[92] —, The statistical analysis of a seiche record, *J. Marine Res.*, vol. 13, pp. 76–100, 1954.
[93] G. Box and G. M. Jenkins, *Time Series Analysis, Forecasting, and Control.* Oakland, CA: Holden-Day, 1970.
[94] H. Akaike, "Power spectrum estimation through autoregressive model fitting," *Ann. Inst. Statist. Math.*, vol. 21, pp. 407–419, 1969.
[95] —, "A new look at statistical model identification," *IEEE Trans. Autom. Contr.*, vol. AC-19, pp. 716–723, 1974.
[96] D. G. Childers, *Modern Spectrum Analysis.* New York: IEEE Press, 1978.
[97] S. Haykin, *Nonlinear Methods of Spectral Analysis.* Berlin, Germany: Springer, 1979.
[98] *Proc. of the RADC Spectrum Estimation Workshop*, Rome Air Development Center, Griffiss Air Force Base, NY, 1979.
[99] *Proc. of the First ASSP Workshop on Spectral Estimation*, McMaster University, Hamilton, Ontario, Canada, 1981.
[100] T. J. Ulrych and T. N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Rev. Geophys.*, vol. 13, pp. 183–200, 1975.
[101] P. R. Gutowski, E. A. Robinson, and S. Treitel, "Spectral estimation: Fact or fiction," *IEEE Trans. Geosci. Electronics*, vol. GE-16, pp. 80–84, 1978. Reprinted in D. G. Childers, *Modern Spectrum Analysis.* New York: IEEE Press, 1978.

Spectral Estimation: An Overdetermined Rational Model Equation Approach

JAMES A. CADZOW, SENIOR MEMBER, IEEE

Abstract—In seeking rational models of time series, the concept of approximating second-order statistical relationships (i.e., the Yule–Walker equations) is often explicitly or implicitly invoked. The parameters of the hypothesized rational model are typically selected so that these relationships "best represent" a set of autocorrelation lag estimates computed from time series observations. One of the objectives of this paper will be that of establishing this fundamental approach to the generation of rational models.

An examination of many popular contemporary spectral estimation methods reveals that the parameters of a hypothesized rational model are estimated upon using a "minimal" set of Yule–Walker equation evaluations. This results in an undesired parameter hypersensitivity and a subsequent decrease in estimation performance. To counteract this parameter hypersensitivity, the concept of using more than the minimal number of Yule–Walker equation evaluations is herein advocated. It is shown that by taking this overdetermined parametric evaluation approach, a reduction in data-induced model parameter hypersensitivity is obtained, and a corresponding improvement in modeling performance results. Moreover, upon adapting a singular value decomposition representation of an extended-order autocorrelation matrix estimate to this procedure, a desired model order determination method is obtained and a further significant improvement in modeling performance is achieved. This approach makes possible the generation of low-order high-quality rational spectral estimates from short data lengths.

I. Introduction

IN A VARIETY of applications such as found in radar Doppler processing, adaptive filtering, speech processing, underwater acoustics, seismology, econometrics, spectral estimation, and array processing, it is desired to estimate the statistical characteristics of a wide-sense stationary time series. More often than not, this required characterization is embodied in the time series' *autocorrelation lag* sequence as specified by

$$r_x(n) = E\{x(n+m)\,\bar{x}(m)\} \tag{1.1}$$

in which E and $^-$ denote the operations of expectation and complex conjugation, respectively. From this definition, the well-known property that the autocorrelation lags are complex conjugate symmetric (i.e., $r_x(-n) = \bar{r}_x(n)$) is readily established. We will automatically assume this property whenever negative lag autocorrelation elements (or their estimates) are required.

The second-order statistical characterization as represented by the autocorrelation sequence may be given an "equivalent" frequency-domain interpretation. Namely, upon taking the Fourier transform of the autocorrelation sequence, that is,

$$S_x(e^{j\omega}) = \sum_{n=-\infty}^{\infty} r_x(n)\, e^{-jn\omega} \tag{1.2}$$

we obtain the associated *power spectral density function* $S_x(e^{j\omega})$ in which the normalized frequency variable ω takes on values in $[-\pi, \pi]$. This function possesses a number of salient properties among which are that it is a positive semi-definite, symmetric (if the time series is real valued), and periodic function of ω. This function is seen to have a Fourier series interpretation in which the autocorrelation lags play the role of the Fourier coefficients. It, therefore, follows that these coefficients may be determined from the power spectral density function through the Fourier series coefficient integral expression

$$r_x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(e^{j\omega})\, e^{j\omega n}\, d\omega. \tag{1.3}$$

Manuscript received May 24, 1982. This work was supported in part by the Signal Processing Section of RADC under Grant AFOSR-81-0250 and by the Statistics and Probability Program of the Office of Naval Research under Contract N00014-82-K-0257.

The author is with the College of Engineering and Applied Sciences, Department of Electrical and Computer Engineering, Arizona State University, Tempe, AZ 85287.

0018-9219/82/0900-0907$00.75

Relationships (1.2) and (1.3) form a Fourier transform pair so that knowledge of the autocorrelation sequence is equivalent to knowledge of the power spectral density function and vice versa. We belabor this point in order to establish the viewpoint that spectral estimation and autocorrelation lag estimation are conceptually equivalent.

In the classical spectral estimation problem, it is desired to effect an estimate of the underlying power spectral density function with this estimate being based on *only* a finite set of time series observations. Typically, these observations will be composed of a set of contiguous data measurements taken at equispaced time intervals T as represented by

$$x(1), x(2), \cdots, x(N) \tag{1.4}$$

where N will be referred to as the data length and we have chosen to suppress the sampling period T. It is apparent that unless some constraints are imposed on the basic nature of the power spectral density function, there exists a fundamental incompatability in seeking an estimate of the infinite parameter function (1.2) (i.e., the infinite set of autocorrelation lag parameters) based on the finite set of observations (1.4). Investigators have often resolved this dilemma by postulating a finite parameter model for the power spectral density function. The time series observations (1.4) are then used to fix the parameters of this parametric model using an appropriate estimation procedure.

Without doubt, the most widely used and studied of finite parametric models are the so-called rational models. When employing a rational model, we are seeking to approximate the generally infinite series expansion (1.2) by a magnitude squared ratio of polynomials in the variable $e^{-j\omega}$, that is,

$$S(e^{j\omega}) = \left| \frac{b_0 + b_1 e^{-j\omega} + \cdots + b_q e^{-jq\omega}}{1 + a_1 e^{-j\omega} + \cdots + a_p e^{-jp\omega}} \right|^2 . \tag{1.5}$$

The finite number of parameters in this model then provides the mechanism for circumventing the aforementioned parameter mismatch dilemma. Namely, if the data length parameter N adequately exceeds this rational function's number of parameters (i.e., $p + q + 1$), then it is feasible to utilize the given time series observations (1.4) to estimate values for these parameters. A few words are now appropriate concerning the adequacy of rational models in representing power spectral density functions. It is well known that if a power spectral density function is continuous in the variable ω, then it may be approximated arbitrarily closely by a rational function of form (1.5) if the order parameters p and q are selected suitably large [41]. Comforted by this knowledge, rational functions have become a standard tool of spectral estimation theoreticians. As an interesting side note, it is ironic that the origin of spectral estimation was in the use of rational models for characterizing time series composed of sinusoids in white noise. Members of this class of time series possess *discontinuous* power spectral density functions and are, therefore, presumably not representable by a rational model. As we will see in Section III, however, it is possible to suitably adapt a specific rational model so as to satisfactorily characterize this class of time series.

This paper is primarily concerned with developing a modeling method which utilizes an overdetermined set of statistical equations for estimating a rational model's parameters. Using this approach, it is found that the resultant modeling performance is generally better than that achieved by other popularly used parametric methods. Although the approach here taken reflects heavily upon the author's previous works [15]–[22], much of this paper will be concerned with formulating many contemporary spectral estimation methods in a common autocorrelation representation setting. It must be emphasized that our main objective is not that of giving an encylopedic coverage of the many available rational spectral estimation techniques. This paper in conjunction with the excellent recent publications [23], [31], [37], however, provides a reasonable complete coverage of parametric methods.

In this section, we shall consider two special classes of rational functions and give a brief historical perspective on their usage in spectral estimation theory. These two classes are commonly referred to as the moving average (MA) and the autoregressive (AR) spectral models. A moving average model is defined to be a rational function (1.5) in which all the a_k parameters are zero (i.e., it has only numerator dynamics) while an autoregressive model is one for which all the b_k parameters are zero except for b_0 (i.e., it has only denominator dynamics). By-in-large, these two classes of rational functions have formed the basic modeling tools in contemporary spectral estimation theory. The more general autoregressive-moving average (ARMA) model (1.5), however, is receiving an increasing amount of attention.

MA Model

Fourier analysis has played a primary role in much of the earlier as well as more recent efforts at spectrally characterizing experimentally collected data. As an example, Schuster developed the periodogram method for detecting hidden periodicities in sun spot activity data at the turn of the century [48]. In a more recent classical work, Blackman and Tukey presented a generalized procedure for effecting spectral estimates [8]. This involved the two-step procedure of i) determining autocorrelation lag estimates $\hat{r}_x(n)$ using the provided data, and ii) taking the Fourier transform of these estimates.[1] The power spectral density estimate which arose when taking this approach then took the form

$$\hat{S}_x(e^{j\omega}) = \sum_{n=-q}^{q} w(n)\, \hat{r}_x(n)\, e^{-j\omega n} \tag{1.6}$$

where $w(n)$ is a symmetric data window that is chosen to achieve various desirable effects such as sidelobe reduction. This window is often selected to be rectangular in which case $w(n) = 1$ although other choices may be more desirable for a given application. A description of some of the more popular choices for the data window may be found in numerour texts (e.g., see [33], [50], [57]).

In the Blackman–Tukey estimate (1.6), it is seen that only a finite number of summand terms (i.e., $2q + 1$) are involved in the spectral estimate. This is a direct consequence of the fact that *only* a finite set of autocorrelation lag estimates are obtainable from the observation set (1.4) if standard lag estimation methods are employed. Due to this finite sum structure, we will now show that the Blackman–Tukey estimation method is a special case of the more general rational moving average (MA) spectral model. In particular, a spectral model is said to be a *moving average* model of order q (i.e., MA(q)) if it may be put into the form

$$\begin{aligned} S_{\mathrm{MA}}(e^{j\omega}) &= |b_0 + b_1 e^{-j\omega} + \cdots + b_q e^{-jq\omega}|^2 \\ &= |B_q(e^{j\omega})|^2 . \end{aligned} \tag{1.7}$$

The $q + 1$ parameters $b_0, b_1, \cdots, b_q$ which identify this

[1] We shall hereafter use the caret symbol (^) to denote a statistical estimate.

MA(q) model are seen to form a qth-order polynomial $B_q(e^{j\omega})$ in the variable $e^{-j\omega}$. A moving average model is then seen to be a special case of the more general rational model (1.5) in which the denominator polynomial has been set equal to the constant one.

If the polynomial constituting the moving average model (1.7) is factored, it is possible to provide additional insight into an MA model's properties. This factorization is seen to give rise to the equivalent representation

$$S_{MA}(e^{j\omega}) = |b_0|^2 \prod_{k=1}^{q} (1 - z_k e^{-j\omega})(1 - \bar{z}_k e^{j\omega}) \quad (1.8)$$

in which the z_k are the roots of the polynomial $B_q(z)$. The zeroes of an MA spectral model are seen to occur in reciprocal pairs. Due to the basic nature of this factorization, moving average models are, therefore, also commonly referred to as *all-zero* models. If any of the roots z_k are close to the unit circle (i.e., $z_k \approx e^{j\omega_k}$), it is clear that $S_{MA}(e^{j\omega})$ will contain sharply defined notches at frequencies in a neighborhood associated with these roots (i.e., $\omega = \omega_k$). It is, therefore, apparent that MA models will be particularly effective when approximating spectra that contain sharply defined notches (zero-like behavior) but not sharply defined peaks. Whenever, a spectrum contains sharply defined peaks, it is possible to simulate their effect at the cost of many additional zeroes (i.e., a high MA order) for an adequate representation. With this in mind, MA models should be normally avoided whenever a peaky type behavior in the underlying spectrum is suspected (as may be made evident from a preliminary Blackman–Tukey estimate).

To establish the fact that the Blackman–Tukey approach to spectral estimation is of a moving average structure, it is possible to give yet another equivalent representation to the MA(q) expression (1.7). This will entail explicitly carrying out the indicated polynomial product $B_q(e^{j\omega})\,\overline{B}_q(e^{j\omega})$ thereby giving

$$S_{MA}(e^{j\omega}) = \sum_{n=-q}^{q} c_n\, e^{-j\omega n} \quad (1.9)$$

in which the complex conjugate symmetric c_n parameters are related to the original b_n parameters according to

$$c_n = \sum_{k=0}^{q} b_k \bar{b}_{k-n}, \qquad -q \leqslant n \leqslant q. \quad (1.10)$$

where b_k is taken to be zero outside the set $[0, q]$. Upon setting the c_n equal to $w(n)\hat{r}_x(n)$, it is apparent that the Blackman–Tukey estimate (1.6) is a special form MA(q) model. This fact is usually overlooked by investigators who have considered the Blackman–Tukey method as well as the periodogram method as nonparametric spectral estimators. When viewed from the approach here taken, however, each of these procedures is recognized as being a realization of an MA parametric model.

AR Model

When we compare the MA(q) spectral model expression (1.9) with the theoretical power spectral density function (1.2) which is being estimated, it is apparent that a serious modeling mismatch can arise whenever the underlying autocorrelation lags are such that the $r_x(n)$ are not approximately equal to zero for $n > q$. For example, this undesirable condition arises when the time series under study is composed of sinusoids in white noise. Conversely, this condition does not arise for broad-band signals. The sinusoid example is mentioned since it forms one of the more interesting special-case time series to which spectral estimation techniques are applied. A special treatment of the sinusoids in white-noise case will be given in Section III.

In recognition of this potential shortcoming of MA models, investigators have examined alternate rational spectral models which do not invoke the unnecessarily harsh requirement of a truncated autocorrelation lag behavior. Undoubtably, the most widely used of such models is the autoregressive (AR) model. Namely, a spectral model is said to be an *autoregressive* model of order p (i.e., AR(p)) if it may be put into the form

$$S_{AR}(e^{j\omega}) = \left| \frac{b_0}{1 + a_1 e^{-j\omega} + a_2 e^{-j2\omega} + \cdots + a_p e^{-jp\omega}} \right|^2 = \frac{|b_0|^2}{|A_p(e^{j\omega})|^2}. \quad (1.11)$$

This AR(p) model has a functional behavior which is completely characterized by its $p + 1$ parameters $b_0, a_1, a_2, \cdots, a_p$. The characteristic pth-order polynomial $A_p(e^{j\omega})$ is seen to influence the frequency behavior of the estimate while the parameter b_0 controls the level.

As in the MA model case, valuable insight into the capabilities of AR modeling is provided upon factoring the polynomial $A_p(e^{j\omega})$. This is found to result in the equivalent representation

$$S_{AR}(e^{j\omega}) = \frac{|b_0|^2}{\prod_{k=1}^{p} (1 - p_k e^{-j\omega})(1 - \bar{p}_k e^{j\omega})} \quad (1.12)$$

where the p_k are the roots of $A_p(z)$. The poles of this AR spectral model are seen to occur in reciprocal pairs. For reasons which are self-evident, the AR(p) spectral model is also commonly referred to as an all-pole model. As such, it is particularly appropriate for modeling spectra which contain sharply defined peaks (pole-like behavior) but not sharply defined notches. If a spectrum does possess notches, however, it is possible to simulate their effect at the cost of many additional poles (i.e., a high-AR order). In terms of parameter parsimony, it is, therefore, prudent to avoid AR models whenever notches in the underlying spectrum are suspected (this may be made evident from a preliminary Blackman–Tukey estimate).

AR models were used by Yule [66] and Walker [63] in forecasting trends of economically based time series. These models were then employed by Burg [13] in 1967 and Parzen [53] in 1968 to achieve spectral estimates which did not possess the aforementioned deficiencies of the MA model. The Burg method is of particular interest since it offered a new insight into spectral modeling and introduced a number of concepts that are now standard tools of spectral estimation. This includes an efficient lattice structured implementation of the Burg method which has since been examined and advanced by many investigators (e.g., see [44]). It is not an exaggeration to say that Burg's method gave rise to a literal explosion in research activity directed towards evolving improved rational modeling methods.

ARMA Models

In many applications, the underlying power spectral density function will contain both notch- and peak-like behavior. As such, neither the MA nor the AR model is the most appropriate model representation from a parameter parsimony view-

point. The more general rational model (1.5), however, is capable of efficiently representing such behavior. This most general rational model is commonly referred to as an autoregressive-moving average model of order (p, q) (i.e., ARMA (p, q)) with its frequency characterization being given by

$$S_{\mathrm{ARMA}}(e^{j\omega}) = \left| \frac{b_0 + b_1 e^{-j\omega} + \cdots + b_q e^{-jq\omega}}{1 + a_1 e^{-j\omega} + \cdots + a_p e^{-jp\omega}} \right|^2 = \left| \frac{B_q(e^{j\omega})}{A_p(e^{j\omega})} \right|^2. \tag{1.13}$$

An ARMA model is seen to have a frequency characterization which is the composite of an MA and an AR model. To further reinforce this interpretation, we have the following equivalent representation upon factoring the polynomials $A_p(e^{j\omega})$ and $B_q(e^{j\omega})$ which characterize its frequency behavior:

$$S_{\mathrm{ARMA}}(e^{j\omega}) = |b_0|^2 \frac{\prod_{k=1}^{q} (1 - z_k e^{-j\omega})(1 - \bar{z}_k e^{j\omega})}{\prod_{k=1}^{p} (1 - p_k e^{-j\omega})(1 - \bar{p}_k e^{j\omega})}. \tag{1.14}$$

An ARMA model is seen to possess q zeroes and p poles, and, as such, it is generally a much more effective model than are its more specialized MA (all-zero) and AR (all-pole) model counterparts. These poles and zeroes are seen to occur in reciprocal pairs.

Although ARMA models are the most preferable choice for various applications, many practitioners have opted to utilize either MA or AR models. There is an increasing awareness, however, of the general superiority of ARMA modeling. This has given rise to a renewed effort to generate computationally efficient ARMA modeling algorithms. A particularly effective approach to ARMA modeling will be presented in this paper.

II. Rational Modeling—Exact Autocorrelation Knowledge

In this section, the theoretical autocorrelation characteristics of MA, AR, and ARMA random processes are examined separately. This characterization will, in turn, enable us to intelligently select the most appropriate rational model which best represents a given set of exact autocorrelation lags

$$r_x(0), r_x(1), \cdots, r_x(s). \tag{2.1}$$

Moreover, a systematic procedure for identifying the selected model's parameters from these given autocorrelation lag values is also developed. Although the assumption here made of exact autocorrelation information is highly idealistic and almost never met in applications, the insight thereby provided is helpful when considering the more practical problem of generating rational model estimates from raw time series observations.

To begin this analysis, it will be hereafter assumed that the time series under examination is generated (or can be adequately modeled) as the response associated with the linear operator

$$x(n) + \sum_{k=1}^{p} a_k x(n-k) = \sum_{k=0}^{q} b_k \epsilon(n-k) \tag{2.2}$$

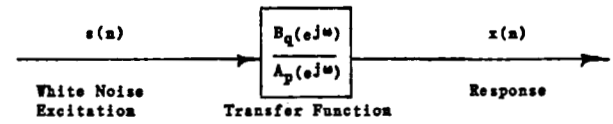

Fig. 1. Model of rational time series.

in which the excitation time series $\{\epsilon(n)\}$ is taken to be a sequence of zero-mean, unit-variance, uncorrelated random variables (i.e., normalized white noise) that is taken to be unobservable. This excitation-response behavior is depicted in Fig. 1. Using standard techniques, it is readily shown that the power spectral density function associated with the response time series is given by the ARMA(p, q) rational form

$$S_x(e^{j\omega}) = \left| \frac{b_0 + b_1 e^{-j\omega} + \cdots + b_q e^{-jq\omega}}{1 + a_1 e^{-j\omega} + \cdots + a_p e^{-jp\omega}} \right|^2.$$

Thus there is an equivalency between an assumed ARMA (p, q) spectral model and the response of the recursive linear operator (2.2) to white noise. In this section, the required rational modeling will be developed through use of the time series description (2.2) and its associated autocorrelation characterization. It is interesting to note that most available rational spectral estimation techniques are based upon a time-domain characterization.

The mechanism for effecting the required rational modeling are the so-called Yule-Walker equations which govern linear relationship (2.2). Namely, upon multiplying both sides of this relationship by $\bar{x}(n-m)$ and then taking expected values, it is found that the Yule-Walker equations

$$\sum_{k=0}^{p} a_k r_x(n-k) = \sum_{i=0}^{q} b_i \bar{h}(i-n) \tag{2.3}$$

arise where $a_0 = 1$. The entity $h(n)$ herein used corresponds to the unit-impulse (i.e., Kronecker delta) response of linear operator (2.2). This unit-impulse response may also be interpreted as being the inverse Fourier transform of the linear operator's transfer function $B_q(e^{j\omega})/A_p(e^{j\omega})$. In what is to follow, it will be assumed that this linear operator is causal thereby implying that $h(n) = 0$ for n negative. Although this assumption is not essential in the analysis which follows, it is here imposed in recognition of the fact that most applications are inherently involved with causal operations. Adaption to the case where noncausal operations are more appropriate is straightforward and will not be given.

The Yule-Walker equations (2.3) take on a particularly simple form when the linear operator (2.2) which they describe is constrained to be an MA or an AR linear operator. To delineate this fact, we shall now examine separately the basic characteristics of the Yule-Walker equations when the underlying linear model is taken to be MA, AR, and ARMA.

MA Time Series

The time series $\{x(n)\}$ is said to be an MA random process if it is generated according to the linear nonrecursive relationship

$$x(n) = \sum_{k=0}^{q} b_k \epsilon(n-k) \tag{2.4}$$

where $\{\epsilon(n)\}$ is the aforementioned normalized white-noise excitation process. According to the general Yule-Walker equations (2.3), the response's autocorrelation sequence is, therefore, specified by

$$r_x(n) = \begin{cases} \sum_{k=0}^{q} b_k \bar{b}_{k-n}, & -q \leqslant n \leqslant q \\ 0, & \text{otherwise} \end{cases} \tag{2.5}$$

where use of the facts that $a_k = 0$ and $h(n) = b_n$ for $0 \leqslant n \leqslant q$ have been incorporated. Thus the autocorrelation sequence associated with a moving average process is seen to be of finite length (i.e., $2q + 1$) with the length identifying the order of the MA(q) process.

We shall now consider the problem of identifying the MA parameters b_k which correspond to a given $2q + 1$ length autocorrelation sequence $r_x(n)$ for $-q \leqslant n \leqslant q$. This identification will be made by examining the spectral density function associated with the autocorrelation sequence. In particular, upon taking the z-transform (in lieu of the Fourier transform) of the given $2q + 1$ length autocorrelation sequence, we have upon using relationship (2.5)

$$\begin{aligned} S_x(z) &= \sum_{n=-q}^{q} r_x(n) z^{-n} \\ &= \sum_{n=-q}^{q} \sum_{k=0}^{q} b_k \bar{b}_{k-n} z^{-n} \\ &= \sum_{k=0}^{q} b_k z^{-k} \sum_{m=0}^{q} \bar{b}_m z^{m}. \end{aligned} \tag{2.6}$$

Since the finite-power series $S_x(z)$ has complex conjugate symmetrical coefficients (i.e., $r_x(-n) = \bar{r}_x(n)$), it follows that the zeroes of this power series must occur in reciprocal pairs. With this in mind, it is, therefore, always possible to factor the power spectral density function as

$$S_x(z) = \alpha^2 \prod_{k=1}^{q} (1 - z_k z^{-1})(1 - \bar{z}_k z) \tag{2.7}$$

where α is a real scalar. Upon comparing expressions (2.6) and (2.7), it is apparent that

$$\sum_{k=0}^{q} b_k z^{-k} = \alpha \prod_{k=1}^{q} (1 - z_k z^{-1}). \tag{2.8}$$

Thus the required b_k parameter identification is achieved by carrying out the right-side multiplications in (2.8) and then equating coefficients of equal powers of z^{-k}. The most critical step of this identification procedure is the factorization of the known power series $S_x(z)$ as given in (2.7).

One point of caution should be raised in following this approach. It arises due to the fact that although the factorization of $S_x(z)$ into its $2q$ first-order product terms is unique, the decomposition (2.7) is certainly not. This is a direct consequence of the appearance of the roots of $S_x(z)$ in reciprocal pairs. Thus the term $(1 - z_1 z^{-1})$ may be replaced by $(1 - z_1^{-1} z^{-1})$ in (2.8) without destroying the required structure (2.6). This replacement, however, will in general lead to a different set of b_k parameters. Since there are typically q different first-order reciprocal pairs in the factorization (2.7), it then follows that there are 2^q different b_n parameter sets which are compatible with the autocorrelation identity (2.5). The one normally chosen corresponds to the so-called *minimum phase* selection in which the z_k roots used in (2.8) are selected so that they all have magnitudes less than or equal to one.

AR Time Series

The time series $\{x(n)\}$ is said to be an AR process of order p if it is generated according to the recursive relationship

$$x(n) + \sum_{k=1}^{p} a_k x(n - k) = b_0 \epsilon(n) \tag{2.9}$$

where $\{\epsilon(n)\}$ is the aforementioned normalized white-noise process. The Yule–Walker equations (2.3) indicate that the AR(p) autocorrelation elements are related by

$$r_x(n) + \sum_{k=1}^{p} a_k r_x(n - k) = \begin{cases} |b_0|^2, & n = 0 \\ 0, & n \geqslant 1 \end{cases} \tag{2.10}$$

where use of the facts that $h(0) = b_0$ and $h(n) = 0$ for $n < 0$ have been made.

In order to effect a direct procedure for identifying the AR(p) model's $p + 1$ parameters $a_1, a_2, \cdots, a_p, b_0$ which best represent the set of autocorrelation lag values (2.1), one may evaluate the first $p + 1$ of the governing Yule–Walker equations. This evaluation when put into a matrix format takes the form

$$\begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-p) \\ r_x(1) & r_x(0) & \cdots & r_x(-p+1) \\ \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots\vdots & \vdots \\ r_x(p) & r_x(p-1) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} |b_0|^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.11a}$$

or more compactly as

$$R\boldsymbol{a} = |b_0|^2 \boldsymbol{e}_1. \tag{2.11b}$$

In this expression, R is the $(p + 1) \times (p + 1)$ AR autocorrelation matrix whose elements are given by

$$R(i, j) = r_x(i - j), \quad 1 \leqslant i \leqslant p + 1, \quad 1 \leqslant j \leqslant p + 1 \tag{2.12}$$

$\boldsymbol{a}$ is the $(p + 1) \times 1$ autoregressive parameter vector with first component equal to one, that is

$$\boldsymbol{a} = [1, a_1, a_2, \cdots, a_p]' \tag{2.13}$$

and $\boldsymbol{e}_1$ is the $(p + 1) \times 1$ standard basis vector whose elements are all zero except for its first which is one. The required parameter identification is then obtained upon solving this system of $p + 1$ linear equations in the $p + 1$ unknowns. Conceptually, this solution may be effected by performing the following computation:

$$\boldsymbol{a} = |b_0|^2 R^{-1} \boldsymbol{e}_1 \tag{2.14}$$

in which the normalizing coefficient b_0 is selected so that the first component of $\boldsymbol{a}$ is one as required in (2.13). In this solution procedure, we are tacitly assuming the invertibility of the autocorrelation matrix R. If matrix R is singular, however, this almost always implies that the underlying time series is an autoregressive process of order less than p. In this case, it will be necessary to decrease the order until R becomes invertible.

Upon examination of (2.11), it is seen that the resultant AR(p) model parameters are totally dependent on the first $p + 1$ given autocorrelation lags $r_x(0), r_x(1), \cdots, r_x(p)$. Although the associated model will have an autocorrelation behavior which perfectly matches these first $p + 1$ lags, it may provide a very poor representation for the remaining given autocorrelation lags $r_x(p + 1), r_x(p + 2), \cdots, r_x(s)$ (which

were not used in the parameter identification). In order to provide a representation for these higher lags by the procedure here taken, it may be necessary to increase the AR model order to s (i.e., $p = s$). In many applications, however, the underlying goal will be that of providing an AR model of relatively low order (i.e., $p \ll s$) which will adequately represent the entire set of autocorrelation lags. A procedure for achieving this objective will be shortly given. Before considering this most relevant objective, let us first outline an elegant method for solving the system of linear equations (2.11).

Levinson-Durbin Algorithm: Although the solution procedure as embodied in (2.14) will result in the desired parameter identification, the evaluation of R^{-1} will entail on the order of p^3 multiplication and addition calculations (i.e., $o(p^3)$) if standard procedures such as Gaussian elimination are used. Fortunately, it is possible to take advantage of the fact that the autocorrelation matrix R is both complex conjugate symmetric (i.e., $R(i,j) = \overline{R}(j,i)$) and Toeplitz (i.e., $R(i,j) = R(i+1, j+1)$) so as to effect a computationally efficient solution procedure. This method was developed by Levinson and is commonly referred to as the Levinson-Durbin algorithm [24], [43]. In this approach, one solves the linear system of equations (2.11) as the AR order parameter p is sequenced through the values $1, 2, 3, \cdots, p_m$ where p_m designates some as yet unknown maximum AR order. In this sequencing scheme, Levinson showed that the parameters for the kth-order AR model solution which are designated by

$$a_1^{(k)}, a_2^{(k)}, \cdots, a_k^{(k)}, b_0^{(k)} \tag{2.15}$$

are related to the $(k-1)$th-order AR model solution as outlined in (2.16a)–(2.17c).

Step 1

$$a_1^{(1)} = -r_x(1)/r_x(0) \tag{2.16a}$$

$$|b_0^{(1)}|^2 = [1 - |a_1^{(1)}|^2]\, r_x(0). \tag{2.16b}$$

Step 2: For $k = 2, 3, 4, \cdots$

$$a_k^{(k)} = -\left[r_x(k) + \sum_{m=1}^{k-1} a_m^{(k-1)} r_x(k-m)\right] \Big/ |b_0^{(k-1)}|^2 \tag{2.17a}$$

$$a_i^{(k)} = a_i^{(k-1)} + a_k^{(k)} a_{k-i}^{(k-1)}, \qquad 1 \leqslant i \leqslant k-1 \tag{2.17b}$$

$$|b_0^{(k)}|^2 = [1 - |a_k^{(k)}|^2]\, |b_0^{(k-1)}|^2. \tag{2.17c}$$

A brief description of this systematic algorithm will now be given.

If one were to solve the linear system of equations (2.11) for the order choice $p = 1$, it would be found that the required first-order AR parameters (with superscript (1) appended) are given in Step 1 (2.16a), (2.16b). Upon setting $p = 2$ in (2.11), a moderate amount of algebraic manipulation will reveal the validity of the solution as given in Step 2 (2.17a)–(2.17c) with $k = 2$ (with superscript (2) appended). Levinson proved that in following the systematic procedure of (2.16a)–(2.17c), the solutions to the Yule-Walker equations (2.11) for order selections $p = 1, 2, 3, \cdots$ are sequentially obtained. Moreover, the number of multiplication and addition computations required in generating the kth AR order parameters from the $(k-1)$th AR order parameters (i.e., Step 2) is seen to be k. Thus the computational complexity of the Levinson-Durbin algorithm for generating a pth-order AR model (and all lower order models as a byproduct) is found to be $o(p^2)$. This is a considerable savings over the computational complexity of $o(p^3)$ required in solving (2.11) using standard techniques.

The Levinson-Durbin algorithm provides not only a computationally efficient method for generating the AR parameters, but it also yields an effective AR model order determination procedure. Specifically, let it be assumed that the autocorrelation lags used in (2.11) correspond to an AR(p) process. If the Levinson-Durbin algorithm were applied to this autocorrelation lag information, by the very nature of this procedure, the AR process parameters would be perfectly identified at the pth iteration (i.e., $a_k^{(p)} = a_k$, $k = 1, 2, \cdots, p$ and $|b_0^{(p)}|^2 = |b_0^2|$). Moreover, if this recursion were continued beyond p, it would be found that $a_i^{(k)} = a_i$ for $1 \leqslant i \leqslant p$, $a_i^{(k)} = 0$ for $p+1 \leqslant i \leqslant k$, and $|b_0^{(k)}|^2 = |b_0|^2$. This is a direct consequence of the fact that $a_{p+1}^{(p+1)}$ must be zero as is evident from (2.17a). From these observations, it is therefore apparent that the nonchanging of the parameters $|b_0^{(k)}|^2$ provides a means for order determination.

When the autocorrelation lags used do not correspond to an AR process, there will be *no* value of k for which $|b_0^{(k)}|^2$ assumes a constant value thereafter. Since the specific high-order coefficients $a_k^{(k)}$ will always have a magnitude which never exceeds one ([5] and [14]), however, it is apparent from (2.17c) that $|b_0^{(k)}|^2 \leqslant |b_0^{(k-1)}|^2$ for all $k \geqslant 1$. Thus the parameters $|b_0^{(k)}|^2$ form a monotonically nonincreasing sequence and this factor can be used in model order determination. In particular, the parameter $|b_0^{(k)}|^2$ may be identified with a "prediction error" associated with a kth-order linear predictor. Once this prediction error becomes satisfactorily small, the associated AR(p) model will form an acceptably good *approximation* to the given autocorrelation sequence (e.g., see [31]). The meaning of "satisfactorily small" is subjective and will depend on the particular application being considered and empirically obtained experience.

The parameters $a_k^{(k)}$ for $k = 1, 2, 3, \cdots$ are also referred to as "reflection coefficients" and are often denoted by $c_k = a_k^{(k)}$. These reflection coefficients have the property that for the truncated sequence $r_x(0), r_x(1), \cdots, r_x(p)$ to be a valid segment of an autocorrelation sequence, it is necessary and sufficient that $|c_k| \leqslant 1$ for $k = 1, 2, \cdots, p$. Moreover, the transfer function

$$A_p(z) = \sum_{n=0}^{p} a_n z^{-n} \tag{2.18}$$

associated with the solution to (2.11) will have all of its roots on or inside the unit circle if and only if the $|c_k| \leqslant 1$ for $k = 1, 2, \cdots, p$.

It is noteworthy that the system of equations (2.11) also arises when solving the optimum one-step predictor problem or when using the maximum entropy principle [31]. In the one-step predictor problem, it is desired to select the p predictor parameters a_k so that the prediction

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k) \tag{2.19}$$

best approximates $x(n)$ in the sense of minimizing the mean-squared prediction error $E\{|x(n) - x(n)|^2\}$. One may readily show that the optimum prediction parameters are found by solving (2.11) in which $|b_0|^2$ plays the role of the minimum mean-squared prediction error. On the other hand, when applying the maximum entropy principle, it is tacitly assumed that the time series $\{x(n)\}$ is a zero-mean Gaussian process. The objective is to then find a power spectral density function

$S_x(e^{j\omega})$ which will maximize the entropy measure

$$\int_{-\pi}^{\pi} \log [S_x(e^{j\omega})]\, d\omega \tag{2.20}$$

subject to the constraint that this function will be consistent with the given set of $p+1$ autocorrelation lags $r_x(0)$, $r_x(1), \cdots, r_x(p)$ through the Fourier transform pair relationship (1.3). It is readily shown that the maximizing power spectral density function is an AR process of order p whose parameters are given by (2.11).

ARMA Time Series

The time series $\{x(n)\}$ is said to be an ARMA process of order (p, q) if it is generated (or can be modeled) according to the recursive relationship

$$x(n) + \sum_{k=1}^{p} a_k x(n-k) = \sum_{k=0}^{q} b_k \epsilon(n-k) \tag{2.21}$$

in which the excitation sequence $\{\epsilon(n)\}$ is the aforementioned normalized white-noise process. Our task is to then determine values for the a_k and b_k parameters of this model which are most compatible with the given autocorrelation lags (2.1). The mechanism for measuring this compatability will be the Yule-Walker equations (2.3) which characterize the above ARMA model. Upon examination of these equations, it is seen that the ARMA parameters appear in a nonlinear fashion through the unit-impulse response $h(n)$. If a best least squares modeling is desired, it is then found that the generation of the optimal a_k, b_k parameters involves the least mean-square solution of the highly nonlinear Yule-Walker equations. This solution will amost always necessitate the use of computationally burdensome nonlinear programming algorithms with the attendant difficulty of initial parameter value selection and the possibilities of convergence to a local extremum or even nonconvergence.

A considerable easing in computational requirements may be achieved if we allow ourselves the luxury of evaluating the a_k and b_k parameters separately. By using this approach, it will be possible to provide for a linear solution procedure for the a_k parameters. Although this approach will be suboptimal in nature, it often provides for a near-optimal modeling. The mechanism for this separate parameter evaluation is obtained upon examining the Yule-Walker equations (2.3) which characterizes the ARMA model (2.21). If this model is taken to be causal, it follows that the Yule-Walker equations assume a particularly simple form for indices $n > q$, that is,

$$\sum_{k=0}^{p} a_k r_x(n-k) = 0, \quad \text{for } n \geq q+1. \tag{2.22}$$

We shall refer to this particular subset of the Yule-Walker equations as the *extended Yule-Walker* equations. The obvious attractiveness of these equations lies in the fact that they are *linear* in the a_k parameters.

To determine the a_k autoregressive parameters which are most compatible with the given set of autocorrelation lags (2.1), we could adopt the approach that characterized many AR and ARMA modeling methods up to as recently as three years ago (e.g., see [26], [28], [35], [38]). This would entail evaluating the first p extended Yule-Walker equations (i.e., $q+1 \leq n \leq q+p$) and then solving the resultant system of p linear equations in the p AR parameters. Although this approach is computationally attractive, it suffers from the obvious drawback that only a subset of the given autocorrelation lags (2.1) are being used in fixing the a_k parameters (i.e., $r_x(n)$ for $q-p < n \leq q+p$). To achieve an ARMA model which better represents the entire set of autocorrelation lags (2.1), it is clearly beneficial to use more than the minimal number (i.e., p) of extended Yule-Walker equation evaluations. The a_k parameters which yield a least squares fit to this overdetermined set of linear equations is then found using a straightforward procedure to be shortly given.

This overdetermined extended Yule-Walker equation approach to ARMA spectral estimation was proposed by the author in 1979 [15]. From a historical perspective, it is to be noted that the idea of using an extended set of model evaluations forms a fundamental concept in system parameter estimation theory (e.g., see [45], [59]). Moreover, the approach here taken can be interpreted as being a generalized application of the Prony procedure in which the autocorrelation lags play the role of the data. With these thoughts in mind, there exists a rich source of evidence justifying the use of an overdetermined set of extended Yule-Walker equations for estimating the ARMA model's AR parameters.

In this overdetermined modeling approach, the extended Yule-Walker equations (2.22) are evaluated for t distinct values of n satisfying $n \geq q+1$. To effect the desired overdeterminacy, the integer t has to be selected to at least equal $p+1$ although larger values will typically yield better model representations. To illustrate this overdetermined approach, let us consider the first t extended Yule-Walker equations (2.22) indexed by $q+1 \leq n \leq q+t$. This particular Yule-Walker equation evaluation gives rise to the following overdetermined system of t linear equations in the p autoregressive parameter unknowns:[2]

$$\begin{bmatrix} r_x(q+1) & r_x(q) & \cdots & r_x(q-p+1) \\ r_x(q+2) & r_x(q+1) & \cdots & r_x(q-p+2) \\ \vdots & \vdots & \vdots\vdots & \vdots \\ r_x(q+t) & r_x(q+t-1) & \cdots & r_x(q-p+t) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.23a}$$

or more compactly as

$$R_1 \boldsymbol{a} = \boldsymbol{\theta}. \tag{2.23b}$$

In this latter expression, $\boldsymbol{\theta}$ denotes the $t \times 1$ zero vector, R_1 is the $t \times (p+1)$ ARMA autocorrelation matrix with Toeplitz-type structure having elements

$$R_1(i,j) = r_x(q+1+i-j), \quad 1 \leq i \leq t, \quad 1 \leq j \leq p+1 \tag{2.24}$$

and $\boldsymbol{a}$ is the $(p+1)$ AR parameter vector whose first component is required to be one

$$\boldsymbol{a} = [1, a_1, a_2, \cdots, a_p]'. \tag{2.25}$$

Examination of (2.23) reveals that the ARMA model's AR

[2] In certain applications, it may be desirable to use an other than contiguous set of extended Yule-Walker equation evaluations.

parameters are obtained upon solving this system of t overdetermined (assuming $t > p$) linear equations. Due to the overdetermined nature of these equation, the fundamental question as to whether a solution exists naturally arises. The following theorem provides an answer to this question and is a direct result of the Yule-Walker equations which governs ARMA processes.

Theorem 2.1: If the autocorrelation lag entries used in matrix R_1 of (2.23) correspond to those of an ARMA (p_1, q_1) process, then the rank of R_1 is p_1 provided that $p \geq p_1, q \geq q_1$.

With this theorem in mind, the existence of a solution to (2.23) will be dependent on the rank of the autocorrelation matrix R_1. We shall now consider separately the cases in which R_1 has full rank and less than full rank.

Rank $[R_1] \leq p$: When matrix R_1 has less than full rank, a nontrivial autoregressive parametric vector solution $\boldsymbol{a}$ will be assured. An interesting algebraic characterization of this solution may be obtained upon premultiplying both sides of (2.23) by the complex conjugate transpose of R_1 as denoted by R_1^* to yield

$$R_1^* R_1 \boldsymbol{a} = \boldsymbol{\theta}. \tag{2.26}$$

Upon examination of this expression, it is clear that the required AR parameter vector may be also identified with a properly normalized eigenvector (i.e., its first component is one) associated with a zero eigenvalue of the $(p+1) \times (p+1)$ matrix $R_1^* R_1$. As such, we may then use standard eigenvector-eigenvalue routines when finding the required ARMA model AR parameters.

Rank $[R_1] = p + 1$: In many cases of interest, however, it will be found that the autocorrelation matrix R_1 will have full rank. This will occur whenever the autocorrelation lag entries used are associated with either a nonrational random time series, an MA process, or a higher order ARMA rational process. Since R_1 has full rank, there then will not exist a nontrivial solution to (2.23). Nonetheless, we still wish to determine an ARMA model which "best fits" these overdetermined extended Yule-Walker equations. Namely, we seek a nonzero AR parameter vector $\boldsymbol{a}$ so that $R_1 \boldsymbol{a}$ most closely equals the required ideal zero vector as specified in (2.23). Although a variety of procedures may be used for accomplishing this selection, the following two approaches typify many spectral estimation algorithms.

i) In the first selection procedure, it is desired to find an AR parameter vector laying on the unit hypersphere which will minimize the Euclidean norm of $R_1 \boldsymbol{a}$. This entails solving the following constrained optimization problem:

$$\min \boldsymbol{a}^* R_1^* R_1 \boldsymbol{a}$$
$$\boldsymbol{a}^* \boldsymbol{a} = 1.$$

Using standard Lagrange multiplier concepts, it is readily shown that the solution to this optimization problem is obtained by selecting that orthonormal eigenvector of the positive definite Hermetian matrix $R_1^* R_1$ associated with its minimum eigenvalue. If x_1 corresponds to that orthonormal eigenvector (i.e., $R_1^* R_1 x_k = \lambda_k x_k$ with $\lambda_k \leq \lambda_{k+1}$ and $x_k^* x_k = 1$, for $k = 1, 2, \cdots, p+1$), the required AR parameter vector with first component of one is obtained by the normalization

$$\boldsymbol{a}^0 = \frac{1}{x_1(1)} x_1 \tag{2.27}$$

where $x_1(1)$ denotes the first component of x_1. This AR parameter vector selection procedure characterizes many spectral algorithms which are variants of the Pisarenko method [55] and is generally not suitable for an efficient computational solution.

ii) In the second selection procedure, we wish to minimize the Euclidean norm of $R_1 \boldsymbol{a}$ over all $(p+1) \times 1$ vectors $\boldsymbol{a}$ with first components equal to one, that is,

$$\min \boldsymbol{a}^* R_1^* R_1 \boldsymbol{a}$$
$$a(1) = 1.$$

Again appealing to the Lagrange multiplier approach, it is found that the solution to this constrained optimization problem is given by solving the following linear system of equations:

$$R_1^* R_1 \boldsymbol{a}^0 = \alpha \boldsymbol{e}_1 \tag{2.28}$$

where the normalizing constant α is selected so that the first component of $\boldsymbol{a}^0$ is one.

In using either of the above two procedures, we are seeking to best satisfy theoretical relationship (2.23) in the least squares sense subject to appropriate constraints.[3] The particular application at hand dictates which AR parameter vector selection procedure provides the best performance. It has been the author's experience that the selection (2.27) has often provided reasonable modeling (also see [10], [12]). In terms of computational efficiency, however, the linear selection (2.28) enjoys a clear superiority due to the availability of efficient adaptive algorithms as outlined in Section X. With this in mind, we shall mainly focus our attention on the linear selection (2.28).

In summary, the ARMA(p, q) model associated with a given set of autocorrelation lags entails an examination of the matrix R_1. If this matrix is not of full rank, the required exact AR parameter vector will be given by solving (2.26). On the other hand, when the matrix has full rank, an appropriate AR parameter vector may be achieved by solving either (2.27) or (2.28). It is important to appreciate the fact that these ARMA results are applicable to the special AR process in which case we simply enter $q = 0$ when forming the ARMA autocorrelation matrix R_1.

Moving Average Parameters

In order to complete the ARMA modeling, it is necessary to determine the model's associated moving average parameters. There are a variety of procedures for achieving this objective. We shall present two such procedures of which the first is the one most often found in the literature while the second possesses a desirable efficient computational implementation.

i) In the first procedure, one conceptually applies the time series $\{x(n)\}$ to the pth-order nonrecursive filter with transfer function $A_p(z)$ whose coefficients correspond to the AR parameters obtained upon solving either (2.26), (2.27), or (2.28). This filtering produces the so-called *residual time* series as specified by

$$s(n) = \sum_{m=0}^{p} a_m x(n-m). \tag{2.29}$$

This filtering causes the residual time series to be a moving

[3] It is possible to generalize the constraints to be a quadratic surface (giving rise to a generalized eigenvector solution) or a hyperplane, respectively [10].

average process of order q with power spectral density function $|B_q(e^{j\omega})|^2$ as is made evident from Fig. 2. This, of course, presumes that $\{x(n)\}$ corresponds to an ARMA processor of order (p, q) or less. A simple analysis indicates that the length $2q+1$ autocorrelation sequence of this residual time series may be computed according to

$$r_s(n) = \begin{cases} \sum_{k=0}^{p} \sum_{m=0}^{p} a_k \bar{a}_m r_x(n+m-k), & -q \leq n \leq q \\ 0, & \text{otherwise.} \end{cases} \quad (2.30)$$

Using these MA(q) autocorrelation lags, it follows from (2.5) that the unknown b_k parameters must be such that

$$r_s(n) = \sum_{k=0}^{q} b_k \bar{b}_{k-n}, \quad -q \leq n \leq q. \quad (2.31)$$

A spectral factorization along the lines mentioned in this section's MA time series subsection will then yield the desired b_k parameters.

ii) If computational requirements are of vital concern, the technique to be now outlined is particularly efficient [15], [16]. It utilizes the Fourier transform of the causal part of the autocorrelation sequence

$$D(e^{j\omega}) = \sum_{n=1}^{\infty} r_x(n)\, e^{-j\omega n}. \quad (2.32)$$

The underlying power spectral density function may be directly determined from this Fourier transform according to

$$S_x(e^{j\omega}) = r_x(0) + 2 \text{ Re } \{D(e^{j\omega})\}. \quad (2.33)$$

A comparison of this expression with (1.13) reveals that the transform $D(e^{j\omega})$ must be of the form

$$D(e^{j\omega}) = \frac{c_1 e^{-j\omega} + c_2 e^{-j2\omega} + \cdots + c_p e^{-jp\omega}}{1 + a_1 e^{-j\omega} + \cdots + a_p e^{-jp\omega}} = \frac{C(e^{j\omega})}{A_p(e^{j\omega})} \quad (2.34)$$

where we are tacitly assuming that the moving average order is not larger than the AR order (i.e., $q \leq p$).

To determine the required c_n coefficients in (2.34), we will first compute the first s impulse response elements of the filter $H(e^{j\omega}) = 1/A_p(e^{j\omega})$. This will entail using the following relationship:

$$h(n) = -\sum_{k=1}^{p} a_k h(n-k), \quad 1 \leq n \leq s \quad (2.35)$$

in which $h(0) = 1$ and $h(n) = 0$ for $n < 0$ are used to initiate the recursion. We next use the time-domain equivalency of (2.34) to conclude that

$$\begin{bmatrix} h(0) & 0 & 0 & \cdots & 0 \\ h(1) & h(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \\ h(p-1) & h(p-2) & \cdots & & h(0) \\ \vdots & \vdots & & & \vdots \\ h(s-1) & h(s-2) & \cdots & & h(s-p) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(s) \end{bmatrix} \quad (2.36a)$$

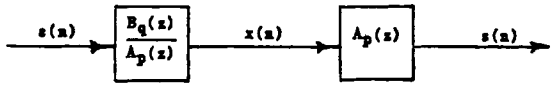

Fig. 2. Generation of residual time series.

or

$$Hc = r. \quad (2.36b)$$

In general, the overdetermined system of (2.36) will not have a solution unless the autocorrelation elements $r_x(n)$ are associated with an ARMA process of order (p, p) or lower. Assuming this not to be the case, we could select the vector c so as to provide a least squares solution to (2.36). This would take the form of solving the consistent system of linear equations

$$c = [H^*H]^{-1}H^*r. \quad (2.37)$$

In order to achieve the aforementioned efficient computational algorithm, the parameter s will be taken to be p which renders the following straightforward method for evaluating the c_n:

$$c_n = \sum_{k=0}^{n-1} a_k r_x(n-k), \quad 1 \leq n \leq p. \quad (2.38)$$

This is basically the approach taken in [15] and [16]. In using (2.38) for evaluating the c_n, we are trading off performance for computational efficiency. It has been the author's experience that the spectral estimates achieved upon using the least squares fit (2.37) do not typically provide a superior performance to those given by the simpler relationship (2.38). In any case, once the c_n parameters have been determined, the Fourier transform (2.34) is used in (2.33) to effect the required power spectral density model. Moreover, if it is desired to evaluate the b_k parameters, we can use the identity

$$|B_q(e^{j\omega})|^2 = A_p(e^{j\omega})\,\overline{C}(e^{j\omega}) + \overline{A}_p(e^{j\omega})\,C(e^{j\omega}) + r_x(0)\,|A_p(e^{j\omega})|^2 \quad (2.39)$$

and a spectral factorization to achieve this objective.

In this section, we have outlined convenient procedures for generating MA, AR, and ARMA spectral models when perfect autocorrelation lag information is available. The principle steps of these procedures are summarized in Table I. Although these results are of primarily theoretical interest, we will subsequently adapt them to evolve effective rational spectral estimation methods for the more practical case where only raw time series observations are used in the modeling.

III. Sinusoids in White-Noise Example

The procedure as developed in the preceding section is applicable to the task of generating rational models for the general class of wide-sense stationary time series. In order to demonstrate the relative effectiveness of MA, AR, and ARMA modeling, the classical problem of the detection and frequency identification of the sinusoids in white noise case will now be considered. Although this does represent a restricted application of rational spectral estimation, it provides a meaningful basis for measuring the relative performance capabilities of MA, AR, and ARMA models. In particular, the time series being now examined is taken to be composed of the sum of m real sinusoids in additive noise as specified by

TABLE I
RATIONAL SPECTRAL MODEL TECHNIQUES EMPLOYING EXACT AUTOCORRELATION LAG INFORMATION

MA Model

$$S_x(e^{j\omega}) = \sum_{n=-q}^{q} w(n)\, r_x(n)\, e^{-j\omega n} \qquad (1.6)$$

AR Model

(i) Form the $(p+1)\times(p+1)$ AR autocorrelation matrix R using expression (2.12)

(ii) Solve $R\underline{a} = |b_0|^2 \underline{e}_1$ (2.11)

where parameter b_0 is selected so that the first component of $\underline{a}$ is one.

(iii) $$S_x(e^{j\omega}) = \left| \frac{b_0}{1 + a_1 e^{-j\omega} + \ldots + a_p e^{-jp\omega}} \right|^2$$

ARMA Model

(i) Form the $(q+1)\times(p+1)$ ARMA autocorrelation matrix R_1 using expression (2.24)

(ii) (a) If Rank $(R_1^* R_1) < p+1$ then solve

$$R_1^* R_1 \underline{a} = \underline{o} \qquad (2.23)$$

(b) If Rank $(R_1^* R_1) = p+1$ then either solve

$$R_1^* R_1 \underline{a} = \alpha \underline{e}_1 \qquad (2.28)$$

where α is selected so that first component of $\underline{a}$ is one.

or

use the minimum eigenvalue-eigenvector yielding selection (2.27).

(iii) $$r_s(n) = \sum_{k=0}^{p} \sum_{m=0}^{p} a_k \bar{a}_m\, r_x(n+m-k) \qquad -q \le n \le q \qquad (2.32)$$

(iv) $$S_x(e^{j\omega}) = \frac{\sum_{n=-q}^{q} r_s(n) e^{-j\omega n}}{\left| 1 + a_1 e^{-j\omega} + \ldots + a_p e^{-jp\omega} \right|^2}$$

$$x(n) = \sum_{k=1}^{m} A_k \sin\left[2\pi f_k n + \theta_k\right] + w(n) \qquad (3.1)$$

in which the θ_k are independent, uniformly distributed random variables on the interval $[-\pi, \pi]$ and $w(n)$ is a zero-mean variance σ^2 white-noise process. It is recalled that the problem of detecting sinusoids in noise originally gave rise to spectral estimation theory. The periodogram method was developed for this very purpose by Schuster in 1898 [58].

The task at hand is to generate MA, AR, and ARMA models from the autocorrelation lags associated with this time series using the procedures outlined in the previous section. It is a simple matter to show that the autocorrelation sequence characterizing time series (3.1) is given by

$$r_x(n) = \sum_{k=1}^{m} 0.5A_k^2 \cos\left[2\pi f_k n\right] + \sigma^2 \delta(n) \qquad (3.2)$$

in which $\delta(n)$ denotes the unit-impulse (Kronecker-delta) sequence. The power spectral density function associated with this process is composed of $2m$ Dirac-delta impulses of amplitudes $0.5A_k^2$ located at frequencies $\pm f_k$ riding on top of a constant value σ^2. As such, this discontinuous power spectral density function may not be associated with a finite-order MA, AR, or ARMA process.

Although the autocorrelation sequence (3.2) is not compatable with a finite-order ARMA model, it is readily shown that this sequence will satisfy the following homogeneous relationships:

$$\sum_{k=0}^{2m} a_k r_x(n-k) = 0, \qquad \text{for } n > 2m \qquad (3.3)$$

where $a_0 = 1$. The a_k parameters required in this expression are obtained by equating coefficients of the following polynomial equivalency:

$$A_{2m}(z) = \sum_{n=0}^{2m} a_n z^{-n} = \prod_{k=1}^{m} \left[1 - 2z^{-1}\cos(2\pi f_k) + z^{-2}\right] \qquad (3.4)$$

where the zeroes of this polynomial (i.e., $e^{\pm j2\pi f_k}$) are identified with the frequencies of the time series' sinusoids (e.g., see [10], [32], [55]).

Upon comparison of (3.3) and (2.22), it might be incorrectly inferred that the autocorrelation sequence (3.2) would be associated with an ARMA process of order $(2m, 2m)$. Upon examination of the Yule–Walker equations for indicies $0 \leqslant n \leqslant 2m$, however, it will be found that an exact correspondence does not result. This simply reflects the fact that the time series (3.1) does not arise from exciting a linear ARMA operator with white noise. Nonetheless, due to the identical forms of (2.22) and (3.3), we may still use the ARMA modeling parameter procedure as outlined in Section II to identify the $2m$ parameters a_k. These parameters would be then, in turn, inserted into (3.4) to identify the frequency parameter f_k upon factorization of the polynomial $A_{2m}(z)$. This spectral behavior can be conveniently displayed in a plot of $|1/A_{2m}(e^{j\omega})|$ versus ω.

Once the f_k frequency parameters have been determined, the associated A_k amplitude parameters may be obtained upon evaluating (3.2) over any set of m or more indices satisfying $n \geqslant 1$. With this in mind, let us evaluate this expression for the contiguous indices $1 \leqslant n \leqslant v$ where the integer $v \geqslant m$. This is found to yield the following overdetermined (if $v > m$) system of consistent linear equations in the A_k unknowns:

$$\begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(v) \end{bmatrix} = \begin{bmatrix} \cos(2\pi f_1) & \cos(2\pi f_2) & \cdots \cos(2\pi f_m) \\ \cos(4\pi f_1) & \cos(4\pi f_2) & \cdots \cos(4\pi f_m) \\ \vdots & \vdots & \vdots \\ \cos(2v\pi f_1) & \cos(2v\pi f_2) & \cdots \cos(2\pi v f_m) \end{bmatrix} \begin{bmatrix} 0.5A_1^2 \\ 0.5A_2^2 \\ \vdots \\ 0.5A_m^2 \end{bmatrix} \qquad (3.5)$$

or equivalently as

$$r = C_p \qquad (3.6)$$

where p is the so-called $m \times 1$ power vector with elements $0.5A_k^2$. If the integer parameter v is selected to be larger than or equal to m, the least squares approximate solution to the overdetermined equations (3.6) is given by

$$p = [C'C]^{-1} C' r \qquad (3.7)$$

where C' designates the transpose of matrix C. In the case of perfect autocorrelation knowledge, we normally set $v = m$ thereby giving the solution $p = C^{-1}r$. In the more practical case in which only raw time series observations are given for the estimate, however, a desirable degree of parameter smoothing is achieved by selecting $v > m$.

Although the time series (3.1) is not compatible with an AR model, such models have also been successfully employed in analyzing such time series. Depending on the underlying signal-to-noise ratios (SNR's)

$$\frac{A_k^2}{2\sigma^2}, \qquad 1 \leqslant k \leqslant m$$

the desired detection and frequency estimation will require that the AR-order parameter p be made *significantly* larger than $2m$. Variants of the Pisarenko method [55] and the SVD approach of Tufts and Kumaresan [42], [61] typically produce satisfactory performance on the sinusoids in the white-noise case. As we will illustrate in Section VIII, the approach taken in this paper will produce a superior performance when an SVD adaption of the ARMA modeling method herein presented is made.

Alternate Method

It is possible to apply the concept of using an overdetermined system of model equation evaluations for achieving high-quality alternative estimates for the frequency parameters appearing in (3.1). This will make use of the observation that homogeneous relationship (3.3) holds for all values of n provided that there is no white noise present (i.e., $\sigma^2 = 0$). Under this restriction, an evaluation of (3.2) with $\sigma^2 = 0$ over the indices $-t + p \leqslant n \leqslant t$ (in which $p = 2m$) is found to result in the following symmetrical relationship:

$$\begin{bmatrix} r_x(-t+p) & r_x(-t+p-1) & \cdots & r_x(-t) \\ r_x(-t+p+1) & r_x(-t+p) & \cdots & r_x(-t+1) \\ \vdots & \vdots & & \vdots \\ r_x(t) & r_x(t-1) & \cdots & r_x(t-p) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.8a}$$

or

$$R_s a = \theta \tag{3.8b}$$

in which t is selected so that $t > 3m/2$ thereby ensuring an overdetermined system of homogeneous relationships.

If the autocorrelation lag entries used in (3.8) correspond to (3.2) with $\sigma^2 = 0$, it then follows that the overdetermined system of (3.8) will have a unique solution for the a_k coefficients. This solution can then be incorporated into (3.4) to obtain estimates for the frequency f_k parameters. In the additive noise case $\sigma^2 \neq 0$, however, this system of equations will generally not have a solution. Since the σ^2 term appears in only $p + 1$ out of the $(2t - p + 1) \times (p + 1)$ entries of matrix R_s (i.e., the $r_x(0)$ entries), it can be argued that so long as $t \gg p$, the effect of the additive noise will be minimal. Based on this premise, it is natural to then seek a vector a such that this inconsistent system of linear equations is best satisfied in a least squares sense. The required least squares solution is then given by solving the system of equations

$$R_s^* W R_s a = \alpha e_1 \tag{3.9}$$

in which α is a normalizing scalar selected to ensure that the first component of a is one. The nonnegative definite diagonal matrix W is typically selected to be equal to the identity matrix. As we will see in Section VIII, the solutions obtained by using (3.9) often provide exceptional estimates so long as $t \gg p$. A paper in preparation will further refine this new approach.

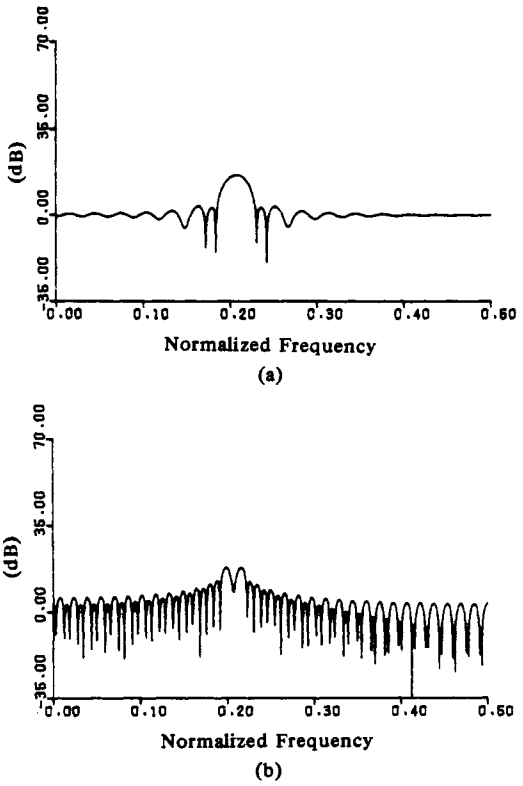

Fig. 3. Moving average (MA) spectral models using (1.6) with $w(n) = 1$ and exact autocorrelation lags. (a) MA(32); $q = 32$. (b) MA(64); $q = 64$.

Numerical Example

In order to illustrate the effectiveness of rational models in resolving sinusoids embedded in white noise, we shall now consider the specific time series

$$x(n) = \sin(0.4\pi n) + \sin(0.43\pi n) + w(n). \tag{3.10}$$

The white-noise series $\{w(n)\}$ will be taken to have a variance of 0.5 thereby creating a 0-dB SNR environment. According to (3.2), the autocorrelation sequence associated with this time series is specified by

$$r_x(n) = 0.5\cos(0.4\pi n) + 0.5\cos(0.43\pi n) + 0.5\delta(n). \tag{3.11}$$

We shall now use these autocorrelation lags along with the concepts developed in Section II to generate appropriate MA, AR, and ARMA models. A brief discussion of the resultant modeling performances in this idealistic situation will now be given.

MA Models: When using the classical spectral modeling expression

$$S_x(e^{j\omega}) = \sum_{n=-q}^{q} r_x(n)\, e^{-j\omega n} \tag{3.12}$$

we are in effect invoking an MA(q) model. Plots of this expression with entries (3.11) for model order selections of $q = 32$ and $q = 64$ are shown in Fig. 3 over the range of normalized

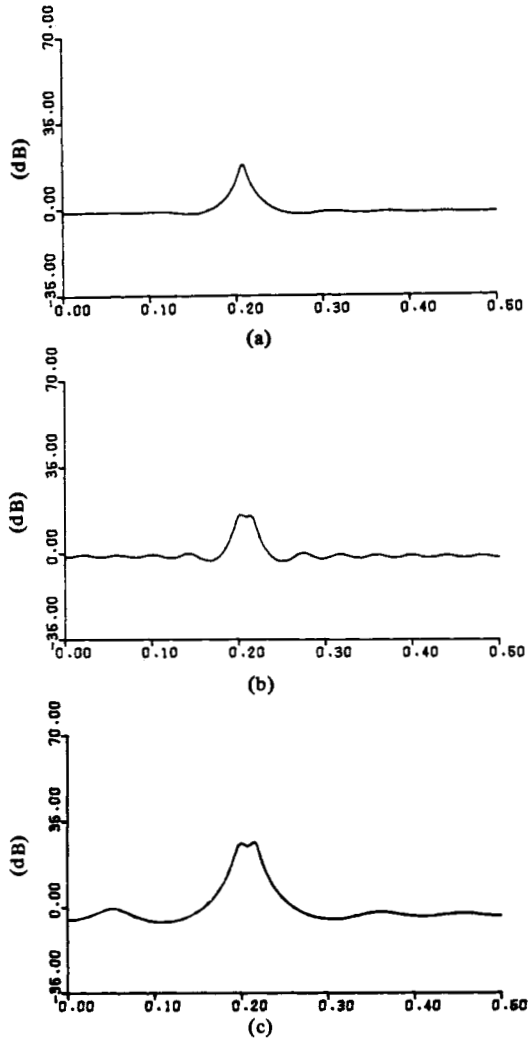

Fig. 4. Autoregressive (AR) spectral models using (2.11) with exact autocorrelation lags. (a) AR(20); $p = t = 20$. (b) AR(24); $p = t = 24$. (c) AR(10); $p = 10$, $t = 100$.

frequencies $0 \leqslant f \leqslant 0.5$. From these results, it is apparent that a resolution of the two equal amplitude sinusoids was not achieved for a 32nd-order MA model, but was achieved for a 64th-order MA model. Thus an artificially high-order MA model was required in order to resolve the two sinusoids when exact autocorrelation lags were used. This example nicely demonstrates the distortions which can result when invoking an MA model if the underlying assumption that $r_x(n) = 0$ for $n > q$ thereby implied is not satisfied (or approximately satisfied). Clearly, the nondamped nature of the autocorrelation sequence (3.2) behavior indicates that the MA modeling of a time series composed of sinusoids in white noise can be inappropriate unless a sufficiently large selection of the MA model order q is made.

AR Models: We next used the same autocorrelation lag information (3.11) to generate AR models of order $p = 20$ and $p = 24$ using (2.11). The resultant spectral estimates $1/|A_p(e^{j\omega})|^2$ are shown in Fig. 4(a) and (b) for these two model order choices. It is apparent that the twentieth-order model was unable to resolve the two sinusoids while the twenty-fourth was just able to achieve the resolution. Since the specific autocorrelation lags $r_x(n)$ for $0 \leqslant n \leqslant p$ were required for generating an AR(p) model, it is apparent that fewer autocorrelation lags were needed to resolve the two sinusoids when using an AR(24) model in comparison to the MA model. This simply gives credance to the previously made suggestion that AR models provide a more effective instrument for representing peak-like spectra than are MA models.

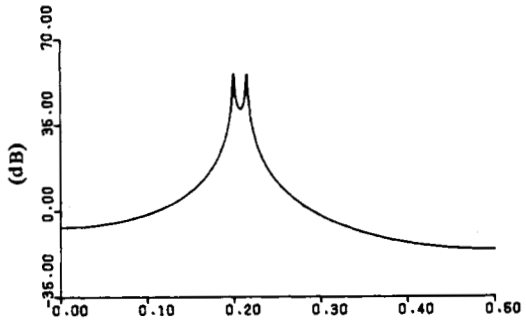

Fig. 5. Autoregressive-moving average (ARMA) spectral model using (2.23) with exact autocorrelation lags and $p = t = 4$.

In order to illustrate the effect of using more than the minimal number of extened Yule–Walker equations (i.e., $t > p$) when generating an AR model, we next used the ARMA modeling equations (2.23) with parameters $p = 10, q = 0$, and $t = 100$. The AR(10) model which results upon solving (2.28) for this choice of order parameters has a spectral behavior as depicted in Fig. 4(c). This AR(10) spectral estimate is seen to be significantly better than that achieved by the higher order AR(24) estimate. Clearly, the process of using 100 (i.e., $t = 100$) extended Yule–Walker equation evaluations instead of the minimal number 10 has produced this significant improvement. This improvement is due to the fact that only the first four of the one hundred extended Yule–Walker equation evaluations are in error due to the imposition of an improper AR model (see (3.3)). By increasing t beyond p, the effect has been to dilute the negative impact of the erroneous first four Yule–Walker equations on the model parameters (i.e., four improper equations and 96 appropriate equations). The reader is urged to fully understand the implications of this result in a more broadly based context.

ARMA Model: We next used the given autocorrelation lag information (3.11) to generate an ARMA model of order $p = 4$ by appealing to (2.23). We here select the variable t to be equal to its minimal value of four and in accordance with this section's discussion take $q = 4$. The resultant ARMA-based spectral $1/|A_4(e^{j\omega})|^2$ without the MA component is plotted in Fig. 5. The two sinusoids are nicely resolved and when the fourth-order polynomial $A_4(e^{j\omega})$ was factored, it was found to have its four roots on the unit circle at $e^{\pm j2\pi f_k}$ for $k = 1,2$ in which $f_1 = 0.2$ and $f_2 = 0.215$. This should not be surprising since it was previously shown in this section that an ARMA-*type* model is perfectly compatible with a sinusoid in white-noise time series (MA and AR models are not compatible). It is noteworthy that only the autocorrelation lags $r_x(n)$ for $1 \leqslant n \leqslant 8$ were required in generating the spectral model depicted in Fig. 5.

Alternative Method: As a final procedure, we used the alternative method as represented by (3.9) in which the parameters were taken to be $p = 10$ and $t = 50$. Using these parameters along with the theoretical autocorrelation lag entries (3.11), a plot of the resultant estimate $1/|A_{10}(e^{j\omega})|^2$ is shown in Fig. 6. The two sinusoids are resolved with well-defined peaks, and the spectral estimates are superior to those achieved by the MA and AR model results but inferior to the ARMA model.

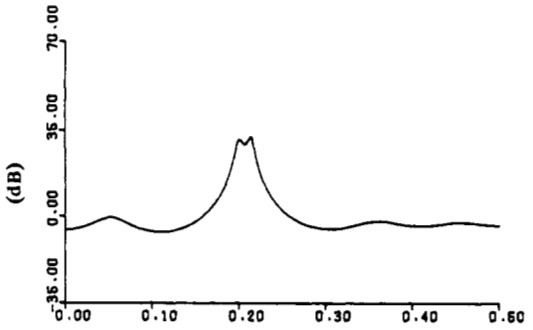

Fig. 6. Alternative method with $p = 10$ and $t = 50$.

IV. MA Modeling—Time Series Observations

From a practical viewpoint, the situation in which exact autocorrelation lag values are given for effecting a spectral estimate almost never arises. More typically, the required spectral estimate is to be generated from a finite set of contiguous time series observations as represented by

$$x(1), x(2), \cdots, x(N). \tag{4.1}$$

In this section, we will be concerned with achieving MA spectral estimates from this observation set. The methods to be presented for this purpose are largely influenced by the theoretical developments found in Section II.

There exist two primary MA spectral estimation procedures that have found favor among users. They are indirect methods based on autocorrelation estimates such as proposed by Blackman and Tukey [8], and direct methods based on the Fourier transform of the time series observations as represented by the periodogram (or the method of averaged periodograms due to Welch [64]). As we will shortly see, the periodogram is a special case of the Blackman–Tukey approach.

Blackman–Tukey Approach

In the Blackman–Tukey method, one first obtains autocorrelation estimates $\hat{r}_x(n)$ from the given observation set (4.1). These autocorrelation estimates are then inserted into (1.2) to effect the required spectral estimate. For a variety of reasons, it is often beneficial to introduce a windowing sequence $w(n)$ to achieve the windowed MA spectral estimate of order q

$$\hat{S}(e^{j\omega}) = \sum_{n=-q}^{q} w(n)\hat{r}_x(n)\, e^{-j\omega n}. \tag{4.2}$$

Considerations to be made in selecting the window sequence are well documented and the reader is referred to [33], [50], [57]. Two of the more popular selections are the rectangular window (i.e., $w(n) = 1$) and the Bartlett triangle window (i.e., $w(n) = (1 - |n|)/(q + 1)$).

The standard unbiased and biased autocorrelation estimates are among the most popular candidates to be used in the spectral estimate (4.2) (e.g., see [33] for a detailed development). The unbiased estimate achieves the required autocorrelation lag estimate according to

$$\hat{r}_x(n) = \frac{1}{N - |n|} \sum_{k=1}^{N} x(k+n)\,\bar{x}(k), \qquad -q \leqslant n \leqslant q \tag{4.3}$$

where the convention of setting to zero any term $x(n)$ in the summand for which $n \notin [1, N]$ is adopted. It is a simple matter to show that $E\{\hat{r}_x(n)\} = r_x(n)$ thereby establishing the unbiased nature of estimate (4.3). Moreover, this unbiased estimate is also consistent so long as the order parameter q is finite.

Notwithstanding the obviously attractive statistical properties possessed by the unbiased estimate (4.3), a number of prominant statisticians have proposed using the standard biased estimate (e.g., see [33], [52], [53])

$$\hat{r}_x(n) = \frac{1}{N} \sum_{k=1}^{N} x(k+n)\,\bar{x}(k), \qquad -q \leqslant n \leqslant q. \tag{4.4}$$

We again adhere to the convention of setting to zero any term $x(n)$ in the summand for which $n \notin [1, N]$. The justification for using the biased estimate is that it is more stable statistically. It must be noted, however, that the relative advantages of unbiased versus biased estimators remains an unsettled issue. With this in mind, the user is cautioned to base his ultimate selection on the particular application being considered. This will undoubtably entail a great deal of empirically based experimentation on the users part.

Periodogram

In the periodogram method, the required spectral estimate is given by the expression

$$\hat{S}_x(e^{j\omega}) = \frac{1}{N} |X_N(e^{j\omega})|^2 \tag{4.5}$$

where $X_N(e^{j\omega})$ is the Fourier transform of the time series observations, that is,

$$X_N(e^{j\omega}) = \sum_{n=0}^{N-1} x(n+1)\, e^{-j\omega n}. \tag{4.6}$$

We here use the subscript N on $X_N(e^{j\omega})$ to explicitly denote its dependency on the observation length parameter. It is readily shown that the periodogram is identical to the Blackman–Tukey approach when the biased estimates (4.4) are used in (4.2) with $q = N - 1$ and $w(n) = 1$.

The primary advantage in using the periodogram approach is computational in nature. Specifically, the values of the periodogram at the N discrete set of uniformly spaced radian frequencies $\omega_k = 2\pi k/N$ for $0 \leqslant k \leqslant N - 1$ is seen to entail evaluation of the entities

$$X_N(e^{j(2\pi k/N)}) = \sum_{n=0}^{N-1} x(n+1)\, e^{-j(2\pi kn/N)}, \qquad 0 \leqslant k \leqslant N - 1. \tag{4.7}$$

These evaluations are readily carried out by use of the N-point fast Fourier transform (FFT) algorithm (e.g., see [50], [57]). With the FFT algorithm, the N quantities (4.7) may be computed in which the required number of complex additions and multiplications is on the order of $N \log_2 N$. The computational savings accrued in using the FFT algorithm for spectral estimates is considerable when it is realized that a direct evaluation of (4.7) is seen to entail N^2 complex additions and multiplications. Due to the computational savings accrued in using the FFT implementation of the periodogram, spectral estimates of long data sequences became feasible with the FFT's development.

Although the FFT algorithm offers a computationally efficient means for numerically evaluating the periodogram (4.5), it possesses a potentially serious drawback. Specifically, as just suggested, this FFT implementation provides a sampled

version of the periodogram in which the frequency samples are separated by $2\pi/N$ radians. For many applications of interest, this sampling may be too coarse in that the detailed continuous frequency behavior of the periodogram (4.5) may be somewhat obscured through the sampling process. An example of this will be given in Section VIII. In order to alleviate this potential difficulty, we may apply the concept of *zero padding*. This simply entails the appending of L zeroes to the given set of time series observations, that is,

$$x(1), x(2), \cdots, x(N), \underbrace{0, 0, \cdots, 0}_{L \text{ zeroes}} \tag{4.8}$$

where L is a yet unspecified positive integer. If we were to take the Fourier transform of this padded time series, we would obtain the same transform (4.6) and the same periodogram function (4.5). On the other hand, if we were to take an $N+L$ point FFT of this padded time series, the following more finely spaced samples of the Fourier transform would be generated:

$$X_N(e^{j(2\pi k/N+L)}) = \sum_{n=0}^{N-1} x(n+1)\, e^{-j(2\pi kn/N+L)}, \qquad 0 \leqslant k < N+L. \tag{4.9}$$

If these sampled values were then substituted into (4.5), we would obtain sampled values of the periodogram at the more finely spaced frequencies $\omega_k = 2\pi/(N+L)$ for $0 \leqslant k < N+L$. The effect of the L zero padding is then seen to result in a reduction of the frequency sampling interval from $2\pi/N$ to $2\pi/(N+L)$. By selecting L suitably large, we can reduce this sampling interval to any degree desirable.

One should not gain the mistaken impression that padding will enable us to achieve any degree of frequency resolution desired. The fundamental unsampled periodogram (4.5) has an *inherent frequency resolution capability* of $\Delta\omega = 2\pi/N$ (or equivalently $\Delta f = 1/N$). When using an N-point FFT implementation of the periodogram, however, it is entirely possible that spectral peaks may lie between the sampled frequencies $\omega_k = 2\pi k/N$. In such cases, the peaks' effect on the sampled periodogram may be seriously diluted even though it would be clearly evident in the unsampled periodogram. Upon padding with L zeroes, we can remove the ambiguity caused by this sampling process and still retain the computational efficiency of an FFT implementation.

V. AR Modeling—Time Series Observations

The task of generating AR spectral models from a set of time series observations has been of primary concern to many investigators over the last few years. Undoubtably, the most widely used AR modeling procedure is the Burg algorithm as first proposed in 1967 [13]. This algorithm not only provided a spectral estimation capability that was theretofore lacking, it also inspired an intense search for improved rational spectral estimation procedures. Much of contemporary spectral estimation theory has been directly influenced by the philosophy contained within the Burg approach. As a matter of fact, many of the more recent rational estimation procedures were developed so as to overcome some of the deficiencies observed in the Burg algorithm as typified by line splitting and biased frequency estimates. Nonetheless, the Burg algorithm still occupies the preeminent position among contemporary AR modeling methods. Since its operational behavior is so well documented, we refer the interested reader to the relevant literature for its detailed development (e.g., see [23], [31]).

TABLE II
Basic Steps in Obtaining an AR (p) Spectral Estimate

Step 1:	Compute Estimates of $R(i,j) = r_x(i-j)$ for $1 \leqslant i,j \leqslant p+1$ to form the $(p+1)\times(p+1)$ autocorrelation matrix estimate $\hat{R}$.
Step 2:	Solve the linear system of equations $\hat{R}\underline{a} = \lvert b_0\rvert^2 \underline{e}_1$ in which the normalizing coefficient b_0 is selected so that the first component of $\underline{a}$ is one.
Step 3:	The required AR(p) spectral estimate is then specified by $\hat{S}_{AR}(e^{j\omega}) = \left\lvert \frac{b_0}{1 + a_1 e^{-j\omega} + \cdots + a_p e^{-jp\omega}} \right\rvert^2$

In this section and in Section IX, we will demonstrate that many of the popularly used AR methods (which includes the Burg algorithm) may be interpreted as providing statistical estimates of the fundamental Yule–Walker equations (2.11) that govern AR processes. These estimates are to be obtained from the set of contiguous time series observations

$$x(1), x(2), \cdots, x(N) \tag{5.1}$$

which are made available through some measurement mechanism. More specifically, it is well known that various contemporary methods either explicitly or implicitly use these observations to generate estimates of the $(p+1)\times(p+1)$ autocorrelation matrix R which appears in the fundamental relationship (2.11). Clearly, the elements of the matrix estimate $\hat{R}$ must be such that

$$\hat{R}(i,j) \text{ is an estimate of } r_x(i-j), \qquad \text{for } 1 \leqslant i,j \leqslant p+1. \tag{5.2}$$

Once these estimates have been computed from the given time series observations, the resultant AR parameter vector estimate is, in accordance with (2.11), obtained by solving the linear system of equations

$$\hat{R}\boldsymbol{a} = \lvert b_0\rvert^2 \boldsymbol{e}_1 \tag{5.3}$$

in which the normalizing parameter b_0 is selected so that the first component of $\boldsymbol{a}$ is one. The steps of this general AR modeling approach are summarized in Table II.

The quality of the AR modeling approach as embodied in (5.3) is critically dependent on the choice of the autocorrelation lag estimation procedure used. For many applications, the standard unbiased autocorrelation estimates as given by

$$\hat{R}(i,j) = \frac{1}{N - \lvert i-j\rvert} \sum_{k=0}^{N} x(k+i-j)\,\bar{x}(k), \qquad \begin{matrix} 1 \leqslant i \leqslant p+1 \\ 1 \leqslant j \leqslant p+1 \end{matrix} \tag{5.4}$$

typically provide the best selection in terms of spectral estimation performance. It is seen that the autocorrelation matrix formed from this set of estimates will be Toeplitz and complex conjugate symmetric; properties shared by the actual autocorrelation matrix being approximated. Moreover, this estimate is *consistent* in the sense that as N approaches infinity, we have $\hat{R} \to R$ under the second-order ergodic assumption on the time series. In view of all of these favorable qualities, it is not surprising that the standard unbiased estimator (5.4) generally provides excellent AR modeling performance. In Section IX, some of the more popularly used adaptive methods of AR spectral estimation will be studied.

VI. ARMA Modeling—Time Series Observations

The methods for generating ARMA models based upon time series observations fall into basically two categories: the a_k and b_k parameters are either evaluated i) simultaneously or ii) separately. In the first category, maximum-likelihood-based techniques are prominent. These include exact-maximum-likelihood approaches (e.g., [6] and [48]), and, least squares methods which approximate the exact-likelihood function (e.g., [3], [9], [29]). Although offering the promise of optimum modeling, these maximum-likelihood methods entail the application of nonlinear programming solution procedures. As such, these solution procedures are computationally inefficient, and they suffer the obvious drawbacks associated with nonlinear programming methods. Other techniques which fall into category i) have been proposed (e.g., see [30], [40], [60]). These methods also entail the utilization of nonlinear programming solution procedures.

In recognition of the obvious shortcomings of nonlinear-programming-based techniques, a number of methods have been proposed which employ a separate evaluation of the AR and MA parameters. By using this approach, it is generally possible to obtain satisfactory modeling while not incurring the drawbacks of a nonlinear programming solution procedure. These techniques typically entailed using the first p extended Yule–Walker equations to obtain the a_k estimates, in a linear fashion (e.g., see [26], [28], [35], [38]). Unfortunately, the utilization of the minimal number of extended Yule–Walker equations (i.e., p) gave rise to an undesirable parameter hypersensitivity. In recognition of this fact, a procedure for using an overdetermined set of Yule–Walker equations evaluations to decrease this hypersensitivity was proposed by the author [15]. This approach has since been adopted successfully by other researchers in estimation applications (e.g., see [7], [12], [36], [51]). With this in mind, we shall now give a detailed development of the overdetermined approach to estimating the a_k parameters of an ARMA model.

AR Parameter Estimation

Although the procedure presented in Section II for generating ARMA models is attractive, one is rarely provided with exact autocorrelation information. The more common situation is one in which the only available information takes the form of a finite set of time series observations

$$x(1), x(2), \cdots, x(N). \tag{6.1}$$

The task at hand is to then use these time series observations to estimate the parameters of a postulated ARMA model. In this parameter estimation, we shall seek to incorporate the philosophy as embodied in the extended Yule–Walker ARMA model equations (2.23) for estimating the model's a_k parameters.

This will effectively entail using the given time series observations to generate an estimate of the $t \times (p+1)$ autocorrelation matrix R_1 which appears in (2.23). Namely, using any of a number of available procedures, we first compute the following autocorrelation lag estimates:

$$\hat{R}_1(i,j) = \text{an estimate of } r_x(q+1+i-j), \quad 1 \leq i \leq t, \quad 1 \leq j \leq p+1. \tag{6.2}$$

Two particularly attractive procedures for effecting these autocorrelation estimates will be detailed at the end of this section and in Section X. Independently of which procedure is eventually used, the net result of this first step will be the generation of a $t \times (p+1)$ autocorrelation matrix estimate $\hat{R}_1$. Due to errors inherent in the autocorrelation estimation process, however, this matrix estimate will generally have full rank (i.e., $\min(p+1, t)$) instead of the theoretical rank p which is possessed by the matrix R_1 being estimated. This being the case, it is, therefore, not generally possible to find an AR parameter vector with first component equal to one which will satisfy the theoretical relationship $R_1 \boldsymbol{a} = \boldsymbol{\theta}$ as given in (2.23). As such, the $t \times 1$ extended Yule–Walker equation error vector as specified by

$$\boldsymbol{e} = \hat{R}_1 \boldsymbol{a} \tag{6.3}$$

will be generated.

A little thought will convince oneself that the elements of this error vector will be composed of a sum of many random variable products (i.e., $x(k+m)\,\bar{x}(m)$) used in formulating the autocorrelation lag estimates. Consequently, an assumption that the error vector elements tend to be distributed in a Gaussian manner is a reasonable one. The joint density function (conditioned on $\hat{R}$) of the extended Yule–Walker equation error vector may be, therefore, approximated by

$$p(\boldsymbol{e}) = \frac{|W|^{1/2}}{(2\pi)^{t/2}} e^{-0.5(\boldsymbol{e}^* W \boldsymbol{e})} \tag{6.4}$$

in which $W^{-1} = E\{\boldsymbol{e}\boldsymbol{e}^*\}$ designates the error covariance matrix which is generally unknown and where the expected value of $\boldsymbol{e}$ is taken to be zero.

With the availability of the error joint density function (6.4), it is now possible to apply the maximum-likelihood concept for estimating the AR parameters. Namely, making use of (6.3) and the joint density function (6.4), it is possible to generate a joint density function for the AR parameter vector $\boldsymbol{a}$ which will be of form

$$p(\boldsymbol{a}) = \gamma e^{-0.5(\boldsymbol{a}^* \hat{R}^* W \hat{R} \boldsymbol{a})}.$$

We now seek that vector $\boldsymbol{a}$ which maximizes this joint density function subject to the constraint that the first component of $\boldsymbol{a}$ be one. Ignoring the effect of the multiplicative term γ, the pseudo-maximum-likelihood selection for $\boldsymbol{a}$ then corresponds to solving the following constrained minimization problem:

$$\min_{a(1)=1} \boldsymbol{a}^* \hat{R}^* W \hat{R} \boldsymbol{a}. \tag{6.5}$$

Using standard Lagrange multiplier techniques, the solution to this constrained minimization problem is obtained by solving the following system of $(p+1) \times (p+1)$ linear equations:

$$\hat{R}_1^* W \hat{R}_1 \boldsymbol{a} = \alpha \boldsymbol{e}_1 \tag{6.6}$$

where α is a normalizing constant selected so that the first component of $\boldsymbol{a}^0$ is one.[4] Expression (6.6) constitutes the so-called *high performance method* of AR parameter selection [15]–[22].

It is to be noted that in minimizing functional (6.5) with respect to the normalization constraint imposed on $\boldsymbol{a}$'s first component, the error vector is being minimized in the least

[4] In those rare cases where the $(p+1) \times (p+1)$ matrix $\hat{R}_1^* W \hat{R}_1$ is singular, the AR parameter vector will correspond to a suitable normalized eigenvector associated with a zero eigenvalue of this matrix.

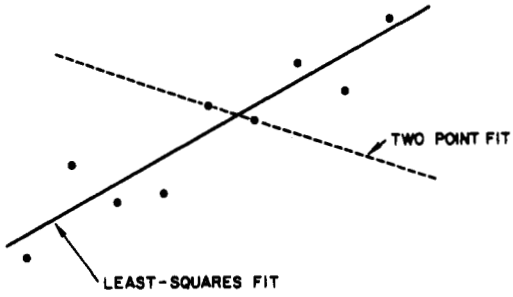

Fig. 7. Linear model fitting.

squares sense. In effect, we are then selecting a so as to best satisfy the theoretical relationship (2.23) given by $R_1 a = \theta$. Using this interpretation, the positive definite matrix W can be alternatively thought of as providing a weighting (instead of being an unknown covariance matrix inverse) in the error functional (6.5). It is, therefore, logical to take W to be a diagonal matrix whose negative diagonal entries w_k for $k = 1,2,\cdots,t$ provide a mechanism for weighting in any desirable fashion the various extended Yule–Walker equation approximations appearing in (6.3). The uniform weighting selection

$$W = I \tag{6.7}$$

where I is the $t \times t$ identity matrix has been found to provide excellent modeling performance when the matrix estimate $\hat{R}_1$ is unbiased.

A few words are now appropriate concerning the selection of the integer t which specifies the number of extended Yule–Walker equations that are being approximated. When t is set equal to its minimal value p, the approach here taken bears a close resemblance to various other ARMA modeling schemes (e.g., see [26], [28], [35], [38]). In this case, the minimal number of p error-contaminated extended Yule–Walker equations are being used in fixing the model's p AR coefficients. A little thought should convince the reader of the potential parameter hypersensitivity which can arise in this situation. To illustrate this point, let us briefly consider the task of finding a line which "best" fits a set of error-contaminated two-tuples (x_k, y_k). Although only two two-tuples are needed to fix the line's two parameters (i.e., its slope and y intercept), it will be generally more desirable to fix these parameters by using more than this minimal number of two-tuples thereby obtaining a more "representative linear fit." This will entail finding the "best least squares linear fit." The benefits generally accrued in using this overdetermined approach are demonstrated in Fig. 7.

With the above in mind, the real advantage of this paper's approach is achieved when the integer t is selected to be larger than p. In this case, more than the minimal number of extended Yule–Walker equation evaluations (i.e., t instead of p) are being used in fixing the model's p AR coefficients. It is then not surprising that a desirable decrease in parameter hypersensitivity is generally realized upon selecting $t > p$. An indication of the benefit accrued by selecting $t > p$ was illustrated in Section III for the case of AR modeling with perfect autocorrelation lag values. A similar advantage will be demonstrated in Section VIII when ARMA models are generated from raw time series observations. In the situation being considered here, the integer parameter t is typically selected to lie within the range

$$p \leqslant t \leqslant N - q - 1 \tag{6.8}$$

with generally *larger* values than the minimum p being preferred for modeling fidelity.

From an overall modeling viewpoint, the standard unbiased estimator has been found to generally provide the best choice for the lag estimates required in (6.2). Specifically, the required autocorrelation lag estimate entries are generated according to

$$\hat{r}_x(n) = \frac{1}{N-n} \sum_{k=1}^{N-n} x(k+n)\,\bar{x}(k), \qquad 0 \leqslant n \leqslant q + t \tag{6.9}$$

where $q + t$ corresponds to the largest autocorrelation lag argument appearing in matrix $\hat{R}_1$. We would, of course, use the property that $r_x(-n) = \bar{r}_x(n)$ to obtain any negative lag autocorrelation entries which may be needed in formulating $\hat{R}_1$. In using this unbiased estimate approach, the resultant autocorrelation matrix estimate will have a desirable Toeplitz structure.

The $(p+1) \times (p+1)$ matrix $\hat{R}_1^* W \hat{R}_1$, which completely characterizes the AR parameter vector solution through (6.6), will have components which are readily computable from the estimates (6.9). Using simple matrix manipulations, it is readily shown that the general (i,j)th element of this matrix is specified by

$$\hat{R}_1^* W \hat{R}_1(i,j) = \sum_{m=1}^{t} w(m)\,\bar{\hat{r}}(q+m+1)\,\hat{r}(q+m+i-j), \qquad \text{for } 1 \leqslant i,j \leqslant p+1 \tag{6.10}$$

where the $w(m)$ correspond to the diagonal elements of the diagonal weighting matrix W. Upon generation of the matrix $\hat{R}_1^* W \hat{R}_1$ according to this expression, the required autoregressive parameter vector is obtained in a straightforward manner by solving the system of linear equations (6.6).

MA Parameter Estimation

To complete the ARMA modeling, it is necessary to compute an estimate for the MA component $|B_q(e^{j\omega})|^2$. It has been the author's experience that independent of which procedure is used, this MA component estimate is almost always of significantly lower quality than the associated AR component

$$|\hat{A}_p(e^{j\omega})|^2 = \left| \sum_{k=0}^{p} \hat{a}_k\, e^{j\omega k} \right|^2 \tag{6.11}$$

in which the $\hat{a}_k$ denote the AR parameter estimates as generated from (6.6). A high-quality low-order MA spectral estimator has yet to be developed. Despite this shortcoming, some reasonably well performing MA estimators will now be briefly discussed.

Many contemporary MA component estimators are based on utilizing the forward and backward residual time series associated with an ARMA time series. In particular, the forward residual time series elements are computed from the given observations (6.1) and the AR parameter estimates (6.6) according to

$$s_f(n) = \sum_{k=0}^{p} \hat{a}_k x(n-k), \quad p+1 \leqslant n \leqslant N. \tag{6.12}$$

Similarly, the backward residuals components are generated using

$$s_b(n) = \sum_{k=0}^{p} \bar{\hat{a}}_k x(n+k), \quad 1 \leqslant n \leqslant N-p. \tag{6.13}$$

As indicated in Section II, each of these residual time series will be governed by the same MA(q) process if the time series $\{x(n)\}$ is an ARMA(p, q) process with AR parameters $\hat{a}_k$. With this in mind, procedures for extracting this MA characterization from the computed forward and backward residuals will now be given.

(i) The most direct procedure for achieving the required MA(q) estimate is to first generate the following estimates of the residual time series' first $q+1$ autocorrelation lags:

$$\hat{r}_s(n) = \left(\frac{1}{N-p-n}\right) \sum_{k=1}^{N-p-n} [s_f(n+p+k)\,\bar{s}_f(p+k) + s_b(n+k)\,\bar{s}_b(k)], \quad 0 \leqslant n \leqslant q. \tag{6.14}$$

If the residual time series do in fact correspond to an MA(q) process, it will be found that the $r_s(n)$ will be approximately zero for $n \geqslant q+1$. This can be used as a convenient test for the appropriateness of the ARMA model, the order selection, and the estimates a_k. In any case, upon taking the Fourier transform of these autocorrelation lags, we obtain the MA(q) spectral estimate component

$$|\hat{B}_q(e^{j\omega})|^2 = \sum_{n=-q}^{q} w(n)\,\hat{r}_s(n)\,e^{-j\omega n} \tag{6.15}$$

in which $w(n)$ is a window sequence and use of the fact that $r_s(-n) = \bar{r}_s(n)$ will be made when evaluating (6.14). The overall ARMA(p, q) spectral estimate is then given by

$$\hat{S}(e^{j\omega}) = \frac{|\hat{B}_q(e^{j\omega})|^2}{|\hat{A}_p(e^{j\omega})|^2} \tag{6.16}$$

where $\hat{A}_p(e^{j\omega})$ is specified by (6.11).

A few words are now appropriate concerning the selection of the window to be used in (6.14). If the rectangular window choice $w(n) = 1$ is made, this estimate will not have the desired property of being guaranteed positive-semidefinite. To achieve this positive-semidefiniteness, we could instead choose the window to be

$$w(n) = \left(\frac{N-p-n}{N-p}\right)\left(\frac{q+1-|n|}{q+1}\right). \tag{6.17}$$

Unfortunately, this selection can give rise to a seriously distorted MA estimate in view of the triangular-like weighting thereby employed. The selection of $w(n)$ is quite important and this choice should be based on the particular application at hand and user experience.

(ii) It is also possible to employ the smoothed periodogram to obtain yet another form of MA(q) estimate. This entails segmenting the computed residuals into blocks of length $q+1$ (overlapping or nonoverlapping) and then averaging the resultant $q+1$ length periodograms for each of these blocks. This procedure has been employed with a moderate degree of success [17]. (iii) Similarly, we could make obvious adaptions of the procedures treated in Section II under the ARMA modeling subsection to achieve alternate MA estimates. For example, if we were to use the procedure as characterized by (2.38), estimates for the c_n parameters would be computed from

$$\hat{c}_n = \sum_{k=0}^{n-1} \hat{a}_k \hat{r}_x(n-k), \quad 1 \leqslant n \leqslant p. \tag{6.18}$$

The required ARMA spectral estimates would then be given by incorporating these estimates into (2.33) to yield

$$\hat{S}_x(e^{j\omega}) = \hat{r}_x(0) + 2\,\mathrm{Re}\,[\hat{D}(e^{j\omega})] \tag{6.19}$$

where $\hat{D}(e^{j\omega})$ is obtained by substituting the $\hat{a}_k$ and $\hat{c}_k$ estimates into expression (2.34).

VII. ARMA Modeling—A Singular Value Decomposition Approach

We have yet to address the important issue of ARMA model order determination. In particular, whether one is provided with exact autocorrelation lags or time series observations for effecting the modeling, how one chooses appropriate values for the order parameters p and q remains an open question. It is recognized that this model order information is implicitly contained in the autocorrelation matrices which characterize ARMA models. In this section, we shall present a procedure for extracting the prerequisite model order values which will make use of a singular value decomposition of an extended autocorrelation matrix. An important byproduct of this procedure will be an adaption of the ARMA modeling procedure of the previous section which provides for a *significant* improvement in spectral estimation performance.

When the ARMA model order parameters are not known *a priori*, it will be judicious to select the initial model order to be much larger than the "anticipated" order. In particular, let us consider the extended order ARMA (p_e, q_e) model for which p_e is selected to be larger (usually much larger) than the eventual model order parameter p to be used. Although we typically do not know p *a priori*, it is generally possible to make an educated guess of p so as to ensure that

$$p_e > p. \tag{7.1}$$

In accordance with (2.23), it then follows that the $t \times (p_e+1)$ extended order autocorrelation matrix associated with this ARMA(p_e, q_e) model may be expressed as

$$R_e = \begin{bmatrix} r_x(q_e+1) & r_x(q_e) & \cdots & r_x(q_e-p_e+1) \\ r_x(q_e+2) & r_x(q_e+1) & \cdots & r_x(q_e-p_e+2) \\ \vdots & \vdots & & \vdots \\ r_x(q_e+t) & r_x(q_e+t-1) & \cdots & r_x(q_e-p_e+t) \end{bmatrix}. \tag{7.2}$$

If the autocorrelation lag entries used in this matrix correspond to an ARMA (p, q) process for which $q_e - p_e \geqslant q - p$, it then follows from the results of Section II that the rank of the $t \times (p_e+1)$ matrix R_e will be p. In arriving at this result, we of course assume that t is selected to at least equal p. To determine the required order parameter p, we then simply set p equal to the rank of R_e for the idealistic case in which exact autocorrelation lag information is available.

To obtain the ARMA model's $(p+1) \times 1$ AR parameter vector $\boldsymbol{a}$ from this extended order autocorrelation matrix, it is

possible to appeal to the theoretical developments of Section II. In particular, let us consider the set of submatrices of R_e formed from any of its $p+1$ contiguous columns. This set of $t \times (p+1)$ matrices is specified by

R_k = [submatrix of R_e composed of its kth through $p+k$th column vectors inclusively],

$$\text{for } 1 \leq k \leq p_e - p + 1. \quad (7.3)$$

In accordance with the ARMA model extended Yule-Walker equation relationships, it is readily established that the required unique AR parameter vector a will satisfy the set of homogeneous relationships

$$R_k a = \theta, \quad \text{for } 1 \leq k \leq p_e - p + 1 \quad (7.4)$$

where the first component of a is constrained to be one. In point of fact, (7.4) provides a matrix representation for the t extended Yule-Walker equations (2.22) defined on the specific indices $q_e + 2 - k \leq n \leq q_e + t + 1 - k$. It is important to note that this conclusion will be valid only if the autocorrelation lag entries used in forming R_e correspond to an ARMA (p, q) process, and the order parameters are suth that $p_e \geq p$ and $q_e - p_e \geq q - p$.

We shall now apply this rank characterization of R_e to the practical problem in which the ARMA modeling is to be based only on the time series observations

$$x(1), x(2), \cdots, x(N) \quad (7.5)$$

and not on actual autocorrelation lag information. In this case, it will be necessary to first compute autocorrelation lag estimates from these observations. These estimates are next substituted into the matrix format (7.2) to, in turn, generate the extended order autocorrelation matrix estimate $\hat{R}_e$. Since the autocorrelation lag estimate entries will be invariably in error, it follows that the matrix $\hat{R}_e$ will normally have full rank (i.e., min $(p_e + 1, t)$) even when the time series under study corresponds to an ARMA (p, q) process. Nonetheless, even though $\hat{R}_e$ will have full rank, its "effective" rank will still tend to be p. To better quantify the vague term "effective" rank, it will be beneficial to introduce the concept of singular value decomposition.

Singular Value Decomposition

In a variety of applications, the primary objective will be that of solving a linear system of equations. The matrix associated with this system of equations not only characterizes the desired solution, but it will also very often convey dynamical property information. With this in mind, it often behooves us to examine the salient properties of this characterizing matrix. The singular value decomposition of a matrix as outlined in the following theorem serves this role particularly well (e.g., see [27] and [39]).

Theorem 7.1: Let A be an $m \times n$ matrix of generally complex valued elements. Then there exists $m \times m$ and $n \times n$ unitary matrices U and V, respectively, such that[5]

$$A = U \Sigma V^* \quad (7.6)$$

where Σ is an $m \times n$ matrix whose elements are zero except possibly along its main diagonal. These nonnegative diagonal elements are ordered such that

$$\sigma_{11} \geq \sigma_{22} \geq \cdots \geq \sigma_{hh} \geq 0$$

where

$$h = \min(m, n).$$

The diagonal elements σ_{kk} are commonly referred to as the singular values of matrix A. It is well known that the nonzero singular values will correspond to the positive square roots of the eigenvalues of the nonnegative Hermitian matrices AA^* and A^*A. Moreover, the columns of U (or V) will correspond to the appropriately ordered orthonormal eigenvectors of the nonnegative Hermitian matrices AA^* (or A^*A).

The singular values σ_{kk} convey valuable information concerning the rank characterization of matrix A. This is readily demonstrated upon considering the problem of finding that $m \times n$ matrix of rank k which will best approximate A in the Frobenious norm sense (this assumes that $k \leq$ rank $[A]$). The Frobenious norm of the $m \times n$ matrix difference $A - B$ is defined to be

$$\|A - B\| = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij} - b_{ij}|^2 \right]^{1/2}. \quad (7.7)$$

We now seek to find that $m \times n$ rank k matrix B which will render this criterion a minimum. The solution to this approximation problem is contained in the following theorem [27]:

Theorem 7.2: The unique $m \times n$ matrix of rank $k \leq$ rank $[A]$ which best approximates the $m \times n$ matrix A in the Frobenious norm sense is given by

$$A^{(k)} = U \Sigma_k V^* \quad (7.8)$$

where U and V are as in (7.6) while Σ_k is obtained from Σ by setting to zero all but its k largest singular values. The quality of this optimum approximation is given by

$$\|A - A^{(k)}\| = \left[\sum_{j=k+1}^{h} \sigma_{jj}^2 \right]^{1/2}, \quad 0 \leq k \leq h. \quad (7.9)$$

The degree to which $A^{(k)}$ approximates A is seen to be dependent on the sum of the $(h - k)$ *smallest singular values* squared. As k approaches h, this sum will become progressively smaller and will eventually go to zero at $k = h$. In order to provide a convenient measure for this behavior which does not depend on the size of matrix A, let us consider the normalized ratio

$$\nu(k) = \frac{\|A^{(k)}\|}{\|A\|} = \left[\frac{\sigma_{11}^2 + \sigma_{22}^2 + \cdots + \sigma_{kk}^2}{\sigma_{11}^2 + \sigma_{22}^2 + \cdots + \sigma_{hh}^2} \right]^{1/2}, \quad 1 \leq k \leq h. \quad (7.10)$$

Clearly, this normalized ratio approaches its maximum value of one as k approaches h. For matrices of low effective rank, the quantity $\nu(k)$ is close to one for values of k significantly smaller than h. On the other hand, matrices for which k must take on high values (i.e., $k \approx h$) to achieve a $\nu(k)$ near one are said to be of high effective rank.

Application of SVD to ARMA Modeling

To determine the required order for an ARMA model, we shall now make an SVD of the $t \times (p_e + 1)$ extended order autocorrelation matrix estimate (7.2), that is,

$$\hat{R}_e = U \Sigma V^* \quad (7.11)$$

where U and V are $t \times t$ and $(p_e + 1) \times (p_e + 1)$ unitary matrices, respectively, and Σ is a $t \times (p_e + 1)$ matrix of the form

[5] The matrices U and V are said to be unitary if $U^{-1} = U^*$ and $V^{-1} = V^*$.

called out in Theorem 7.1. The required AR order p is obtained by examining the normalized ratio $\nu(k)$. Namely, p is set equal to the smallest value of k for which $\nu(k)$ is deemed "adequately" close to one. The terminology "adequately close to one" is subjective and will depend on the particular application under consideration as well as user experience gained through empirical experimentation. In any case, the net result of this step will be a rank p optimum approximation of the $t \times (p_e + 1)$ extended order autocorrelation matrix, that is,

$$\hat{R}_e^{(p)} = U \Sigma_p V^*. \tag{7.12}$$

A simple matrix manipulation reveals that this rank p approximation may be equivalently represented as

$$\hat{R}_e^{(p)} = \sum_{n=1}^{p} \sigma_{nn} u_n v_n^* \tag{7.13}$$

where u_k and v_k are the kth column vectors of the $t \times t$ and $(p_e + 1) \times (p_e + 1)$ unitary matrices U and V, respectively. We shall now provide two separate procedures for using this rank p approximation for effecting autoregressive parameter estimates.

Method I: ARMA (p_e, q_e) Model

In this approach, the rank p approximation (7.12) is interpreted as providing an improved estimate of the underlying extended autocorrelation matrix. It will be convenient to decompose this rank p approximation as follows:

$$\hat{R}_e^{(p)} = [\hat{r}_1^{(p)} : \hat{R}_a^{(p)}] \tag{7.14}$$

where $\hat{r}_1^{(p)}$ is the leftmost $t \times 1$ column vector of $\hat{R}_e^{(p)}$ and $\hat{R}_a^{(p)}$ is a $t \times p_e$ matrix composed of the p_e rightmost $t \times 1$ column vectors of $\hat{R}_e^{(p)}$. We now seek a $(p_e + 1) \times 1$ AR parameter vector a with first component equal to one that will satisfy the theoretical relationship

$$R_e^{(p)} a = \theta.$$

Since the rank of $\hat{R}_e^{(p)}$ is less than full, there will exist an infinity of solutions to this problem. We shall select the minimum norm solution as specified by

$$\begin{bmatrix} a_1^0 \\ a_2^0 \\ \vdots \\ a_{p_e}^0 \end{bmatrix} = -[\hat{R}_a^{(p)}] \# \hat{r}_1^{(p)}$$

in which the superscript notation # denotes the operation of generalized (pseudo) matrix inversion. This AR parameter selection procedure has provided to be particularly effective in low SNR environments. It is readily shown that this minimum norm solution can be simplified to

$$a^0 = \frac{e_1 - \sum_{k=1}^{p} \bar{v}_k(0)\, v_k}{1 - \sum_{k=1}^{p_e} |v_k(0)|^2} = \frac{\sum_{k=p+1}^{p_e} \bar{v}_k(0)\, v_k}{\sum_{k=p+1}^{p_e} |v_k(0)|^2} \tag{7.15}$$

where the v_k correspond to the column vectors of the unitary matrix V appearing in the SVD representation (7.12).

Method II: ARMA (p, q) Model

The best rank p approximation matrix (7.12) contains within its column structure the characteristics required to estimate AR parameters of a lower order ARMA(p, q) model. In particular, the submatrices of $\hat{R}_e^{(p)}$ composed of its columns k through $p + k$ inclusively yield rank p approximations of the $t \times (p + 1)$ autocorrelation matrices R_k for $1 \leq k \leq p_e - p + 1$ as specified by (7.3). We shall denote these rank p approximations by $\hat{R}_k^{(p)}$. Due to the SVD operation and errors inherent in generating $\hat{R}_e$, there will generally not exist a unique AR parameter vector with first component equal to one which will satisfy all of the $p_e - p + 1$ estimates of (7.4). Nonetheless, it is still desirable to find an AR parameter vector for which each of these relationships are almost satisfied. A functional that measures the degree to which this is accomplished is given by

$$f(a) = a^* S^{(p)} a \tag{7.16a}$$

where

$$S^{(p)} = \sum_{k=1}^{p_e - p + 1} \hat{R}_k^{(p)*} \hat{R}_k^{(p)}. \tag{7.16b}$$

The $(p + 1) \times (p + 1)$ matrix $S^{(p)}$ is nonnegative definite Hermitian and may be conveniently computed using the relationship

$$S^{(p)} = \sum_{n=1}^{p} \sum_{k=1}^{p_e - p + 1} \sigma_{nn}^2 v_n^k v_n^{k*} \tag{7.17}$$

in which v_n^k denotes the $(p + 1) \times 1$ vector as specified by

$$v_n^k = [v_n(k), v_n(k+1), \cdots, v_n(k+p)]', \quad 1 \leq k \leq p_e - p + 1, \quad 1 \leq n \leq p. \tag{7.18}$$

This vector is seen to be a windowed segment of the nth column vector (i.e., v_n) of the unitary matrix V that in part identifies the SVD representation (7.11). Moreover, due to the simple shift relationship between the vectors v_n^k and v_n^{k+1}, it is possible to devise an iterative procedure for updating the $(p + 1) \times (p + 1)$ matrices $v_n^k v_n^{k*}$ as k evolves. This will entail $(p + 1)$ computations for each value of k.

Upon generating the $(p + 1) \times (p + 1)$ matrix $S^{(p)}$, we next wish to select that AR parameter vector a with first component of one so as to minimize quadratic functional (7.16). This constrained minimization will result in the best least squares approximation of the theoretical relationships (7.4). Using standard procedures, the required optimum AR parameter vector is found by solving the following linear system of equations:[6]

$$S^{(p)} a = \alpha e_1 \tag{7.19}$$

in which the normalizing constant α is selected so that the first component of a is one. It will be shown in the next section that these SVD versions of ARMA modeling procedures can lead to a significant improvement in modeling performance.

The concept of SVD representation has been also in-

[6] In those cases where $S^{(p)}$ is singular, the required AR parameter vector is set equal to an appropriately normalized eigenvector associated with a zero eigenvalue of $S^{(p)}$.

corporated with success in effecting AR models [42] and [61]. Incorporation of an SVD AR model was there shown to produce an increase in spectral resolution capabilities. More recently, the SVD representation was used in ARMA modeling where impressive results were reported [22]. Undoubtably, the impact which SVD representations will ultimately have on spectral estimation (and in other applications) is only beginning to be realized.

VIII. Numerical Examples

In this section, we shall investigate the comparative spectral estimation performance of the ARMA modeling procedures as developed in Sections VI and VII with those of popularly used alternatives. The first example will treat the problem of effecting a rational spectral estimate from a set of observations of an ARMA(4,4) process. In the second example, we shall examine the modeling performance for the special case of sinusoids in white noise.

Example 1: In this example, we shall examine the time series as characterized by (see [11])

$$x(n) = x_1(n) + x_2(n) + 0.5\epsilon(n) \tag{8.1a}$$

which is composed of the two AR(2) time series generated according to

$$x_1(n) = 0.4x_1(n-1) - 0.93x_1(n-2) + \epsilon_1(n) \tag{8.1b}$$

$$x_2(n) = -0.5x_2(n-1) - 0.93x_2(n-2) + \epsilon_2(n) \tag{8.1c}$$

where $\epsilon(n)$, $\epsilon_1(n)$, and $\epsilon_2(n)$ are mutually uncorrelated Gaussian zero-mean white-noise processes with variance one. A simple analysis indicates that the power spectral density function associated with time series (8.1) is given by

$$S_x(\omega) = |1 - 0.4\,e^{-j\omega} + 0.93\,e^{-j2\omega}|^{-2} + |1 + 0.5\,e^{-j\omega} + 0.93\,e^{-j2\omega}|^{-2} + 0.25 \tag{8.2}$$

and is plotted in Fig. 8(a).

Using the time series description (8.1), twenty statistically independent realizations each of length 125 were next generated. These 20 realizations were then used to compare the modeling effectiveness of this paper's method with the Box–Jenkin's maximum-likelihood method. The twenty (one for each realization) superimposed ARMA(4,4) spectral estimates obtained using the Box–Jenkin's iterative method are shown in Fig. 8(b). The number of iterations required to achieve these estimates ranged from 10 to 700 with 50 being a typical requirement. Next, this paper's high-performance method as represented by expression (6.6) with unbiased autocorrelation lag estimates and $W = I$ was used to obtain the ARMA (4, 4) model's AR coefficients. Relationship (6.15) with the window selection (6.17) was used in forming the MA component of the spectral estimates. The twenty superimposed ARMA (4, 4) spectral estimates thereby obtained are shown in Fig. 8(c)–(e) for various choices of t. From these plots, it is apparent that progressively improved estimates are achieved upon increasing t from its minimal value of 4, to 8, and then to 20. Moreover, these spectral estimates were of higher quality than those obtained with the maximum-likelihood method which exhibited a larger variance in estimate.

Example 2: In this example, we shall investigate the comparative spectral estimation performances of various widely used methods on the classical sinusoids in additive white-noise problem. The particular time series to be considered is given by

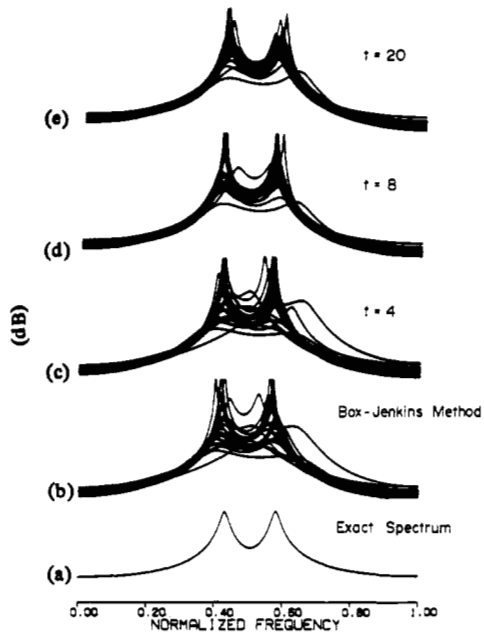

Fig. 8. ARMA spectral estimates of order (4,4). (a) Exact. (b) Box–Jenkins maximum-likelihood method. (c) Paper's method for $t = 4$. (d) Paper's method for $t = 8$. (e) Paper's method for $t = 20$.

$$x(n) = \sin(2\pi f_1 n) + \sin(2\pi f_2 n) + w(n), \quad 1 \leq n \leq N \tag{8.3}$$

$$f_1 = 0.2, \quad f_2 = 0.215, \quad \sigma_w^2 = 0.5.$$

This time series was previously examined in Section III where different rational models were generated from the "exact" autocorrelation lags associated with it. This is a particularly appropriate time series for testing the resolution capabilities of spectral estimators because of the closeness of the sinusoidal frequencies (i.e., $f_2 - f_1 = 0.015$) and the prevailing low SNR of 0 dB (individual sinusoid power to total noise power).

In order to gain a reasonable good statistical basis for comparison, ten statistically independent realizations of the time series (8.3) were generated with each realization being of length 128 (i.e., $N = 128$). Using these ten different sets of time series observations, ten spectral estimates were made for various widely used rational spectral estimators. The resultant ten spectral estimates for each estimator were then plotted in Figs. 9–14 in a superimposed fashion (except for the periodogram) so as to depict consistency of estimate. The ideal estimate would, of course, be two sharply defined peaks at frequencies 0.2 and 0.215. A brief description of the different estimators and their performance on these test samples is now given.

MA Estimates: The periodogram as implemented by the FFT was first used in generating spectral estimates for each of the ten different 128 data length realizations. Specifically, (4.7) with $N = 128$ was incorporated into the MA spectral estimator (4.5) to generate the sampled periodogram estimate

$$\hat{S}_x(e^{j(2\pi k/N)}) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n+1)\,e^{-j(2\pi k/N)}\right|^2, \quad 0 \leq k \leq N-1. \tag{8.4}$$

It was found that each of the ten periodograms produced remarkably similar results. A typical 128-point FFT periodogram

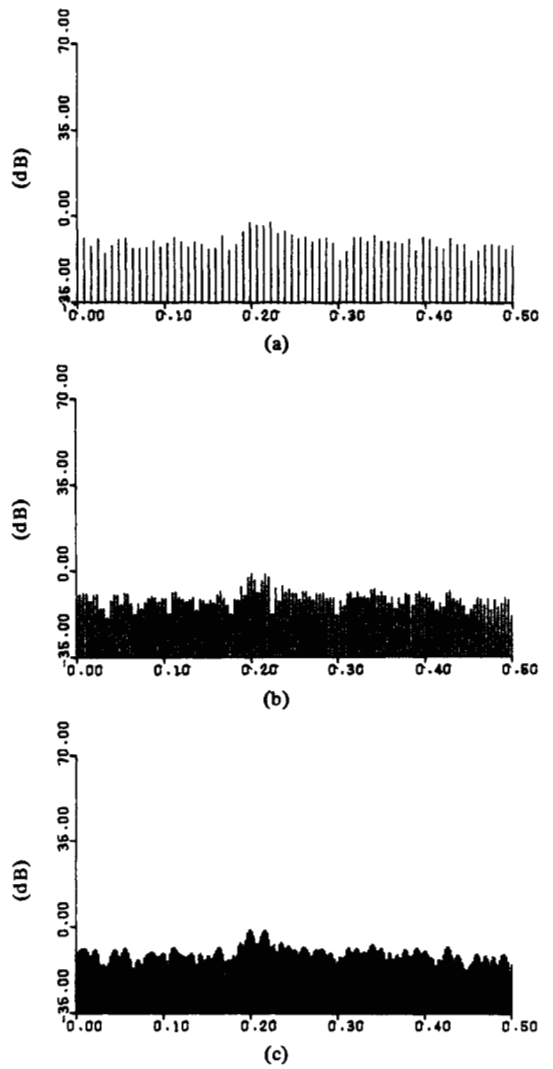

Fig. 9. Moving average (MA) spectral estimates using the FFT implementation of the periodogram with 128 time series observations. (a) $N = 128$; no zero padding. (b) $N = 256$; 128 zero padding. (c) $N = 512$; 384 zero padding.

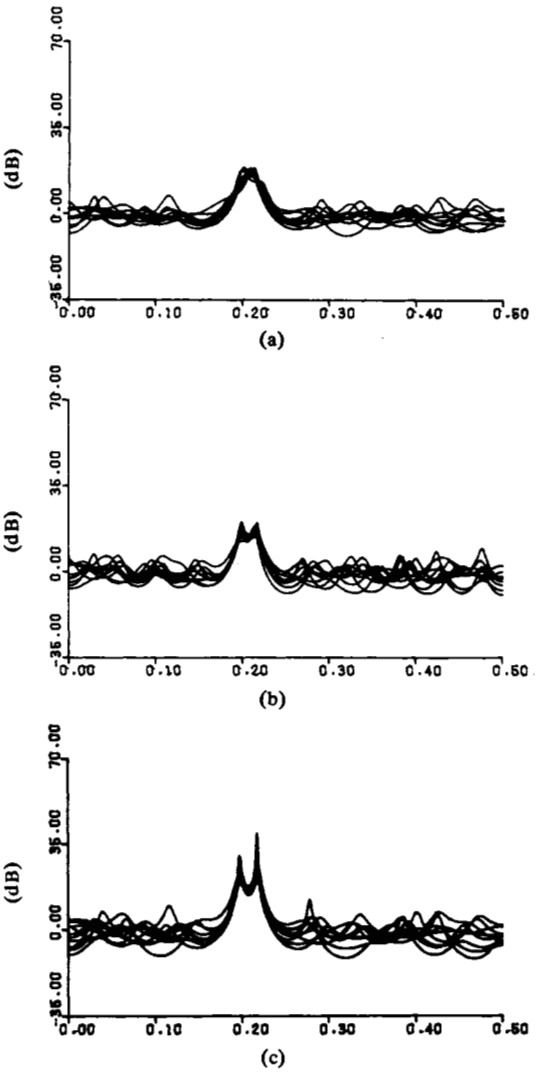

Fig. 10. Autoregressive (AR) spectral estimates from 128 time series observations. (a) AR(20); $p = 20$, Burg estimate. (b) AR(24); $p = 24$, Burg estimate. (c) AR(20), this paper's method (6.6) with $q = 0$, $p = 20$, $t = 50$.

for one of these trials is shown in Fig. 9(a). From this plot (and the nine others not shown), it was not possible to unambiguously detect the presence of two spectral peaks at frequencies 0.2 and 0.215.

In order to ease the potential ambiguity created by the finite-frequency sampling of the periodogram (i.e., $\Delta\omega = 2\pi/N$), the concept of *padding* as described in Section IV was next incorporated. Using this approach, the original time series observation of length 128 was next appended with 128 zeroes. The resultant 256-point padded FFT periodogram is shown in Fig. 9(b). In this padded case, we are able to unambiguously detect the presence of the two spectral peaks at 0.2 and 0.215. A further padding of 256 zeroes is found to result in the 512-point padded FFT periodogram shown in Fig. 9(c). The prerequisite spectral resolution is again achieved.

AR Estimates: In AR modeling, the most widely used procedure is the Burg algorithm. With this in mind, the Burg algorithm was next used to generate spectral estimates for each of the aforementioned ten observation sets of length 128. The ten superimposed Burg AR(20) estimates which resulted are depicted in Fig. 10(a). Although a detection of spectral energy in the region about $f = 0.2$ is evident, the appearance of two spectral peaks is absent. The ordering selection $p = 20$ was evidently not sufficient for the required resolution. Upon increasing the AR order to $p = 24$, however, the Burg AR(24) estimates produced two reasonably well defined peaks about $f = 0.2$ and $f = 0.215$ in nine out of the ten estimates. These estimates are plotted in superimposed fashion in Fig. 10(b). It was further determined that more sharply defined peaks are achieved in all ten estimates when the order was increased to forty. The Burg algorithm is then seen to provide a satisfactory resolution performance for the time series under study provided that the AR order is selected to be larger than 24.

In order to demonstrate the effect of using more than the minimal number of extended Yule-Walker equations in arriving at an AR model (the Burg algorithm uses the minimal number), the ARMA modeling technique as embodied in (6.6) with $W = I$ and unbiased autocorrelation lag estimates was next used with $p = 20$, $q = 0$, and $t = 50$. The resultant ten AR(20) spectral estimates which arose when using this approach are shown in Fig. 10(c). A resolution of the two sinusoids was achieved in all ten estimates. It is significant that the lower order AR(20) spectral estimates as generated in using one of this paper's methods provided more sharply defined peaks than

TABLE III
STATISTICS OF SVD ARMA (4,4) ESTIMATES

k	$\bar{f}_k$	$\bar{\bar{f}}_k$	σ_{f_k}	$\overline{\lvert p_k \rvert}$	$\sigma_{\lvert p_k \rvert}$
1	0.20	0.1998	0.0012	0.9944	0.0062
2	0.215	0.2159	0.0011	0.9974	0.0080

the higher order Burg AR(24) spectral estimates. This is primarily due to the fact that fifty extended Yule–Walker equations were used in specifying the 20 AR parameters. The degree of smoothing achieved in applying this approach is evident from this numerical example.

ARMA Estimates: The ARMA modeling procedure as represented by (6.6) with $W = I$ and unbiased autocorrelation lag estimate entries was next used to generate estimates of the AR coefficients of an ARMA(p, p) model for $p = 8$ and 12. In accordance with the results of Section III, plots of $|A_p(e^{j\omega})|^{-2}$ were then made so as to reveal the required spectral information for the sinusoids in the white-noise case (i.e., the zeroes are not used). In Fig. 11(a), the ten AR(8,8) spectral estimates which arose for a choice of $t = 70$ are shown superimposed. Although spectral energy in the neighborhood of $f = 0.2$ is detected, the presence of the required two spectral peaks is not. Clearly, the order selection $p = 8$ was not sufficient to achieve the desired resolution. Upon increasing the order to ARMA(12, 12) and retaining $t = 70$, however, the resultant ten spectral estimates shown in Fig. 11(b) each achieved the desired spectral resolution with two sharply defined peaks about $f = 0.2$ and $f = 0.215$. These spectral estimates have been obtained with but twelve AR parameters, and are seen to be *significantly* superior to the Burg AR(24) estimates which required twenty-four AR parameters. In terms of spectral estimation, fidelity, and parameter parsimony (i.e., effective use of parameters), it is clear that the ARMA modeling method herein developed has provided a superior performance for the problem at hand.

A truly significant increase in spectral estimation performance is achieved upon adopting the SVD approaches to ARMA modeling as outlined in Section VII. Namely, after setting $p_e = q_e = 14$ and $t = 50$, it was found that the effective rank of the extended order autocorrelation matrix estimate R_e was four. Next, letting $p = 4$ in (7.19), the ten SVD derived lower order ARMA(4, 4) spectral estimates which arose are shown superimposed in Fig. 11(c). In all ten estimates, the two sinusoids were detected and the spectrum was approximately at the theoretical 0-dB level for most other frequencies. These spectral estimates are not only of uniformly high quality, but, they represent the *lowest* order rational model which is compatible with the two sinusoids in white-noise case. Moreover, the quality of the peak frequency estimates and associated pole magnitude (theoretically equal to one) estimates is exceptional as shown in Table III. The quantities $\bar{f}_k(\bar{p}_k)$ and $\sigma^2_{\bar{f}_k}(\sigma^2_{\bar{p}_k})$ for $k = 1, 2$ represent the sampled means and variances, respectively, of the peak frequencies (pole magnitudes) as determined from the ten spectral estimates.

To demonstrate the worth of singular values in model order determination when using the SVD approach, the fifteen singular values which characterized the extended order autocorrelation matrix estimate R_e for one of the ten observation sets are now given

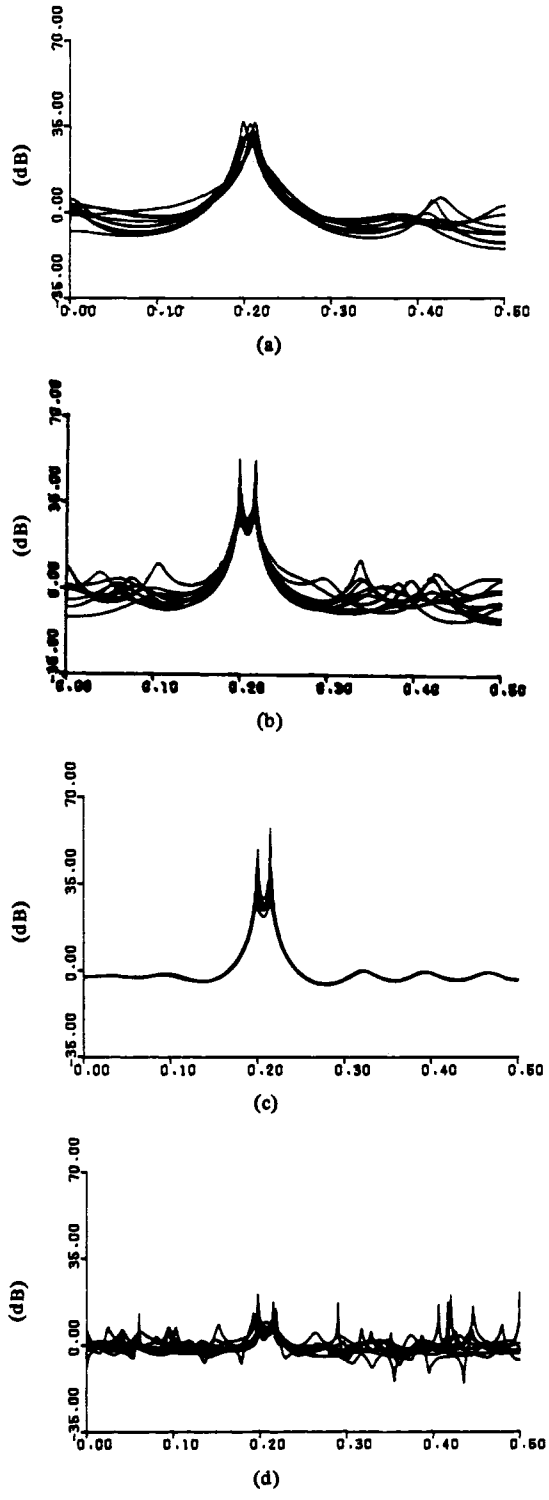

Fig. 11. Autoregressive-moving average (ARMA) estimates from 128 time series observations. (a) High-performance method ARMA(8,8); $p = q = 8$, $t = 70$. (b) High-performance method ARMA(12,12); $p = q = 12$, $t = 70$. (c) ARMA(4,4); SVD method with $p_e = q_e = 14$, $t = 50$ yielding $p = 4$. (d) Box–Jenkins ARMA(15,15) model.

$$\sigma_{11} = 18.3, \quad \sigma_{22} = 18.2, \quad \sigma_{33} = 5.30, \quad \sigma_{44} = 4.69$$
$$\sigma_{55} = 0.85, \quad \sigma_{66} = 0.78, \quad \cdots, \quad \sigma_{15,15} = 0.21.$$

It is apparent that the first four singular values are dominant

(i.e., $\nu(4) = 0.995$) thereby indicating that the effective rank of R_e is four. Thus the correct selection of ARMA order $p = q = 4$ is made upon examination of the singular values behavior.

The Box–Jenkins maximum likelihood method (using 150 iterations per run) was next applied to each of the sample time series to generate ten ARMA (15,15) spectral estimates. A plot of these spectra is shown in Fig. 11(d) where a relatively poor performance is in evidence. In addition, the computational expense of the Box–Jenkins method far exceeds that of the techniques which lead to the results shown in Fig. 11(a)–(c).

Comparison with the Kumaresan–Tufts Method: We shall now consider a time series of form (8.3) in which the relevant parameters are given by

$$f_1 = 0.2, \quad f_2 = 0.21, \quad \sigma_w^2 = 1.778.$$

This particular parameter choice provides a more challenging test of resolution capability in that the frequency spacing $f_2 - f_1 = 0.01$ is smaller and the SNR of −5 dB is lower than that of time series (8.3). Again, ten statistically sample runs each of length 128 were used for testing four AR-type models. In the first AR model, (7.2) with choices of $q_e = -1$, $p_e = 35$, $t = 90$ (giving 90 YW equation approximations) were made. Unbiased autocorrelation estimates were then used to form the 90×36 autocorrelation matrix estimate $\hat{R}_e$. Finally, the optimum AR parameter estimates were generated upon using (7.15). The resultant ten AR(35) spectral estimates are shown in superimposed plots in Fig. 12(a) where resolution was achieved in each of the ten runs. Next, the extended autocorrelation matrix model (7.2) with $q_e = -1$, $p_e = 96$, $t = 96$, and unbiased autocorrelation lags was tested. Expression (7.15) with $p = 4$ was then used to generate the a_k estimates of the AR(96) model. A plot of the resultant spectra is shown in Fig. 12(b) where resolution was achieved for each of the ten runs.

The Kumaresan–Tufts method, which provides a near maximum-likelihood performance, was next tested on these same ten sample runs [61]. The resultant AR type 35th and 96th (the optimum KT order choice) order spectra are shown plotted in Fig. 12(c) and (d), respectively. The 35th-order model was unable to resolve the sinusoids in any of the ten runs while the 96th-order model achieved a resolution in each case. For this example, it is apparent that the overextended modeling approach advocated in this paper has outperformed the pseudo-maximum-likelihood Kumaresan–Tufts method. Moreover, the computational efficiency of this paper's overextended modeling method (7.15) is far superior as will be documented in a forthcoming paper.

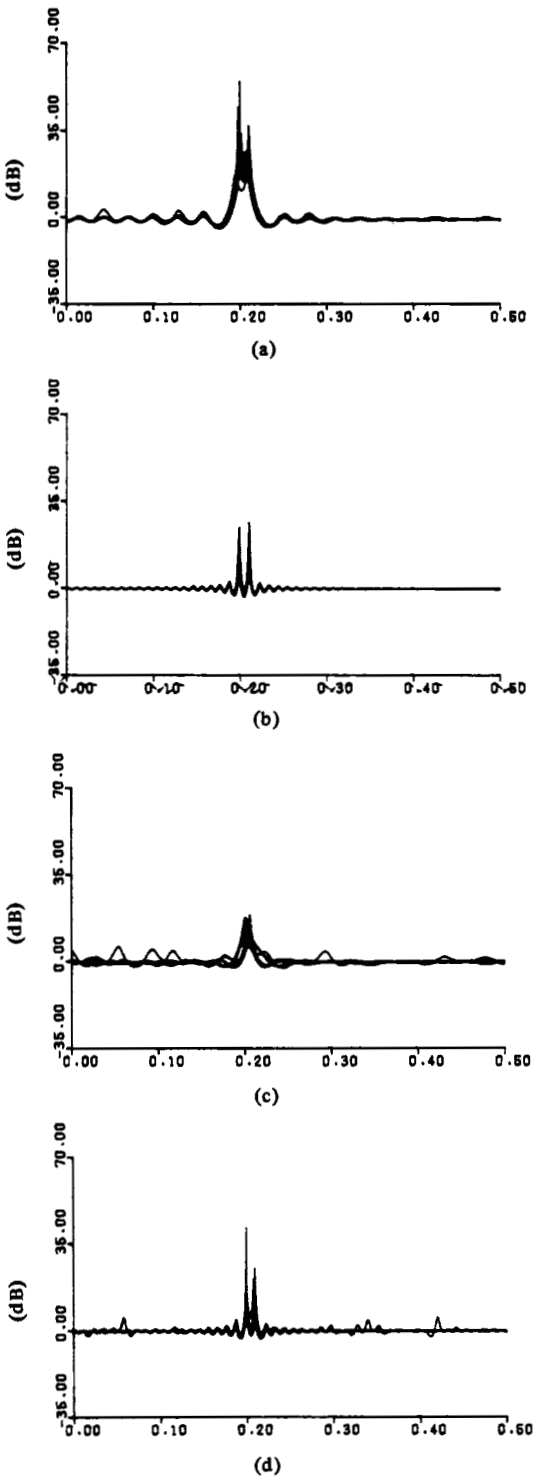

Fig. 12. AR type models. (a) Paper's method (7.15) of order 35. (b) Paper's method (7.15) of order 96. (c) Kumaresan–Tufts model of order 35. (d) Kumaresan–Tufts model of optimum order 96.

Adaptive ARMA Modeling

As a final example, the adaptive ARMA modeling procedure to be developed in Section X was applied to the time series (8.3) in which the covariance mode ($k_1 = 40$, $k_2 = 1$) was selected with ARMA order $p = 12$. The spectral estimates of five independent runs at data length $N = 128$, $N = 256$, and $N = 1024$ are shown superimposed in Fig. 13. From these plots, it is apparent that the twelfth-order ARMA model detects the presence of spectral energy in the neighborhood of $f = 0.2$ at data length $N = 128$, but the resolution of two spectral peaks is somewhat unsatisfactory. As the ARMA model adapts to the data, however, two well defined spectral peaks appear at $N = 256$. The model has, therefore, adapted to the signal using less than 256 time series observations.

To illustrate the performance of this adaptive ARMA approach relative to popularly used alternatives, the classical adaptive AR covariance method to be developed in Section IX was next used on the same set of time series observations. The five spectral plots which arose for an AR(22) model are shown

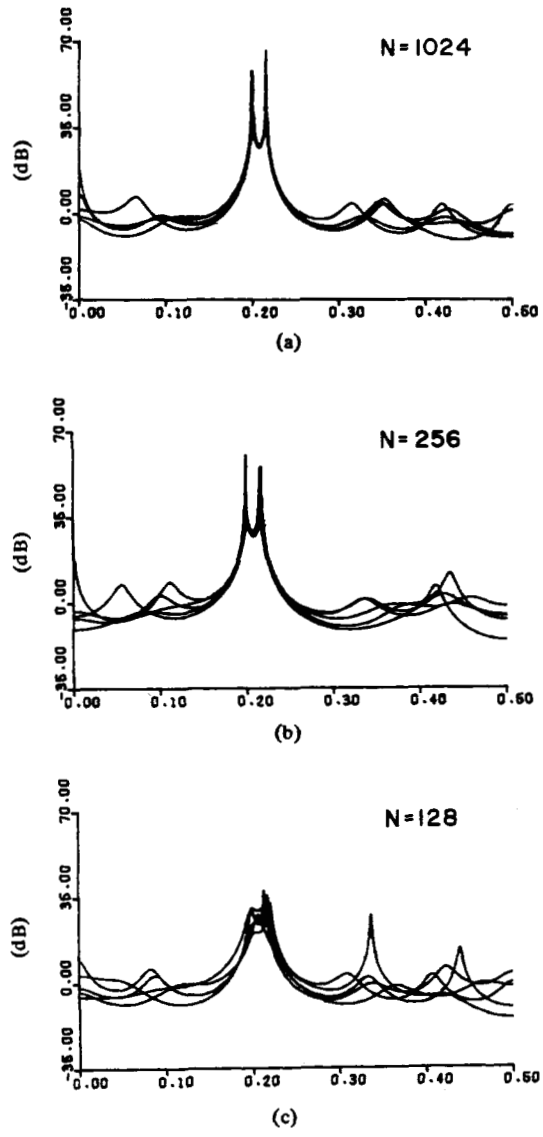

Fig. 13. Adaptive ARMA (12,12) spectral estimates with $k_1 = 40$, $k_2 = 1$. (a) $N = 1024$. (b) $N = 256$. (c) $N = 128$.

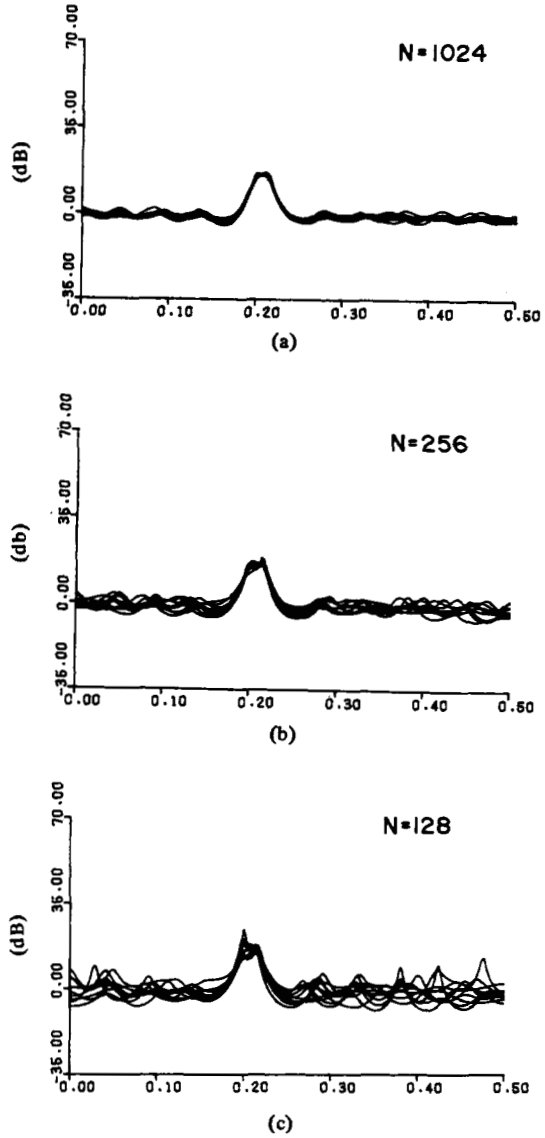

Fig. 14. Adaptive AR(12) spectral estimates using the covariance method. (a) $N = 1024$. (b) $N = 256$. (c) $N = 128$.

superimposed in Fig. 14 at $N = 128$, $N = 256$, and $N = 1024$. Clearly, the higher order covariance AR model was unable to satisfactorily resolve the two sinusoids even at data length $N = 1024$. Thus the lower order ARMA (12, 12) covariance adaptive model significantly outperformed the higher order AR(22) covariance adaptive model. This is indeed noteworthy when it is realized that some of the more widely used adaptive filters utilize the AR covariance model. This includes the fast LMS algorithm of Morf [25], [46], [47] and the approximating gradient approach of Widrow [65].

IX. AR Modeling—Adaptive Implementation

In Section V, a general procedure for effecting an AR model which represents a set of given time series observations

$$x(1), x(2), \cdots, x(N) \tag{9.1}$$

was presented. It was shown there that the required modeling entailed using these time series observations to generate estimates for the entries of the $(p + 1) \times (p + 1)$ AR autocorrelation matrix as specified by

$$R(i, j) = r_x(i - j), \qquad 1 \leqslant i, j \leqslant p + 1. \tag{9.2}$$

From a performance viewpoint, the unbiased autocorrelation estimate (5.4) was suggested as a logical choice for estimating these entries. In any case, once estimates for the $R(i, j)$ elements have been made, the required AR(p) model parameters are obtained by solving the linear system of equations

$$\hat{R}\boldsymbol{a} = |b_0|^2 \, \boldsymbol{e}_1 \tag{9.3}$$

where it will be recalled that the parameter b_0 is chosen so that the first component of $\boldsymbol{a}$ is one.

In applications requiring a continuous updating of the AR model parameters as new time series observations become available (i.e., $x(N + 1)$, $x(N + 2)$, $\cdots$), however, the standard unbiased estimator approach poses a serious computational burden. To overcome this difficulty, it behooves us to seek alternate autocorrelation estimators which are more ameanable

to an adaptive solution. With this objective in mind, we shall now consider the *adaptive class of autocorrelation estimators* as defined by

$$\hat{R}(i,j) = \left(\frac{1}{N+k_2-k_1}\right) \sum_{k=k_1}^{N+k_2-1} \bar{x}(k+1-i)\, x(k+1-j),$$
$$1 \leq i \leq p+1$$
$$1 \leq j \leq p+1 \quad (9.4)$$

in which the convention of setting to zero any summand terms $x(n)$ whose argument lie outside the observation set $1 \leq n \leq N$ has again been adopted. Although this expression might initially appear to be unduly contrived, it does provide us with an autocorrelation estimate of $r_x(i-j)$ as called out for in (9.2). More importantly, however, this estimator will be shortly shown to have a most convenient matrix product representation.

The integer constants k_1 and k_2 which characterize the autocorrelation lag estimator rule (9.4) are to be selected so that the number of lag products there used (i.e., $N+k_2-k_1$) at least equals $p+1$. This requirement will generally ensure the invertibility of the autocorrelation matrix estimate $\hat{R}$ associated with the estimates (9.4). In most cases of interest, these constants are further confined to the range $1 \leq k_1, k_2 \leq p+1$ although other choices are permissable. It then follows that each member of the adaptive autocorrelation estimator class will be identified by a specific two-tuple (k_1, k_2).[7] Moreover, each such estimator will provide a generally different set of autocorrelation estimates from the given set of time series observations (9.1).

Members of the adaptive class of autocorrelation estimators have a particularly convenient algebraic representation which we shall employ when effecting the promised adaptive implementation. Specifically, the $(p+1) \times (p+1)$ autocorrelation matrix estimate that arises upon using the estimates (9.4) as entries can be always expressed in the following *data matrix product* format

$$\hat{R} = \left(\frac{1}{N+k_2-k_1}\right) X_N^* X_N \quad (9.5)$$

in which X_N is the $(N+k_2-k_1) \times (p+1)$ data matrix whose individual elements are specified by

$$X_N(i,j) = x(k_1+i-j), \quad 1 \leq i \leq N+k_2-k_1$$
$$1 \leq j \leq p+1. \quad (9.6)$$

We have here appended the subscript N to the data matrix so as to explicitly recognize its dependency on the data length. The incorporation of this subscript will be also useful when obtaining the promised adaptive implementation. A straightforward analysis will demonstrate the equivalency of (9.4) and (9.5). The data matrix is seen to have elements whose entries are the given time series observations (9.1) as well as zeroes which appear whenever the time index argument (k_1+i-j) falls outside the observation set $1 \leq n \leq N$.

It is possible to provide a revealing visual interpretation to the concept of data windowing for this class of estimators. In particular, let us consider the following $(N+p) \times (p+1)$ *kernel* Toeplitz-type matrix which contains, as submatrices, all of the data matrices associated with the adaptive class of autocorrelation estimators.

$$\mathfrak{X}_N = \begin{bmatrix} x(1) & & \mathbf{0} \\ \vdots & \ddots & \\ x(p+1) & \cdots & x(1) \\ \vdots & & \vdots \\ x(N) & \cdots & x(N-p) \\ & \ddots & \vdots \\ \mathbf{0} & & x(N) \end{bmatrix} \begin{matrix} \longleftarrow k_1 = 1 \\ \text{prewindowing} \\ \longleftarrow k_1 = p+1 \\ \longleftarrow k_2 = 1 \\ \text{postwindowing} \\ \longleftarrow k_2 = p+1 \end{matrix} \quad (9.7)$$

Upon examination of the data matrix definition (9.6), it is apparent that X_N may be identified with that submatrix of $\mathfrak{X}_N$ composed of its k_1th through $(N+k_2-1)$th rows, inclusively. Thus corresponding to each adaptive autocorrelation estimator (i.e., pair (k_1, k_2)), we may obtain the associated data matrix using this row identification scheme.

The zeroes which appear in the upper right corner of the kernel matrix $\mathfrak{X}_N$ are there due to the implicit assumption that $x(n) = 0$ for $-p+1 \leq n \leq 0$. This rather *unrealistic* assumption concerning an unobserved segment of the time series is commonly referred to as a *prewindowing* of the data. It is seen that a degree of prewindowing is incorporated whenever the constant k_1 is selected such that $1 \leq k_1 \leq p$. Normally, such choices are to be avoided since they will generally lead to relatively poor AR modeling due to the unrealistic prewindow assumption thereby being made on the time series. As k_1 ranges over the integers 1 to $p+1$, the degree of prewindowing incorporated varies from full at $k_1 = 1$ to none at $k_1 = p+1$. This prewindowing behavior is conveniently depicted in (9.7).

In a similar fashion, the zeroes which appear in the lower left corner of matrix $\mathfrak{X}_N$ are there due to the implicit *postwindow* assumption that $x(n) = 0$ for $N+1 \leq n \leq N+p$. This equally unrealistic assumption concerning an unobserved segment of the time series is to be generally avoided. A degree of postwindowing is incorporated whenever the index k_2 is chosen to lie in the range $2 \leq k_2 \leq p+1$. The largest value of k_2 for which the postwindow assumption is avoided is seen to be $k_2 = 1$. Thus as the index k_2 ranges from 1 to $p+1$, the degree of postwindowing incorporated varies from none at $k_2 = 1$ to full at $k_2 = p+1$.

The four most widely used of the adaptive autocorrelation estimator methods are listed in Table IV (e.g., see [25], [46], [47]). Each of the methods shown there are seen to entail combinations of maximum windowing and no windowing. In the *covariance* method, the characteristic constants are chosen to be $k_1 = p+1$ and $k_2 = 1$. This particular choice is seen to provide the *largest* number of lag products in the autocorrelation estimates (9.4) over which no data windowing is involved. As might be expected, the covariance method generally provides the best AR modeling and spectral resolution performance when compared with the remaining members of the adaptive autocorrelation estimator class. With this in mind, unless special considerations dictate otherwise, the covariance method is the most preferable choice for an adaptive implementation.

In the three remaining methods listed in Table IV, it is seen that a maximum amount of prewindowing, postwindowing, or

[7] As we will shortly see, the four most widely used members of this class are the covariance method $(k_1 = p+1, k_2 = 1)$, the autocorrelation method $(k_1 = 1, k_2 = p+1)$, the prewindow method $(k_1 = 1, k_2 = 1)$, and, the postwindow method $(k_1 = p+1, k_2 = p+1)$.

TABLE IV
ADAPTIVE AR AUTOCORRELATION ESTIMATION METHODS

METHOD	CONSTANT k_1	CONSTANT k_2	STATISTICAL PROPERTIES OF $\hat{R}$
1. Covariance (No windowing)	p+1	1	(i) unbiased (ii) consistent
2. Full Prewindowing No Postwindowing	1	1	(i) biased (ii) consistent
3. Full Postwindowing No Prewindowing	p+1	p+1	(i) biased (ii) consistent
4. Autocorrelation (Full pre and postwindowing)	1	p+1	(i) biased (ii) consistent (iii) Toeplitz

both are being employed. It is then not surprising that each of these methods will generally provide relatively poor modeling performance. This will be particularly true for data lengths N which are not significantly larger than the AR-order parameter p. As the data length N increases so that $N \gg p$, however, each of the four methods will provide comparable modeling performance. This is due to the fact that the windowed portions of the data matrix will play a proportionately smaller role in the estimate $\hat{R}$ as N increases. An appreciation for this behavior is readily obtained upon examination of the kernel matrix (9.7).

As suggested earlier, the primary reason for preferring the adaptive autocorrelation estimator (9.4) over the standard unbiased estimator (5.4) is that the former may be used to effect a computationally efficient adaptive AR modeling method. To gain an insight as to why this is so, let us first substitute the autocorrelation matrix estimate (9.5) into the fundamental AR modeling expression (9.3). The required parameters of the AR(p) model are then found by solving the resultant system of *normal* equations

$$X_N^* X_N \boldsymbol{a}_N = (N + k_2 - k_1)\,|b_0|^2\, \boldsymbol{e}_1 \tag{9.8}$$

in which the normalizing parameter b_0 is to be selected so that the first component of $\boldsymbol{a}_N$ is one. The data matrix product $X_N^* X_N$ in this expression is seen to *completely* characterize the desired AR parameter vector $\boldsymbol{a}_N$ associated with the N time series observations (9.1).

As the time index N is incremented by one (i.e., the $(N+1)$st time series observation $x(N+1)$ becomes available), it is seen that a new system of normal equations of form (9.8) will arise in which the index N is replaced by $N+1$. The resultant data matrix product $X_{N+1}^* X_{N+1}$ which characterizes this new system of equations will, in turn, give rise to the updated AR parameter vector $\boldsymbol{a}_{N+1}$. We can continue this systematic procedure to generate the updated AR parameter vectors $\boldsymbol{a}_{N+2}$, $\boldsymbol{a}_{N+3}$, etc., as the new time series observations $x(N+2)$, $x(N+3)$, etc., become available. The ability to evolve an adaptive solution procedure when using this approach will be then dependent on our obtaining an effective method for updating the data matrix products $X_N^* X_N$ as N evolves.

Adaptive Algorithm: $k_2 = 1$

The adaptive expression relating the successive data matrix products will be considerably eased if the constant k_2 is selected so as to provide either no or full postwindowing. To illustrate this point let us first examine the case of *nonpostwindowing* for which $k_2 = 1$ while allowing k_1 to take on any value in $[1, p+1]$. From examination of the defining expression (9.6), it is seen that the data matrix X_{N+1} may be obtained by appending a row vector to the bottom of the data matrix X_N. This results in the following recursion on the data matrix products:

$$X_{N+1}^* X_{N+1} = X_N^* X_N + \boldsymbol{x}_{N+1}^* \boldsymbol{x}_{N+1}, \qquad N \geq p+1 \tag{9.9}$$

in which $\boldsymbol{x}_{N+1}$ is the above mentioned appended $1 \times (p+1)$ row vector

$$\boldsymbol{x}_{N+1} = [x(N+1), x(N), \cdots, x(N+1-p)]. \tag{9.10}$$

It is important to note that this data matrix product recursive expression commences at $N = p+1$ which corresponds to the first time index at which $X_N^* X_N$ has its full form. Thus the matrix $X_{p+1}^* X_{p+1}$ serves the role of initializing the above recursive relationship. The elements of this initializing matrix are obtained from (9.4) upon setting $k_2 = 1$ and $N = p+1$, that is,

$$X_{p+1}^* X_{p+1}(i,j) = \sum_{k=k_1}^{p+1} \bar{x}(k+1-i)\, x(k+1-j), \qquad 1 \leq i \leq p+1, \quad 1 \leq j \leq p+1. \tag{9.11}$$

It is interesting to note that although each member of the nonpostwindow class (as identified by $k_2 = 1$ and $k_1 \in [1, p+1]$) will be governed by the *same* recursion (9.9), they will each give rise to a generally different set of autocorrelation estimates. This is due to the fact that the initializing matrix (9.11) will be generally different for various choices of k_1.

From recursive expression (9.9), it is seen that successive data matrix products differ by the rank one matrix $\boldsymbol{x}_{N+1}^* \boldsymbol{x}_{N+1}$. This simple interrelationship will, in turn, enable us to obtain a recursive expression for the data matrix product inverses $[X_N^* X_N]^{-1}$. We are interested in these inverses since they will be ultimately used when solving (9.8) for the AR model parameters. This required matrix inverse recursion will make use of (9.9) and the following well-known lemma:

Lemma 9.1: Let A and $A + \boldsymbol{u}^* \boldsymbol{v}$ each be nonsingular $s \times s$ matrices where $\boldsymbol{u}$ and $\boldsymbol{v}$ are $1 \times s$ vectors, then

TABLE V
ADAPTIVE AR MODELING ALGORITHM—COVARIANCE METHODS ($N_0 = 2p + 1$) AND PREWINDOW ($N_0 = p + 1$) METHODS

Step 0: Input Data: $x(N+1)$, $[X_N^* X_N]^{-1}$
Step 1: Compute $[X_{N+1}^* X_{N+1}]^{-1}$ using recursion (9.13)
Step 2: Let $\underline{c} = [X_{N+1}^* X_{N+1}]^{-1} \underline{a}_1$
Step 3: $\underline{a}_{N+1} = c(1)^{-1} \underline{c}$ where $c(1)$ is the first component of $\underline{c}$.

$$[A + u^* v]^{-1} = A^{-1} - \frac{[A^{-1}u^*][vA^{-1}]}{(1 + vA^{-1}u^*)}. \quad (9.12)$$

Upon setting $A = X_N^* X_N$ and $u = v = x_{N+1}$ in this lemma, the required recursive matrix inverse expression is found to be

$$[X_{N+1}^* X_{N+1}]^{-1} = [X_N^* X_N]^{-1} - \frac{y_{N+1}^* y_{N+1}}{1 + y_{N+1} x_{N+1}^*}, \quad \text{for } N \geq p + k_1 \quad (9.13)$$

where

$$y_{N+1} = x_{N+1} [X_N^* X_N]^{-1}.$$

In using this matrix inverse recursion, it is important to note that it is only applicable for time indices $N \geq p + k_1$. This is a direct consequence of the fact that the data matrix products $X_N^* X_N$ are singular for all time indices $N < p + k_1$ in the nonpostwindowing case $k_2 = 1$. To use this recursive approach, it is therefore necessary to first compute the initializing matrix inverse $[X_N^* X_N]^{-1}$ for $N = p + k_1$ using a standard matrix inversion routine such as Gaussian elimination. Subsequent matrix inverses for $N > p + k_1$ may be then efficiently obtained upon using recursion formula (9.13).

To complete the adaptive AR modeling procedure, we next incorporate the data matrix product inverse routine (9.13) into the AR modeling equations (9.8). A little thought will convince the reader that the simple three-step procedure outlined in Table V will provide the required adaptive AR parameter vector procedure. The second step is seen to yield the unnormalized solution to (9.8) with N replaced by $N + 1$, while the third step ensures that the first component of $\underline{a}_N$ is one. In terms of computational complexity, an examination of (9.13) indicates that $2(p + 1)^2$ operations will be required for updating $[X_N^* X_N]^{-1}$. The resultant AR parameter vector solution as represented by Steps 2 and 3 of Table V will require an additional $(p + 1)$ operation. Thus the computational complexity of the nonpostwindowing adaptive algorithm (i.e., $k_2 = 1$) is then $o(p^2)$. This algorithmic approach is applicable for any selection of the constant k_1 with the most likely choices being from the range $[1, p + 1]$. The most useful implementation of this adaptive algorithm corresponds to the particular selection $k_1 = p + 1$. In this case, the covariance method as specified by $k_1 = p + 1$, $k_2 = 1$ is obtained. As pointed out previously, this choice normally provides the best adaptive AR modeling performance behavior.

Adaptive Algorithm $k_2 = p + 1$

Using similar reasoning, it is also possible to evolve an efficient adaptive algorithm for the full postwindowing case $k_2 = p + 1$. In this situation, it is readily found that the data matrix products are recursively related according to

$$X_{N+1}^* X_{N+1} = X_N^* X_N + D_{N+1}, \quad N \geq p + 1 \quad (9.14)$$

in which D_N is a $(p + 1) \times (p + 1)$ Toeplitz conjugate symmetric matrix with elements

$$D_N(i, j) = \begin{cases} \bar{x}(N+1)\ x(N+1+i-j), & i \leq j \\ x(N+1)\ \bar{x}(N+1+i-j), & j \geq i. \end{cases} \quad (9.15)$$

Due to the Toeplitz conjugate symmetric property of the perturbation matrix D_N, it will be possible to evolve an efficient adaptive method for inverting the data matrix products $[X_N^* X_N]$. The computational complexity of this matrix inversion routine will be also $o(p^2)$. The details of this routine are rather involved and will be therefore not given here due to space limitations. Since the covariance method is the most preferable choice for the adaptive class of autocorrelation estimators, however, this omission is not serious in any case. It is to be noted that efficient adaptive lattice structured algorithms may also be employed for updating the AR parameters [46], [47].

Forward-Backward Approach

In some applications, it is possible to achieve a degree of improvement in the AR spectral models by using the concept of data time reversal. Namely, it makes use of the observation that if $\{x(n)\}$ represents a wide-sense stationary process, then its time transposed conjugated image as specified by

$$y(n) = \bar{x}(s - n) \quad (9.16)$$

will also be wide-sense stationary for any choice of the fixed shift variable s. Moreover, the autocorrelation sequence of this time transposed conjugated image is readily found to be identical to that of the original time series, that is,

$$r_y(n) = r_x(n). \quad (9.17)$$

It is now possible to use this time transpose property to effect a new autocorrelation estimation scheme. In particular, upon selecting $s = N + 1$, the original observation set (9.1) is seen to give rise to the following set of time transposed conjugated elements:

$$y(n) = \bar{x}(N + 1 - n), \quad 1 \leq n \leq N. \quad (9.18)$$

If these time reversed observations are incorporated into (9.4), it will be generally found that a new set of autocorrelation estimates will result. In particular, the overall *backward* autocorrelation matrix estimate will take the form

$$\hat{R} = \left(\frac{1}{N + k_2 - k_1}\right) Y_N^* Y_N \quad (9.19)$$

in which the elements of the $(N+k_2-k_1)\times(p+1)$ matrix Y_N are given by

$$Y_N(i,j)=\bar{x}(N+1-k_1-i+j), \quad 1\leqslant i\leqslant N+k_2-k_1, \quad 1\leqslant j\leqslant p+1. \quad (9.20)$$

Although the forward and backward autocorrelation matrix estimates (9.5) and (9.19) will be generally different (except for the autocorrelation choice $k_1=1$, $k_2=p+1$), they are each seeking to estimate the same underlying autocorrelation matrix R. It then follows the so-called forward-backward estimate as specified by

$$\hat{R}=\left(\frac{1}{2(N+k_2-k_1)}\right)[X_N^*X_N+Y_N^*Y_N] \quad (9.21)$$

will generally provide an additional improvement in autocorrelation estimation fidelity. This is due to the fact that each of the entities $X_N^*X_N$ and $Y_N^*Y_N$ will contain lag products not found in the other.

The additional autocorrelation estimation fidelity achieved in using this time transposition approach typically results in a marginal improvement in spectral estimation performance. Fortunately, this improvement is not accrued at the cost of an excessive increase in computational complexity. This is due to the fact that the matrices X_k and Y_k which form R are Toeplitz type. It is, therefore, possible to devise efficient algorithms that will solve the system of equations

$$[X_N^*X_N+Y_N^*Y_N]\,a_N=\alpha e_1 \quad (9.22)$$

in which the computational complexity is $o(p^2)$.

AR Model Order Determination

One of the primary considerations in obtaining AR models from raw time series observations is that of model order selection. It has been observed that when p is selected too low, there will be generally too few model poles to adequately represent the underlying spectrum. On the other hand, too high of a choice for p will typically result in spurious effects (e.g., false peaks) in the spectral estimate. With these thoughts in mind, investigators have proposed various order selection procedures. Three of the more widely used techniques are Akiake's final prediction error method as well as his information criterion [1], [2], [4], and Parzen's "criterion autoregressive transfer" function [54]. Although these procedures typically work well, they can yield unsatisfactory performance in some cases (e.g., see [34] and [62]). The user is, therefore, cautioned to use discretion in applying the above and other model order determination procedures. The method to be ultimately used should be determined through empirical experimentation based on time series related to the specific application under consideration.

It is possible to apply a conceptionally straightforward procedure for model order selection which does not possess many of the drawbacks alluded to above. It is based on the observation that in the case where perfect AR autocorrelation lag values are given, the $(p+1)\times(p+1)$ autocorrelation matrix R with elements

$$R(i,j)=r_x(i-j), \quad 1\leqslant i,j\leqslant p+1 \quad (9.23)$$

will have rank $p+1$ so long as p is less than or equal to the order of the underlying AR process (hereafter taken to be p_1). For all values of p greater than p_1, however, the rank of the $(p+1)\times(p+1)$ autocorrelation matrix will be p_1+1. Thus to determine the proper rank selection in the idealistic case of perfect autocorrelation lag information, we simply increase the parameter p until the rank of R is less than full (i.e., less than $p+1$). This will occur at $p=p_1+1$, thereby giving us the appropriate order selection. It should be noted that when the autocorrelation lags being used do not correspond to an AR process, then the matrix R will be generally of full rank for *all* $p\geqslant 1$.

In the more realistic case in which raw time series are used to form the autocorrelation matrix estimate $\hat{R}$, the rank of this matrix will be typically full for *all* values of p. This will be true even when the time series is an AR process. This seeming contradiction arises due to statistical errors inherent in any autocorrelation lag estimation procedure that might be used in forming $\hat{R}$. Nonetheless, even though R will have full rank, it will be generally found that when $p>p_1$, this matrix will have $(p-p_1)$ of its eigenvalues "close" to zero. Thus an order selection procedure which has provided satisfactory performance is one entailing examination of the eigenvalue behavior of the autocorrelation matrix estimate $\hat{R}$ as a function of p. The appropriate order choice will be that value of p, denoted by p_1, for which $\hat{R}$ has $(p-p_1)$ of its eigenvalues sufficiently close to zero for all $p>p_1$. A particularly attractive method (i.e., the SVD method) for implementing this procedure was given in Section VII.

X. ARMA Modeling—Adaptive Implementation

When an adaptive implementation of the ARMA modeling methods as described in Sections VI and VII is desired, it will be necessary to incorporate autocorrelation lag estimate procedures which are compatible with an adaptive implementation (the unbiased estimator is not compatible). In particular, we shall now examine a class of estimators which provides an adaptive mechanism for estimating the elements of the autocorrelation matrix estimate $\hat{R}_1$ as required in (6.6). This class of adaptive estimators will be governed by the relationship

$$\hat{R}_1(i,j)=\left(\frac{1}{N+k_2-k_1-q-1}\right)\cdot\sum_{k=k}^{N-q-2+k_2}\bar{x}(k+1-i)\,x(k+q+2-j), \quad 1\leqslant i\leqslant t, \quad 1\leqslant j\leqslant p+1. \quad (10.1)$$

It is apparent that this expression provides an estimate for the lag element $r_x(q+1+i-j)$ which is the (i,j)th element of the autocorrelation matrix R_1 as defined in (6.2). The fixed constants k_1 and k_2 which characterize this estimator are normally selected so that the number of lag products there used (i.e., $N+k_2-k_1-q-1$) equals or exceeds $p+1$. This choice will generally ensure the invertibility of matrix $\hat{R}_1^*W\hat{R}_1$ and thereby a unique solution for the AR parameter when using (6.6). For reasons which will be shortly made apparent, these constants are usually further constrained to satisfy $1\leqslant k_1\leqslant t$ and $1\leqslant k_2\leqslant p+1$ although other choices are possible.

Each autocorrelation estimator in the adaptive class (10.1) will be identified by a particular choice of the two-tuple (k_1,k_2). Moreover, each estimator in this class will provide a generally different set of autocorrelation lag estimates from the set of time series observations $x(n)$ for $1\leqslant n\leqslant N$. Clearly,

our ultimate desire is to select that estimator which generally provides the best ARMA modeling. The *covariance* estimator as identified by $k_1 = t$ and $k_2 = 1$ furnishes an obvious choice. Before treating specific estimators, however, let us first examine the general adaptive estimator (10.1).

The primary reason as to why estimator (10.1) lends itself to an adaptive implementation is due to the algebraic structure implicitly contained within its definition. Namely, the autocorrelation matrix estimate as formed from the entries (10.1) may be representable in the convenient matrix product format

$$\hat{R}_1 = \left(\frac{1}{N + k_2 - k_1 - q - 1}\right) Y_N^* X_N \qquad (10.2)$$

in which the $(N + k_2 - k_1 - q - 1) \times (p + 1)$ data matrix X_N has its elements specified by

$$X_N(i, j) = x(k_1 + q + 1 + i - j), \quad 1 \leqslant i \leqslant N + k_2 - k_1 - q - 1, \quad 1 \leqslant j \leqslant p + 1 \qquad (10.3)$$

while the $(N + k_2 - k_1 - q - 1) \times t$ data matrix Y_N has elements

$$Y_N(i, j) = x(k_1 + i - j), \quad 1 \leqslant i \leqslant N + k_2 - k_1 - q - 1, \quad 1 \leqslant j \leqslant t. \qquad (10.4)$$

A simple matrix manipulation will prove the equivalencies of (10.1) and (10.2). We again adopt the convention of setting to zero any element entries of X_N or Y_N for which $x(n)$ lies outside the observation interval $1 \leqslant n \leqslant N$, and we also attach the subscript N to these data matrices so as to explicitly recognize their dependency on data length.

As in the AR modeling case, the parameters k_1 and k_2 that identify the autocorrelation estimator (10.1) can give rise to data windowing. To see why this is so, let us consider two kernel Toeplitz-type matrices which contain, as submatrices, all of the data matrices associated with the adaptive class of ARMA autocorrelation estimators. These kernel matrices are specified by

Upon examination of (10.3) and (10.4), it is readily established that the data matrix X_N (or Y_N) may be identified with that submatrix of the kernel matrix $\mathfrak{X}_N$ (or $\mathfrak{Y}_N$) composed of its k_1th through $(N - q - 2 + k_2)$th rows inclusively. Thus corresponding to each adaptive autocorrelation estimator (i.e., choice of pair (k_1, k_2)), there will be an associated pair of data matrices obtained by using this row identification scheme.

The zeroes which appear in the upper right corner of kernel matrix $\mathfrak{Y}_N$ are there due to the implicit prewindow assumption that $x(n) = 0$ for $2 - t \leqslant n \leqslant 0$. This unrealistic restriction on an unobserved segment of the time series is to be normally avoided. It is to be noted from the representation for Y_N that a selection of $k_1 \geqslant t$ will avoid any data prewindowing. On the other hand, a degree of prewindowing is incorporated whenever k_1 is such that $1 \leqslant k_1 \leqslant t - 1$. Thus as k_1 ranges over the integers 1 to t, the amount of prewindowing varies from full at $k_1 = 1$ to none at $k_1 = t$.

In a similar fashion, the zeroes which arise in the lower left corner of kernel matrix $\mathfrak{X}_N$ are there due to the implicit postwindow assumption that $x(n) = 0$ for $N + 1 \leqslant n \leqslant N + p$. This contrived assumption on an unobserved segment of the time series is also to be avoided. Upon examination of the kernel matrix $\mathfrak{X}_N$, it is apparent that postwindowing may be avoided by selecting $k_2 \leqslant 1$. It is also clear that the degree of postwindowing varies from none at $k_2 = 1$ to full at $k_2 = p + 1$.

The four most appealing choices for adaptive estimators are identified in Table VI in which it is noted that each involves combinations of maximum windowing and no windowing. The *covariance* method entails that particular combination of no prewindowing (i.e., $k_1 = t$) and no postwindowing (i.e., $k_2 = 1$). This method is seen to provide the *largest* number of lag products (i.e., $N - q - t$) in estimator (10.1) for which no data windowing is involved. As might be expected, the *covariance* method typically provides the best modeling performance from the ARMA adaptive class of autocorrelation estimators.

The three other methods listed in Table VI are seen to employ either full prewindowing, full postwindowing, or both. It is clear that the modeling performance capabilities of each of these three methods will tend to be relatively poor when the

$$\mathfrak{X}_N = \begin{matrix} \\ k_2 = 1 \\ \text{postwindowing} \\ k_2 = p + 1 \end{matrix} \begin{matrix} \\ \uparrow \\ \\ \downarrow \end{matrix} \begin{bmatrix} x(q+2) & \cdots & x(q-p+2) \\ \vdots & & \\ x(t+q+1) & \cdots & x(t+q-p+1) \\ \vdots & & \vdots \\ x(N) & \cdots & x(N-p) \\ \mathbf{0} & \ddots & \vdots \\ & & x(N) \end{bmatrix}$$

$$\mathfrak{Y}_N = \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ k_1 = 1 \\ \uparrow \\ \text{prewindowing} \\ \downarrow \\ k_1 = t \\ \\ \end{matrix} \begin{bmatrix} x(1) & & \mathbf{0} \\ \vdots & \ddots & \\ x(t) & \cdots & x(1) \\ \vdots & & \vdots \\ x(N-q-1) & \cdots & x(N-q-t) \\ \vdots & & \vdots \\ x(N+p-q-1) & \cdots & x(N+p-q-t) \end{bmatrix} \qquad (10.5)$$

TABLE VI
FOUR ARMA ADAPTIVE AUTOCORRELATION ESTIMATOR METHODS

METHOD	CONSTANT k_1	CONSTANT k_2	STATISTICAL PROPERTIES OF R_N
1. Covariance (No windowing)	t	1	(i) unbiased (ii) consistent
2. Full Prewindowing No Postwindowing	1	1	(i) biased (ii) consistent
3. Full Postwindowing No Prewindowing	t	p+1	(i) biased (ii) consistent
4. Autocorrelation (Full pre and postwindowing)	1	p+1	(i) biased (ii) consistent (iii) Toeplitz

data length N is only marginally larger than the ARMA order parameter p or the parameter t. On the other hand, for the case in which N is much larger than either p or t, each of the methods listed in Table VI will provide comparable modeling performance. This is a consequence of the fact that the windowed portions of the data matrices X_N and Y_N play a proportionately smaller role in the estimate of R_1 as N increases. In any case, unless special considerations dictate otherwise, the covariance method is the most preferable choice for an adaptive implementation.

In order to provide the reason as to why members of the adaptive class of autocorrelation estimators are amenable to an adaptive implementation, let us substitute the matrix product representation for R_1 as given by (10.2) into the basic ARMA modeling equation (6.6). The resultant AR parameter vector is then obtained by solving the normal equations

$$X_N^* Y_N Y_N^* X_N a_N = \alpha e_1 \tag{10.6}$$

where the weighting matrix W has been set equal to the identity matrix while the normalizing constant α is selected so that the first component of a_N is one. From this expression it is apparent that the *data matrix product* $X_N^* Y_N Y_N^* X_N$ completely identifies the ARMA model's AR parameter vector. In order to compute a_N, it will be then necessary to evaluate the data matrix product's inverse for each value of N in which the AR parameter vector is required. This can be a particularly imposing computational task if real-time signal processing is to be achieved.

Adaptive Alogrithm: $k_2 = 1$

When the AR parameter vector is required at each time index N, it will be beneficial to effect an adaptive method for updating the data matrix product in (10.6). This adaptive implementation is readily achieved for the nonpostwindowing case $k_2 = 1$ in which k_1 may take on any appropriate value (e.g., $1 \leqslant k_1 \leqslant t$). Namely, upon examination of (10.5) with $k_2 = 1$, it is seen that the data matrices X_{N+1} and Y_{N+1} are obtained by appending appropriate row vectors to the bottom of data matrices X_N and Y_N, respectively. Using this property, the following recursion on the data matrix product is obtained:

$$Y_{N+1}^* X_{N+1} = Y_N^* X_N + y_N^* x_N, \qquad N \geqslant t \tag{10.7a}$$

where x_N and y_N are the above mentioned $1 \times (p+1)$ and $1 \times t$ row vectors, that are appended to X_N and Y_N, respectively. These vectors are specified by

$$x_N = [x(N+1), x(N), \cdots, x(N+1-p)] \tag{10.7b}$$

$$y_N = [x(N-q), x(N-q-1), \cdots, x(N+1-q-t)]. \tag{10.7c}$$

It is to be noted that recursive expression (10.7a) holds only for time indices $N \geqslant t$ since t is the first time index where $Y_N^* X_N$ takes on its full algebraic form. With this in mind, $Y_t^* X_t$ then serves the role of an initializing matrix for this recursion. Although the perturbation matrix $Y_N^* X_N$ in this recursion does not depend on the parameter k_1, the initializing matrix $Y_t^* X_t$ does. As such, the sequence of matrix products as generated by (10.7a) will be different for various choices of k_1.

The full data matrix product as required in (10.6) may be readily obtained from (10.7) and takes on the following recursive form:

$$X_{N+1}^* Y_{N+1} Y_{N+1}^* X_{N+1} = X_N^* Y_N Y_N^* X_N + z_N^* z_N + x_N^* z_N + (y_N y_N^*) x_N^* x_N, \qquad N \geqslant t \tag{10.8a}$$

where z_N is the $1 \times (p+1)$ vector given by

$$z_N = y_N Y_N^* X_N. \tag{10.8b}$$

An examination of this recursive expression indicates that $t(p+1)$ operations are required to compute z_N while another $2(p+1)^2$ operations are expended in updating the full matrix product (10.8a). In arriving at this computational requirement measure, it has been tacitly assumed that the matrix products $Y_N^* X_N$ and $X_N^* Y_N Y_N^* X_N$ are available.

When updates of the AR parameter vectors a_N are not required at each time index N, we could then use recursions (10.7) and (10.8) to compute the data matrix products $Y_N^* X_N$ and $X_N^* Y_N Y_N^* X_N$ in a computationally efficient manner. At those time instants at which the evaluation of a_N is required, we would then simply solve the ARMA modeling equations (10.6). If standard procedures are used, this solution will entail on the order of $(p+1)^3$ computations.

In various applications, however, it may be necessary to compute the AR parameter vector at each (or nearly each)

TABLE VII
ADAPTIVE ARMA ALGORITHM FOR COMPUTING a_{N+1}

STEP 0	The input to commence the algorithm at $N = q+p+k_1+1$ is $Y_N^* X_N$ and $[X_N^* Y_N Y_N^* X_N]^{-1}$
STEP 1	$N = q+p+k_1+1$
STEP 2	Compute $Y_{N+1}^* X_{N+1}$ using expression (10.7)
STEP 3	$\underline{z}_N = \underline{y}_N Y_N^* X_N$
STEP 4	$\underline{u}_1 = \underline{z}_N$, $\underline{v}_1 = \underline{z}_N$, $A_1^{-1} = [X_N^* Y_N Y_N^* X_N]^{-1}$ Compute $[A_1 + \underline{u}_1^* \underline{v}_1]^{-1}$ using Lemma 9.1
STEP 5	$\underline{u}_2 = \underline{x}_N^*$, $\underline{v}_2 = \underline{z}_N$, $A_2^{-1} = [A_1 + \underline{u}_1^* \underline{v}_1]^{-1}$ Compute $[A_2 + \underline{u}_2^* \underline{v}_2]^{-1}$ using Lemma 9.1
STEP 6	$\underline{u}_3 = (Y_N Y_N^*) \underline{x}_N$, $\underline{v}_3 = \underline{z}_N$, $A_3^{-1} = [A_2 + \underline{u}_2^* \underline{v}_2]^{-1}$ Compute $[A_3 + \underline{u}_3^* \underline{v}_3]^{-1} = [X_{N+1}^* Y_{N+1} Y_{N+1}^* X_{N+1}]^{-1}$ using Lemma 9.1
STEP 7	$\underline{c} = [X_{N+1}^* Y_{N+1} Y_{N+1}^* X_{N+1}]^{-1} \underline{a}_1$ $\underline{a}_{N+1} = c(1)^{-1} \underline{c}$ where $c(1)$ is the first component of $\underline{c}$
STEP 8	Let $N = N+1$, GO TO STEP 2

value of time N. For such cases, it would be much more advantageous to replace the recursion (10.8) by a recursion for the inverse matrix $[X_N^* Y_N Y_N^* X_N]^{-1}$. To effect this recursion, we note from (10.8) that the matrix products $X_N^* Y_N Y_N^* X_N$ at two contiguous time indicies (i.e., N and $N+1$) differ by the sum of three rank-one matrices. Using this fact, it is then possible to apply Lemma 9.1 successively three times to effect the desired matrix product inverse recursion. The main steps of this recursive inversion are listed in Table VII. It is important to note that this recursion commences at $N = q + p + k_1 + 1$ which corresponds to the first time instant at which the matrix product $X_N^* Y_N Y_N^* X_N$ is generally invertible. Steps 3 through 6 provide the mechanism for this matrix product inversion while Step 7 gives the required solution to the ARMA modeling equations (10.6). In terms of computational complexity, it is readily shown that the number of multiplication and addition operations required to implement this algorithm is of order $p(t + 3p)$ for each data point update.

The adaptive algorithm described in Table VII is for the particular nonpostwindowing selection $k_2 = 1$ wherein the parameter k_1 will be typically selected to satisfy $1 \leqslant k_1 \leqslant t$. As suggested earlier, the covariance method identified by $k_1 = t$ and $k_2 = 1$ generally provides the best overall modeling performance for the class of adaptive estimators. We may, therefore, use this adaptive ARMA modeling method to provide an efficient procedure for recursively implementing the desirable covariance method. As a final note, although it is possible to effect adaptive implementations for other choices of k_1 (i.e., $k_1 \neq 1$), the resultant algorithm is of a much more complex nature. Since the covariance method is almost invariably used, however, we shall be content with the nonpostwindowing algorithm.

It is also possible to provide a lattice implementation of the adaptive algorithm here developed [19] and [49]. This will entail restricting $t = p$ thereby imparting a decrease in spectral estimation performance. The advantage accrued by using the lattice implementation is computational in nature. In particular, the number of operations to update the lattice network is found to be $o(p \log_2 p)$ for each new time series observation.

XI. CONCLUSIONS

A philosophy directed towards the rational modeling of wide-sense stationary time series has been presented. It is explicitly based upon the Yule–Walker equations which characterize the autocorrelation sequence associated with the rational time series being modeled. In particular, the key concept is that of using an overdetermined set of Yule–Walker equation evaluations for estimating the parameters of a postulated rational model. This approach has been found to reduce the data-induced hypersensitivity of the parameter estimates in comparison to many of the more popular parametric approaches which invoke a minimal set of evaluations for obtaining the parameter estimates. These latter methods include the Burg algorithm, many LMS methods, and the one-step predictor. Comparative examples illustrating this reduced hypersensitivity have been given in which the modeling is based on both exact autocorrelation lag information, and raw time series observations.

The method of singular value decomposition was next introduced and was used to obtain an effective rational model order determination procedure as well as providing a novel rational modeling procedure whose performance has been empirically found to typically exceed that of existing techniques. Studies are currently under way to more effectively use this SVD adaption for achieving yet further performance improvements.

ACKNOWLEDGMENT

The author would like to acknowledge the many contributions of B. Baseghi who prepared the numerical examples, helped proofread, and participated in a valuable manner in the manuscript's preparation.

REFERENCES

[1] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statist. Math.*, vol. 21, pp. 243–247, 1969.
[2] ——, "Power spectrum estimation through autoregressive model fitting," *Ann. Inst. Statist. Math.*, vol. 21, pp. 407–419, 1969.
[3] ——, "Maximum likelihood identification of Gaussian autoregres-

sive moving-average model's," *Biometrica*, vol. 60, no. 2, pp. 255-265, Aug. 1973.
[4] ——, "A new look at the statistical model identificant," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716-723, Dec. 1974.
[5] N. O. Anderson, "On the calculation of filter coefficients for maximum entropy spectral analysis," *Geophys.*, vol. 39, pp. 69-72, Feb. 1974.
[6] G. F. Ansley, "An algorithm for the exact likelihood of a mixed autoregressive moving-average process," *Biometrika*, vol. 66, no. 1, pp. 59-65, Apr. 1979.
[7] A. A. Beex and L. L. Scharf, "Covariance sequence approximation for parametric spectrum modeling," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 5, pp. 1042-1052, Oct. 1981.
[8] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*. New York: Dover 1959.
[9] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (revised ed.). San Francisco, CA: Holden Day, 1976.
[10] T. P. Bronez and J. A. Cadzow, "An algebraic approach to super-resolution array processing," in *1981 ICASSP Proc.*, pp. 302-305 (Atlanta, GA, 1981) and accepted for publication in the *IEEE Trans. Aerosp. Electron. Syst.*
[11] S. Bruzzone and M. Kaveh, "On some suboptimum ARMA spectral estimators," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 753-754, Dec. 1980.
[12] K. J. Bry and J. LeRoux, "Comparison of some algorithms for identifying autoregressive signals in the presence of observation noise," in *1982 ICASSP* (Paris, France), pp. 224-227, May 1982.
[13] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meet. Society of Exploration Geophysicists* (Oklahoma City, OK, Oct. 31, 1967).
[14] ——, "Maximum entropy spectral analysis," Ph.D. dissertation, Dep. Geophysics, Stanford Univ., Stanford, CA, May 1975.
[15] J. A. Cadzow, "ARMA spectral estimation: An efficient closed form procedure," in *Proc. RADC Spectrum Estimation Workshop*, pp. 81-97, Oct. 1979.
[16] ——, "High performance spectral estimation—A new ARMA method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 524-529, Oct. 1980.
[17] ——, "Autoregressive-moving average spectral estimation: A model equation error procedure," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-19, pp. 24-28, Jan. 1981.
[18] J. A. Cadzow and K. Ogino, "Two-dimensional spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 396-401, June 1981.
[19] J. A. Cadzow and R. L. Moses, "An adaptive ARMA spectral estimator, parts 1 and 2," presented at the First ASSP Workshop on Spectral Estimation, McMaster Univ., Hamilton, Ont., Canada, Aug. 1981.
[20] J. A. Cadzow, "ARMA modeling of time series," *IEEE Trans. Patt. Analys. Mach. Intel.* (Special Issue on Digital Signal and Waveform Analysis), vol. PAMI-4, no. 2, pp. 124-128, Mar. 1982.
[21] ——, "Rational time series modeling: An effective method," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-18, Aug. 1982.
[22] J. A. Cadzow and B. Baseghi, "Data adaptive ARMA modeling of time series," in *1982 ICASSP* (Paris, France), pp. 256-261, May 1982.
[23] D. G. Childers, Ed., *Modern Spectral Analysis*. New York: IEEE Press, 1978.
[24] J. Durbin, "The fitting of time series models," *Rev. Inst. Int. de Stat.*, vol. 28, pp. 233-244, 1960.
[25] 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 Appl.*, vol. 27, pp. 31-60, 1979.
[26] W. Gersh, "Estimation of the autoregressive parameters of mixed autoregressive moving-average time series," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 583-588, Oct. 1970.
[27] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *J. SIAM Numer. Anal.* (Ser. B), vol. 2, no. 2, pp. 205-224, 1965.
[28] 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.
[29] N. K. Gupta and R. K. Mehra, "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations," *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, pp. 774-783, Dec. 1974.
[30] P. R. Gutowski, E. A. Robinson, and S. Treitel, "Sepctral estimation: Fact or fiction?," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-16, pp. 80-84, Apr. 1978.
[31] S. S. Haykin, Ed., *Nonlinear Methods of Spectral Analysis*. New York: Springer, 1979.
[32] T. L. Henderson, "Geometric methods for determining system poles from transient response," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 5, pp. 982-988, Oct. 1981.
[33] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*. San Francisco, CA: Holden-Day, 1968.
[34] R. L. Kashyap, "Inconsistency of the AIC rule for estimating the order of autoregressive models," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 996-998, Oct. 1980.
[35] M. Kaveh, "High resolution spectral estimation for noisy signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 286-287, June 1979.
[36] S. M. Kay, "A new ARMA spectral estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 585-588, Oct. 1980.
[37] S. M. Kay and S. L. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380-1419, Nov. 1981.
[38] J. F. Kinkel, J. Perl, L. Scharf, and A. Stubberub, "A note on covariance-invariant digital filter design and autoregressive-moving average spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 200-202, Apr. 1979.
[39] V. C. Klema and A. J. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 164-176, Apr. 1980.
[40] I. S. Konralinka and M. R. Matausek, "Simultaneous estimation of poles and zeroes in speech analysis and ITIF iterative inverse filtering algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, no. 5, pp. 485-492, Oct. 1979.
[41] L. M. Koopmans, *The Spectral Analysis of Time Series*. New York: Academic Press, 1974.
[42] R. Kumaresan and D. W. Tufts, "Singular value decomposition and spectral analysis," presented at the First ASSP Workshop on Spectral Estimation, McMasters Univ., Hamilton, Ont., Canada, Aug. 1981.
[43] N. Levinson, "The Wiener (root mean square) error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261-278, 1947.
[44] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 423-428, Oct. 1977.
[45] R. K. Mehra, "On-line identification of linear dynamic systems with applications to Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. AC-16, no. 1, pp. 12-22, Feb. 1971.
[46] M. Morf, T. Kailath, and L. Ljung, "Fast algorithms for recursive identification," in *Proc. 1976 IEEE Conf. Decision and Control* (Clearwater, FL), pp. 916-921, Dec. 1976.
[47] M. Morf, B. Dickinson, T. Kailath, and A. Vieira, "Efficient solutions of covariance equations for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 429-433, Oct. 1979.
[48] P. Newbold, "The exact likelihood function of a mixed autoregressive moving-average process," *Biometrika*, vol. 61, no. 3, pp. 423-426, Dec. 1974.
[49] K. Ogino, "Computationally fast algorithms for ARMA spectral estimation," Ph.D. dissertation, Viriginia Polytech. Inst., Blacksburg, VA, 1981.
[50] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[51] Y. Pao and D. T. Lee, "Performance characteristics of the Cadzow modified direct ARMA method for spectrum estimation," presented at the First IEEE-ASSP Workshop on Spectral Estimation, McMaster Univ., Hamilton, Ontario, Canada, Aug. 1981.
[52] E. Parzen, "Mathematical considerations in the stimation of spectra," *Technometrics*, vol. 3, pp. 167-190, May 1961.
[53] ——, "Statistical spectral analysis (single channel case) in 1968," Dept. Statistics, Stanford Univ., Stanford, CA, Tech. Rep. 11, June 1968.
[54] ——, "Some recent advances in time series modeling," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 723-730, Dec. 1974.
[55] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Royal Astron. Soc.*, pp. 347-366, 1973.
[56] R. Prony, "Essai experimentale et analytique," *Paris J. l'Ecole Polytechnique*, vol. 1, pp. 24-76, 1795.
[57] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[58] 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.
[59] J. L. Shanks, "Recursion filters for digital processing," *Geophys.*, vol. 32, pp. 33-51, Feb. 1967.
[60] S. A. Tretter and K. Steiglitz, "Power spectrum identifications in terms of rational models," *IEEE Trans. Automat. Contr.*, vol.

AC-12, pp. 185-188, Apr. 1967.
[61] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," this issue, pp. 975-989.
[62] T. J. Ulrych and R. W. Clayton, "Time series modelling and maximum entropy," *Phys. Earth Planetary Interiors*, vol. 12, pp. 188-200, Aug. 1976.
[63] G. Walker, "On periodicity in series of related terms," *Proc. Roy. Soc. London, Ser. A.*, vol. 131, pp. 518-532, 1931.
[64] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70-73, June 1970.
[65] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1694-1716, Dec. 1975.
[66] G. U. Yule, "On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers," *Philosoph. Trans. Roy. Soc. London, Ser. A.*, vol. 226, pp. 276-298, July 1927.

On The Rationale of Maximum-Entropy Methods

EDWIN T. JAYNES

Invited Paper

Abstract—We discuss the relations between maximum-entropy (MAXENT) and other methods of spectral analysis such as the Schuster, Blackman-Tukey, maximum-likelihood, Bayesian, and Autoregressive (AR, ARMA, or ARIMA) models, emphasizing that they are not in conflict, but rather are appropriate in different problems. We conclude that:

1) "Orthodox" sampling theory methods are useful in problems where we have a known model (sampling distribution) for the properties of the noise, but no appreciable prior information about the quantities being estimated.

2) MAXENT is optimal in problems where we have prior information about multiplicities, but no noise.

3) The full Bayesian solution includes both of these as special cases and is needed in problems where we have both prior information and noise.

4) AR models are in one sense a special case of MAXENT, but in another sense they are ubiquitous in all spectral analysis problems with discrete time series.

5) Empirical methods such as Blackman-Tukey, which do not invoke even a likelihood function, are useful in the preliminary, exploratory phase of a problem where our knowledge is sufficient to permit intuitive judgments about how to organize a calculation (smoothing, decimation, windows, prewhitening, padding with zeroes, etc.) but insufficient to set up a quantitative model which would do the proper things for us automatically and optimally.

I. Introduction

THIS PAPER concerns what is in one sense a small detail in the context of the vast amount of work done on spectral analysis. But in another sense, we are concerned with the general principles underlying all scientific inference, in which context spectral analysis is only one specialized application. Our aim is to clarify some currently puzzling questions about rationale and method.

There are many different spectral analysis problems, corresponding to different kinds of prior information about the phenomenon being observed, different kinds of data, different kinds of perturbing noise, and different objectives. It is, therefore, quite meaningless to pass judgment on the merits of any proposed method unless one specifies clearly: "In what class of problems is this method intended to be used?"

Most of the current confusion on these questions is, in the writer's opinion, the direct result of failure to define the problem explicitly enough. Today, programming and running a computer is much easier than actually thinking about a problem, so one may program an algorithm appropriate to one kind of problem, and then feed it the data of an entirely different problem. If the result is unsatisfactory, there is an undertandable tendency to blame the algorithm and the method that produced it, rather than the faulty application.

The maximum-entropy (MAXENT) method is particularly vulnerable in this respect, because its rationale is so different from that of "orthodox" statistics that it seems new and mysterious to many (although historically it dates back to Boltzmann, 1877). To compound the confusion, the MAXENT spectral estimate is, for one particular kind of data, identical in analytical form with that of an AR model, as found by Burg [1].

If that were not enough, any MAXENT solution also defines a particular model for which the predictive distribution using the maximum-likelihood estimates of the parameters, is identical with the MAXENT distribution. This is essentially the

Manuscript received March 1, 1982; revised May 21, 1982.
The author is with Arthur Holly Compton Laboratory of Physics, Washington University, St. Louis, MO 63130.

0018-9219/82/0900-0939$00.75