) \tag{2.120}$$

where $D(z)$ is a polynomial of second degree that reflects the initial conditions. It has the characteristic polynomial $1 - 2\cos\Omega z^{-1} + z^{-2}$, or equivalently $z^2 - 2\cos\Omega z + 1$, with roots $z_1 = \exp(j2\pi f\Delta t)$ and $z_2 = z_1^* = \exp(-j2\pi f\Delta t)$. The roots are of unit modulus, $|z_1| = |z_2| = 1$, and the sinusoidal frequency in hertz is determined from the roots as follows:

$$f_i = [\tan^{-1}(\operatorname{Im}\{z_i\}/\operatorname{Re}\{z_i\})]/2\pi\Delta t, \quad \text{for } i = 1, 2. \tag{2.121}$$

Note that $f_1 = -f_2$. Observe that (2.119) is the limiting case of an AR(2) process in which the driving noise variance tends to zero and the poles tend to the unit circle. Also, with only two coefficients and knowledge of two samples, (2.119) makes it possible to perfectly predict the sinusoidal process for all time.

In general, a $2p$th-order difference equation of real coefficients of the form

$$x_n = -\sum_{m=1}^{2p} a_m x_{n-m} \tag{2.122}$$

can represent a deterministic process consisting of p real sinusoids of the form $\sin(2\pi f_i \Delta t)$. In this case, the $\{a_m\}$ are coefficients of the polynomial

$$z^{2p} + a_1 z^{2p-1} + \cdots + a_{p-1}z^{p+1} + a_p z^p + a_{p+1}z^{p-1} + \cdots + a_{2p-1}z + a_{2p} = \sum_{i=1}^{p}(z - z_i)(z - z_i^*) = 0 \tag{2.123}$$

with unit modulus roots that occur in complex conjugate pairs of the form $z_i = \exp(j2\pi f_i \Delta t)$, where the f_i are arbitrary frequencies such that $-1/2\Delta t \leq f_i < 1/2\Delta t$, and $i = 1, \cdots, p$. For this purely harmonic process, it can be shown that $a_i = a_{2p-i}$ for $i = 0, \cdots, p$.

For sinusoids in additive white noise w_n, the observed process is

$$y_n = x_n + w_n = -\sum_{m=1}^{2p} a_m x_{n-m} + w_n \tag{2.124}$$

where $E[w_n w_{n+k}] = \sigma_w^2 \delta_k$, $E[w_n] = 0$, and $E[x_n w_m] = 0$ since the noise is assumed to be uncorrelated with the sinusoids. Substituting $x_{n-m} = y_{n-m} - w_{n-m}$ into (2.124), it is possible to rewrite (2.124) as

$$\sum_{m=0}^{2p} a_m y_{n-m} = \sum_{m=0}^{2p} a_m w_{n-m} \tag{2.125}$$

where $a_0 = 1$ by definition. Expression (2.125), first developed by Ulrych and Clayton [251], represents the sinusoids in white-noise process in terms of the noise w_n and the noisy observations y_n; it has the structure of an ARMA(p, p). However, this ARMA has a special symmetry in which the AR parameters are identical to the parameters of the MA portion of the model.

If the autocorrelation function of y_n is known, the ARMA parameters can be found as the solution to an eigenequation, as is now shown. An equivalent matrix expression for (2.125) is [67]

$$\boldsymbol{Y}^T\boldsymbol{A} = \boldsymbol{W}^T\boldsymbol{A} \tag{2.126}$$

where

$$\boldsymbol{Y}^T = [y_n \quad y_{n-1} \quad \cdots \quad y_{n-2p}]$$
$$\boldsymbol{A}^T = [1 \quad a_1 \quad \cdots \quad a_{2p-1} a_{2p}]$$
$$\boldsymbol{W}^T = [w_n \quad w_{n-1} \quad \cdots \quad w_{n-2p}].$$

Premultiplying both sides of (2.148) by the vector $\boldsymbol{Y}$ and taking the expectation yields

$$E[\boldsymbol{Y}\boldsymbol{Y}^T]\boldsymbol{A} = E[\boldsymbol{Y}\boldsymbol{W}^T]\boldsymbol{A}. \tag{2.127}$$

Defining

$$\boldsymbol{X}^T = [x_n \cdots x_{n-2p}]$$

then

$$E[\boldsymbol{Y}\boldsymbol{Y}^T] = \boldsymbol{R}_{yy} = \begin{bmatrix} R_{yy}(0) & \cdots & R_{yy}(-2p) \\ \vdots & \ddots & \vdots \\ R_{yy}(2p) & \cdots & R_{yy}(0) \end{bmatrix} \tag{2.128}$$

$$E[\boldsymbol{Y}\boldsymbol{W}^T] = E[(\boldsymbol{X}+\boldsymbol{W})\boldsymbol{W}^T] = E[\boldsymbol{W}\boldsymbol{W}^T]$$
$$= \sigma_w^2 \boldsymbol{I}. \tag{2.129}$$

$\boldsymbol{R}_{yy}$ is the Toeplitz autocorrelation matrix for the observed process and $\boldsymbol{I}$ is the identity matrix. The fact that $E[\boldsymbol{X}\boldsymbol{W}^T] = 0$ follows from the assumption that the sinusoids are uncorrelated with the noise. Expression (2.127) is then rewritten as

$$\boldsymbol{R}_{yy}\boldsymbol{A} = \sigma_w^2\boldsymbol{A} \tag{2.130}$$

which is an eigenequation where the noise variance (σ_w^2) is an eigenvalue of the autocorrelation matrix $\boldsymbol{R}_{yy}$. The ARMA parameter vector $\boldsymbol{A}$ is the eigenvector associated with the eigenvalue σ_w^2, scaled so that the first element is unity. Equation (2.130) will yield the ARMA parameters when the lags are known. Knowledge of the noise variance σ_w^2 is not required. It may be shown [149, app. C] for a process consisting of p real sinusoids in additive white noise that σ_w^2 corresponds to the *minimum* eigenvalue of $\boldsymbol{R}_{yy}$ when the dimension of $\boldsymbol{R}_{yy}$ is $(2p+1) \times (2p+1)$ or greater (the minimum eigenvalue is repeated if the dimension is greater than $2p+1$).

Equation (2.130) forms the basis of a harmonic decomposition procedure developed by Pisarenko [195]. This procedure gives the exact frequencies and powers of p real sinusoids in white noise assuming exact knowledge of $2p+1$ autocorrelation lags, including the zero lag. Since only the autocorrelation lags are assumed known, phase information about each sinusoid is lost. Pisarenko noted the applicability of a trigonometric theorem of Caratheodory [82] for developing a method to find not only the frequencies $\Omega_i = 2\pi f_i \Delta t$, but also the powers $P_i = A_i^2/2$ and the noise PSD $\sigma_w^2 \Delta t$, from only knowledge of $2p+1$ values of the autocorrelation function. For sinusoids in white noise, the autocorrelation function is

$$R_{yy}(0) = \sigma_w^2 + \sum_{i=1}^{p} P_i$$

$$R_{yy}(k) = \sum_{i=1}^{p} P_i \cos(2\pi f_i k \Delta t), \quad \text{for } k \neq 0. \tag{2.131}$$

Noting that white noise only affects the zero lag term, Pisarenko was led to the eigenequation (2.130) by the approach of Caratheodoy's theorem, rather than by the approach presented here.

Once the ARMA coefficients a_i are found, the roots z_n of the polynomial

$$z^{2p} + a_1 z^{2p-1} + \cdots + a_{2p-1} z + a_{2p} = 0 \tag{2.132}$$

formed from the coefficients will yield the sinusoid frequencies, since the roots are of unit modulus with

$$z_n = \exp(j 2\pi f_n \Delta t). \tag{2.133}$$

See the discussion leading to (2.123). Due to the structure of (2.128), it turns out that (2.132) must be symmetrical, that is, $a_i = a_{2p-i}$.

No recursive technique is known for solving the eigenequation (2.130) for order p based on knowledge of the solution for order $(p-1)$. If the number of sinusoids is unknown, but the autocorrelation lags are exactly known, then independent solutions of (2.130) for successively higher orders must be computed until a point is reached where the minimum eigenvalue does not change from one order to the next higher order. This is an indication that the correct order has been reached. At this point, the minimum eigenvalue is the noise variance. In practice, only autocorrelation estimates are available, so that one must choose the number of sinusoids p as that order in which the minimum eigenvalue of (2.130) changes little from the minimum eigenvalue at order $p-1$. The computational requirements for solution of (2.130) can be reduced somewhat by utilizing the Toeplitz structure of the matrix $\boldsymbol{R}_{yy}$. The minimum eigenvalue and associated eigenvectors may be found by the classical power method in which the sequence of vectors

$$\boldsymbol{A}(k+1) = \boldsymbol{R}_{yy}^{-1}\boldsymbol{A}(k), \quad \text{for } k = 0, 1, \cdots \tag{2.134}$$

converges in the limit to the eigenvector of the minimum eigenvalue, for some initial guess $\boldsymbol{A}(0)$. Equation (2.134) can be rewritten as

$$\boldsymbol{R}_{yy}\boldsymbol{A}(k+1) = \boldsymbol{A}(k) \tag{2.135}$$

which can be solved for the unknown vector $\boldsymbol{A}(k+1)$, given $\boldsymbol{A}(k)$. Gaussian elimination type techniques require $o(p^3)$ operations. However, algorithms are available [63], [277] that use the Toeplitz structure of $\boldsymbol{R}_{yy}$ to solve (2.135) in $o(p^2)$ operations. A good starting vector is the all unity vector $\boldsymbol{A}^T(0) = [1, \cdots, 1]$, and the eigenvector $\boldsymbol{A}(\infty)$ is usually obtained after only a few iterations. Once $\boldsymbol{A}$ is found, the minimum eigenvalue $\lambda_{\min}$ (and therefore the noise variance estimate) is given by

$$\lambda_{\min} = \sigma_w^2 = \frac{\boldsymbol{A}^T\boldsymbol{R}_{yy}\boldsymbol{A}}{\boldsymbol{A}^T\boldsymbol{A}}. \tag{2.136}$$

Note that $\boldsymbol{A}(k)$ is rescaled for use in (2.134) each time by the Rayleigh quotient of (2.136) until convergence to $\lambda_{\min}$ is achieved.

Once the frequencies have been determined from the polynomial rooting of $\boldsymbol{A}$, the sinusoid powers can be determined. The autocorrelation lags $R_{yy}(1)$ to $R_{yy}(p)$ may be expressed in matrix form, based on (2.131), as

$$\boldsymbol{F}\boldsymbol{P} = \boldsymbol{r} \tag{2.137}$$

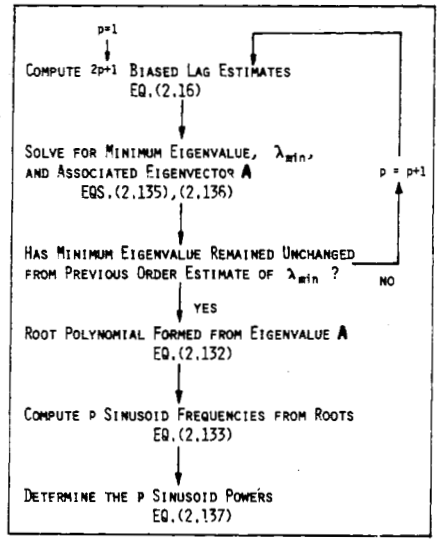

Fig. 12. Summary of Pisarenko spectral line decomposition procedure.

where

$$\boldsymbol{F} = \begin{bmatrix} \cos(2\pi f_1 \Delta t) & \cdots & \cos(2\pi f_p \Delta t) \\ \vdots & & \vdots \\ \cos(2\pi f_1 p\Delta t) & \cdots & \cos(2\pi f_p p\Delta t) \end{bmatrix}$$

$$\boldsymbol{P} = \begin{bmatrix} P_1 \\ \vdots \\ P_p \end{bmatrix} \quad \text{and} \quad \boldsymbol{r} = \begin{bmatrix} R_{yy}(1) \\ \vdots \\ R_{yy}(p) \end{bmatrix}.$$

The matrix $\boldsymbol{F}$ is composed of terms that depend upon the sinusoid frequencies as determined from polynomial rooting. The sinusoid powers are found by solving the simultaneous equation set (2.137) for the power vector $\boldsymbol{P}$. The noise power can also be determined from

$$\sigma_w^2 = R_{yy}(0) - \sum_{i=1}^{p} P_i.$$

Fig. 12 is a summary of the PHD technique.

Since the order is usually unknown, determining order by checking the minimum eigenvalue involves several solutions of eigenequation (2.130). This is not only computationally expensive, but it is also not often clear when the minimum eigenvalue has been reached since estimated lags, rather than known lags, are normally used. If the selected order is too high, p sinusoids for eigenequation order $2p$ will be computed, even though fewer than p sinusoids really exist. Thus spurious components will be introduced. If the order is too low, then the spectral components that are found tend to appear at incorrect frequencies. The use of the biased autocorrelation lag estimates guarantees a positive-definite Toeplitz autocorrelation matrix. However, the implied triangular windowing of the biased autocorrelation estimate, as discussed in Section II-B, yields significantly inaccurate frequency and power estimates for actual signals present. It will also introduce spurious components into the spectral decomposition. Unbiased lag estimates like (2.14) could be used, but the autocorrelation matrix is not guaranteed to be positive definite, as required in order to perform the Pisarenko decomposition. This can lead to negative eigenvalues and meaningless frequency estimates. Non-Toeplitz positive-definite autocorrelation matrix forms such as $\boldsymbol{X}_1^H \boldsymbol{X}_1$ from Section II-E have been tried in place of $\boldsymbol{R}_{yy}$ in (2.130), but they produce nonunit modulus roots and do not seem to improve the results. For colored noise that contributes to a finite number of lags beyond lag zero, a modification of the Pisarenko technique can be used [214].

The eigenanalysis approach of the Pisarenko technique can be generalized to the idea of extracting the most information concerning a signal by processing for the largest eigenvalues and corresponding eigenvectors of an estimated correlation matrix [135], [185], [252].

J. Prony's Energy Spectral Density [36], [49], [101], [102], [156], [160], [199], [200], [217], [253], [261]

Prony's method, a technique for modeling data of equally spaced samples by a linear combination of exponentials, is not a spectral estimation technique in the usual sense, but a spectral interpretation is provided in this section. Gaspard Riche, Baron de Prony [202], was led to believe that laws governing expansion of various gases could be represented by sums of exponentials. He proposed a method for providing interpolated data points in his measurements by fitting an exponential model to the measured points and computing the interpolated values by evaluation of the exponential model at these points. The modern version of Prony's method bears little resemblance to his original approach due to evolutionary changes that have been made. The original procedure exactly fitted an exponential curve having p exponential terms (each term has two parameters—an amplitude A_i and an exponent α_i where $A_i \exp(\alpha_i t)$) to $2p$ data measurements. This approach is discussed in Hildebrand [95]. For the case where only an *approximate* fit with p exponentials to a data set of N samples is desired, such that $N > 2p$, a least squares estimation procedure is used. This procedure is called the extended Prony method.

The model assumed in the extended Prony method is a set of p exponentials of arbitrary amplitude, phase, frequency, and damping factor. The discrete-time function

$$\hat{x}_n = \sum_{m=1}^{p} b_m z_m^n, \quad \text{for } n = 0, \cdots, N-1 \qquad (2.138)$$

is the model to be used for approximating the measured data $x_0, \cdots, x_{N-1}$. For generality, b_m and z_m are assumed complex and

$$b_m = A_m \exp(j\theta_m)$$
$$z_m = \exp[(\alpha_m + j2\pi f_m)\Delta t] \qquad (2.139)$$

where A_m is the amplitude, θ_m is the phase in radians, α_m is a damping factor, f_m is the oscillation frequency in hertz, and Δt represents the sample interval in seconds. Finding $\{A_m, \theta_m, \alpha_m, f_m\}$ and p that minimize the squared error

$$\mathcal{E} = \sum_{n=0}^{N-1} |x_n - \hat{x}_n|^2 \qquad (2.140)$$

is a difficult nonlinear least squares problem. The solution involves an iterative process in which an initial guess of the unknown parameters is successively improved. McDonough and Huggins [157] and Holtz [97] provide such iterative schemes for the solution of (2.140). An alternative suboptimum solution that does not minimize (2.140) but still provides satisfactory results, is based on Prony's technique. Prony's method solves two sequential sets of linear equations with an intermediate polynomial rooting step that concentrates

the nonlinearity of the problem in the polynomial rooting procedure.

The key to the Prony technique is to recognize that (2.138) is the homogeneous solution to a constant coefficient linear difference equation, the form of which is found as follows. Define the polynomial $\Psi(z)$ as

$$\Psi(z) = \prod_{k=1}^{p} (z - z_k) = \sum_{i=0}^{p} a_i z^{p-i}, \quad a_0 = 1. \quad (2.141)$$

Thus $\Psi(z)$ has the complex exponentials z_k of (2.139) as its roots and complex coefficients a_i when multiplied out. Based on (2.138), one way of expressing $\hat{x}_{n-m}$ is

$$\hat{x}_{n-m} = \sum_{l=1}^{p} b_l z_l^{n-m} \quad (2.142)$$

for $0 \leqslant n - m \leqslant N - 1$. Multiplying (2.142) by a_m and summing over the past $p + 1$ products yields

$$\sum_{m=0}^{p} a_m \hat{x}_{n-m} = \sum_{l=1}^{p} b_l \sum_{m=0}^{p} a_m z_l^{n-m} \quad (2.143)$$

defined for $p \leqslant n \leqslant N - 1$. If in (2.143) the substitution $z_l^{n-m} = z_l^{n-p} z_l^{p-m}$ is made, then

$$\sum_{m=0}^{p} a_m \hat{x}_{n-m} = \sum_{l=1}^{p} b_l z_l^{n-p} \sum_{m=0}^{p} a_m z_l^{p-m} = 0. \quad (2.144)$$

The zero result in (2.144) follows by recognizing that the final summation above is just the polynomial $\Psi(z_l)$ of (2.141), evaluated at one of its roots. Expression (2.144) then yields the recursive difference equation

$$\hat{x}_n = -\sum_{m=1}^{p} a_m \hat{x}_{n-m} \quad (2.145)$$

defined for $p \leqslant n \leqslant N - 1$. Compare this with (2.122) of the PHD procedure. Thus the exponential parameters are found by rooting polynomial (2.141) using the a_m coefficients.

To set up the extended Prony method, first define the difference between the actual measured data x_n and the approximation $\hat{x}_n$ to be e_n, so that

$$x_n = \hat{x}_n + e_n \quad (2.146)$$

defined for $0 \leqslant n \leqslant N - 1$. Substituting (2.145),

$$x_n = -\sum_{m=1}^{p} a_m \hat{x}_{n-m} + e_n$$

$$= -\sum_{m=1}^{p} a_m x_{n-m} + \sum_{m=0}^{p} a_m e_{n-m} \quad (2.147)$$

defined for $p \leqslant n \leqslant N - 1$, where $\hat{x}_{n-m} = x_{n-m} - e_{n-m}$ has been used. Based on (2.147), an alternative model to the sum of exponentials plus additive noise model is that of an ARMA model with identical AR and MA parameters driven by the noise process e_n. Unlike the Pisarenko technique, the a_i coefficients are not constrained to produce polynomial roots of unit modulus (no damping). Although the true least squares estimate of the parameters is obtained by minimizing

$$\sum_{n=p}^{N-1} |e_n|^2$$

this leads to a set of nonlinear equations that are difficult to solve. An alternative procedure, termed the extended Prony approach [266], defines

$$\epsilon_n = \sum_{m=0}^{p} a_m e_{n-m}, \quad \text{for } n = p, \cdots, N - 1 \quad (2.148)$$

so that

$$x_n = -\sum_{m=1}^{p} a_m x_{n-m} + \epsilon_n. \quad (2.149)$$

One then minimizes $\Sigma_{n=p}^{N-1} |\epsilon_n|^2$, rather than $\Sigma_{n=p}^{N-1} |e_n|^2$. Thus the extended Prony parameter estimation procedure reduces to that of an AR parameter estimation for which the least square covariance algorithm of (2.69) with $\boldsymbol{X}_1^H \boldsymbol{X}_1$ may be used. Note that the nonwhite random input process ϵ_n is derived from a MA process driven by the approximation error e_n, as indicated by (2.148). Also, ϵ_n is the difference between x_n and its linear prediction based on p past data samples, whereas e_k is the difference between x_n and its exponential approximation. The number of exponentials p is determined using the AR order selection techniques discussed in Section II-E. An alternate scheme to determine p involves an eigenanalysis of the $\boldsymbol{X}_1^H \boldsymbol{X}_1$ matrix [252] and bears a close relationship to the non-Toeplitz Pisarenko eigenanalysis discussed in Section II-H.

Once the z_i have been determined from the polynomial rooting, expression (2.138) reduces to a set of linear equations in the unknown b_m parameters, expressible in matrix form as

$$\boldsymbol{\Phi}\, \boldsymbol{B} = \hat{\boldsymbol{X}} \quad (2.150)$$

where

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_p \\ \vdots & \vdots & & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_p^{N-1} \end{bmatrix}$$

$$\boldsymbol{B} = [b_1 \ \cdots \ b_p]^T$$

$$\hat{\boldsymbol{X}} = [\hat{x}_0 \ \cdots \ x_{N-1}]^T.$$

Note that $\boldsymbol{\Phi}$ is a Van der Monde matrix similar to (2.24), except that the z_i terms have damping and arbitrary frequency assignments instead of a harmonic relationship. A least squares minimization of $\Sigma(x - \hat{x})^2$ yields the well-known solution

$$\boldsymbol{B} = [\boldsymbol{\Phi}^H \boldsymbol{\Phi}]^{-1} \boldsymbol{\Phi}^H \boldsymbol{X}. \quad (2.151)$$

A useful relationship that reduces the computational burden of (2.151) is

$$\boldsymbol{\Phi}^H \boldsymbol{\Phi} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1p} \\ \vdots & & \vdots \\ \gamma_{p1} & \cdots & \gamma_{pp} \end{bmatrix}$$

where

$$\gamma_{ij} = \frac{(z_i^* z_j)^N - 1}{(z_i z_j) - 1}. \quad (2.152)$$

Determining the a_i parameters by a least squares estimation, rooting the polynomial, and then solving for the b_j parameters (or residues) constitute the extended Prony method. To obtain the amplitude A_i, phase θ_i, damping factor α_i, and fre-

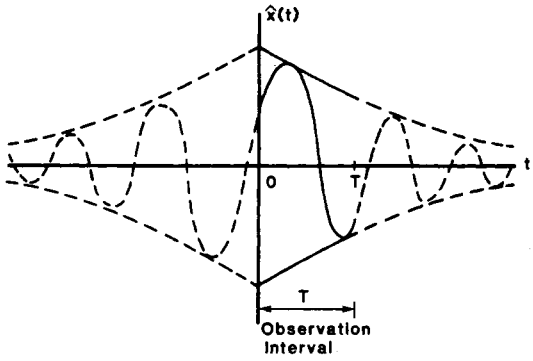

Fig. 13. Symmetric envelope exponential model.

quency f_i from the z_i and b_i estimates, simply compute

$$A_i = |b_i|$$
$$\theta_i = \tan^{-1}\,[\mathrm{Im}\,(b_i)/\mathrm{Re}\,(b_i)]$$
$$\alpha_i = \ln|z_i|/\Delta t$$
$$f_i = \tan^{-1}\,[\mathrm{Im}\,(z_i)/\mathrm{Re}\,(z_i)]/2\pi\Delta t. \qquad (2.153)$$

Normally, the Prony method is completed with the computation of the exponential parameters given in (2.153). As such, Prony's method has found most of its application in transient analysis, such as finding resonant modes in electromagnetic pulse problems [196]. However, a "spectrum analysis" can be performed in the following manner. Although many different spectra could be defined, one "spectrum" found to be useful makes the assumption that the model of the process has symmetry as illustrated for one damped real sinusoid in Fig. 13. The assumed approximation function becomes

$$\hat{x}(t) = \sum_{m=1}^{p} A_m \exp(\alpha_m|t|)\exp(j[2\pi f_m t + \theta_m]) \qquad (2.154)$$

defined for $-\infty < t < \infty$. For $x(t)$ real, complex conjugate pairs like $\exp j(2\pi f_m t + \theta_m)$ and $\exp -j(2\pi f_m + \theta_m)$ in (2.154) are required. It is further assumed that all the damping factors are negative, so that decaying exponentials are obtained. One motivation for the selection of a symmetric envelope is that for $\alpha = 0$, $\hat{x}(t)$ will have undamped sinusoidal components which are defined over $-\infty < t < \infty$. As a result, unwindowed sinusoids are accurately modeled by this approach.

Since (2.154) is a finite energy, deterministic expression, its ESD based on the Fourier transform of (2.154) is

$$\hat{\mathcal{S}}_{\mathrm{PRONY}}(f) = |\hat{X}(f)|^2 \qquad (2.155)$$

where

$$\hat{X}(f) = \sum_{m=1}^{p} A_m \exp(j\theta_m)\frac{2\alpha_m}{[\alpha_m^2 + (2\pi[f - f_m])^2]}. \qquad (2.156)$$

This then constitutes one possible Prony "spectrum." Note that the spectral estimate (2.155) maintains peaks that are linearly proportional to the energy, unlike AR spectra peaks, which are nonlinearly related to power [136]. The Prony spectrum has the ability to produce narrow-band or wide-band spectral shapes, the shapes being a function of the size of the damping factor (illustrated in Fig. 14). The bell shaped curves have bandwidths (to the -3 dB points) of α/π Hz, so resolution varies as a function of damping. Note that for selection of model order p, the Prony spectrum requires $2p$ parameters to characterize the spectrum, which is twice that required for the

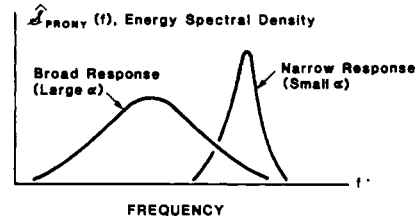

Fig. 14. Narrow-band and wide-band Prony spectral responses.

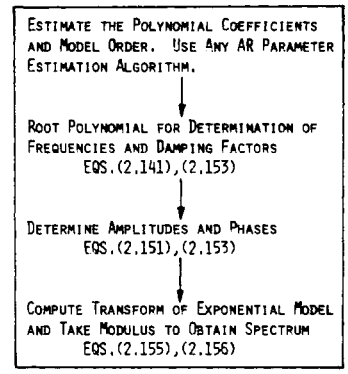

Fig. 15. Prony spectrum estimation procedure.

AR spectral estimate. Also, the Prony method yields phase information not available with AR spectral estimation. Fig. 15 summarizes the Prony spectrum estimation procedure.

The Prony technique is a data adaptive procedure in the sense that it adjusts the parameters of a damped exponential model of varying frequency, phase, amplitude, and damping to fit the data. The periodogram, in contrast, uses a fixed number of undamped sinusoids of fixed frequencies.

There are several problems of which one should be aware when applying Prony's method. The problem of determinating the number of exponential terms is similar to the problem of model order selection in AR estimation, so that the same considerations apply. However, since $2p$ parameters are computed, the maximum order is limited to be $p \leqslant N/2$, whereas $p > N/2$ is possible with AR spectral estimation (although not advisable). Noise impacts the accuracy of the Prony pole estimates greatly in some situations [252]-[254]. Noise also can cause the damping factors to be too large.

K. Prony Spectral Line Estimation

For a process consisting of p real undamped ($\alpha = 0$) sinusoids in noise, a special variant of Prony's method has been developed. The basic approach was described by Hildebrand [95]. In this case, (2.138) may be expressed as

$$\hat{x}_n = \sum_{m=1}^{p}[b_m z_m^n + b_m^* z_m^{*n}] = \sum_{m=1}^{p} A_m \cos(2\pi f_m n\Delta t + \theta_m) \qquad (2.157)$$

where $b_m = A_m \exp(j\theta_m)/2$ and $z_m = \exp(j2\pi f_m \Delta t)$. Note that the z_m are roots of unit modulus with arbitrary frequencies and occur in complex conjugate pairs as long as $f_m \neq 0$ or $1/2\Delta t$. Thus one must solve (2.141) for the roots of the polynomial

$$\Psi(z) = \prod_{i=1}^{p}(z - z_i)(z - z_i^*) = \sum_{k=0}^{2p} a_k z^{2p-k} = 0 \qquad (2.158)$$

with $a_0 = 1$ and the a_i being real coefficients. Since the roots are of unit modulus and occur in complex conjugate pairs, then (2.158) must be invariant under the substitution z^{-1} for z,

$$z^{2p}\Psi\left(\frac{1}{z}\right) = z^{2p}\sum_{k=0}^{2p} a_k z^{k-2p} = \sum_{k=0}^{2p} a_k z^k = 0. \quad (2.159)$$

Comparing (2.158) and (2.159), one may conclude that $a_j = a_{2p-j}$ for $j = 0$ to p, with $a_0 = a_{2p} = 1$. Thus the requirement for complex conjugate root pairs of unit modulus is implemented by constraining the polynomial coefficients to be symmetric about the center element. Based on order $2p$, a linear prediction error similar to (2.149) can be rewritten as

$$\epsilon_n = \sum_{m=0}^{p} a_m(x_{n+m} + x_{n-m}) \quad (2.160)$$

which reduces the number of coefficients required by one-half. All the least squares minimization approaches apply, except now the data matrix $\boldsymbol{X}$ that represents (2.160) is a data matrix of Toeplitz plus Hankel structure, rather than just Toeplitz as in the AR case. For example, using ϵ_n ranging from $n = p$ to N, we have

$$\begin{bmatrix} \epsilon_{p+1} \\ \vdots \\ \epsilon_{N-p} \end{bmatrix} = \boldsymbol{X}\boldsymbol{A} \quad (2.161)$$

where

Toeplitz Data Matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_{p-1} \\ \frac{1}{2}a_p \end{bmatrix}, \quad \boldsymbol{X} = \boldsymbol{T} + \boldsymbol{H}, \quad \boldsymbol{T} = \begin{bmatrix} x_{p+1} & \cdots & x_1 \\ \vdots & \ddots & \vdots \\ x_{N-2p} & & x_{p+1} \\ \vdots & \ddots & \vdots \\ x_{N-p} & \cdots & x_{N-2p} \end{bmatrix}$$

Hankel Data Matrix

$$\boldsymbol{H} = \begin{bmatrix} x_{p+1} & \cdots & x_{2p+1} \\ \vdots & \ddots & \vdots \\ x_{2p+1} & & x_{N-p} \\ \vdots & \ddots & \vdots \\ x_{N-p} & \cdots & x_N \end{bmatrix}$$

so that the minimum of the squared error $\Sigma_{n=p+1}^{N-p} |\epsilon_n|^2$ determines the real coefficients $a_1, \cdots, a_{p-1}, a_p/2$ analogous to those solutions in AR batch estimation. Note that the last coefficient is $a_p/2$ rather than a_p. The factor of half is due to a symmetry in $\boldsymbol{X}$ that counts the last factor a_p twice. These coefficients are used to set up the order $2p + 1$ symmetrical coefficient polynomial (2.158). Although the unit modulus roots give rise to a symmetrical polynomial, the converse is not necessarily true. Symmetric coefficients only guarantee that if a root z_i occurs, then so does its reciprocal z_i^{-1}; to have $|z_i| = 1$ is not required. In practice [226], nonunit modulus roots are only observed rarely, but when they do occur the roots are usually at $f_i = 0$ or $f_i = 1/2\Delta t$. The algorithm is completed with the determination of amplitude and phase as given by (2.151), which can be reduced in size by one-half by combining related complex pairs. The spectrum will then consist of delta functions, representing the sinusoids, and damped exponentials for those rare cases of nonunit modulus root pairs.

The Prony harmonic decomposition technique described above has several performance advantages over the PHD procedure. For one, autocorrelation lag estimates are not required with the Prony method. The Prony method appears from experiments to yield fewer spurious spectral lines than the Pisarenko approach since the order can be better determined by monitoring the residual squared error of the special Prony method. Also, the frequency and power estimates are less biased than those obtained from the Pisarenko method [153], [226]. See Fig. 16 in the summary section for a comparison of the spectral lines given by each approach. The Prony method requires only the solution of two sets of simultaneous linear equations and a polynomial rooting. The Pisarenko approach requires a more computationally complex eigenequation solution.

L. Maximum Likelihood (Capon) Spectral Estimation [42], [137], [203]

In maximum likelihood spectral estimation (MLSE), originally developed for seismic array frequency-wave number analysis [42], one estimates the PSD by effectively measuring the power out of a set of narrow-band filters [136]. MLSE is actually a misnomer in that the spectral estimate is not a true maximum likelihood estimate of PSD. MLSE is sometimes referenced as the Capon spectral estimate [92]. The name MLSE is retained here only for historic reasons. The difference between MLSE and conventional BT/periodogram spectral estimation is that the shape of the narrow-band filters in MLSE are, in general, different for each frequency whereas they are fixed with the BT/periodogram procedures. The filters adapt to the process for which the PSD is sought. In particular, the filters are finite impulse response (FIR) types with p weights (taps),

$$\boldsymbol{A} = [a_0 a_1 \cdots a_{p-1}]^T. \quad (2.162)$$

The coefficients are chosen so that at the frequency under consideration, f_0, the frequency response of the filter is unity (i.e., an input sinusoid at that frequency would be undistorted at the filter output) and the variance of the output process is minimized. Thus the filter should adjust itself to reject components of the spectrum not near f_0 so that the output power is due mainly to frequency components close to f_0. To obtain the filter, one minimizes the output variance σ^2, given by

$$\sigma^2 = \boldsymbol{A}^H \boldsymbol{R}_{xx} \boldsymbol{A} \quad (2.163)$$

subject to the unity frequency response constraint (so that the sinusoid of frequency f_0 is filtered without distortion)

$$\boldsymbol{E}^H \boldsymbol{A} = 1 \quad (2.164)$$

where $\boldsymbol{R}_{xx}$ is the covariance matrix of x_n, and $\boldsymbol{E}$ is the vector

$$\boldsymbol{E} = [1 \exp(j2\pi f_0 \Delta t) \cdots \exp(j2\pi[p-1] f_0 \Delta t)]^T$$

and H denotes the complex conugate transpose. The solution for the filter weights is easily shown to be [203]

$$\boldsymbol{A}_{\text{OPT}} = \frac{\boldsymbol{R}_{xx}^{-1}\boldsymbol{E}}{\boldsymbol{E}^H \boldsymbol{R}_{xx}^{-1}\boldsymbol{E}} \quad (2.165)$$

and the minimum output variance is then

$$\sigma^2_{\text{MIN}} = \frac{1}{\boldsymbol{E}^H \boldsymbol{R}_{xx}^{-1}\boldsymbol{E}}. \quad (2.166)$$

It is seen that the frequency response of the optimum filter is unity at f_0 and that the filter characteristics change as a

function of the underlying autocorrelation function. Since the minimum output variance is due to frequency components near f_0, then $\sigma^2_{MIN}\Delta t$ can be interpreted as a PSD estimate. Thus, the MLSE PSD is defined as

$$\hat{\mathcal{P}}_{ML}(f_0) = \frac{\Delta t}{E^H R_{xx}^{-1} E}. \tag{2.167}$$

To compute the spectral estimate, one only needs an estimate of the autocorrelation matrix.

In practice, the MLSE exhibits more resolution than the periodogram and BT spectral estimators, but less than an AR spectral estimator [136]. When the autocorrelation function must be estimated, it has been observed and verified analytically for large data records that the MLSE exhibits less variance than the AR spectral estimate [13]. It should be noted that for a narrow-band process, in which the autocorrelation function is known, the peak of the AR spectrum is proportional to the square of the power of the process, while for the MLSE the peak is proportional to the power [136], [137]. Also, the AR spectral estimate power can be found by determining the area under the peak, while the area under a MLSE peak is proportional to the square root of the power.

The MLSE and ARSE have been related analytically as follows [40]:

$$\frac{1}{\hat{\mathcal{P}}_{ML}(f)} = \frac{1}{p}\sum_{m=1}^{p} \frac{1}{\hat{\mathcal{P}}_{AR}^{(m)}(f)} \tag{2.168}$$

where $\hat{\mathcal{P}}_{AR}^{(m)}(f)$ is the AR PSD for an mth order model and $\hat{\mathcal{P}}_{ML}(f)$ is the MLSE PSD, both based upon a known autocorrelation matrix of order p [40], [203]. Thus the lower resolution of the MLSE can be explained by the "parallel resistor network averaging" effect of combining the low-order AR spectra of least resolution with the high order AR spectra of highest resolution. Also of interest is the fact that the inverse Fourier transform of $\hat{\mathcal{P}}_{ML}(f)$, which yields the estimated autocorrelation function, is not identical to the autocorrelation function used to obtain the PSD. The inverse Fourier transform of the AR PSD, on the other hand, yields the identical autocorrelation functions over the known range of lag values, as indicated by (2.61).

III. Summary of Techniques

Table II provides a summary of eleven of the more commonly used spectral estimation techniques presented in this paper. A brief overview of key properties, equation references for computing each spectral estimate, and a list of key references will aid the reader to readily implement any of the techniques.

Fig. 16 illustrates typical spectra of the eleven techniques described in Table II. Each spectral estimate is based on the same 64-point real sample sequence from a process consisting of three sinusoids and a colored noise process obtained by filtering a white Gaussian process. Table III is a list of the data samples used. The true PSD is shown in Fig. 16(a). The frequency axis ranges from 0.0 to 0.5 and represents the fraction of the sampling frequency. The three sinusoids are at fractional frequencies of 0.10, 0.20, and 0.21 and have SNR's of +10, +30, and +30 dB, respectively, where SNR is defined as the ratio of the sinusoid power to the total power in the passband noise process. The noise process passband is centered at 0.35. This particular signal was selected to demonstrate how each spectral estimation technique performs against both narrow-band and wide-band processes. Fig. 16 is intended to illustrate properties of each technique, especially for short data records, rather than to serve as a basis for comparing relative performance among the techniques.

A periodogram based on the 64 data samples of Table III is shown in Fig. 16(b). The periodogram was generated with an FFT that had been double padded with 64 zeros. The nominal resolution in Hz of a 64-point sequence is 0.015 625 times the sampling frequency, so that the sinusoids at 0.20 and 0.21 are closer than the resolution width. Indeed, the periodogram shown here is unable to resolve these two sinusoidal components. The weaker sinusoid can be seen among the sidelobes (no data windowing was used). The presence of the colored noise is also indicated by the discrete spectral lines in the upper part of the frequency band. Fig. 16(c) illustrates the BT spectrum, based on 16 autocorrelation lag estimates. The number of lags was around 20 percent of the number of data samples, as recommended by Blackman and Tukey.

Several AR PSD estimates are pictured in Figs. 16(d)–(f). Although all are AR spectral estimates, differing only in the manner that the AR coefficients are estimated, the resulting AR spectra are quite different. Using the 64 samples, sixteen coefficients were computed for the AR and all the remaining techniques to be discussed. The Yule–Walker AR approach, which requires estimation of lags, does not resolve the two closely spaced sinusoids (it has the least resolution of all AR methods) and does not give much insight into the spectrum on either side of the main response. The AR PSD estimate based on the Burg algorithm shown in Fig. 16(e) provides sharp responses at the three sinusoid frequencies, although the one at .1 is barely visible on the scale shown. It also shows power is present at the high frequency end of the spectrum, although it is not a smooth, broad spectrum as it should be. This illustrates the "peaky" nature of AR spectra. A more accurate response for the three sinusoids frequencies is obtained with the forward–backwards (or least squares) technique for AR spectral estimation, as shown in Fig. 16(f). Otherwise, AR spectra Figs. 16(e) and (f) are comparable.

The MA PSD estimate is depicted in Fig. 16(g). It is identical to the BT spectrum since only autocorrelation lag estimates were used. The broad-band response of the MA spectrum stands in contrast to the sharp narrow-band response of the AR spectra. It is unable to resolve the two close sinusoids; the response around 0.1 is as broad as the response at the high frequency end of the spectrum, making it difficult to detect narrow-band components in a wide-band response. One ARMA (8,8) PSD estimate is illustrated in Fig. 16(h), based on the modified Yule-Walker approach with biased lag estimates computed from the data samples. It is not a very good spectral estimate, although an ARMA (16,16) spectrum not shown here was able to separate the three sinusoid components.

The Pisarenko spectral line decomposition of Fig. 16(i) is, like the FFT periodogram, a discrete spectrum. The two close sinusoids are resolved, but the frequencies and powers are grossly inaccurate. The lower level sinusoid at 0.10 has a spectral line near this frequency, but there are many other spectral lines, making selection of actual signals from spurious components difficult. The broad portion of the spectrum has been modeled by placing several spectral lines in the area of the broad-band process spectrum. Thus, the Pisarenko method does not model the broad-band processes well, though it shows there is power in this frequency region.

The energy spectral density based on the extended Prony method yields the spectrum shown in Fig. 16(j). The three

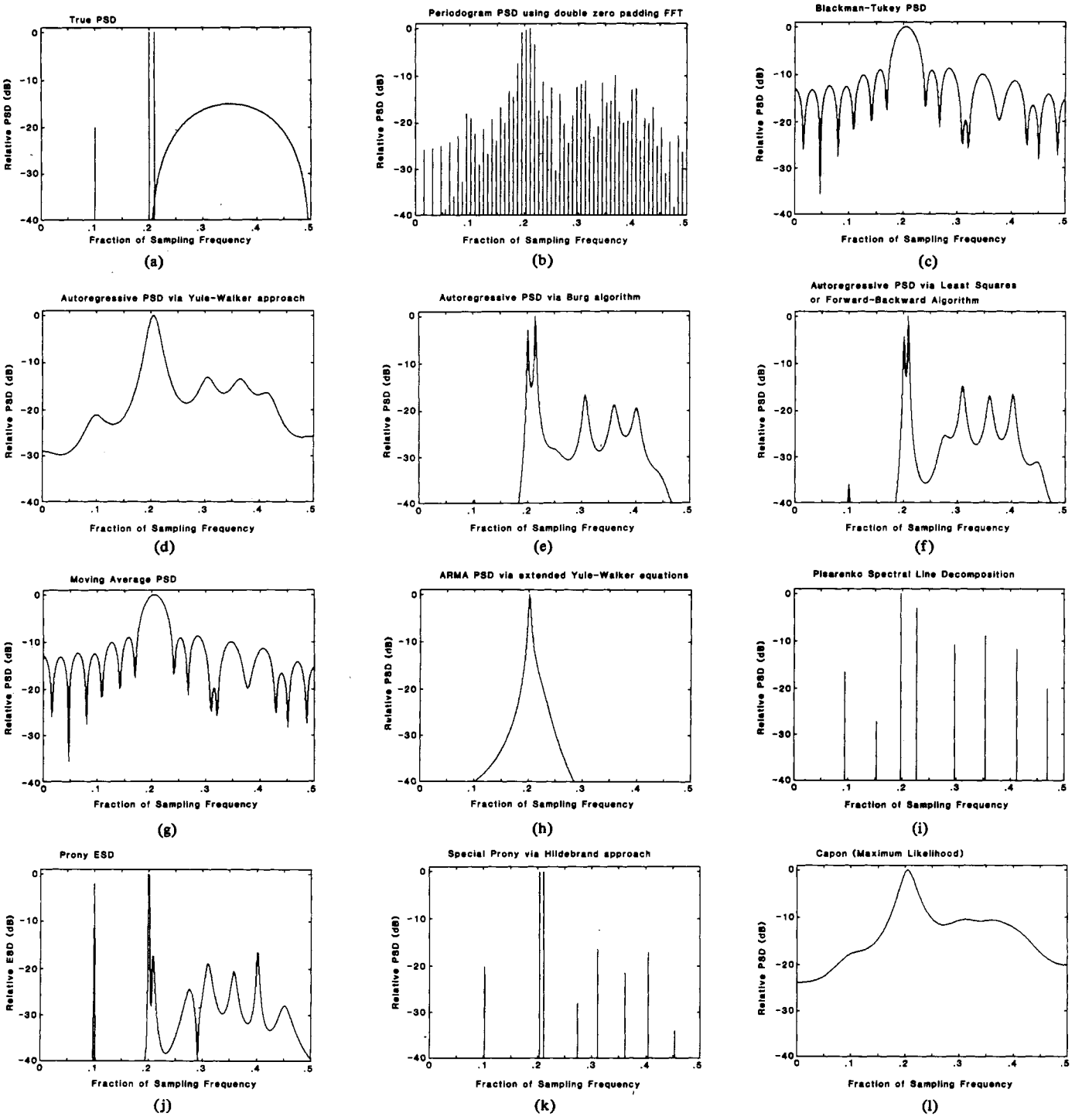

Fig. 16. Illustration of various spectra for the same 64-point sample sequence.

sinusoid components have very sharp responses at the sinusoid frequencies, with a broad response at the higher end of the spectrum. Table IV shows the actual parameter estimates obtained with the Prony method. The actual amplitudes for the sinusoids of frequencies 0.1, 0.2, and 0.21 were 0.1, 1., and 1., respectively. The most accurate estimates of the three sinusoid powers and frequencies is provided by the spectral line decomposition variant of the Prony method, pictured in Fig. 16(k). This is no surprise since this technique is the least squares approach that assumes a sinusoidal model. It is a discrete spectrum so that the broad-band process is not well modeled, although several lines are present to indicate spectral power in this region. Table V lists the actual parameter estimates obtained with this procedure.

The maximum likelihood spectrum, shown in Fig. 16(l) has a smooth spectrum. It cannot resolve the two closely spaced sinusoidal components. The smooth nature of the MLSE spectrum, being the equivalent of an average of all the AR spectra from order 1 to 16, is typical of this method.

If more accurate frequency estimation of noisy sinusoids

TABLE II
SUMMARY OF MODERN SPECTRAL ESTIMATION METHODS

TECHNIQUE	KEY REFS.	DISCRETE LINE OR CONTINUOUS SPECTRUM	ALGORITHM PROCEDURE	ROUGH COMPUTATIONAL COMPLEXITY (ADDS / MULTS)	MODEL(S)	ADVANTAGES AND DISADVANTAGES	REMARKS
Periodogram* FFT Version	24 46 207 218 262	D	Eqn. (2.18)	$N \log_2 N$	Sum of harmonically related sinusoids	Output directly proportional to power Most computationally efficient Resolution roughly the reciprocal of the observation interval Performance poor for short data records Leakage distorts spectrum and masks weak signals	Harmonic least squares fit Requires some type of frequency domain statistical averaging to stabilize spectrum (e.g., Welch method) Windowing can reduce sidelobes at expense of resolution
Blackman-Tukey (BT)**	25 33 204	C	Eqns. (2.13) & (2.16)	Lag Ests.: NM PSD Est.*: MS	Identical to MA with windowing of the lags	Most computationally efficient if $M<N$ Resolution roughly the reciprocal of the observation interval Leakage distorts spectrum and masks weak signals	Negative PSD values in spectrum may result with some window weightings and autocorrelation estimates (e.g., unbiased)
Autoregressive (AR) Yule-Walker Version	59 108 250 276	C	Fig. 6**	Lag. Ests.: NM AR Coeffs.: M^3 PSD Est.*: MS	Autoregressive (all-pole) process	Model order must be selected Better resolution than FFT or BT, but not as good as other AR methods Spectral line splitting occurs Implied windowing distorts spectrum No sidelobes	Model applicable to seismic, speech, radar clutter data Minimum-phase (stable) linear prediction filter guaranteed if biased lag estimates computed AR related to linear prediction analysis and adaptive filtering Models peaks in spectrum better than valleys
Autoregressive (AR) Burg-Algorithm Version	* 9 15 37 41 57 64 65	C	Fig. 7	AR Coeffs.: $NM + M^2$ PSD Est.*: MS	Autoregressive (all-pole) process	High resolution for low noise levels Good spectral fidelity for short data records Spectral line splitting can occur Bias in the frequency estimates of peaks No implied windowing No sidelobes Must determine order	Stable linear prediction filter guaranteed Adaptive filtering applicable Uses constrained recursive least squares approach
Autoregressive (AR) Least squares or forward-backward linear prediction version	154 227	C	See Ref. 154 for flowchart	AR Coeffs.: $MN + M^2$ PSD Est.*: MS	Autoregressive (all-pole) process	Sharper response for narrowband processes than other AR estimates No spectral line splitting observed Bias reduced in the frequency estimates Must determine order No sidelobes	Stable linear prediction filter not guaranteed, though stable filter results in most instances Based on exact recursive least squares solution with no constraint
Moving Average (MA)	47 236	C	Eqns. (2.104) & (2.35)	MA Coeffs.: Nonlinear Simult. Eqn. Set Lag Ests.: NM PSD Est.*: MS	Moving average (all-zero) process	Broad spectral responses (low resolution) Must determine order Has sidelobes	Generalized form of BT technique
ARMA (Yule Walker Version)	31 97	C	High Order YW Eqns. (2.108)	Lag Ests.: NM Coeff. Computation: M^3 PSD Est.*: MS	ARMA process (Rational Transfer Function) (MA order ≠ AR order)	Must determine AR & MA orders	Models all rational transfer function processes Requires accurate lag estimates to obtain good results
Pisarenko Harmonic Decomposition (PHD)	195 251	D	Fig. 12	Lag Ests.: NM Eigen eqn.: M^2 to M^3 Poly. Rooting: Dependent on Root Algorithm Powers: M^3	Special ARMA with equal MA and AR coefficients Sum of nonharmonically undamped sinusoids in additive white noise	Must determine order Does not work well in high noise levels Eigenequation and rooting are computationally inefficient	Requires accurate lag estimates to obtain good results Spurious spectral lines if order selected too high
Prony's Method (Extended)	95 97 252	C	Fig. 15	AR Coeffs.: $M^2 + NM$ Poly. Rooting: Dependent on Root Algorithm Amp. Coeffs.: M^3 PSD Est.*: MS	Sum of nonharmonically related damped exponentials ARMA with equal MA and AR coefficients and equal orders (p=q)	Must determine order Output linearly proportional to power Requires a polynominal rooting Resolution as good as AR techniques, sometimes better No sidelobes	Uses least squares estimates to obtain exponential parameters First step same as AR least squares estimation
Prony Spectral Line Decomposition	95 226	D	Eqns. (2.161) (2.158) (2.151) (2.153)	Coeffs.: M^3 Rooting: Function of root algorithm used Amp. Coeffs.: M^3	Sum of nonharmonically related sinusoids	Must determine order Output linearly proportional to power Requires a polynominal rooting Resolution as good as AR techqniues, sometimes better No sidelobes	Uses least squares estimation
Capon Maximum Likelihood (MLSE)	40 136 193 203	C	Eqns. (2.167) & (2.16)	Lag Ests.: NM Matrix Inversion M^3 PSD Ests.*: MS	Forms an optimal bandpass filter for each spectral component	Resolution better than BT; not as good as AR Statistically less variability in MLSE spectra than AR spectra	MLSE is related to AR spectra (see Eqn. (2.160))

** Computer programs may be found in Programs for Digital Signal Processing, edited by Digital Signal Processing Committee of the IEEE ASSP Society, IEEE Press, 1979.
* FFT could be used to generate $S = 2^{\nu}$ values of the PSD.
N=Number of data samples
S=Number of Spectral Samples Computed (usually S>>M)
M=Order of Model (or number of Autocorrelation Lags)

TABLE III
LIST OF DATA SAMPLES

X(1)=	1.291061	X(33)=	0.309840
X(2)=	-2.086368	X(34)=	1.212892
X(3)=	-1.691316	X(35)=	-0.119985
X(4)=	1.243138	X(36)=	-0.441686
X(5)=	1.641072	X(37)=	-0.879733
X(6)=	-0.008688	X(38)=	0.306181
X(7)=	-1.659390	X(39)=	0.795431
X(8)=	-1.111467	X(40)=	0.189598
X(9)=	0.985908	X(41)=	-0.342332
X(10)=	1.991979	X(42)=	-0.328700
X(11)=	-0.046613	X(43)=	0.197881
X(12)=	-1.649269	X(44)=	0.071179
X(13)=	-1.040810	X(45)=	0.185931
X(14)=	1.054665	X(46)=	-0.324595
X(15)=	1.055816	X(47)=	-0.366892
X(16)=	-0.951182	X(48)=	0.368467
X(17)=	-1.476495	X(49)=	-0.191935
X(18)=	-0.212242	X(50)=	0.519116
X(19)=	0.780282	X(51)=	0.008320
X(20)=	1.416083	X(52)=	-0.425946
X(21)=	0.195282	X(53)=	0.651470
X(22)=	-2.027026	X(54)=	-0.639978
X(23)=	-0.483577	X(55)=	-0.344389
X(24)=	1.664913	X(56)=	0.814130
X(25)=	0.614114	X(57)=	-0.385168
X(26)=	-0.791469	X(58)=	0.064218
X(27)=	-1.195311	X(59)=	-0.388888
X(28)=	0.119881	X(60)=	-0.163088
X(29)=	0.887635	X(61)=	1.188961
X(30)=	0.895236	X(62)=	0.114286
X(31)=	-0.012734	X(63)=	-0.667626
X(32)=	-1.763842	X(64)=	-0.814997

TABLE IV
LIST OF PRONY METHOD PARAMETER ESTIMATES

Signal No.	Parameter Estimates							
1	AMP=	0.0924844	PHASE=	127.8651	DAMP=	0.0001263	FREQ=	0.100033
2	AMP=	1.2756225	PHASE=	150.5185	DAMP=	0.0006981	FREQ=	0.201151
3	AMP=	0.8447961	PHASE=	198.7896	DAMP=	0.0092935	FREQ=	0.298949
4	AMP=	0.2363858	PHASE=	288.9569	DAMP=	-0.0592385	FREQ=	0.275976
5	AMP=	0.5834477	PHASE=	51.5229	DAMP=	-0.0363819	FREQ=	0.318288
6	AMP=	0.2562685	PHASE=	181.3553	DAMP=	-0.0328243	FREQ=	0.357724
7	AMP=	0.2442874	PHASE=	183.9528	DAMP=	-0.0119484	FREQ=	0.402838
8	AMP=	0.1313714	PHASE=	51.7846	DAMP=	-0.0738351	FREQ=	0.451616

TABLE V
LIST OF PRONY SPECTRAL LINE METHOD PARAMETER ESTIMATES

Signal No.	Parameter Estimates							
1	AMP=	0.1033464	PHASE=	125.2777	DAMP=	0.0000000	FREQ=	0.100039
2	AMP=	1.0255654	PHASE=	156.4714	DAMP=	0.0000000	FREQ=	0.200267
3	AMP=	1.0396414	PHASE=	178.5962	DAMP=	0.0000000	FREQ=	0.299599
4	AMP=	0.0564418	PHASE=	36.3485	DAMP=	0.0000000	FREQ=	0.270887
5	AMP=	0.2186561	PHASE=	65.7856	DAMP=	0.0000000	FREQ=	0.309460
6	AMP=	0.1186532	PHASE=	174.4837	DAMP=	0.0000000	FREQ=	0.350735
7	AMP=	0.1971185	PHASE=	95.5934	DAMP=	0.0000000	FREQ=	0.402381
8	AMP=	0.0278794	PHASE=	32.9251	DAMP=	0.0000000	FREQ=	0.450818

and also improved resolution are the most important aspects of spectral estimation, rather than spectral shape, then some recent research by Tufts and Kumaresan [244], [245], [134], [135] has addressed this problem. They consider improvements in linear prediction, eigen-analysis, and maximum likelihood approaches to reduce the frequency estimation variance and increase the resolution. However, these deal strictly with the sinusoids in noise process.

IV. OTHER APPLICATIONS OF SPECTRAL ESTIMATION METHODS

A. Introduction

The preceding sections have discussed the theory and application of modern spectral estimation. Much of the underlying theory presented, however, has been applied to areas other than spectral estimation. Since these further applications are of sufficient interest to researchers in many fields, this section summarizes some of these applications. The topics to be discussed are not meant to be an all inclusive listing of these additional applications, but only a representative sampling of the more common areas.

B. Time Series Extrapolation and Interpolation

The theoretical foundations of modern spectral estimation have led to other applications. An obvious one is that of extrapolation of a time series of unknown PSD. If the time series is an AR(p) process, for example, then the optimum linear predictor parameters are the AR parameters. The latter are estimated from the data as discussed in Section II-E. If the

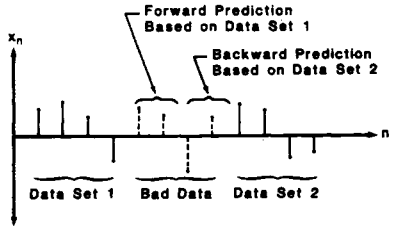

Fig. 17. Interpolation of bad data points.

process is not an AR process, but an AR model is used, then the number of linear prediction parameters of the optimal predictor is, in general, infinite. Theoretically, as the number of predictor parameters increases, the extrapolation error will decrease. When one is limited to a finite data set from which to estimate the predictor parameters, the prediction error power will be minimized by choosing a predictor with as large an order as possible, subject to the constraint that the predictor parameters can be accurately estimated [51].

Although the techniques described in Section II-E can only yield a one step predictor, one can use the predicted sample as if it were part of the original data set and continue the extrapolation to the next sample [27], [29]. It has even been proposed to use the enlarged set of original and extrapolated data with a conventional periodogram or a BT spectral estimator to improve the resolution [68]. In addition to extrapolation, interpolation may be performed by using a forward and backward predictor as shown in Fig. 17. This is valuable for replacing bad data points [176].

C. Prewhitening Filters

A prewhitening filter is a natural use of the parameters obtained from a spectral estimate. For example, in AR spectral analysis $\hat{A}(z)$ of (2.91) is a whitening filter. The output of the filter (the prediction error) is white noise if the observed process is AR(p) and the predictor coefficients estimated are the actual AR parameters. In the event that the time series is not an AR process, the output time series power spectral density will still be flatter than the input and "approximately" white. This property is particularly valuable in the design of detectors for signals in colored noise of unknown spectral shape. The detection of target returns in a background of clutter is an example. The optimal detector is a prewhitener followed by a matched filter, matched to the signal at the prewhitener output [256]. Since the clutter spectrum usually is time varying, the whitening filter parameters and matched filter must be updated. The success of the prewhitening scheme will depend upon the time variation of the clutter spectrum and the ability to estimate the parameters of the spectrum before they change [30].

A closely related concept is that of prewhitening a time series to reduce the bias of conventional spectral estimators. It may be shown that [107]

$$E[\hat{\mathcal{P}}_x(f)] = \int_{-1/2\Delta t}^{1/2\Delta t} \mathcal{P}_x(\nu)\, W(f-\nu)\, d\nu \qquad (4.1)$$

where

$\hat{\mathcal{P}}_x(f)$ is a BT type spectral estimate,
$\mathcal{P}_x(f)$ is the true PSD,
$W(f)$ is a spectral window required to reduce the variance of the estimate [$\int_{-1/2\Delta t}^{1/2\Delta t} W(f)\, df = 1$].

If $\mathcal{P}_x(f)$ is nearly constant with frequency, then

$$E[\hat{\mathcal{P}}_x(f)] \approx \mathcal{P}_x(f) \int_{-1/2\Delta t}^{1/2\Delta t} W(f-\nu)\, d\nu = \mathcal{P}_x(f). \qquad (4.2)$$

Thus, to reduce the bias, one should attempt to prewhiten the data [213], [232]. Following the whitening, a BT estimate $\mathcal{P}_e(f)$ of the filtered time series e_n is found. Since e_n is the output of the prewhitening filter $\hat{A}(z)$, its PSD is $\mathcal{P}_e(f) = |\hat{A}(\exp[j2\pi f\Delta t])|^2\, \mathcal{P}_x(f)$. The spectral estimate of x_n is then given as

$$\hat{\mathcal{P}}_x(f) = \frac{\hat{\mathcal{P}}_e(f)}{|\hat{A}(\exp[j2\pi f\Delta t])|^2} \qquad (4.3)$$

where an all-zero prewhitener is assumed. It is interesting to note that this approach yields a spectral estimate that is the standard AR spectral estimate, with the white noise PSD $\sigma_n^2\Delta t$ replaced by the PSD estimate of the residual time series, $\hat{\mathcal{P}}_e(f)$.

D. Bandwidth Compression

An important problem in speech research is that of bandwidth compression. If the redundancy of speech can be reduced, then more speech signals can be transmitted through a fixed bandwidth channel or stored in some mass storage. One common technique is differential pulse code modulation (DPCM) [104]. The basis of DPCM is to transmit only information that cannot be predicted, often termed the innovations of the process [112]. In fact, if the speech waveform were perfectly predictable from a set of previous samples, then the receiver, once it had those samples, could perfectly reconstruct the entire waveform (assuming no channel noise). Transmission could be halted! In practice using DPCM, speech samples are analyzed at the transmitter to determine the predictor parameters. Then, only the prediction error time series and the predictor parameters are transmitted. The speech signal is reconstructed at the receiver. The bandwidth reduction is possible because the variance of the prediction error time series is less than that of the speech waveform [251], i.e.,

$$\sigma_e^2 = R_{xx}(0) \prod_{i=1}^{p} (1 - K_i^2) \leqslant R_{xx}(0). \qquad (4.4)$$

Thus fewer quantizer levels are necessary to code the residual time series. Note that for maximum bandwidth compression, the predictor parameters must be continually updated as the statistical character of speech changes, i.e., voiced to unvoiced and vice versa.

The most dramatic technique for bandwidth reduction is linear predictor coding (LPC), in which AR modeling is used to represent the speech waveform [145]. Assuming speech can be accurately modeled as the output of an all-pole filter driven by white noise for unvoiced speech, or driven by an impulse train for voiced speech, the speech waveform may be reduced to a small set of parameters. Thus, only the model parameters and the period of the impulse train need be transmitted or stored. Speech synthesis is then accomplished by employing the appropriate model for each speech sound.

E. Spectral Smoothing

Conventional periodogram and BT analysis lead to spectral estimates that are characterized by many "hills and valleys," since the Fourier transform of a zero mean random process

is itself a zero mean random process. Autocorrelation lag windowing or spectral window smoothing will substantially reduce the fluctuations but not eliminate them. An AR spectral estimator can be used to smooth these fluctuations since a pth-order AR spectral estimate is constrained to have p or less peaks (or troughs). For p small, a smoothed spectral estimate will result. It is now shown that the AR spectral model accurately represents the peaks of a periodogram but not the valleys [26], [76], [145]. Consider the estimate of the AR parameters found by minimizing the error criterion ((2.64) with the use of $\boldsymbol{X}_2^H\boldsymbol{X}_2$, the autocorrelation normal equations)

$$\mathcal{E} = \frac{1}{N}\sum_{n=-\infty}^{\infty} |e_n|^2 \tag{4.5}$$

where it is assumed x_n, $n = 0, 1, \cdots, N-1$, is available and $x_n = 0$ outside this interval. Then, by Parseval's theorem

$$\mathcal{E} = \frac{1}{N\Delta t}\int_{-1/2\Delta t}^{1/2\Delta t} |E(\exp[j2\pi f\Delta t])|^2\, df$$

where

$$E(z) = \Delta t \sum_{n=-\infty}^{\infty} e_n z^{-n}.$$

Since

$$X(z) = \Delta t \sum_{n=-\infty}^{\infty} x_n z^{-n}$$

then E becomes

$$\begin{aligned} E &= \frac{1}{N\Delta t}\int_{-1/2\Delta t}^{1/2\Delta t} |X(\exp[j2\pi f\Delta t])|^2 \\ &\quad \cdot |\hat{A}(\exp[j2\pi f\Delta t])|^2\, df \\ &= \hat{\sigma}^2\Delta t \int_{-1/2\Delta t}^{1/2\Delta t} \frac{1}{N\Delta t}|X(\exp[j2\pi f\Delta t])|^2 \Big/ \hat{\mathcal{P}}_x(f)\, df \end{aligned} \tag{4.6}$$

where $|X(\exp[j2\pi f\Delta t])|^2/N\Delta t$ is the periodogram and $\hat{\mathcal{P}}_x(f) = \hat{\sigma}^2\Delta t/|\hat{A}(\exp[j2\pi f\Delta t])|^2$. Thus when

$$\frac{1}{N\Delta t}|X(\exp[j2\pi f\Delta t])|^2$$

is large, $\hat{\mathcal{P}}_x(f)$ should match the periodogram to reduce $\mathcal{E}$. For $|X(\exp[j2\pi f\Delta t])|^2/N\Delta t$ small, there is only a small contribution to the error, so that matching is not necessary. The result is that the AR spectral estimate matches the peaks, but not the valleys, of the periodogram. If one wishes to represent the peaks of a spectrum, then one need only take an inverse Fourier transform to find the first $p+1$ autocorrelation lags, which are then used to find the AR(p) model. An example is shown in Fig. 18 for the periodogram of a speech signal.

F. Beamforming

In beamforming, one is interested in obtaining an estimate of the spatial structure of a random spatial field. If one samples in space a random field using a line (linear) array of

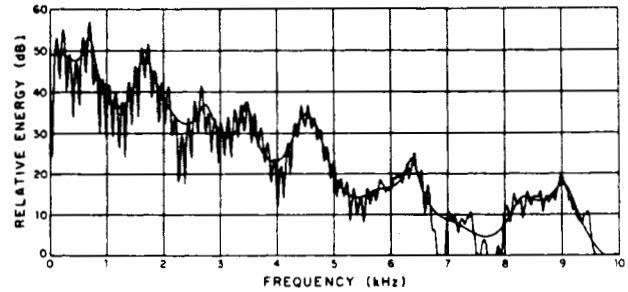

Fig. 18. Periodogram PSD and smoothed AR(28) PSD estimates of speech data (from Makhoul [144]).

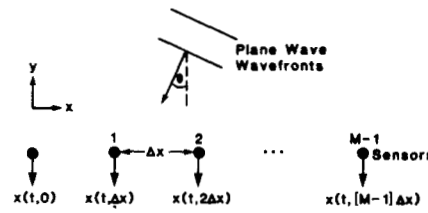

Fig. 19. Line array geometry.

sensors, then a vector time series, $\{x(t, 0), x(t, \Delta x), \cdots, x(t, [M-1]\,\Delta x)\}$, is obtained, where $x(t, i\Delta x)$ is the continuous waveform at the ith sensor, $0 \leq i \leq M-1$, and M is the number of sensors as shown in Fig. 19. The field can be expressed as [42]

$$x(t, i\Delta x) = \iint \Psi(f, k_x)\exp[j2\pi(ft - k_x i\Delta x)]\, df\, dk_x \tag{4.7}$$

which represents the field as the sum of an infinite number of monochromatic plane waves with random amplitudes $\Psi(f, k_x)$. The temporal frequency is denoted by f, while the spatial frequency along the x direction is denoted by k_x. The wavenumber component k_x is the reciprocal of the wavelength of a monochromatic plane wave along the x direction. Since $k_x = (f/c)\sin\theta$, where θ is the angle indicated in Fig. 19, then $E[|\Psi(f, k_x)|^2]$ is the power at frequency f arriving from the θ direction. From (4.7) the inverse Fourier transform is

$$\begin{aligned} \Psi(f, k_x) &= \sum_{i=0}^{M-1}\left(\int x(t, i\Delta x)\exp(-j2\pi ft)\, dt\right)\exp(j2\pi k_x i\Delta x) \\ &= \sum_{i=0}^{M-1} X_f(i\Delta x)\exp(j2\pi k_x i\Delta x). \end{aligned} \tag{4.8}$$

Expression (4.8) is a Fourier transform relationship between a spatial "time series" $X_f(i\Delta x)$, where Δx is the distance between the samples, and its spectrum $\Psi(f, k_x)$. If the spatial field is assumed homogeneous, i.e.,

$$E[x(t, i\Delta x)\, x^*(t, j\Delta x)] = f(t, [i-j]\,\Delta x) \tag{4.9}$$

then the spatial "time series" is wide sense stationary and the estimation of $E(|\Psi(f, k_x)|^2)$ for all k_x at a given temporal frequency f is analogous to the one-dimensional temporal power spectral estimation. Any of the techniques described in this paper are then applicable if the time data record is replaced by the spatial data record $\{x(t_0, 0), x(t_0, \Delta x), \cdots, x(t_0, [M-1]\,\Delta x)\}$ at some time t_0. Note that some extra averaging is afforded in the spatial case that is not available in the temporal case. For instance, the spatial "autocorrelation"

estimate could be chosen to be

$$\hat{R}_x(k\Delta x) = \frac{1}{N(M-k)} \sum_{n=0}^{N-1} \sum_{i=0}^{M-k-1} x^*(n\Delta t, i\Delta x) \cdot x(n\Delta t, (i+k)\Delta x) \qquad k \geqslant 0 \qquad (4.10)$$

where $x(n\Delta t, i\Delta x)$ is assumed to be temporally stationary over the interval $0 \leqslant n\Delta t \leqslant (N-1)\Delta t$. This estimate includes an extra time averaging operation.

G. Lattice Filters

The minimum phase lattice filter described in Section II-E has the property that its coefficients are bounded by one in magnitude. This is very desirable when one must quantize the coefficients for transmission or storage [257]. The lattice structure may be used to synthesize stable minimum-phase FIR filters, stable all-pole IIR filters, or stable pole-zero IIR filters that have zeros within the unit circle of the z-plane.

H. Other Applications

Other applications of the techniques described in this paper are

1) equalization for digital communications [147]
2) transient analysis [252], [254]
3) digital filter deisgn [183]
4) predictive deconvolution [209]
5) cepstral analysis [139].

The interested reader may consult the references for further details.

V. Conclusions

Modern spectral estimation techniques are based upon modeling of the data by a small set of parameters. When the model is an accurate representation of the data, spectral estimates can be obtained whose performance exceed that of the classical periodogram or BT spectral estimators. The improvement in performance is manifested by higher resolution and a lack of sidelobes. It should also be emphasized that in addition to an accurate model of the data, one must base the spectral estimator on a good estimator of the model parameters. Usually, this entails a maximum likelihood parameter estimator. If the model is inappropriate, as in the case of an AR model for an AR process with additive observation noise, poor (biased) spectral estimates will result. If the model is accurate but a poor statistical estimator of the parameters is employed, as in the case of the ARMA spectral estimate using the modified Yule-Walker equations, poor (inflated variance) spectral estimates will also result.

Computationally efficient procedures exist for maximum-likelihood AR spectral estimation. These techniques generally do not require substantially more computation than conventional Fourier spectral estimators. However, maximum likelihood ARMA spectral estimation involves the solution of nonlinear equations so that no efficient computational procedures now exist. Since ARMA spectral estimators are more desirable than AR spectral estimators when the data characteristics are unknown, due to their robustness, future research is being directed at computationally efficient maximum likelihood ARMA spectral estimation.

A multitude of modern spectral estimation algorithms have been proposed, with only a small but representative subset described in this tutorial. Unfortunately few algorithms, if any, have been analyzed statistically for finite data records. Comparisons among various competing algorithms have been based on limited computer simulations, which can be misleading. Therefore, future research should also be directed at providing more complete statistical descriptions of modern spectral estimators.

In summary, modern spectral estimation techniques, when used properly, are extremely valuable for data analysis. It has been the intent of the authors to present the various techniques in a unified modeling framework and with common nomenclature. Hopefully, this approach will aid users in the selection of the spectral estimation method appropriate to their application.

References

References marked with * may be found in *Modern Spectrum Analysis*, D. G. Childers, Ed., IEEE Press selected Reprint Series, New York, 1978.

[1] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statist. Math.*, vol. 21, pp. 243–247, 1969.
[2] —, "Power spectrum estimation through autoregression model fitting," *Ann. Inst. Statist. Math.*, vol. 21, pp. 407–419, 1969.
[3] —, "On a semi-automatic power spectrum estimation procedure," in *Proc. 3rd Hawaii Int. Conf. System Sciences*, Part 2, pp. 974–977, 1970.
[4] —, "Statistical predictor identification," *Ann. Inst. Statist. Math.*, vol. 22, pp. 203–217, 1970.
[5] —, "Autoregressive model fitting for control," *Ann. Inst. Statist. Math.*, vol. 23, pp. 163–180, 1971.
[6] —, "Use of an information theoretic quantity for statistical model identification," in *Proc. 5th Hawaii Int. Conf. System Sciences*, pp. 249–250, Jan. 11–13, 1972.
*[7] —, "A new look at the statistical model identification," *IEEE Trans. Autom. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.
*[8] M. A. Alam, "Adaptive spectral estimation," in *Proc. Joint Automat. Contr. Conf.*, vol. 1, San Francisco, CA, June 22–24, 1977, pp. 105–112.
*[9] N. O. Andersen, "On the calculation of filter coefficients for maximum entropy spectral analysis," *Geophys.*, vol. 39, pp. 69–72, Feb. 1974.
[10] —, "Comments on the performance of maximum entropy algorithms," *Proc. IEEE*, vol. 66, pp. 1581–1582, Nov. 1978.
[11] T. W. Anderson, "Estimation by maximum likelihood in autoregressive moving average models in the time and frequency domains," Dep. Statistics, Tech. Rep. 20, Contract NR-042-034, Stanford University, Stanford, CA, June 1975.
[12] H. Babic and G. C. Temes, "Optimum low-order windows for discrete Fourier transform systems," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-24, pp. 512–517, Dec. 1976.
*[13] A. B. Baggeroer, "Confidence intervals for regression (MEM) spectral estimates," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 534–545, Sept. 1976.
[14] J. F. Banas, Comments on "Harmonic analysis via Kalman filtering technique," *Proc. IEEE*, vol. 61, pp. 1759–1760, Dec. 1973.
[15] T. E. Barnard, "The maximum entropy spectrum and the Burg technique," Tech. Rep. 1, Advanced Signal Processing, Texas Instruments, prepared for Office of Naval Research, ALEX (03)-TR-75-01, June 25, 1975.
[16] I. Barrodale and R. E. Erickson, "Algorithms for least-squares linear prediction and maximum entropy spectral analysis—Part I: Theory and Part II: FORTRAN Program," *Geophys.*, vol. 45, pp. 420–446, Mar. 1980.
[17] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, pp. 1–16, June 1950.
[18] M. S. Bartlett and J. Medhi, "On the efficiency of procedures for smoothing periodograms from time series with continuous spectra," *Biometrika*, vol. 42, pp. 143–150, 1955.
[19] G. D. Bergland, "A guided tour of the fast Fourier transform," *IEEE Spectrum*, vol. 6, pp. 41–52, July 1969.
[20] A. J. Berkhout and P. R. Zaanen, "A comparison between Weiner filtering, Kalman filtering, and deterministic least squares estimation," *Geophysical Prospecting*, vol. 24, pp. 141–197, Mar. 1976.
[21] J. G. Berryman, "Choice of operator length for maximum entropy spectral analysis," *Geophys.*, vol. 43, pp. 1384–1391, Dec. 1978.

[22] S. Bertram, "On the derivation of the fast Fourier transform," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 55–58, Mar. 1970.
[23] ——, "Frequency analysis using the discrete Fourier transform," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 495–500, Dec. 1970.
[24] C. Bingham, M. D. Godfrey, and J. W. Tukey, "Modern techniques of power spectrum estimation," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 56–66, June 1967.
[25] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra From the Point of View of Communications Engineering*. New York: Dover, 1959.
[26] S. F. Boll, "*A priori* digital speech analysis," Ph.D. dissertation, Dep. Elec. Eng., Utah Univ., Mar. 1973.
[27] S. B. Bowling, "Linear prediction and maximum entropy spectral analysis for radar applications," M.I.T. Lincoln Lab., Project Rep. RMP-122 (ESD-TR-77-113), May 24, 1977.
[28] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*. New York: Wiley, 1976.
[29] S. B. Bowling and S. Lai, "The use of linear prediction for the interpolation and extrapolation of missing data prior to spectral analysis," in *Proc. RADC Workshop on Spectrum Estimation*, pp. 39–50, Oct. 3–5, 1979.
[30] D. E. Bowyer, P. K. Rajasekaran, and W. W. Gebhart, "Adaptive clutter filtering using autoregressive spectral estimation," *IEEE Trans. Aerospace Elec. Syst.*, vol. AES-15, pp. 538–546, July 1979.
[31] G.E.P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1970.
[32] E. O. Brigham and R. E. Morrow, "The fast Fourier transform," *IEEE Spectrum*, vol. 4, pp. 63–70, Dec. 1967.
[33] E. Oran Brigham, *The Fast Fourier Transform*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[34] D. R. Brillinger, "Fourier analysis of stationary processes," *Proc. IEEE*, vol. 62, pp. 1628–1643, Dec. 1974.
[35] S. Bruzzone and M. Kaveh, "On some suboptimum ARMA spectral estimates," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-28, pp. 753–755, Dec. 1980.
[36] H. P. Bucker, "Comparison of FFT and Prony algorithms for bearing estimation of narrow-band signals in a realistic ocean environment," *J. Acoust. Soc. Amer.*, vol. 61, pp. 756–762, Mar. 1977.
[37] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meeting Society of Exploration Geophysicists* (Oklahoma City, OK), Oct. 31, 1967.
[38] ——, "A new analysis technique for time series data," NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands, Aug. 12–23, 1968.
[39] ——, "New concepts in power spectral estimation," in *Proc. 40th Annu. Int. Society of Exploration Geophysicists (SEG) Meeting* (New Orleans, LA) Nov. 11, 1970.
[40] ——, "The relationship between maximum entropy and maximum likelihood spectra," *Geophys.*, vol. 37, pp. 375–376, Apr. 1972.
[41] ——, "Maximum entropy spectral analysis," Ph.D. dissertation, Dep. Geophysics, Stanford Univ., Stanford, CA, May 1975.
[42] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," in *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
[43] G. C. Carter and A. H. Nuttall, "On the weighted overlapped segment averaging method for power spectral estimation," *Proc. IEEE*, vol. 68, pp. 1352–1354, Oct. 1980.
[44] ——, "A brief summary of a generalized framework for power spectral estimation," *Signal Processing*, vol. 2, pp. 387–390, Oct. 1980.
[45] W. Y. Chen and G. R. Stegen, "Experiments with maximun entropy power spectra of sinusoids," *J. Geophysical Res.*, vol. 79, pp. 3019–3022, July 10, 1974.
[46] D. Childers and A. Durling, *Digital Filtering and Signal Processing*. St. Paul, MN: West Publishing, 1975.
[47] J. C. Chow, "On the estimation of the order of a moving-average process," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 386–387, June 1972.
[48] ——, "On estimating the orders of an autoregressive moving-average process with uncertain observations," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 707–709, Oct. 1972.
[49] C. W. Chuang and D. L. Moffatt, "Natural resonances of radar targets via Prony's method and target discrimination," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-12, pp. 583–589, Sept. 1976.
[50] J. F. Claerbout, *Fundamentals of Geophysical Data Processing*. New York: McGraw-Hill: 1976.
[51] W. S. Cleveland, "Fitting time series models for prediction," *Technometrics*, pp. 713–723, Nov. 1971.
[52] W. T. Cochran *et al.*, "What is the fast Fourier transform?," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 45–55, June 1967.
[53] J. W. Cooley and J. W. Tukey, "An algorithm for machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, Apr. 1965.
[54] J. W. Cooley, P.A.W. Lewis, and P. D. Welch, "The fast Fourier transform algorithm: Programming considerations in the calculation of sine, cosine, and Laplace transforms," *J. Sound Vibration*, vol. 12, pp. 315–337, July 1970.
[55] ——, "The application of the fast Fourier transform algorithm to the estimation of spectra and cross-spectra," *J. Sound Vibration*, vol. 12, pp. 339–352, July 1970.
[56] G. Cybenko, "Round-off error propagation in Durbin's, Levinson's, and Trench's Algorithms," *Rec. 1979 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 498–501.
[57] A. K. Datta, Comments on "The complex form of the maximum entropy method for spectral estimation," *Proc. IEEE*, vol. 65, pp. 1219–1220, Aug. 1977.
[58] W. J. Done, "Estimation of the parameters of an autoregressive process in the presence of additive white noise," Computer Science Dep., Univ. Utah, Rep. UTEC-CSC-79-021, Dec. 1978 (order NTIS no. ADA 068749).
*[59] R. E. Dubroff, "The effective autocorrelation function of maximum entropy spectra," *Proc. IEEE*, vol. 63, pp. 1622–1623, Nov. 1975.
[60] J. Durbin, "The fitting of time series models," *Rev. Inst. Int. de Stat.*, vol. 28, pp. 233–244, 1960.
[61] A. Eberhard, "An optimal discrete window for the calculation of power spectra," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 37–43, Feb. 1973.
*[62] J. A. Edward and M. M. Fitelson, "Notes on maximum entropy processing," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 232–234, Mar. 1973.
[63] D. C. Farden, "Solution of a Toeplitz set of linear equations," *IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 906–908, Nov. 1976.
[64] P. F. Fougere, E. J. Zawalick, and H. R. Radoski, "Spontaneous line splitting in maximum entropy power spectrum analysis," *Physics Earth Planetary Interiors*, vol. 12, pp. 201–207, Aug. 1976.
[65] P. F. Fougere, "A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis," *J. Geophysical Res.*, vol. 82, pp. 1051–1054, Mar. 1, 1977.
[66] B. Friedlander, M. Morf, T. Kailath, and L. Ljung, "New inversion formulas for matrices classified in terms of their distance from Toeplitz matrices," *Linear Algebra and Its Applications*, vol. 27, pp. 31–60, 1979.
[67] O. L. Frost, "Power spectrum estimation," in *Proc. 1976 NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*, Portovener (LaSpezia), Italy, Aug. 30–Sept. 11, 1976.
[68] O. L. Frost and T. M. Sullivan, "High resolution two dimensional spectral analysis," in *Conf. Rec. 1979 ICASSP*, pp. 673–676, 1979.
[69] J. Fryer, M. E. Odegard, and G. H. Sutton, "Deconvolution and spectral estimation using final prediction error," *Geophys.*, vol. 40, pp. 411–425, June 1975.
[70] W. F. Gabriel, "Spectral analysis and adaptive array superresolution techniques," *Proc. IEEE*, vol. 68, pp. 654–666, June 1980.
[71] W. Gersch, "Spectral analysis of EEG's by autoregressive decomposition of time series," *Math. Biosci.*, vol. 7, pp. 205–222, 1970.
[72] ——, "Estimation of the autoregressive parameters of a mixed autoregressive moving-average time series," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 583–588, Oct. 1970.
[73] W. Gersch and D. R. Sharpe, "Estimation of power spectra with finite-order autoregressive models," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 367–369, Aug. 1973.
[74] W. Gersch and J. Yonemoto, "Automatic classification of EEG's: A parametric model new features for classification approach," in *Proc. 1977 Joint Automatic Control Conf.* (San Francisco, CA), June 22–24, 1977, pp. 762–769.
[75] J. Gibson, S. Haykin, and S. B. Kesler, "Maximum entropy (adaptive) filtering applied to radar clutter," in *Rec. 1979 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 166–169.
[76] J. D. Gibson *et al.*, "Digital speech analysis using sequential estimation techniques," *IEEE Trans. Acoustics, Speech, Signal Process.*, pp. 362–369, Aug. 1975.
[77] T. H. Glisson, C. I. Black, and A. P. Sage, "The digital computation of discrete spectra using the fast Fourier transform," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 271–287, Sept. 1970.
[78] J. Grandell, M. Hamrud, and P. Toll, "A remark on the correspondence between the maximum entropy method and the autoregressive models," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 750–751, Nov. 1980.
[79] D. Graupe, D. J. Krause, and J. B. Moore, "Identification of

autoregressive moving-average parameters of time series," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 104–107, Feb. 1975.
[80] A. H. Gray, Jr. and D. Y. Wong, "The Burg algorithm for LPC speech analysis/synthesis," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-28, pp. 609–615, Dec. 1980.
[81] R. M. Gray, "Toeplitz and circulant matrices: A review," Information Systems Laboratory, Center for Systems Research, Stanford University, Tech. Rep. No. 6502-1, June 1971.
[82] O. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley, CA: Univ. California Press, 1958.
[83] L. J. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-23, pp. 207–222, Apr. 1975.
[84] L. J. Griffiths and R. Prieto-Diaz, "Spectral analysis of natural seismic events using autoregressive techniques," *IEEE Trans. Geosci. Electron.*, vol. GE-15, pp. 13–25, Jan. 1977.
[85] L. J. Griffiths, "Adaptive structures for multiple-input noise canceling applications," in *Conf. Rec. ICASSP 79*, pp. 925–928.
[86] P. R. Gutowski, E. A. Robinson, and S. Treitel, "Novel aspects of spectral estimation," in *Proc. 1977 Joint Automatic Control Conf.*, vol. 1 (San Francisco, CA), June 22–24, 1977, pp. 99–104.
[87] ——, "Spectral estimation: Fact or fiction?," *IEEE Trans. Geosci. Electron.*, vol. GE-16, pp. 80–84, Apr. 1978.
[88] C. S. Hacker, "Autoregressive and transfer function models of mosquito populations," in *Time Series and Ecological Processes*, H. M. Shugat, Ed. Boulder, CO: SIAM, 1978, pp. 294–303.
[89] E. Hannan, *Multiple Time Series Analysis*. New York: Wiley, 1970.
[90] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, pp. 51–83, Jan. 1978.
[91] S. Haykin and S. Kesler, "The complex form of the maximum entropy method for spectral estimation," *Proc. IEEE*, vol. 64, pp. 822–823, May 1976.
[92] S. S. Haykin, Ed., *Nonlinear Methods of Spectral Analysis*. New York: Springer-Verlag, 1979.
[93] S. Haykin and J. Reilly, "Mixed autoregressive-moving average modeling of the response of a linear array antenna to incident plane waves," *Proc. IEEE*, vol. 68, pp. 622–623, May 1980.
[94] R. W. Herring, "The cause of line splitting in Burg maximum-entropy spectral analysis," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-28, pp. 692–701, Dec. 1980.
[95] F. B. Hildebrand, *Introduction to Numerical Analysis*. New York: McGraw-Hill, 1956, ch. 9.
[96] S. Holm and J. M. Hovem, "Estimation of scalar ocean wave spectra by the maximum entropy method," *IEEE J. Ocean. Eng.*, vol. OE-4, pp. 76–83, July 1979.
[97] H. Holtz, "Prony's method and related approaches to exponential approximation," Aerospace Corp., Rep. ATR-73(9990)-5, June 15, 1973.
[98] M. Hsu, "Maximum entropy principle and its application to spectral analysis and image reconstruction," Ph.D. dissertation, Ohio State Univ., 1975.
[99] F. M. Hsu and A. A. Giordano, "Line tracking using autoregressive spectral estimates," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-25, pp. 510–519, Dec. 1977.
[100] M. Huzii, "On a spectral estimate obtained by an autoregressive model fitting," *Ann. Inst. Statist. Math.*, vol. 29, pp. 415–431, 1977.
[101] L. B. Jackson and F. K. Soong, "Observations on linear estimation," in *Rec. 1978 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 203–207.
[102] L. B. Jackson, D. W. Tufts, F. K. Soong, and R. M. Rao, "Frequency estimation by linear prediction," in *Rec. 1978 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 352–356.
[103] A. K. Jain and S. Ranganath, "Extrapolation algorithms for discrete signals with a application in spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 830–845, Aug. 1981.
[104] N. S. Jayant, "Digital coding of speech waveforms: PCM, DPCM, and DM quantizers," *Proc. IEEE*, pp. 611–632, May 1974.
[105] G. M. Jenkins, "General considerations in the analysis of spectra," *Technometrics*, vol. 3, pp. 133–166, May 1961.
[106] ——, "A survey of spectral analysis," *Appl. Stat.*, vol. 14, pp. 2–32, 1965.
[107] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.
[108] S. J. Johnsen and N. Andersen, "On power estimation in maximum entropy spectral analysis," *Geophys.*, vol. 43, pp. 681–690, June 1978.
[109] R. H. Jones, "A reappraisal of the periodogram in spectral analysis," *Technometrics*, vol. 7, pp. 531–542, November 1965.
[110] ——, "Identification and autoregressive spectrum estimation," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 894–898, Dec. 1974.
[111] ——, "Autoregression order selection," *Geophys.*, vol. 41, pp. 771–773, Aug. 1976.
[112] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Informa. Theory*, vol. IT-20, pp. 146–181, Mar. 1974.
[113] R. P. Kane and N. B. Trivedi, "Effects of linear trend and mean value on maximum entropy spectral analysis," Institute de Pesquisas Espaciais, Rep. INPE-1568-RPE/069, Sao Jose dos Compos, Brasil (Order NTIS no. N79-33949).
[114] R. P. Kane, "Maximum entropy spectral analysis of some artificial samples," *J. Geophys. Res.*, vol. 84, pp. 965–966, Mar. 1, 1979.
[115] R. L. Kashyap, "Inconsistency of the AIC rule for estimating the order of autoregressive models," *IEEE Trans. Automa. Contr.*, vol. AC-25, pp. 996–998, Oct. 1980.
[116] M. Kaveh and G. R. Cooper, "An empirical investigation of the properties of the autoregressive spectral estimator," *IEEE Trans. Informa. Theory*, vol. IT-22, pp. 313–323, May 1976.
[117] M. Kaveh, "High resolution spectral estimation for noisy signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, pp. 286–287, June 1979.
[118] S. M. Kay, "Improvement of autoregressive spectral estimates in the presence of noise," in *Rec. 1978 Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 357–360.
[119] ——, "Maximum entropy spectral estimation using the analytical signal," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-26, pp. 467–469, Oct. 1978.
[120] S. M. Kay and S. L. Marple, Jr., "Sources of and remedies for spectral line splitting in autoregressive spectrum analysis," in *Rec. 1979 Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 151–154.
[121] S. M. Kay, "The effects of noise on the autoregressive spectral estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, pp. 478–485, Oct. 1979.
[122] ——, "Autoregressive spectral analysis of narrowband processes in white noise with application to sonar," Ph.D. dissertation, Georgia Institute of Technology, Mar. 1980.
[123] ——, "Noise compensation for autoregressive spectral estimates," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 292–303, June 1980.
[124] R. J. Keeler, "Uncertainties in adaptive maximum entropy frequency estimators," NOAA Tech. Rep. ERL-105-WPL53, Feb. 1979 (Order NTIS no. N79-32489).
[125] S. B. Kesler and S. S. Haykin, "The maximum entropy method applied to the spectral analysis of radar clutter," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 269–272, Mar. 1978.
[126] ——, "Maximum entropy estimation of radar clutter spectra," in *Natl. Telecommunications Conf. Rec.* (Birmingham, AL), pp. 18.5.1–18.5.5, Dec. 3–6, 1978.
[127] A. Ya. Khinchin, "Korrelationstheorie der Stationären Stochastischen Prozesse," *Math. Annalen*, vol. 109, pp. 604–615, 1934.
[128] W. R. King, "Maximum entropy spectral analysis in the spatial domain," Rome Air Development Center, Tech. Rep. RADC-TR-78-160, July 1978 (Order NTIS no. ADA 068558).
[129] J. F. Kinkel *et al.*, "A note on covariance-invariant digital filter design and autoregressive moving average spectrum analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 200–202, Apr. 1979.
[130] H. Kobatake *et al.*, "Linear predictive coding of speech signals in a high ambient noise environment," in *Conf. Rec. 1978 ICASSP*, pp. 472–475.
[131] I. S. Konvalenka *et al.*, "Simultaneous estimation of poles and zeros in speech analysis and ITIF—Iterative inverse fitting algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 485–491, Oct. 1979.
[132] L. H. Koopmans, *The Spectral Analysis of Time Series*. New York: Academic Press, 1974.
[133] F. Kozin and F. Nakajima, "The order determination problem for linear time-varying AR models," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 250–257, Apr. 1980.
[134] R. Kumaresan and D. W. Tufts, "Improved spectral resolution III: Efficient realization," *Proc. IEEE*, vol. 68, pp. 1354–1355, Oct. 1980.
[135] ——, "Data-adaptive principal component signal processing," in *Proc. 19th IEEE Conf. Decision and Control* (Albuquerque, NM), Dec. 10–12, 1980, pp. 949–954.
[136] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36. pp. 661–675, Aug. 1971.
[137] ——, "Autoregressive and maximum likelihood spectral analysis methods," in *Proc. 1976 NATO Conf. Signal Processing.*
[138] T. E. Landers and R. T. Lacoss, "Some geophysical applications of autoregressive spectral estimates," *IEEE Trans. Geosci. Electron.*, vol. GE-15, pp. 26–32, Jan. 1977.
[139] T. Landers, "Maximum entropy cepstral analysis," in *Proc. 1978 RADC Spectral Estimation Workshop*, pp. 245–258.
[140] S. W. Lang and J. H. McClellan, "A simple proof of stability for all-pole linear prediction models," *Proc. IEEE*, vol. 67, pp. 860–861, May 1979.
[141] ——, "Frequency estimation with maximum entropy spectral

estimators," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 716–724, Dec. 1980.
[142] N. Levinson, "The Wiener (root mean square) error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.
[143] J. S. Lim, "All pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 197–209, June 1978.
[144] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.
[145] ——, "Spectral linear prediction: Properties and applications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 283–296, June 1975.
[146] ——, "Stable and efficient lattice methods for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 423–428, Oct. 1977.
[147] J. Makhoul and R. Viswanathan, "Adaptive lattice methods for linear prediction," in *Conf. Rec. 1978 ICASSP*, pp. 83–86.
[148] J. D. Markel, "FFT pruning," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 305–311, Dec. 1971.
[149] S. L. Marple, Jr., "Conventional Fourier, autoregressive, and special ARMA methods of spectrum analysis," Engineer's Dissertation, Stanford Univ., Stanford, CA, Dep. Elec. Eng., Dec. 1976.
[150] ——, "Resolution of conventional Fourier, autoregressive and special ARMA methods of spectral analysis," in *Rec. 1977 Int. Conf. Acoustics, Speech and Signal Processing* (Hartford, CT), May 9–11, 1977, pp. 74–77.
[151] ——, "High resolution autoregressive spectrum analysis using noise power cancellation," in *Rec. 1978 IEEE Int. Conf. Acoustics, Speech and Signal Processing.*, pp. 345–348.
[152] ——, "Frequency resolution of high-resolution spectrum analysis techniques," in *Proc. 1978 RADC Spectrum Estimation Workshop*, pp. 19–35.
[153] ——, "Spectral line analysis by Pisarenko and Prony methods," in *Rec. 1979 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 159–161.
[154] ——, "A new autoregressive spectrum analysis algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 441–454, Aug. 1980.
[155] D. Q. Mayne *et al.*, "Linear estimation of ARMA systems," Dep. Computing and Control, Imperial College of Science and Technology, London, England, Research Rep. 77113.
[156] R. N. McDonough, "Representations and analysis of signals. Part XV. Matched exponents for the representation of signals," Ph.D. dissertation, Dep. Elec. Eng., Johns Hopkins Univ., Baltimore, MD, Apr. 30, 1963.
[157] R. N. McDonough and W. H. Huggins, "Best least-squares representation of signals by exponentials," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 408–412, Aug. 1968.
[158] R. N. McDonough, "Maximum entropy spatial processing of array data," *Geophysics*, vol. 39, pp. 843–851, Dec. 1974.
[159] J. M. Melsa and J. D. Tomcik, "Linear predictive coding with additive noise for application to speech digitization," in *Proc. 14th Allerton Conf. Circuits and Systems Theory*, pp. 500–508, Sept. 27, 1976.
[160] R. Mittra and L. W. Pearson, "A variational method for efficient determination of SEM poles," *IEEE Trans. Antennas Propagat.*, vol. AP-26, pp. 354–358, Mar. 1978.
[161] A. Mohammed and R. G. Smith, "Data windowing in spectral analysis," Defense Research Establishment Atlantic, Dartsmouth, Nova Scotia, Rep. 7512, June 1975.
[162] D. R. Moorcroft, "Maximum entropy spectral analysis of radio-auroral signals," Department of Physics and Centre for Radio Science, Univ. Western Ontario, London, Ontario, Canada.
[163] M. Morf, "Fast algorithms for multivariable systems," Ph.D. dissertation, Stanford University, Stanford, CA, Dep. Elec. Eng., Aug. 1974.
[164] M. Morf, T. Kailath, and L. Ljung, "Fast algorithms for recursive identification," in *Proc. 1976 IEEE Conf. Decision and Control* (Clearwater, FL), Dec. 1–3, 1976, pp. 916–921.
[165] M. Morf, D. T. Lee, J. R. Nickolls, and A. Vieira, "A classification of algorithms for ARMA models and ladder realizations," in *Rec. 1977 IEEE Int. Conf. Acoustics, Speech, and Signal Processing.* (Hartford, CT), May 9–11, 1977, pp. 13–19.
[166] M. Morf, A. Vieira, D. T. Lee, and T. Kailath, "Recursive multichannel maximum entropy method," in *Proc. 1977 Joint Automatic Control Conf.* (San Francisco, CA), pp. 113–117, June 22–24, 1977.
[167] M. Morf, B. Dickinson, T. Kailath, and A. Vieira, "Efficient solution of covariance equations for linear prediction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, pp. 429–433, Oct. 1977.
[168] M. Morf, A. Vieira, and D. T. Lee, "Ladder forms for identification and speech processing," in *Proc. 1977 Conf. Decision and Control* (New Orleans, LA), pp. 1074–1078, Dec. 1977.
[169] M. Morf, A. Vieira, D. T. Lee, and T. Kailath, "Recursive multichannel maximum entropy spectral estimation," *IEEE Trans. Geosci. Electron.*, vol. GE-16, pp. 85–94, Apr. 1978.
[170] P. O. Neudorfer, "Alternate methods of harmonic analysis," *Proc. IEEE*, vol. 61, pp. 1661–1662, Nov. 1973.
*[171] W. I. Newman, "Extension to the maximum entropy method," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 89–93, Jan. 1977.
[172] ——, "Extension to the maximum entropy method II," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 705–708, Nov. 1979.
[173] R. Nitzberg, "Spectral estimation: An impossibility?" *Proc. IEEE*, vol. 67, pp. 437–439, Mar. 1979.
[174] B. Noble and J. W. Daniel, *Applied Linear Algebra*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1977.
[175] A. H. Nuttall, "Spectral estimation by means of overlapped fast Fourier transform processing of windowed data," NUSC Tech. Rep. 4169, New London, CT, Oct. 13, 1971.
[176] ——, "Spectral analysis of a univariate process with bad data points, via maximum entropy, and linear predictive techniques," Naval Underwater Systems Center, Tech. Rep. 5303, New London, CT, Mar. 26, 1976.
[177] ——, "FORTRAN program for multivariate linear predictive spectral analysis, employing forward and backward averaging," Naval Underwater Systems Center, Tech. Document 5419, New London, CT, May 19, 1976.
[178] ——, "Multivariate linear predictive spectral analysis employing weighted forward and backward averaging: A generalization of Burg's algorithm," Naval Underwater Systems Center, Tech. Rep. 5501, New London, CT, Oct. 13, 1976.
[179] ——, "Probability distribution of spectral estimates obtained via overlapped FFT processing of windowed data," Naval Underwater Systems Center, Tech. Rep. 5529, New London, CT, Dec. 3, 1976.
[180] ——, "Positive definite spectral estimate and stable correlation recursion for multivariate linear predictive spectral analysis," NUSC Tech. Rep. 5729, Naval Underwater Systems Center, New London, CT, Nov. 14, 1977.
[181] A. H. Nuttall and G. C. Carter, "A generalized framework for power spectral estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 334–335, June 1980.
[182] A. H. Nuttall, "Some windows with very good sidelobe behavior," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, pp. 84–89, Feb. 1981.
[183] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.
[184] R. K. Otnes and L. Enochson, *Digital Time Series Analysis.* New York: Wiley, 1972.
[185] N. L. Owsley, "Adaptive data orthogonalization," in *Proc. 1978 IEEE ICASSP* (Tulsa, OK), pp. 109–112, April 1978.
[186] M. Pagano, "Estimation of models of autoregressive signal plus white noise," *Ann. Statistics*, vol. 2, pp. 99–108, 1974.
[187] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 1965.
[188] E. Parzen, "Mathematical considerations in the estimation of spectra," *Technometrics*, vol. 3, pp. 167–190, May 1961.
[189] ——, "Statistical spectral analysis (single channel case) in 1968," Dep. Statistics, Stanford Univ., Stanford, CA, Tech. Rep. 11, June 10, 1968.
[190] ——, "Multiple time series modelling," Dep. Statistics, Stanford, Univ., Stanford, CA, Tech. Rep. 12, July 8, 1968 (Contract Nonr-225-80).
[191] ——, "Multiple time series modelling," *Multivariate Analysis II*, P. R. Krishnaiah, Ed. New York: Academic Press, pp. 389–409, 1969.
*[192] ——, "Some recent advances in time series modeling," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 723–730, Dec. 1974.
[193] J. V. Pendrel and D. E. Smylie, "The relationship between maximum entropy and maximum likelihood spectra," *Geophysics*, vol. 44, pp. 1738–1739, Oct. 1979.
[194] V. F. Pisarenko, "On the estimation of spectra by means of nonlinear functions of the covariance matrix," *Geophysical J. Royal Astronomical Soc.*, vol. 28, pp. 511–531, 1972.
[195] ——, "The retrieval of harmonics from a covariance function," *Geophysical J. Royal Astronomical Soc.*, vol. 33, pp. 347–366, 1973.
[196] A. J. Poggio, M. L. Van Blaricum, E. K. Miller, and R. Mittra, "Evaluation for a processing technique for transient data," *IEEE Trans. Antennas Propagat.*, vol. AP-26, pp. 165–173, Jan. 1978.
[197] J. Ponnusarny *et al.*, "Identification of complex autoregressive processes," in *Rec. 1979 ICASSP*, pp. 384–387.
[198] G. Prado and P. Moroney, "Linear predictive spectral analysis for Doppler sonar applications," Charles Stark Draper Laboratory, Tech. Rep. R-1109, Sept. 1977.
[199] ——, "The application of linear predictive methods of spectral analysis to frequency estimation," in *Rec. 1978 NAECON.*
[200] ——, "The accuracy of center frequency estimators using linear predictive methods," in *Rec. 1978 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 361–364.
[201] J. F. Prewitt, "Amplitude bias in the Fourier transforms of noisy

signals," *IEEE Trans. Antennas Propagat.*, vol. AP-26, pp. 730-731, Sept. 1978.
[202] G.R.B. Prony, "Essai experimental et analytique, etc.," *Paris, J. de L'Ecole Polytechnique*, vol. 1, cahier 2, pp. 24-76, 1795.
[203] L. C. Pusey, "High resolution spectral estimates," Lincoln Laboratory, M.I.T., Tech. Note 1975-7, Jan. 21, 1975.
[204] C. M. Rader, "An improved algorithm for high speed autocorrelation with applications to spectral estimation," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 439-442, Dec. 1970.
[205] H. R. Radoski, P. F. Fougere, and E. J. Zawalick, "A comparison of power spectral estimates and applications of the maximum entropy method," *J. Geophysical Res.*, vol. 80, pp. 619-625, Feb. 1, 1975.
[206] H. R. Radoski, E. J. Zawalick, and P. F. Fougere, "The superiority of maximum entropy power spectrum techniques applied to geomagnetic micropulsations," *Phys. Earth Planetary Interiors*, vol. 12, pp. 208-216, Aug. 1976.
[207] P. I. Richards, "Computing reliable power spectra," *IEEE Spectrum*, vol. 4, pp. 83-90, Jan. 1967.
[208] D. C. Rife and G. A. Vincent, "Use of the discrete Fourier transform in the measurement of frequencies and levels of tones," *Bell Syst. Tech. J.*, vol. 49, pp. 197-228, Feb. 1970.
[209] E. A. Robinson, "Predictive decomposition of time series with application to seismic exploration," *Geophysics*, vol. 32, pp. 418-484, June 1967.
[210] —, *Multichannel Time Series Analysis with Digital Computer Programs.* San Francisco, CA: Holden-Day, 1967.
[211] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communication and Control.* New York: McGraw-Hill, 1971, ch. 6.
[212] H. Sakai, "Statistical properties of AR spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 402-409, Aug. 1979.
[213] M. R. Sambur, "A preprocessing filter for enhancing LPC analysis/synthesis of noisy speech," in *Rec. 1979 ICASSP*, pp. 971-974.
[214] E. H. Satorius and J. T. Alexander, "High resolution spectral analysis of sinusoids in correlated noise," in *Rec. 1978 ICASSP* (Tulsa, OK), pp. 349-351, Apr. 10-12, 1978.
[215] E. H. Satorius and J. R. Zeidler, "Maximum entropy spectral analysis of multiple sinusoids in noise," *Geophysics*, vol. 43, pp. 1111-1118, Oct. 1978.
[216] J. H. Sawyers, "Applying the maximum entropy method to adaptive digital filtering," in *Rec. 12th Asilomar Conf. Circuits, Systems and Computers*, pp. 198-202.
[217] D. H. Scahubert, "Application of Prony's method to time-domain reflectometer data and equivalent circuit synthesis," *IEEE Trans. Antennas Propagat.*, vol. AP-27, pp. 180-184, Mar. 1979.
[218] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terrestrial Magnetism*, vol. 3, pp. 13-41, Mar. 1898.
[219] —, "The periodogram of magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871-1895," *Trans. Cambridge Philosophical Soc.*, vol. 18, pp. 107-135, 1899.
[220] K.L.S. Sharma and A. K. Mahalanabis, "Harmonic analysis via Kalman filtering technique," *Proc. IEEE*, pp. 391-392, Mar. 1973.
[221] D. E. Smylie, G.K.C. Clarke, and T. J. Ulrych, "Analysis of irregularities in the earth's rotation," in *Methods in Computational Physics*, vol. 13. B. A. Bolt, Ed. New York: Academic Press, 1973, pp. 391-430.
[222] M. D. Srinath and M. M. Viswanathan, "Sequential algorithm for identification of parameters of an autoregressive process," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 542-546, Aug. 1975.
[223] G. G. Stokes, "Note on the search for periodic inequalities," *Proc. Royal Soc. London*, vol. 29, pp. 122-123, May 29, 1879.
[224] N. R. Strader, II, "Effects of subharmonic frequencies on DFT coefficients," *Proc. IEEE*, vol. 68, pp. 285-286, Feb. 1980.
[225] O. N. Strand, "Multichannel maximum entropy spectral analysis," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 634-640, Aug. 1977.
[226] T. M. Sullivan, O. L. Frost, and J. R. Treichler, "High resolution signal estimation," ARGO Systems, Internal Rep. June 16, 1978.
[227] D. N. Swingler, "A comparison between Burg's maximum entropy method and a nonrecursive technique for the spectral analysis of deterministic signals," *J. Geophysical Res.*, vol. 84, pp. 679-685, Feb. 10, 1979.
[228] —, "A modified Burg algorithm for maximum entropy spectral analysis," *Proc. IEEE*, vol. 67, pp. 1368-1369, Sept. 1979.
[229] —, "Frequency errors in MEM processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 257-259, Apr. 1980.
[230] —, "Comments on maximum entropy spectral estimation using the analytical signal," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 259-260, Apr. 1980.
[231] D. J. Thomson, M. F. Robbins, C. G. Maclennan, and L. J. Lanzerotti, "Spectral and windowing techniques in power spectral analyses of geomagnetic data," *Phys. Earth Planetary Interiors*, vol. 12, pp. 217-231, Aug. 1976.
[232] D. J. Thomson, "Spectrum estimation techniques for characterization and development of WT4 waveguide-I," *Bell Syst. Tech. J.*, vol. 56, pp. 1769-1815, Nov. 1977.
[233] T. Thorvaldsen, A. T. Waterman Jr., and R. W. Lee, "Maximum entropy angular response patterns of microwave transhorizon signals," *IEEE Trans. Antennas Propagat.*, vol. AP-28, pp. 722-724, Sept. 1980.
[234] K. Toman, "The spectral shift of truncated sinusoids," *J. Geophysical Res.*, vol. 70, pp. 1749-1750, Apr. 1, 1965.
*[235] H. Tong, "Autoregressive model fitting with noisy data by Akaike's information criterion," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 476-480, July 1975.
[236] —, "Fitting a smooth moving average to noisy data," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 493-496, July 1976.
*[237] —, "More on autoregressive model fitting with noisy data by Akaike's information criterion," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 409-410, May 1977.
[238] H. Tong, "Final prediction error and final interpolation error: A paradox?" *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 758-759, Nov. 1979.
[239] J. P. Toomey, "High-resolution frequency measurement by linear prediction," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-16, pp. 517-525, July 1980.
[240] J. Treichler, "γ-LMS and its use in a noise-compensating adaptive spectral analysis technique," in *Rec. 1979 ICASSP*, pp. 933-936.
[241] W. Trench, "An algorithm for the inversion of finite Hankel matrices," *J. Soc. Indust. Appl. Math.*, vol. 13, no. 4, pp. 1102-1107, Dec. 1965.
[242] S. A. Tretter and K. Steiglitz, "Power spectrum identification in terms of rational models," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 185-188, Apr. 1967.
[243] D. W. Tufts, "Adaptive line enhancement and spectrum analysis," *Proc. IEEE*, vol. 65, pp. 169-173, Jan. 1977.
[244] D. W. Tufts and R. Kumaresan, "Improved spectral resolution," *Proc. IEEE*, vol. 68, pp. 419-420, Mar. 1980.
[245] —, "Improved spectral resolution II," in *Proc. 1980 ICASSP* (Denver, CO), vol. 2, pp. 592-597, Apr. 9-11, 1980.
[246] T. J. Ulrych, "Maximum entropy power spectrum of long period geomagnetic reversals," *Nature*, vol. 235, pp. 218-219, Jan. 28, 1972.
[247] —, "Maximum entropy power spectrum of truncated sinusoids," *J. Geophysical Research*, vol. 77, pp. 1396-1400, March 10, 1972.
[248] T. J. Ulrych, D. E. Smylie, O. G. Jensen, and G.K.C. Clarke, "Predictive filtering and smoothing of short records by using maximum entropy," *J. Geophysical Res.*, vol. 78, pp. 4959-4964, Aug. 10, 1973.
[249] T. J. Ulrych and O. G. Jensen, "Cross-spectral analysis using maximum entropy," *Geophysics*, vol. 39, pp. 353-354, June 1974.
*[250] T. J. Ulrych and T. N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Rev. Geophysics Space Phys.*, vol. 13, pp. 183-200, Feb. 1975.
[251] T. J. Ulrych and R. W. Clayton, "Time series modelling and maximum entropy," *Phys. Earth Planetary Interiors*, vol. 12, pp. 188-200, Aug. 1976.
[252] M. L. Van Blaricum and R. Mittra, "Problems and solutions associated with Prony's method for processing transient data," *IEEE Trans. Antennas Propagat.*, vol. AP-26, pp. 174-182, Jan. 1978.
[253] M. L. Van Blaricum, "A review of Prony's method techniques for parameter estimation," in *Proc. Rome Air Development Center Spectrum Estimation Workshop*, Griffis Air Force Base, May 24-26, 1978. pp. 125-139.
[254] M. L. Van Blaricum and R. Mittra, Correction to "Problems and solutions associated with Prony's method for processing transient data," *IEEE Trans. Antennas Propagat.*, vol. AP-28, p. 949, Nov. 1980.
**[255] A. VanDenBos, "Alternative interpretation of maximum entropy spectral analysis," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 493-494, July 1971.
[256] H. L. Van Trees, *Detection, Estimation, and Modulation Theory.* New York: McGraw-Hill, 1968, vol. I.
[257] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 309-321, June 1975.
[258] G. Walker, "On periodicity in series of related terms," *Proc. Roy. Soc. London, Series A*, vol. 131, pp. 518-532, 1931.
[259] C. Webb, "Practical use of the fast Fourier transform (FFT) algorithm in time series analysis," ARL, Univ. Texas, Austin, TX, ARL-TR-70-22, June 22, 1970.

[260] R. J. Webster, "Leakage regulation in the discrete Fourier transform spectrum," *Proc. IEEE*, vol. 68, pp. 1339-1341, Oct. 1980.
[261] L. Weiss and R. N. McDonough, "Prony's method, Z-transforms, and Pade approximation," *SIAM Rev.*, vol. 5, pp. 145-149, Apr. 1963.
[262] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70-73, June 1967.
[263] P. D. Welch, "On the variance of time and frequency averages over modified periodograms," in *Rec. 1977 IEEE Int. Conf. Acoustics, Speech, and Signal Processing* (Hartford, CT), May 9-11, 1977, pp. 58-62.
[264] S. J. Wernecke and L. R. D'Addario, "Maximum entropy image reconstruction," *IEEE Trans. Comput.*, vol. C-26, pp. 351-364, Apr. 1977.
[265] S. J. Wernecke, "Two-dimensional maximum entropy reconstruction of radio brightness," *Radio Sci.*, vol. 12, pp. 831-844, Sept.-Oct. 1977.
[266] E. C. Whitman, "The spectral analysis of discrete time series in terms of linear regressive models," Naval Ordnance Labs Rep. NOLTR-70-109, White Oak, MD, June 23, 1974.
[267] P. Whittle, "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," *Biometrika*, vol. 50, pp. 129-134, 1963.
[268] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and application," *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.
[269] N. Wiener, "Generalized harmonic analysis," *Acta Mathematica*, vol. 55, pp. 117-258, 1930.
[270] R. A. Wiggins and E. A. Robinson, "Recursive solution to the multi-channel filtering problem," *J. Geophysical Res.*, vol. 70, pp. 1885-1891, Apr. 1965.
[271] A. S. Willsky, "Relationship between digital signal processing and control and estimation theory," *Proc. IEEE*, vol. 66, pp. 996-1017, Sept. 1978.
[272] ——, *Digital Signal Processing and Control and Estimation Theory: Points if Tangency, Areas of Intersection, and Parallel Directions.* Cambridge, MA: M.I.T. Press, 1980.
[273] T. H. Wonnacott, "Spectral analysis combining a Bartlett window with an associated inner window," *Technometrics*, vol. 3, pp. 235-243, May 1961.
[274] C. K. Yuen, "A comparison of five methods for computing the power spectrum of a random process using data segmentation," *Proc. IEEE*, vol. 65, pp. 984-986, June 1977.
[275] ——, "Comments on modern methods for spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 298-299, June 1979.
[276] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Trans. Royal Soc. London, Series A*, vol. 226, pp. 267-298, July 1927.
[277] S. Zohar, "The solution of a Toeplitz set of linear equations," *J. Assoc. Comput. Mach.*, vol. 21, pp. 272-276, Apr. 1974.
[278] ——, "FORTRAN subroutines for the solution of Toeplitz sets of linear equations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 656-658, Dec. 1979.

A Comprehensive Survey of Digital Transmultiplexing Methods

HELMUT SCHEUERMANN AND HEINZ GÖCKLER

Abstract—With this survey, an attempt is made to describe the great majority of all known methods of digital transmultiplexing (i.e., conversion) of time-division-multiplex (TDM) to frequency-division-multiplex (FDM) signals, and vice versa. To this end, the individual transmultiplexer approaches are classified into four categories according to the underlying algorithm: Bandpass filter bank, low-pass filter bank, Weaver structure method, and multistage modulation method. Finally, the overall performance of the various transmultiplexer approaches are compared with each other by means of different criteria [1], such as stability under looped conditions, absolute value of the group delay, computational and control complexity, modularity, potential of intelligible crosstalk, absence of an additional analog frequency conversion, and the impact of out-of-band signaling. For a more profound understanding of the individual digital transmultiplexer approaches, the main chapter is preceded by an introductory discussion on analog and digital generation of single-sideband signals. In this context, the associated problems of sample rate alteration and multirate filtering arising from digital signal processing are dealt with.

Manuscript received February 8, 1980; revised July 30, 1981. This work was supported in part by the German Federal Ministry of Research and Technology.

The authors are with the Advanced Development Department, AEG-Telefunken Kommunikationstechnik, P.O.B. 1120, D-7150 Backnang, Germany.

I. Introduction

ANALOG TRANSMISSION and switching facilities for telephony signals are nowadays to a growing extent being expanded and replaced by digital facilities. Thereby the conventional multiple utilization of transmission paths by frequency-division-multiplex (FDM) is being substituted by time-division-multiplex (TDM) techniques. The chief advantages of digital TDM transmission as compared with analog FDM transmission are as follows:

1) no generation of additive noise on the transmission path;
2) within certain limits, no occurrence of interference through crosstalk;
3) possibility for concentration of switching and transmission facilities.

The widespread use and high investment outlays of the installed facilities will require that analog and digital technologies coexist well into the foreseeable future. This will lead to an increasing extent to interfaces between analog and digital sections of the toll communication network. Interconnection

0018-9219/81/1100-1419$00.75

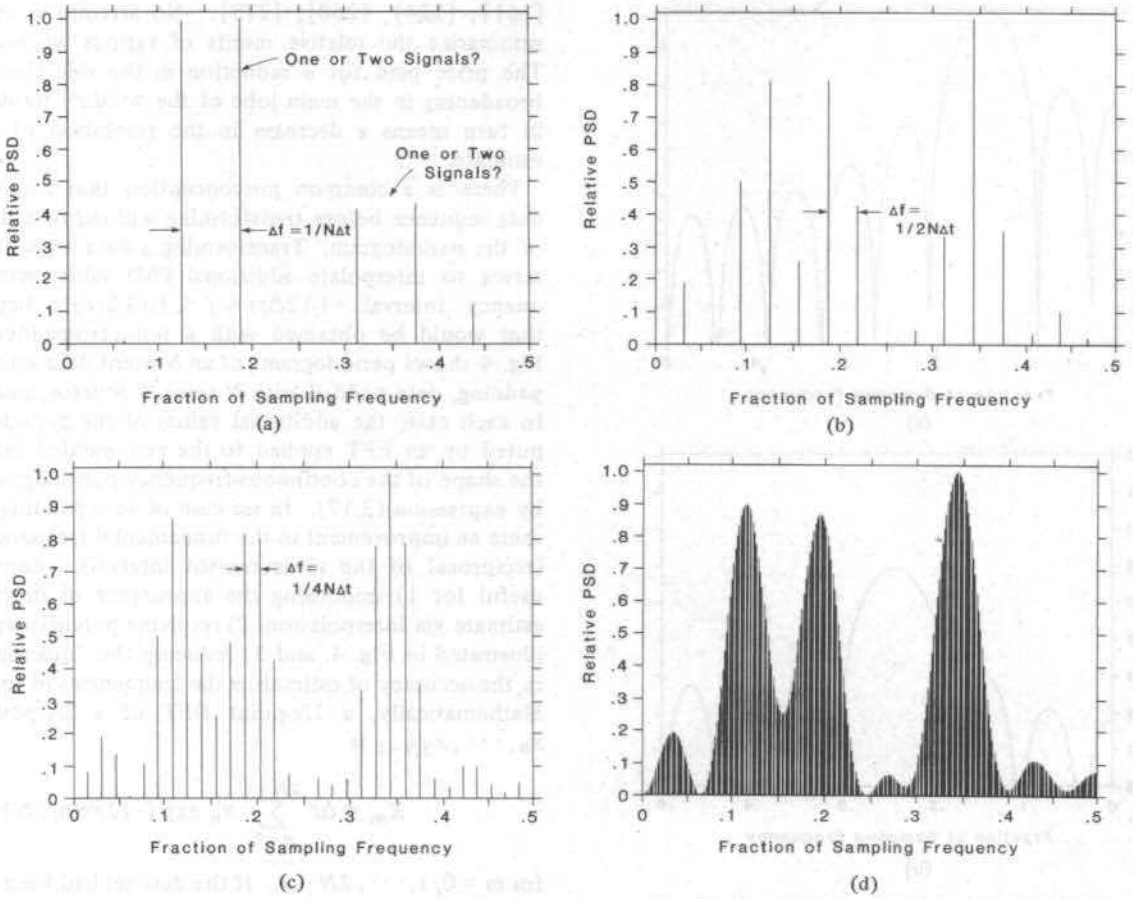

Fig. 4. Impact of zero padding the periodogram to interpolate the spectral shape and to resolve ambiguities. The spectra were estimated using the same 16 samples of a process consisting of three sinusoids of fractional sampling frequencies 0.1335, 0.1875, 0.3375 and initial phases 0°, 90°, 0°, respectively. (a) No zero padding; ambiguities are present in the spectrum. (b) Double padding; ambiguities resolved. (c) Quadruple padding; smoothest spectrum seen. (d) 32-times pad-