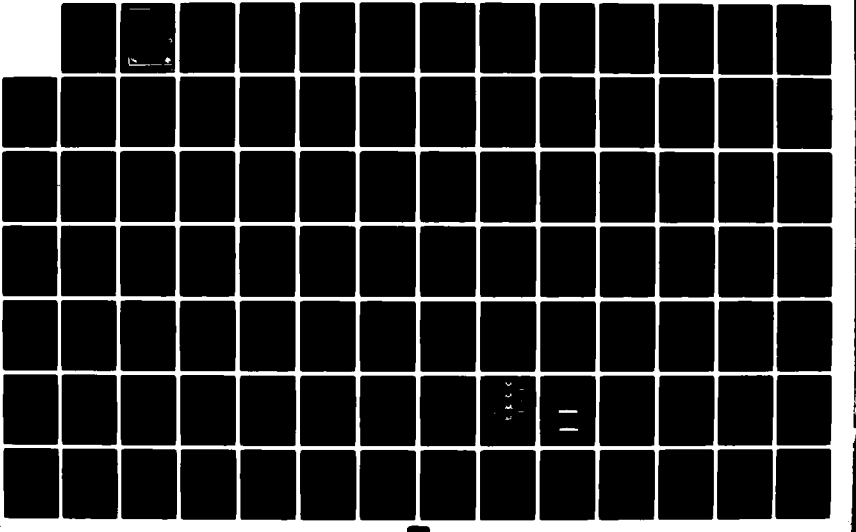
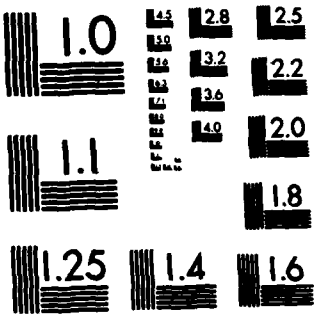
AD-A123 122 SPECTRAL ESTIMATION: AN OVERDETERMINED RATIONAL MODEL EQUATION APPROACH..(U) ARIZONA STATE UNIV TEMPE DEPT OF ELECTRICAL AND COMPUTER ENGI.. J A CADZOW 15 SEP 82 1/2

UNCLASSIFIED AFOSR-TR-82-1050 N00014-82-K-0257 F/G 12/1 NL





AFOSR-TR- 82-1050

(12)

electrical and computer engineering

SPECTRAL ESTIMATION: AN OVERDETERMINED RATIONAL MODEL EQUATION APPROACH

FINAL REPORT

GRANT AFOSR-81-0250

Submitted to

Air Force Office of Scientific Research
Building 410
Boiling AFB, D.C. 20332

Principal Investigator: Dr. James A Cadzow
September 15, 1982

College of Engineering & Applied Sciences
Arizona State University
Tempe, Arizona 85287

QR-

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 82-1050	2. GOVT ACCESSION NO. AD-A123122	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Spectral Estimation: An Overdetermined Rational Model Equation Approach		5. TYPE OF REPORT & PERIOD COVERED Final Report 9/30/81 to 8/15/82 6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) James A. Cadzow		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0250
9. PERFORMING ORGANIZATION NAME AND ADDRESS Arizona State University Tempe, Arizona 85287		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2305/B1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Building 410 Boiling AFB, D.C. 20332		12. REPORT DATE Sept. [illegible], 1982 13. NUMBER OF PAGES 107
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) ~~Rome Air Development Center Griffiss Air Force Base, N.Y. 13441~~		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

~~Unlimited~~

18. SUPPLEMENTARY NOTES

RADC project Engineer - Paul Van Etten.
A portion of the research reported upon in this report was sponsored by the Office of Naval Research under Contract N00014-82-K-0257

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Rational Spectral Estimation, ARMA model, AR model, MA model, Spectrum, Singular Value Decomposition, Adaptive Implementation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

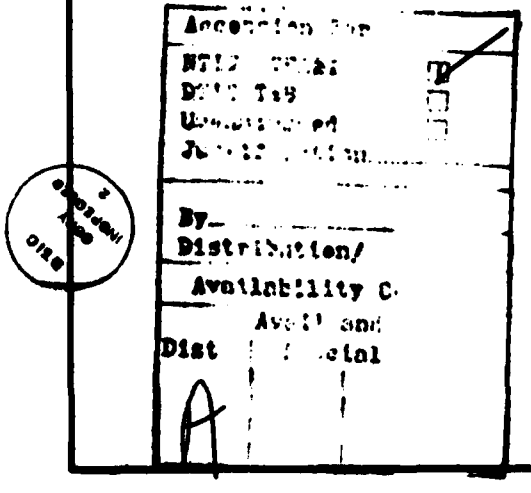
In seeking rational models of time series, the concept of approximating second order statistical relationships (i.e., the Yule-Walker equations) is often explicitly or implicitly invoked. The parameters of the hypothesized rational model are typically selected so that these relationships 'best represent'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

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

lishing 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 over-determined parametric evaluation approach, a reduction in data induced model parameter hypersensivity is obtained, and, a corresponding improvement in modeling performance results. Moreover, upon adapting a singular value decomposition representation of an extended order aurocorrelation 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.



Spectral Estimation: An Overdetermined Rational Model Equation Approach

James A. Cadzow
Department of Electrical and Computer Engineering
Arizona State University
Tempe, Arizona 85287

SPECTRAL ESTIMATION: AN OVERDETERMINED RATIONAL MODEL EQUATION APPROACH

TABLE OF CONTENTS

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

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

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., $r_x(-n) = \overline{r}_x(n)$) 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 $S_x(ej^{\omega})$ 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 semidefinite, 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

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

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 <u>only</u> 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

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

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 lags $r_x(n)$) 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

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

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 ironical 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 <u>discontinuous</u> 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 the remainder of 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.

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 applied the periodogram method for detecting hidden periodicities in sun spot activity data at the turn of the century [58]. 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 side lobe 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 numerous texts (e.g., see refs. [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 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

---

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

$$S_{MA}(e^{j\omega}) = | b_0 + b_1 e^{-j\omega} + \ldots + b_q e^{-jq\omega} |^2$$
$$= | B_q(e^{j\omega}) |^2 \tag{1.7}$$

The q+1 parameters $b_0, b_1, \ldots, b_q$ which identify this MA(q) model are seen to form a q^{th} 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 $B_q(e^{j\omega})$ constituting the moving average model (1.7) is factored, it is possible to provide additional insight into a MA model's properties. This factorization is seen to give rise to the equivalent representation

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

in which the z_k are the roots of the polynomial $B_q(e^{j\omega})$. The zeroes of a 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 <u>all-zero</u> 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, do not contain 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 $B_q(e^{j\omega})\,\bar{B}_q(e^{j\omega})$ thereby giving

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

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

$$c_n = \sum_{k=0}^{q} b_k\, \bar{b}_{k-n} \qquad -q \leq n \leq q \tag{1.10}$$

Upon setting the c_n equal to $w(n)r_x(n)$, 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 as nonparametric spectral estimators. When viewed from the approach here taken, however, each of these procedures is recognized as being a realization of a 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 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

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

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

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

where the p_k are the roots of $A_p(e^{j\omega})$. 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, do not contain 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).

Autoregressive 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 ref. [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

$$S_{ARMA}(e^{j\omega}) = \left| \frac{b_o + b_1 e^{-j\omega} + \ldots + b_q e^{-jq\omega}}{1 + a_1 e^{-j\omega} + \ldots + a_p e^{-jp\omega}} \right|^2$$

$$= \left| \frac{B_q(e^{j\omega})}{A_p(e^{j\omega})} \right|^2 \qquad (1.13)$$

An ARMA model is seen to have a frequency characterization which is the composite of a 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

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

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 most 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

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

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

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

in which the excitation time series $\{\varepsilon(n)\}$ 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 Figure 2.1. 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

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

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.

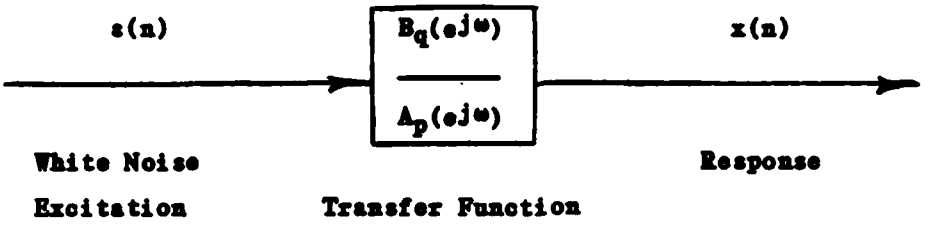


Figure 2.1. Model of rational time series.

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

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

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 a 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 a moving average random process if it is generated according to the linear nonrecursive relationship

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

where $\{\varepsilon(n)\}$ 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

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

where use of the facts that $a_k = 0$, and, $h(n) = b_n$ for $0 \leq n \leq q$ 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 $-q \leq n \leq q$. 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 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 expression (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 equation (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 expression (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 compatable with the autocorrelation identity (2.5). The one normally chosen corresponds to the so-called <u>minimum delay</u> selection in which the z_k roots used in expression (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 autoregressive (AR) process of order p if it is generated according to the recursive relationship

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

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

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

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 $a_1, a_2, \ldots, a_p, b_0$ 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

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

or more compactly as

$$R\,\underline{a} = |b_0|^2\,\underline{e}_1 \qquad (2.11b)$$

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

$$R(i,j) = r_x(i-j) \qquad 1 \leq i \leq p+1 \qquad 1 \leq j \leq p+1 \qquad (2.12)$$

$\underline{a}$ is the (p+1)x1 autoregressive parameter vector with first component equal to one, that is

$$\underline{a} = [1, a_1, a_2, \ldots, a_p]' \qquad (2.13)$$

and $\underline{e}_1$ is the (p+1)x1 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

$$\underline{a} = |b_0|^2\,R^{-1}\,\underline{e}_1 \qquad (2.14)$$

in which the normalizing coefficient b_0 is selected so that the first component of $\underline{a}$ is one as required in expression (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 expression (2.11), it is seen that the resultant AR(p) model parameters are totally dependent on the first p+1 given autocorrelation lags $r_x(0)$, $r_x(1)$, ..., $r_x(p)$. 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 $r_x(p+1)$, $r_x(p+2)$, ..., $r_x(s)$ (which were not used in the parameter identification). In order to provide a represention 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., $p<<s$) 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 expression (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, ..., p_m where p_m designates some as yet unknown maximum AR order. In this sequencing scheme, Levinson showed that the parameters for the k^{th} order AR model solution which are designated by

$$a_1^{(k)},\ a_2^{(k)},\ \ldots,\ a_k^{(k)},\ b_o(k) \tag{2.15}$$

are related to the $(k-1)^{th}$ order AR model solution as outlined in Table 2.1. A brief description of this systematic algorithm will now be given.

Step 1 $a_1^{(1)} = -r_x(1)/r_x(0)$	(2.16a)
$\lvert b_o^{(1)}\rvert^2 = [1 - \lvert a_1^{(1)}\rvert^2]\ r_x(0)$	(2.16b)
Step 2 For $k = 2, 3, 4, \ldots$	
$a_k^{(k)} = -\ r_x(k) + \sum_{m=1}^{k-1} a_m^{(k-1)}\ r_x(k-m)\ /\lvert b_o^{(k-1)}\rvert^2$	(2.17a)
$a_i^{(k)} = a_i^{(k-1)} + a_k^{(k)}\ a_{k-i}^{(k-1)} \quad 1 \le i \le k-1$	(2.17b)
$\lvert b_o^{(k)}\rvert^2 = [1 - \lvert a_k^{(k)}\rvert^2]\lvert b_o^{(k-1)}\rvert^2$	(2.17c)

Table 2.1. Levinson-Durbin Algorithm for Recursively Solving Expression (2.11)

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 of Table 2.1. Upon setting p = 2 in expression (2.11), a moderate amount of algebraic manipulation will reveal the validity of the solution as given in Step 2 of Table 2.1 with k = 2 (with superscript (2) appended). Levinson proved that in following the systematic procedure of Table 2.1, the solutions to the Yule-Walker equation (2.11) for order selections
p = 1, 2, 3, ... are sequentially obtained. Moreover, the number of multiplication (and addition) computations required in generating the k^{th} AR order parameters from the $k-1^{st}$ AR order parameters (i.e., Step 2) is seen to be k. Thus, the computational complexity of the Levinson-Durbin algorithm for generating a p^{th} 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 expression (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 expression (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 p^{th} iteration (i.e., $a_k^{(p)} = a_k$ k = 1, 2, ..., p and $|b_o^{(p)}|^2 = |b_o^2|$). Moreover, if this recursion were continued beyond p, it would be found that $a_i^{(k)} = a_i$ for $1 \leq i \leq p$, $a_i^{(k)} = 0$ for $p+1 \leq i \leq k$, and, $|b_o^{(k)}|^2 = |b_o|^2$. This is a direct consequence of the fact that $a_{p+1}^{(p+1)}$ must be zero as is evident from expression (2.17a). From these observations, it is therefore apparent that the nonchanging of the parameters $|b_o^{(k)}|^2$ 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 $|b_o^{(k)}|^2$ 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 expression (2.17c) that $|b_o^{(k)}|^2 \leq |b_o^{(k-1)}|^2$ for all $k \geq 1$.

Thus, the parameters $|b_o^{(k)}|^2$ form a monotonically nonincreasing sequence and this factor can be used in model order determination. In particular, the parameter $|b_o^{(k)}|^2$ may be identified with a 'prediction error' associated with a k^{th} 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 ref. [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, ... 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 $r_x(0)$, $r_x(1)$, ..., $r_x(p)$ to be a valid segment of an autocorrelation sequence, it is necessary and sufficient that $|c_k| \leq 1$ for k = 1, 2, ..., p. Moreover, the transfer function

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

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

It is noteworthy that the system of equations (2.11) also arise 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

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

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 expression (2.11) in which $|b_o|^2$ 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

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

subject to the constraint that this function will be consistent with the given set of p+1 autocorrelation lags $r_x(0)$, $r_x(1)$, ..., $r_x(p)$ 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 expression (2.11).

ARMA Time Series

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

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

in which the excitation sequence {e(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 compatable 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 the 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 will almost 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 extrema 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 take to be causal, it follows that the Yule-Walker equations assume a particularly simple form for indices $n > q$, that is

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

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 compatable with the given set of autocorrelation lags (2.21), we could adopt the approach that characterized extended Yule-Walker equation AR and ARMA modeling methods up to as recently as three years ago (e.g., see refs [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 auto-regressive 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 a 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 refs. [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 autoregressive 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[1]

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

or more compactly as

$$R_1 \underline{a} = \underline{0} \qquad (2.23b)$$

In this latter expression, $\underline{0}$ denotes the tx1 zero vector, R_1 is the tx(p+1) ARMA autocorrelation matrix with Toeplitz type structure having elements

$$R_1(i,j) = r_x(q+1+i-j) \qquad 1 \leq i \leq t \qquad 1 \leq j \leq p+1 \qquad (2.24)$$

---

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

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

$$\underline{a} = [1, a_1, a_2, \ldots, a_p]' \tag{2.25}$$

Examination of relationships (2.23) reveals that the ARMA model's autoregressive 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 entires used in matrix R_1 of expression (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 relationship (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 the rank of matrix R_1 has less than full rank, a nontrival autoregressive parameteric vector solution $\underline{a}$ will be assured. An interesting algebraic characterization of this solution may be obtained upon premultiplying both sides of relationship (2.23) by the complex conjuage transpose of R_1 as denoted by R_1^* to yield

$$R_1^* R_1 \underline{a} = \underline{0} \tag{2.26}$$

Upon examination of this expression, it is clear that the required autoregressive 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)x(p+1) matrix $R_1^* R_1$. As such, we may then use standard eigenvector-eigenvalue routines when finding the required ARMA model autoregressive 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, with a higher order ARMA rational process. Since R_1 has full rank, there then will not exist a nontrivial solution to relationship (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 autoregressive parameter vector $\underline{a}$ so that $R_1\underline{a}$ 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 autoregressive parameter vector lying on the unit hypersphere which will minimize the Euclidean norm of $R_1\underline{a}$. This entails solving the following contrained optimization problem

$$\min \underline{a}^* R_1^* R_1 \underline{a}$$
$$\underline{a}^* \underline{a} = 1$$

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 $x_k^* x_k = 1$), the required autoregressive parameter vector with first component of one if obtained by the normalization.

$$\underline{a}^o = \frac{1}{x_1(1)} \underline{x}_1 \tag{2.27}$$

where $x_1(1)$ denotes the first component of $\underline{x}_1$. This autoregressive parameter vector selection procedure characterizes many spectral algorithms which are varients 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 $R_1\underline{a}$ over all $(p+1)x1$ vectors $\underline{a}$ with first components equal to one, that is

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

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

$$R_1^* R_1 \underline{a}^o = \alpha \underline{e}_1 \tag{2.28}$$

where the normalizing constant α is selected so that the first component of $\underline{a}^o$ 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[1]. The particular application at hand dictates which autoregressive 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 ref. [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 autoregressive lags entails an examination of the matrix R_1. If this matrix is not of full rank, the required exact autoregressive parameter vector will be given by solving expression (2.26). On the other hand, when the matrix has full rank, an appropriate autoregressive parameter vector may be achieved by solving either expression (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.

---

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

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 p^{th} order nonrecursive filter with transfer function $A_p(z)$ whose coefficients correspond to the autoregressive parameters obtained upon solving either expression (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 average process of order q with power spectral density function $|B_q(e^{j\omega})|^2$ as is made evident from Figure 2.2. 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

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

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

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

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

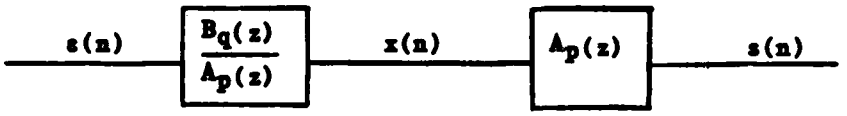


Figure 2.2. Generation of residual time series.

(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

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

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

$$S_x(e^{j\omega}) = r_x(0) + 2\mathrm{Re}\{D(e^{j\omega})\} \tag{2.33}$$

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

$$D(e^{j\omega}) = \frac{c_1e^{-j\omega}+c_2e^{-j2\omega} + \ldots + c_pe^{-jp\omega}}{1+a_1e^{-j\omega} + \ldots + a_pe^{-jp\omega}}$$

$$= \frac{C(e^{j\omega})}{A_p(e^{j\omega})} \tag{2.34}$$

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

To determine the required c_n coefficients in expression (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

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

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 relationship (2.34) to conclude that

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

or

$$H\,\underline{c} = \underline{r} \tag{2.36b}$$

In general, the overdetermined system of equations (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 $\underline{c}$ so as to provide a least squares solution to expression (2.36). This would take the form of solving the consistent system of linear equations

$$\underline{c} = [H^*H]^{-1}H^*\underline{r} \tag{2.37}$$

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

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

This is basically the approach taken in references [15] and [16]. In using expression (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 expression (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

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

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 2.2. 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.

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)x(p+1) AR autocorrelation matrix R using expression (2.12)

(ii) Solve $\quad R\underline{a} = |b_o|^2\, \underline{e}_1 \qquad (2.11)$

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

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

ARMA Model

(i) Form the tx(p+1) ARMA autocorrelation matrix R_1 using expression (2.24)

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

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

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

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

Table 2.2. Rational spectral model techniques employing exact autocorrelation lag information.

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 very narrow application of rational spectral estimation techniques, it provides a meaningful basis for understanding the relative performance capabilities of MA, AR, 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

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

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 values 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

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

in which $\delta(n)$ denotes the unit-impule (Kronecker delta) sequence. The power spectral density function associated with this process is composed of 2m dirac delta impulses of amplitude $0.5\, A_k^2$ 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

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

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

$$A_{2m}(z) = \sum_{n=0}^{2m} a_n \, z^{-n}$$

$$= \prod_{k=1}^{m} [1 - 2 \, z^{-1} \cos(2\pi f_k) + z^{-2}] \tag{3.4}$$

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 refs. [10],[32],[55]).

Upon comparison of relationships (3.3) and (2.22), it might be incorrectly infered 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 $0 \leq n \leq 2m$, 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 equations (2.22) and (3.3), we may still use the ARMA modeling autoregressive parameter procedure as outlined in Section II to identify the 2m parameters a_k. These parameters would be then in turn inserted into relationship (3.4) to identify the frequency parameters 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 expression (3.2) over any set of m or more indices satisfying $n \geq 1$. With this in mind, let us evaluate this expression for the contiguous indices $1 \leq n \leq v$ where the

integer $v \geq m$. This is found to yield the following overdetermined (if $v>m$) system of consistent linear equations in the A_k unknowns

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

or equivalently as

$$\underline{r} = C\underline{p} \quad (3.6)$$

where $\underline{p}$ is the so-called mx1 power vector with elements $A_{k/2}^2$. If the integer parameter v is selected to be larger than or equal to m, the least square approximate solution to the overdetermining equations (3.6) is given by

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

where C' designates the transpose of matrix C. In the case of perfect autocorrelation knowledge, we normally set v = m thereby giving the solution $\underline{p} = C^{-1}\underline{r}$. 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 sinusoids in white noise time series (3.1) is not compatable with an AR model, AR models have also been successfully employed in analyzing such time series. Depending on the underlying signal to noise ratios

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

the desired detection and frequency estimation will require that the AR order parameter p be made <u>significatly</u> 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 white noise case. As we will illustrate in Section VIII, the approach taken in this paper will also

produce exceptional 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 evaluations for achieving high quality alternative estimates for the frequency parameters appearing in expression (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 expression (3.2) with $\sigma^2 = 0$ over the indices $-t + 2p \leq n \leq t$ (in which $p = 2m$) is found to result in the following symmetrical relationship

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

or

$$R_s \, \underline{a} = o \qquad (3.8b)$$

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

If the autocorrelation lag entries of expression (3.8) correspond to (3.2) with $\sigma^2 = 0$, it then follows that the overdetermined system of equations (3.8) will have a unique solution for the a_k coefficients. This solution can then be incorporated into equation (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)x(p+1) entries of matrix R_s (i.e., the $r_x(0)$ entries), it can be argued that so long as t>>p, the effect of the additive noise will be minimal. Based on this premise, it is natural to then seek a vector $\underline{a}$ 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

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

in which α is a normalizing scalar selected to ensure that the first component of $\underline{a}$ is one. The nonnegative 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 expression (3.9) often provide exceptional estimates so long as t>>p. A paper in preparation will further refine this new approach.

Numerical Example

In order to illustrate the effectiveness of the three 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 zero dB signal-to-noise ratio (SNR) environment. According to relationship (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 a MA(q) model. Plots of this expression with entries (3.11) for model order selections of q = 32 and q = 64 are shown in Figure 3.1 over the range of normalized frequencies $0 \leq f \leq 0.5$. From these results, it is apparent that a resolution of the two equal amplitude sinusoids was not achieved for a thirty second order MA model, but, was achieved for a sixty fourth 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 a 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.

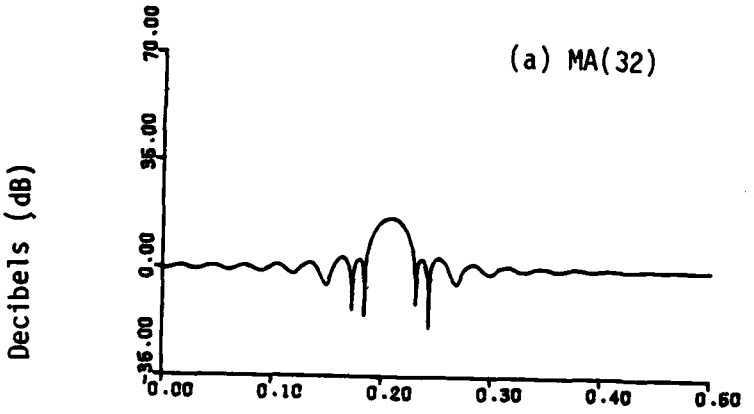
AR Models: We next used the same autocorrelation lag information (3.11) to generate AR models of order p = 20 and p = 24 using expression (2.11). The resultant spectral estimates $1/|A_p(e^{j\omega})|^2$ are shown in Figure 3.2a and b for these two model order choices. It is apparent that the twentieth 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 $0 \leq n \leq p$ 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.

In order to illustrate the effect of using more than the minimal number of extended 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 equations (2.23) for this choice of order parameters has a spectral behavior as depicted in Figure 3.2c. 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 equation (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 expression (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 model $1/|A_4(e^{j\omega})|^2$ without the MA component is plotted in Figure 3.3. 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 <u>type</u> model is perfectly compatible with a sinusoids in white noise time series (MA and AR models are not compatible). It is noteworthy that only the autocorrelation lags $r_x(n)$ for $1 \leq n \leq 8$ were required in generating the spectral model depicted in Figure 3.3.

Alternative Method: As a final procedure, we used the alternative method as represented by relationship (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 Figure 3.4. 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.



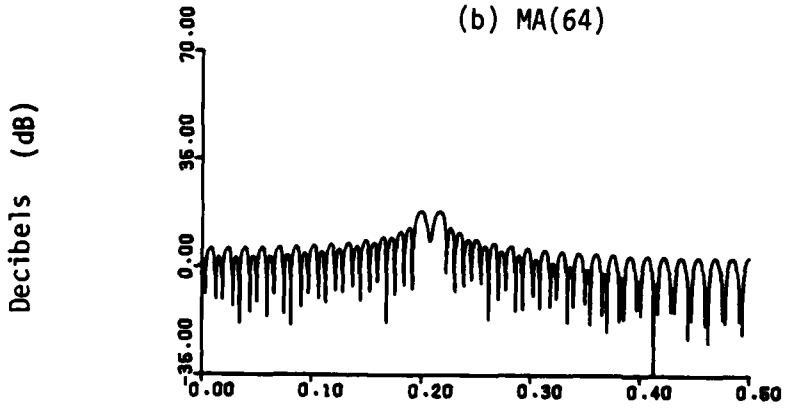


Fig. 3.1 Moving average (MA) spectral models using expression (1.6) with w(n) = 1 and exact autocorrelation lags (a) q = 32, (b) q = 64.

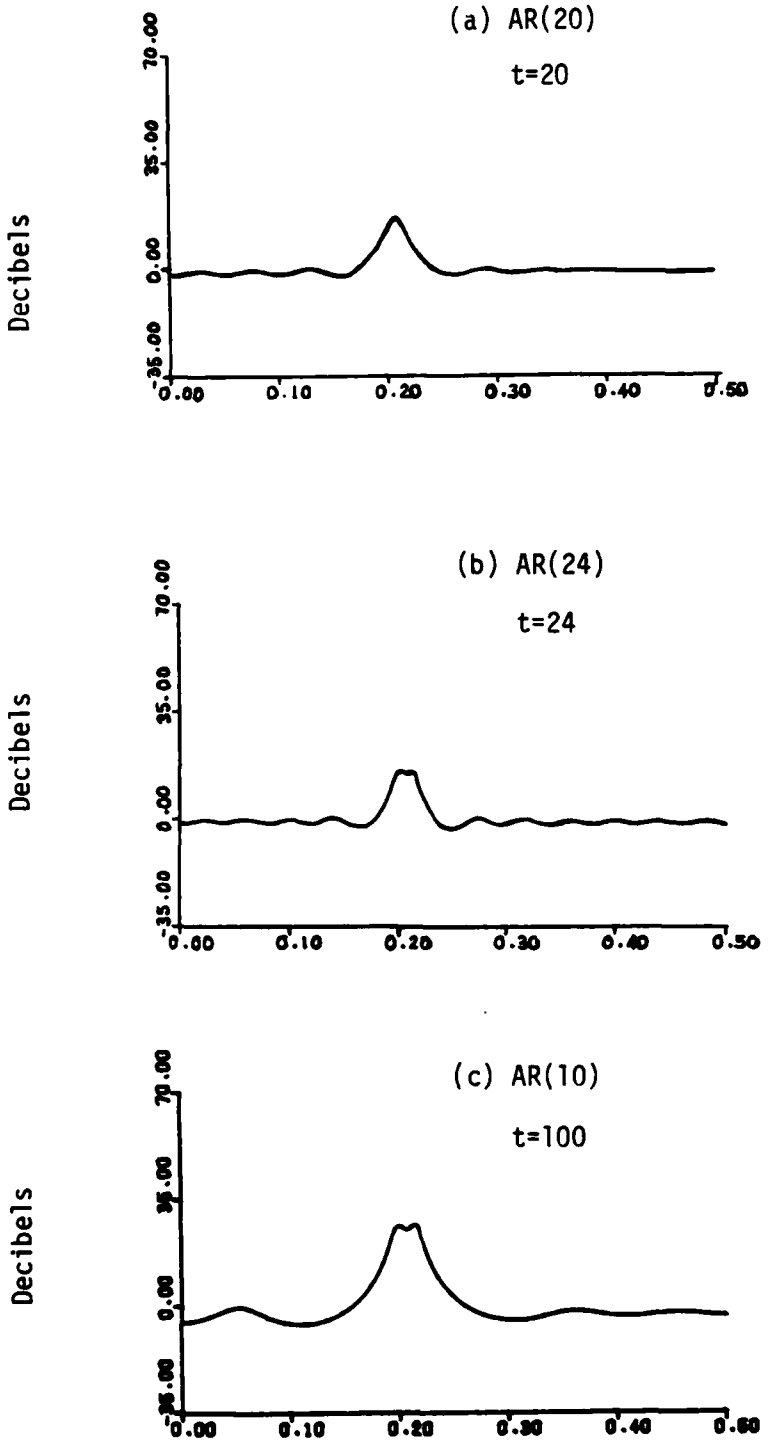


Fig. 3.2 Autoregressive (AR) spectral models using expression (2.11) with exact autocorrelation lags (a) p=t=20, (b) p=t=24, (c) p=10, t=100.

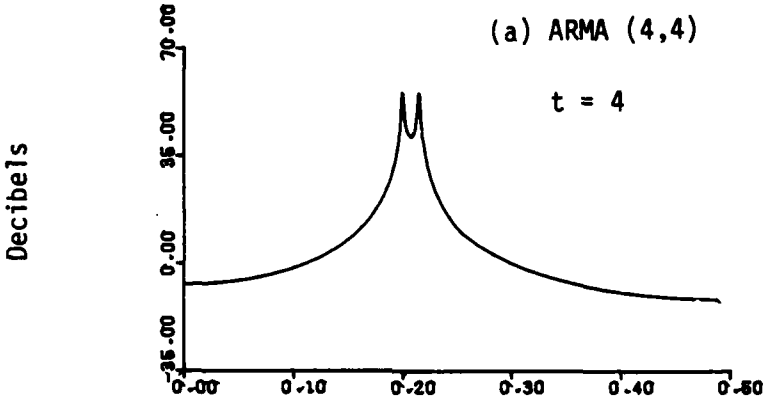


Fig. 3.3 Autoregressive-moving average (ARMA) spectral models using expression (2.23) with exact autocorrelation lags and p=t=4.

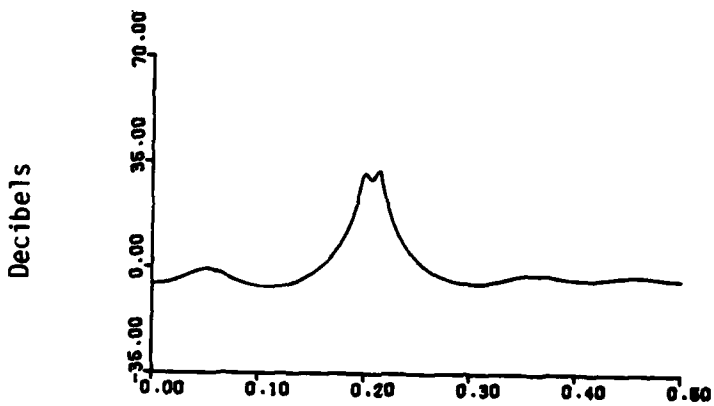


Fig. 3.4 Alternative method with
p = 10 and t = 50.

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

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

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 autocor-relation estimates such as proposed by Blackman and Tukey [8], and, direct methods based on the Fourier transform of the time series observations and widely known as 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 expression (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

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

Considerations to be made in selecting the window sequence are well documented and the reader is referred to references [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 ref. [33] for a detailed development). The unbiased estimate achieves the required autocorrelation lag estimate according to

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

where the convention of setting to zero any term x(n) in the summand for which n ∉ [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 refs. [33],[52],[53].

$$\hat{r}_x(n) = \frac{1}{N} \sum_{k=1}^{N} x(k+n)\overline{x}(k) \qquad -q \leq n \leq q \qquad (4.4)$$

We again adhere to the convention of setting to zero any term x(n) in the summand for which n ∉ [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 vs. 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

$$S_x(e^{j\omega}) = \frac{1}{N} |X_N(e^{j\omega})|^2 \qquad (4.5)$$

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

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

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 expression (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 $0 \leq k \leq N-1$ is seen to entail evaluation of the entities

$$X_N(e^{j\frac{2\pi k}{N}}) = \sum_{n=0}^{N-1} x(n+1)e^{-j\frac{2\pi kn}{N}}, \quad 0 \leq k \leq N-1 \qquad (4.7)$$

These evaluations are readily carried out by use of the N point fast Fourier transform (FFT) algorithm (e.g., see refs. [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 expression (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 <u>zero padding</u>. This simply entails the appending of L zeroes to the given set of time series observations, that is

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

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 a N+L point FFT of this padded time series, the following more finely spaced samples of the Fourier transform would be generated

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

If these sampled values were then substituted into expression (4.5), we would obtain sampled values of the periodogram at the more finely spaced frequencies $\omega_k = 2\pi/(N+L)$ for $0 \leq k < N+L$. 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 <u>inherent frequency resolution capability</u> of $\Delta\omega = 2\pi/N$ (or equivalently $\Delta f = 1/N$). When using a 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 pre-eminent 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 refs. [23],[31]).

In this section and 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

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

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 R must be such that

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

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

$$R\,\underline{a} = |b_o|^2\,\underline{e}_1 \tag{5.3}$$

in which the normalizing parameter b_o is selected so that the first component of $\underline{a}$ is one. The steps of this general AR modeling approach are summarized in Table 5.1.

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

Table 5.1 Basic steps in obtaining an AR(p) spectral estimate.

The quality of the AR modeling approach as embodied in expression (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

$$\hat{R}(i,j) = \frac{1}{N-|i-j|} \sum_{k=1}^{N} x(k+i-j)\overline{x}(k) \qquad \begin{matrix} 1 \leq i \leq p+1 \\ 1 \leq j \leq p+1 \end{matrix} \qquad (5.4)$$

typically provides 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 symmetric; properties shared by the actual autocorrelation matrix being approximated. Moreover, this estimate is <u>consistent</u> in the sense that as N approaches infinity, we have $\hat{R} \longrightarrow R$ 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 also be studied.

VI. ARMA Modeling: Time Series Observations

The methods for generating ARMA models based upon times 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 form one of the most widely used of such methods. These include exact maximum likelihood approaches (e.g., refs [6] and [48]), and, least square methods which approximate the exact likelihood function (e.g., refs [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 characteristic of nonlinear programming methods. Other nonmaximum likelihood methods which fall into category (i) have been proposed (e.g., see refs. [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 estiamtes, in a linear fashion, for the AR parameters (e.g., see refs [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 a overdetermined set of Yule-Walker equation evaluations to decrease this hypersensitivity was proposed [15]. This approach has since been adopted by other researchers in spectral estimation applications with success (e.g., see refs. [7],[12],[36],[51]). With this in mind, we shall now give a detailed development of the overdetermined approach to estimating the AR 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

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

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 tx(p+1) autocorrelation matrix R_1 which appears in expression (2.24). Namely, using any of a number of available procedures, we first compute the following autocorrelation lag estimates

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

Two particularly attractive procedures for effecting these autocorrelation estimtaes will be detailed at the end of this section and in Section X. Independent of what procedure is eventually used, the net result of this first step will be the generation of a tx(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 autoregressive parameter vector with first component equal to one which will satisfy the theoretical relationship $R_1\underline{a} = \underline{0}$ as given in equation (2.23). As such, the tx1 extended Yule-Walker equation error vector as specified by

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

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)x(m)) used in formulating the autocorrelation lag estimates. Consequently, an assumption that the error vector elements tend to be Gaussianly distributed is a reasonable one. The joint density function of the extended Yule-Walker equation error vector may be therefore approximated by

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

in which $W^{-1} = E[\underline{e}\underline{e}^*]$ designates the error covariance matrix which is generally unknown and where the expected value of $\underline{e}$ 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 autoregressive parameters. Namely, making use of relationship (6.3) and the joint density function (6.4), it is possible to generate a joint density function for the autorgressive parameter vector $\underline{a}$ which will be of form

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

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

$$\begin{aligned} &\min \ \underline{a}^* \hat{R}^* W \hat{R} \underline{a} \\ &a(1)=1 \end{aligned} \tag{6.5}$$

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

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

where α is a normalizing constant selected so that the first component of $\underline{a}^o$ is one.[1] Expression (6.6) constitutes the so-called <u>high performance method</u> of autoregressive parameter selection [15]-[20].

It is to be noted that in minimizing functional (6.5) with respect to the normalization constraint imposed on $\underline{a}$'s first component, the error vector is being minimized in the least squares sense. In effect, we are then selecting $\underline{a}$ so as to best satisfy the theoretical relationship (2.23) given by $R_1 \underline{a} = \underline{e}$. Using this interpretation, the positive definite matrix W can be

---

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

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 nonnegative diagonal entries w_k for k = 1,2, ..., 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 txt identity matrix has been found to provide excellent modeling performance when the matrix estimate 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 refs. [26],[28],[35],[38]). In this case, the minimal number of p error contaminated extended Yule-Walker equation evaluations are being used in fixing the model's p autoregressive coefficients. A little thought should convince oneself 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 Figure 6.1.

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 autoregressive 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

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

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 expression (6.2). Specifically, the required autocorrelation lag estimate entries are generated according to

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

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)x(p+1) matrix $\hat{R}_1^* W \hat{R}_1$, which completely characterizes the autoregressive parameter vector solution through expression (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

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

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 autocorrelation parameter vector is straightforwardly obtained by solving the system of linear equations (6.6). A Fortran program listing of an implementation of this procedure is given in the appendix where the flexibility of using the standard unbiased or the standard biased (i.e., divisor N-n in equation (6.9) is replaced by N) autocorrelation estimate is available.

MA Parameter Estimation

To complete the ARMA modeling, it is necessary to compute an estimate for the moving average 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

$$|\hat{A}_p(e^{j\omega})|^2 = \left| \sum_{k=0}^{p} a_k e^{j\omega k} \right|^2 \qquad (6.11)$$

in which a_k denote the autoregressive parameter estimates as generated from expression (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 autoregressive parameter estimates (6.6) according to

$$s_f(n) = \sum_{k=0}^{p} a_k \, x(n-k) \qquad p+1 \leq n \leq N \qquad (6.12)$$

Similarly, the backward residuals component are generated using

$$s_b(n) = \sum_{k=0}^{p} a_k \, x(n+k) \qquad 1 \leq n \leq N-p \qquad (6.13)$$

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 autoregressive parameters a_k. With this in mind, a procedure for extracting this MA characterization from the computed forward and backward residuals will now be given.

The most direct procedure for achieving the required MA(q) estimate is to first generate the following estimates of the residual time series' first

It is also possible to employ the smoothed periodogram to obtain another form of MA(q) estimate. This entails segmenting the computed residuals in blocks of length q+1 (overlapping or not overlapping) 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]. 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 expression (2.38), estimates for the c_n parameters would be computed from

$$c_n = \sum_{k=0}^{n-1} a_k \hat{r}_x(n-k) \qquad 1 \leq n \leq p \tag{6.18}$$

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

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

where $D(e^{j\omega})$ is obtained by substituting the a_k and c_k estimates into form (2.34).

q+1 autocorrelation lags

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

If the residual time series do in fact correspond to a MA(q) process, it will be found that the $r_s(n)$ will be approximately zero for $n \geq q+1$. 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

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

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

A few words are now appropriate concerning the selection of the window to be used in estimate (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.

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 <u>significant</u> improvement in spectral estimation performance.

When the ARMA model order parameters are not known apriori, 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 apriori, 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 expression (2.23), it then follows that the $tx(p_e+1)$ 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 $q_e-p_e \geq q-p$, it then follows from the results of section II that the rank of the $tx(p_e+1)$ 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)x1 autoregressive parameter vector $\underline{a}$ 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 p+1 contiguous columns. This set of tx(p+1) matrices is specified by

$$R_k = [\text{submatrix of } R_e \text{ composed of its } k^{th} \text{ through } p+k^{th} \text{ column vectors inclusively}] \quad \text{for} \quad 1 \leq k \leq p_e - p + 1 \tag{7.3}$$

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

$$R_k \underline{a} = \underline{0} \quad \text{for} \quad 1 \leq k \leq p_e - p + 1 \tag{7.4}$$

where the first component of $\underline{a}$ is constrained to be one. In point of fact, expression (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 such that $p_e \geq p$ and $q_e - p_e > q - p$.

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

$$x(1), x(2), \ldots, x(N) \tag{7.5}$$

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 principle of singular value decomposition.

Singular Value Decomposition

In a variety of applications, the ultimate 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 ref. [27] and [39]).

> **Theorem 7.1:** Let A be a mxn matrix of generally complex valued elements. Then there exists mxm and nxn unitary matrices U and V, respectively, such that[1]
>
> $$A = U \Sigma V^* \tag{7.6}$$
>
> where Σ is a mxn 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

---

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

the problem of finding that mxn matrix of rank k which will best approximate A in the Frobenious norm sense (this assumes that $\sigma_{kk} > 0$ with $k \leq h$). The Frobenious norm of the mxn matrix difference A-B is defined to be

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

We now seek to find that mxn 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 mxn matrix of rank $k \leq$ Rank [A] which best approximates the mxn matrix A in the Frobenious norm sense is given by
>
> $$A^{(k)} = U \Sigma_k V^* \qquad (7.8)$$
>
> where U and V are as in expression (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
>
> $$||A - A^{(k)}|| = \left[\sum_{j=k+1}^{h} \sigma_{jj}^2\right]^{1/2} \quad 0 \leq k \leq h \qquad (7.9)$$

The degree to which $A^{(k)}$ approximates A is seen to be dependent on the sum of the (h-k) <u>smallest singular values</u> 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

$$\gamma(k) = \frac{||A^{(k)}||}{||A||}$$

$$= \left[\frac{\sigma_{11}^2 + \sigma_{22}^2 + \ldots + \sigma_{kk}^2}{\sigma_{11}^2 + \sigma_{22}^2 + \ldots + \sigma_{hh}^2}\right]^{1/2} \qquad 1 \leq k \leq h \qquad (7.10)$$

Clearly, this normalized ratio approaches its maximum value of one as k approaches h. For matrices of low effective rank, the quantity $\gamma(k)$ 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 h) to achieve a (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 a SVD of the $tx(p_e+1)$ extended order autocorrelation matrix estimate R_e, that is

$$R_e = U\Sigma V^* \tag{7.11}$$

where U and V are txt and $(p_e+1)x(p_e+1)$ unitary matrices, respectively and Σ is a $tx(p_e+1)$ matrix of the form called out in Theorem 7.1. The required autoregressive order p is obtained by examining the normalized ratio $\gamma(k)$. Namely, p is set equal to the smallest value of k for which $\gamma(k)$ 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 $tx(p_e+1)$ extended order autocorrelation matrix, that is

$$R_e^{(p)} = U\,\Sigma_p\,V^* \tag{7.12}$$

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

$$R_e(p) = \sum_{n=1}^{p} \sigma_{nn}\,\underline{u}_n\,\underline{v}_n^* \tag{7.13}$$

where $\underline{u}_k$ and $\underline{v}_k$ are the k^{th} column vectors of the txt and $(p_e+1)x(p_e+1)$ 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

$$R_e(p) = [r_1(p) \vdots R_a(p)] \tag{7.14}$$

where $r_1(p)$ is the left most tx1 column vector of $R_e(p)$ and $R_a(p)$ is a txp_e matrix composed of the p_e right most tx1 column vectors of $R_e(p)$. We now seek a $(p_e+1)x1$ autoregressive parameter vector $\underline{a}$ with first component equal to one that will satisfy the theoretical relationship

$$R_e(p)\,\underline{a} = \underline{0}$$

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

$$\begin{bmatrix} a_1^o \\ a_2^o \\ \cdot \\ \cdot \\ \cdot \\ a_{p_e}^o \end{bmatrix} = -\,[R_a(p)]^{\#}\,\underline{r}_1(p)$$

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

$$\underline{a}^{o} = \frac{\underline{e}_1 - \sum_{k=1}^{p} \bar{v}_k(o)\, \underline{v}_k}{1 - \sum_{k=1}^{p_e} |v_k(o)|^2}$$

$$= \frac{\sum_{k=p+1}^{p_e} v_k(o)\underline{v}_k}{\sum_{k=p+1}^{p_e} |v_k(o)|^2} \qquad (7.15)$$

where the $\underline{v}_k$ 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 autoregressive parameters of a lower order ARMA(p,q) model. In particular, the submatrices of $R_e(p)$ composed of its columns k through p+k inclusively yield rank p approximations of the tx(p+1) autocorrelation matrices R_k for $1 \leq k \leq p_e - p+1$ as specified by expression (7.3). We shall denote these rank p approximations by $R_k^{(p)}$. Due to the SVD operation and errors inherent in generating R_e, there will generally not exist a unique autoregressive parameter vector with first component equal to one which will satisfy all of the p_e-p+1 estimates of relationships (7.4). Nonetheless, it is still

desirable to find an autoregressive 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

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

where

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

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

$$S(p) = \sum_{n=1}^{p} \sum_{k=1}^{p_e-p+1} \underline{v}_n{}^k \underline{v}_n{}^{k*} \tag{7.17}$$

in which $\underline{v}_n{}^k$ denotes the $(p+1)x1$ vector as specified by

$$\underline{v}_n{}^k = [v_n(k), v_n(k+1), \ldots, v_n(k+p)]' \tag{7.18}$$

$$1 \leq k \leq p_e-p+1$$

$$1 \leq n \leq p$$

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

Upon generating the $(p+1)x(p+1)$ matrix $S^{(p)}$, we next wish to select that autoregressive parameter vector $\underline{a}$ with first component of one so as to minimize quadratic functional (7.16). This constrained minimization will result in the best least square approximation of the theoretical relationships (7.4). Using standard procedures, the required optimum

autoregressive parameter vector is found by solving the following linear system of equations[1]

$$S^{(p)}\underline{a} = \alpha\,\underline{e}_1 \tag{7.19}$$

in which the normlizing constant α is selected so that the first component of $\underline{a}$ is one. It will be shown in the next section that this SVD version of the ARMA modeling procedure can lead to a significant improvement in modeling performance.

The concept of a SVD representation has been previously incorporated 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 will ultimately have on spectral estimation (and in other applications) is only beginning to be appreciated.

---

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

VIII. Numerical Examples

In this section, we shall investigate the comparative spectral estimation performance of the rational 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 ref. [11])

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

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

$$x_1(n) = 0.4\, x_1(n-1) - 0.93\, x_1(n-2) + \varepsilon_1(n) \tag{8.1b}$$

$$x_2(n) = -0.5\, x_2(n-1) - 0.93\, x_2(n-2) + \varepsilon_2(n) \tag{8.1c}$$

where $\varepsilon(n)$, $\varepsilon_1(n)$ and $\varepsilon_2(n)$ are mutually uncorrelated Gaussian zero mean white noise processes with variance one. A simply analysis indicates that the power spectral density function associated with time series (8.1) is given by

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

and is plotted in Figure 7.1a.

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-Jenkins iterative method are shown in Figure 8.1b. The number of iterations required to achieve these estimates ranged from 10 to 700 with 50 being a typical requirement. Next, this paper's 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 autoregressive 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 Figures 8.1c, 8.1d, and 8.1e 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

$$x(n) = \sin(2\pi f_1 n) + \sin(2\pi f_2 n) + w(n), \qquad 1 \leq n \leq N \qquad (8.3)$$

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

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 signal-to-noise ratio of zero 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 Figures 8.2 to 8.7 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 fast Fourier transform was first used in generating spectral estimates for each of the ten different 128 data length realizations. Specifically, expression (4.7) with N = 128 was incorporated into the MA spectral estimator (4.5) to generate the sampled periodogram estimate

$$S_x\left(e^{\frac{j2\pi k}{N}}\right) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n+1)\, e^{-j\frac{2\pi k}{N}}\right|^2 \qquad 0 \le k \le N-1 \qquad (8.4)$$

It was found that each of the ten periodograms produced remarkably similar results. A typical 128 point FFT periodogram for one of these trials is shown in Figure 8.2a. 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 <u>padding</u> 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 Figure 8.2b. 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 Figure 8.2c. 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 Figure 8.3a. Although a detection of spectral energy in the region about $f = 0.2$ is evident, the appearance of two spectral peaks is not. 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 Figure 8.3b. 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 on the order of 24 or more.

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 expression (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 Figure 8.3c. 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 using this paper's method provided more sharply defined peaks than 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 autoregressive 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 expression (6.6) with W = I and unbiased autocorrelation lag estimate entries was next used to generate estimates of the autoregressive 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 white noise case (i.e., the zeroes are not used). In Figure 8.4a, 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 Figure 8.4b 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 autoregressive parameters, and, are seen to be significantly superior to the Burg AR(24) estimates which required twenty-four autoregressive 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. Setting p=4 and using relationship (7.15), the ten ARMA(14,14) spectral estimates which arose are shown in Figure 8.4c. Next, letting p=4

in expression (7.17), the ten SVD derived ARMA(4,4) spectral estimates which arose are shown superimposed in Figure 8.4c. These spectral estimates are not only of uniformly high quality, but, they represent the lowest order rational model which is compatable 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 8.1. The quantities $\bar{f}_k(\bar{p}_k)$ and $\sigma_{\bar{f}_k}^2$ ($\sigma_{\bar{p}_k}^2$) for k = 1,2 represent the sampled means and variances, respectively, of the peak frequencies (pole magnitudes) as determined from the ten spectral estimates.

k	f_k	$\bar{f}_k$	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

Table 8.1: Statistics of SVD ARMA (4,4) 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

$$\sigma_{11} = 18.3 \quad , \sigma_{22} = 18.2 \quad , \sigma_{33} = 5.30 \quad , \sigma_{44} = 4.69$$

$$\sigma_{55} = 0.85 \quad , \sigma_{66} = 0.78 \quad , \quad \ldots \quad , \sigma_{15,15} = 0.21$$

It is apparent that the first four singular values are dominant (i.e., (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.

Alternative Method: In Section III, an alternate method for detecting and estimating the frequencies of sinusoids in white noise was proposed. This

method is represented by the least squares solution (3.9) to an overdetermined system of linear equations. Using this expression with a selection of p = 14, t = 50, W = I, and, unbiased autocorrelation lag estimate entries for R_s, spectral estimates for each of the ten time series of length 128 were generated. The results of these estimates are depicted in superimposed style in Figure 8.5 in plots of $1/|A_{14}(e^{j\omega})|^2$. Two sharply resolved peaks are achieved in each of the ten estimates. It is noteworthy that this procedure provided good estimates for a low order choice. A paper now in preparation will demonstrate the exceptional performance of this new procedure for a more general class of deterministic signals in white noise. Improvement is there achieved by making an estimate of the white noise variance σ^2 using expression (3.2) at n=0, then subtracting this estimate from the $r_x(0)$ term and then using an SVD. Initial empirical evidence suggests that this new approach provides significantly better performance than the Pisarenko method [55] and it varients, and, the Kumaresan-Tufts approach [42],[61].

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 -5dB 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, expression (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 x 36 matrix estimate R_e. Finally, the optimum autoregressive parameter estimates were generated upon using expression (7.15). The resultant ten AR(35) spectral estimates are shown in superimposed plots in Fig. 8.6a 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. 8.6b 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 Figs. 8.6c and 8.6d, 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.

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 lengths N = 128, N = 256 and N = 1024 are shown superimposed in Figure 8.7. 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 methods, 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 superimposed in Figure 8.8 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].

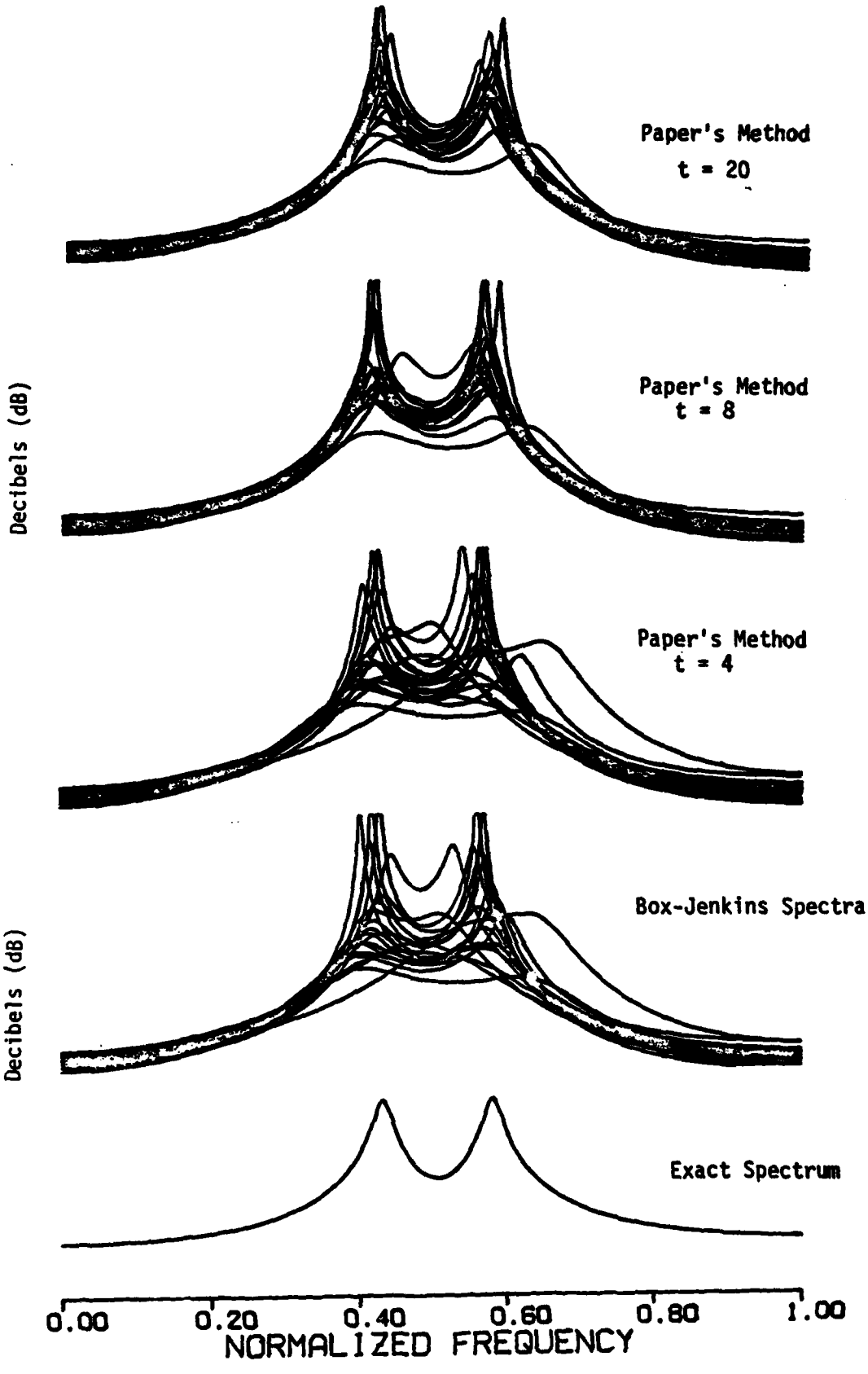


Fig. 8.1 ARMA spectral estimates of order (4,4). (a)Exact. (b) Box-Jenkins maximum-likelihood method. (c) Paper's method for t=4. (d) Paper's method for t=8. (e) Paper's Method for t=20.

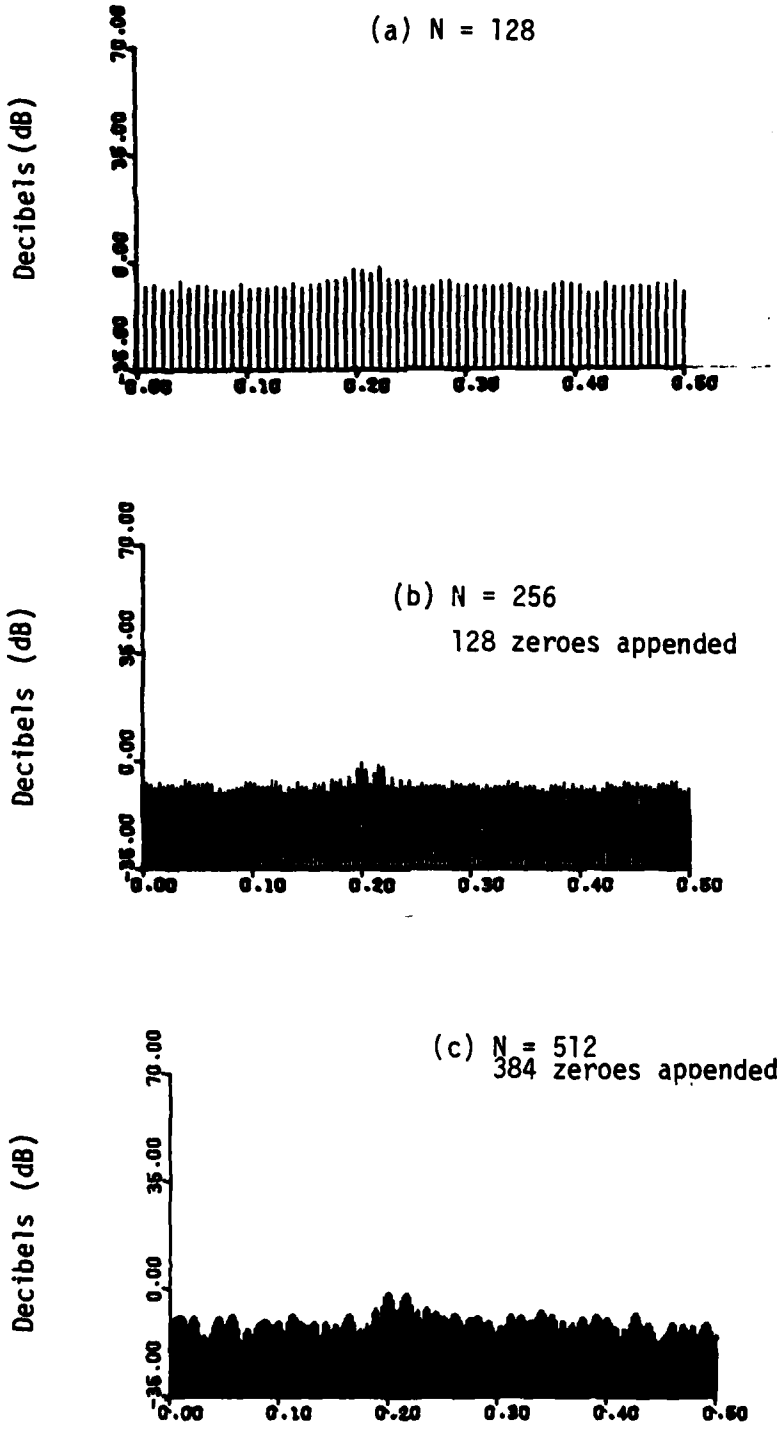


Fig. 8.2 Moving average (MA) spectral estimates using the FFT implementation of the periodogram with 128 time series observations (a) no zero padding, (b) 128 zero padding, (c) 384 zero padding.

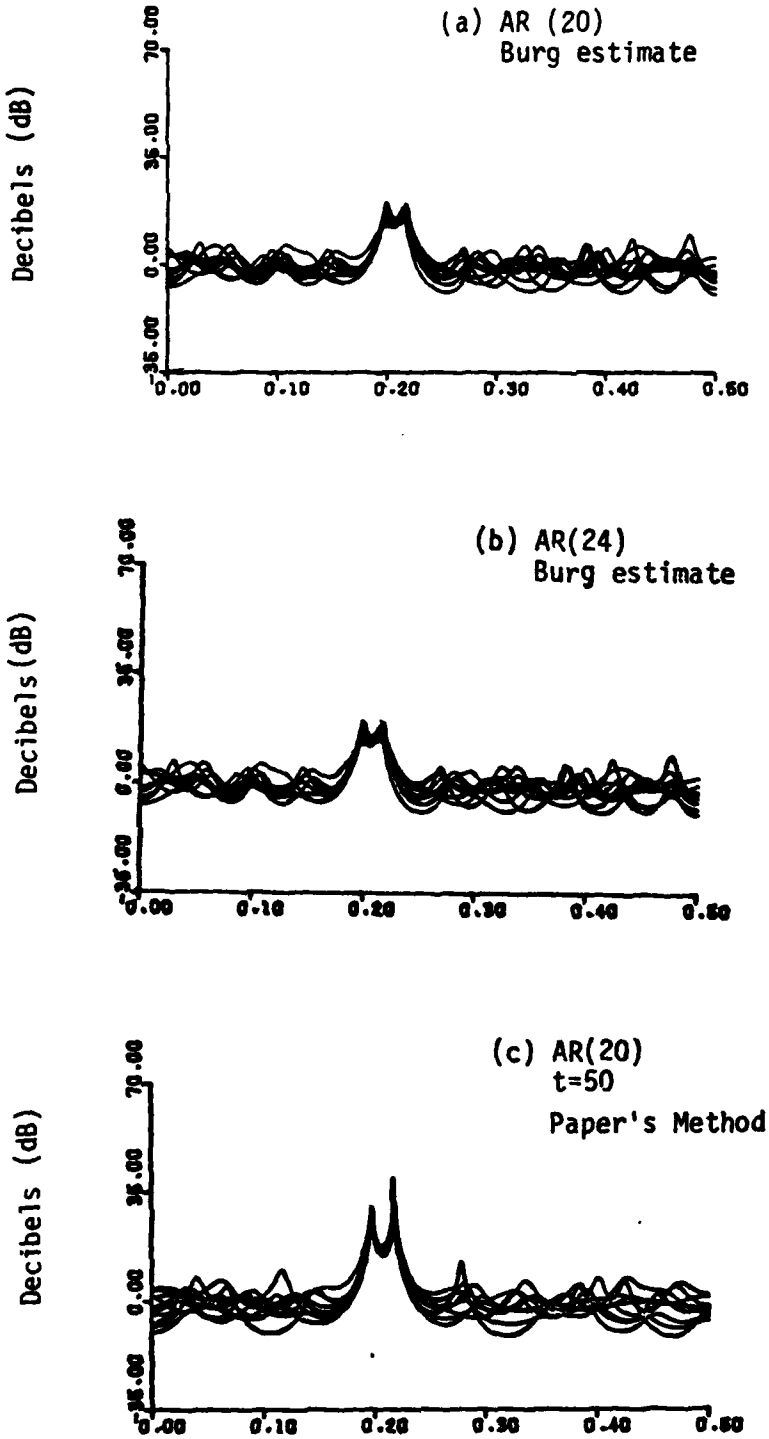
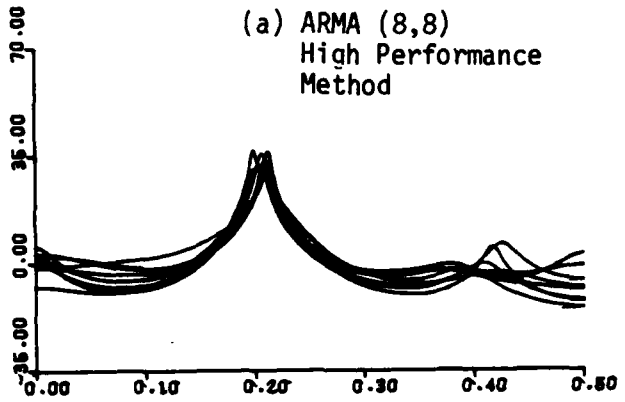
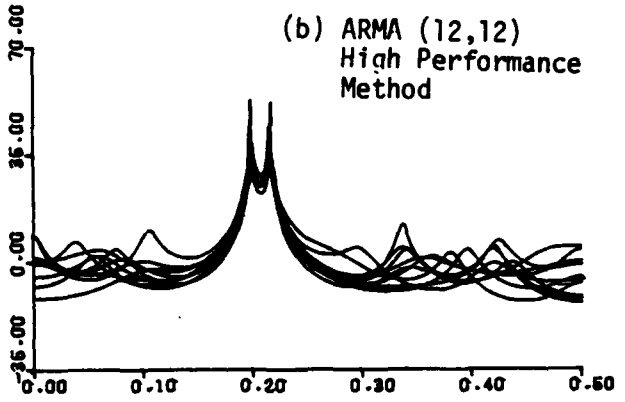
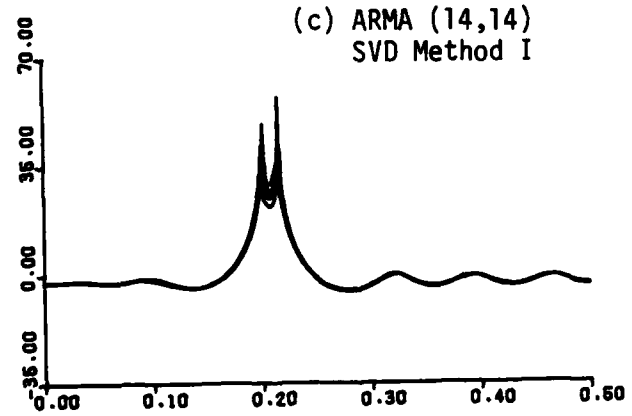


Fig. 8.3 Autoregressive (AR) spectral estimates from 128 time series observations (a) p = 20 Burg estimate, (b) p = 24 Burg estimate (c) This paper's method (6.6) with q = 0, p = 20, t = 56 estimate using expression (6.6).







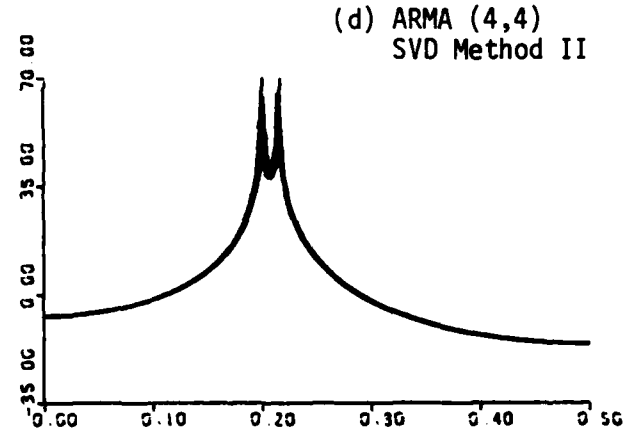


Fig. 8.4 Autoregressive-moving average (ARMA) estimates from 128 time series observations (a) p=q=8, t=70, (b) p=q=12, t=70, (c) SVD method I with $p_e=q_e=14$, t=50 yielding p=4, (d) SVD method II with $p_e=q_e=14$, t=50 yielding p=4.

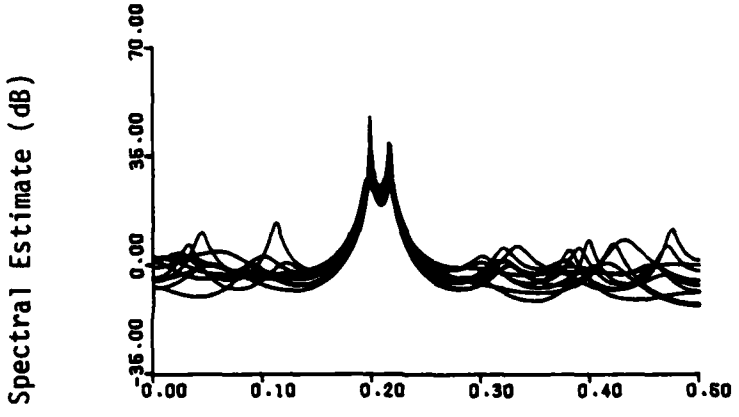


Fig. 8.5 Spectral estimates using alternative method (3.9) with p = 14, t = 50 and W = I.

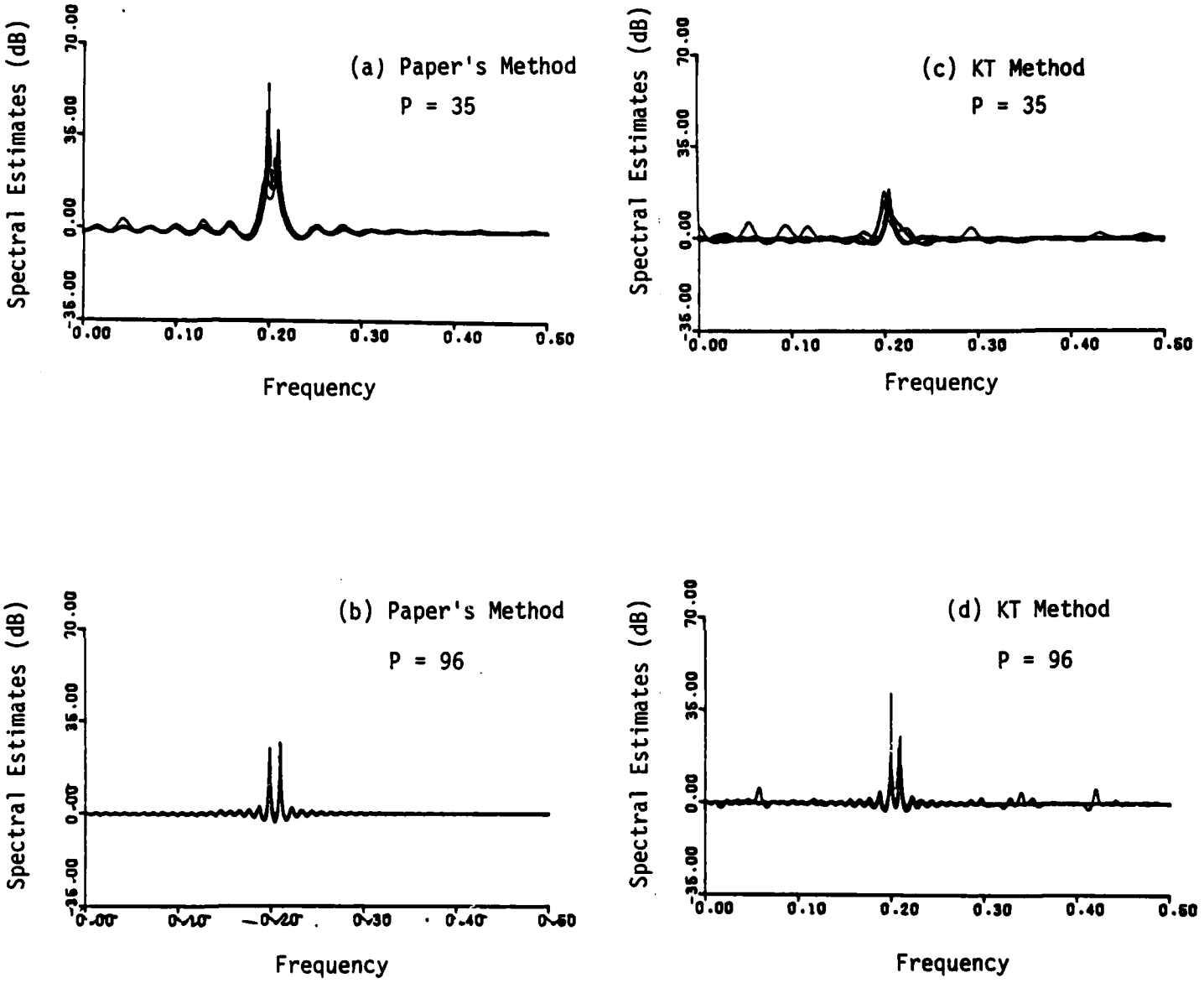


Fig. 8.6 AR Type Models

(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.

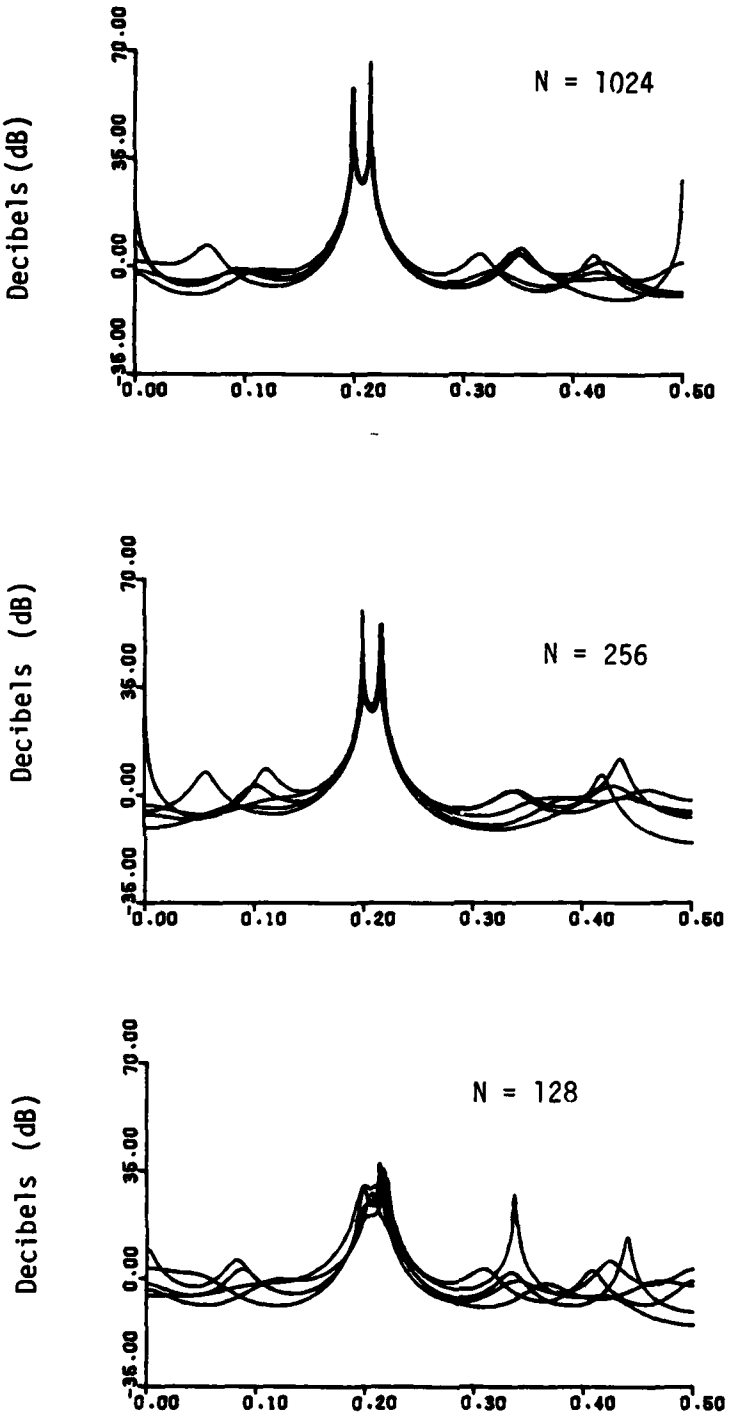


Fig. 8.7 Adaptive ARMA (12,12) spectral estimates with k_1=40, k_2=1. (a) N = 128, (b) N = 256, (c) N = 1024.

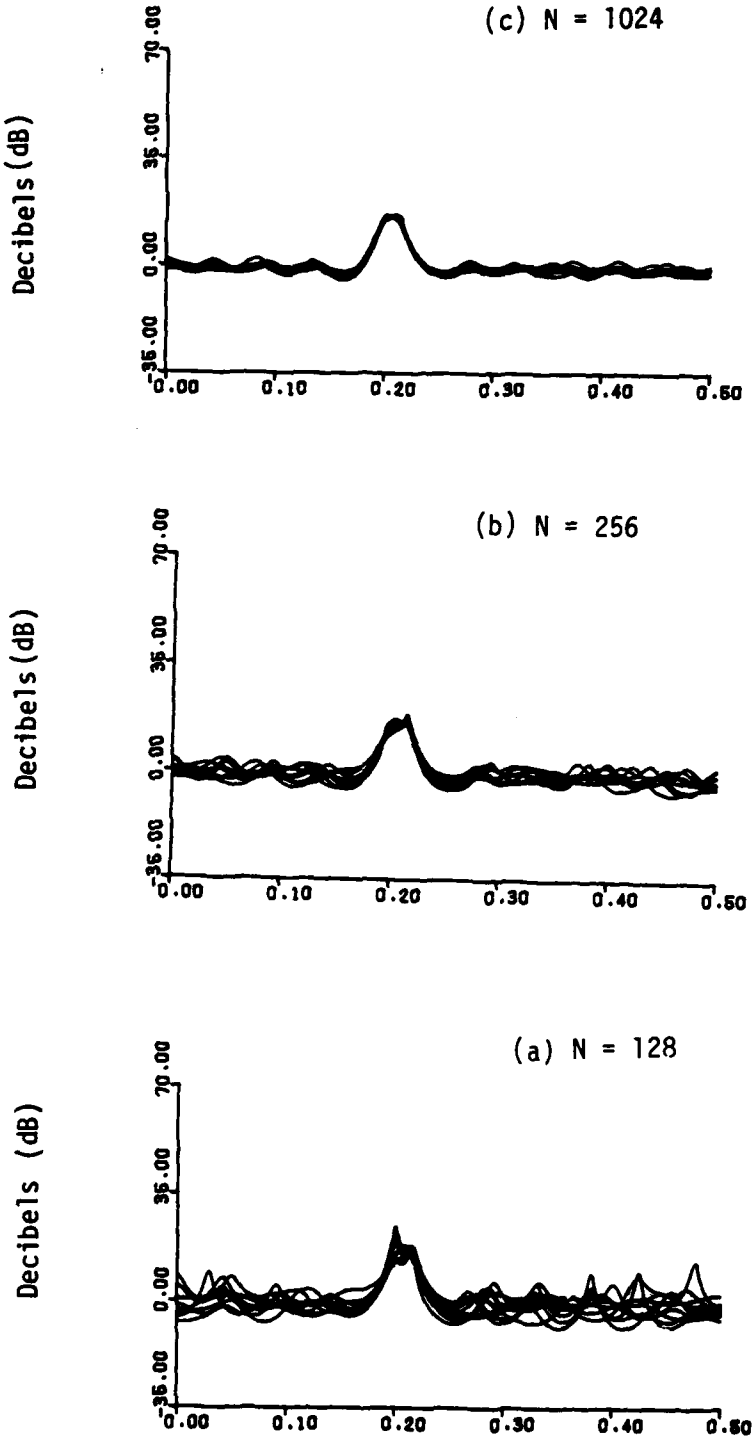


Fig. 8.8 Adaptive AR(12) spectral estimates using the covariance method. (a) N = 128, (b) N = 256, (c) N = 1024.

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

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

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

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

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

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

where it will be recalled that the parameter b_o is chosen so that the first component of $\underline{a}$ 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), ...), 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 amenable to an adaptive solution. With this objective in mind, we shall now consider the <u>adaptive class of autocorrelation estimators</u> as defined by

$$\hat{R}(i,j) = \frac{1}{N+k_2-k_1} \sum_{k=k_1}^{N+k_2-1} \bar{x}(k+1-i)x(k+1-j) \qquad \begin{matrix} 1\leq i\leq p+1 \\ 1\leq j\leq p+1 \end{matrix} \tag{9.4}$$

in which the convention of setting to zero any summand terms x(n) whose argument lies 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 expression (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)[1]. 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 $(p+1)x(p+1)$ autocorrelation matrix estimate that arises upon using the estimates (9.4) as entries can be always expressed in the following <u>data matrix product</u> format

$$\hat{R} = \frac{1}{N+k_2-k_1} X_N^* X_N \tag{9.5}$$

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

$$X_N(i,j) = x(k_1+i-j) \qquad \begin{matrix} 1 \leq i \leq N+k_2-k_1 \\ 1 \leq j \leq p+1 \end{matrix} \tag{9.6}$$

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 expressions (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 $(N+p)x(p+1)$ <u>kernel</u> Toeplitz type matrix which

---

1As 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 ($k_1=p+1, k_2=p+1$).

contains, as submatrices, all of the data matrices associated with the adaptive class of autocorrelation estimators.

$$\chi_N = \begin{bmatrix} x(1) & & \\ \vdots & \ddots & \\ & & \\ x(p+1) & \cdots & x(1) \\ \vdots & & \vdots \\ x(N) & \cdots & x(N-p) \\ & \ddots & \vdots \\ & & x(N) \end{bmatrix} \begin{matrix} \\ \\ \\ k_1 = 1 \\ \text{Prewindowing} \\ k_1 = p+1 \\ \\ k_2 = 1 \\ \text{Postwindowing} \\ k_2 = p+1 \end{matrix} \qquad (9.7)$$

Upon examination of the data matrix definition (9.6), it is apparent that X_N may be identified with that submatrix of χ_N composed of its k_1^{st} through $(N+k_2-1)^{st}$ 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 χ_N are there due to the implicit assumption that $x(n) = 0$ for $-p+1 \leq n \leq 0$. This rather <u>unrealistic</u> assumption concerning an unobserved segment of the time series is commonly referred to as a <u>prewindowing</u> 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 expression (9.7).

In a similar fashion, the zeroes which appear in the lower left corner of matrix χ_N are there due to the implicit <u>postwindow</u> 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 smallest 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 9.1 (e.g., see refs. [25],[46],[47]). Each of the methods there shown 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 9.1, it is seen that a maximum amount of prewindowing, postwindowing, or, 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 $N>>p$, 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

$$X_N^* X_N \underline{a}_N = (N+k_2-k_1)\,|b_o|^2\,\underline{e}_1 \qquad (9.8)$$

in which the normalizing parameter b_o is to be selected so that the first component of $\underline{a}_N$ is one. The data matrix product $X_N^* X_N$ in this expression is

METHOD	CONSTANT k_1	CONSTANT k_2	STATISTICAL PROPERTIES OF 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

Table 9.1 Adaptive AR Autocorrelation Estimation Methods

seen to completely characterize the desired autoregressive parameter vector $\underline{a}_N$ 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 autoregressive parameter vector $\underline{a}_{N+1}$. We can continue this systematic procedure to generate the updated autoregressive parameter vectors $\underline{a}_{N+2}$, $\underline{a}_{N+3}$, 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 data matrix X_N. This results in the following recursion on the data matrix products

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

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

$$\underline{x}_{N+1} = [x(N+1),\ x(N),\ \ldots,\ x(N+1-p)] \tag{9.10}$$

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 expression (9.4) upon setting $k_2=1$ and N=p+1, that is

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

It is interesting to note that although each member of the nonpostwindow class (as identified by $k_2=1$ and $k_1 \varepsilon [1,p+1]$) 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 $\underline{x}_{N+1}\ \underline{x}_{N+1}^*$. 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 expression (9.8) for the AR model parameters. This required matrix inverse recursion will make use of expression (9.9) and the following well known lemma

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

$$[A + \underline{u}^* \underline{v}]^{-1} = A^{-1} - \frac{[A^{-1}\underline{u}^*][\underline{v}A^{-1}]}{(1 + \underline{v}\, A^{-1}\, \underline{u}^*)} \tag{9.12}$$

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

$$[X_{N+1}^* X_{N+1}]^{-1} = [X_N^* X_N]^{-1} - \frac{\underline{y}_{N+1}^* \underline{y}_{N+1}}{1 + \underline{y}_{N+1} \underline{x}_{N+1}^*} \qquad \text{for } N \geq p+k_1$$

where (9.13)

$$\underline{y}_{N+1} = \underline{x}_{N+1}^* [X_N^* X_N]^{-1}$$

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 oneself that the simple three step procedure outlined in Table 9.2 will provide the required adaptive autoregressive parameter vector procedure. The second step is seen to yield the unnormalized solution to expression (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 equation (9.13) indicates that $2(p+1)^2$ operations will be required for updating $[X_N^* X_N]^{-1}$. The resultant autoregressive parameter vector solution as represented by steps 2 and 3 of Table 9.2 will require an additional (p+1) operations. 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 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.

Step 0: Input Data: $x(N+1)$, $[X_N^* X_N]^{-1}$
Step 1: Compute $[X_{N+1}^* X_{N+1}]^{-1}$ using recursion (5.14)
Step 2: Let $\underline{c} = [X_{N+1}^* X_{N+1}]^{-1} \underline{e}_1$
Step 3: $\underline{a}_{N+1} = c(1)^{-1} \underline{c}$ where $c(1)$ is the first component of $\underline{c}$.

Table 9.2: Adaptive AR Modeling Algorithm—Covariance Methods ($N_0=2p+1$) and Prewindow ($N_0=p+1$) methods.

Adaptive Algorithm $k_2 = p+1$

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

$$X_{N+1}^* X_{N+1} = X_N^* X_N + D_{N+1} \qquad N \geq p+1 \tag{9.14}$$

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

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

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

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

will also be wide-sense stationary for any choice of the 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

$$r_y(n) = r_x(n) \tag{9.17}$$

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

$$y(n) = x(N+1-n) \qquad 1 \leq n \leq N \tag{9.18}$$

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

$$\hat{R} = \frac{1}{N+k_2-k_1} Y_N^{*} Y_N \tag{9.19}$$

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

$$Y_N(i,j) = x(N+1-k_1-i+j) \qquad \begin{matrix} 1 \leq i \leq N+k_2-k_1 \\ 1 \leq j \leq p+1 \end{matrix} \tag{9.20}$$

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

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

will 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

$$[X_N^* X_N + Y_N^* Y_N] \underline{a}_N = \alpha \, \underline{e}_1 \tag{9.22}$$

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

AR Model Order Determination

One of the principal 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 ref. [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)x(p+1) autocorrelation matrix R with elements

$$R(i,j) = r_x(i-j) \qquad 1 \leq i,j \leq p+1 \tag{9.23}$$

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)x(p+1) autocorrelation matrix will be p_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 don't correspond to an AR process, then the matrix R will be generally of full rank for <u>all</u> $p \geq 1$.

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 <u>all</u> 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 R. Nonetheless, even though $\hat{R}$ 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 behaviour 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 X.

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 R_1 as required in expression (6.6). This class of adaptive estimators will be governed by the relationship

$$\hat{R}_1(i,j) = \frac{1}{N+k_2-k_1-q-1} \sum_{k=k_1}^{N-q-2+k_2} x(k+1-i)x(k+q+2-j) \qquad \begin{matrix} 1 \le i \le t \\ 1 \le j \le p+1 \end{matrix} \qquad (10.1)$$

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 equation (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 autoregressive parameter when using expression (6.6). For reasons which will be shortly made apparent, these constants are usually further constrained to satisfy $1 \le k_1 \le t$ and $1 \le k_2 \le p+1$ although other choices are possible.

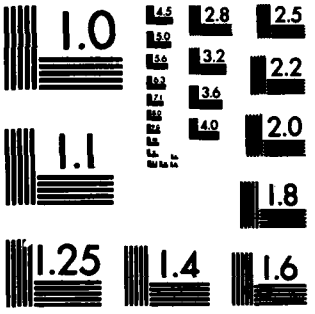
Each autocorrelation estimator in the adaptive class (10.1) will be identified by a particular choice of the two-tuple (k_1,k_2). Moreover, each estimator in this class will provide a generally different set of autocorrelation lag estimates from the set of time series observations $x(n)$ for $1 \le n \le N$. Clearly, our ultimate desire is to select that estimator which generally provides the best ARMA modeling. The <u>covariance</u> 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 the adaptive 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

AD-A123 122 SPECTRAL ESTIMATION: AN OVERDETERMINED RATIONAL MODEL EQUATION APPROACH..(U) ARIZONA STATE UNIV TEMPE DEPT OF ELECTRICAL AND COMPUTER ENGI.. J A CADZOW 15 SEP 82 2/2

UNCLASSIFIED AFOSR-TR-82-1050 N00014-82-K-0257 F/G 12/1 NL

END
DATE FILMED
2 83
DTIC



as formed from the entries (10.1) may be always 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 $(N+k_2-k_1-q-1)x(p+1)$ data matrix X_N has its elements specified by

$$X_N(i,j) = x(k_1+q+1+i-j) \qquad \begin{matrix} 1 \le i \le N+k_2-k_1-q-1 \\ 1 \le j \le p+1 \end{matrix} \qquad (10.3)$$

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

$$Y_N(i,j) = x(k_1 + i-j) \qquad \begin{matrix} 1 \le i \le N+k_2-k_1-q-1 \\ 1 \le j \le t \end{matrix} \qquad (10.4)$$

A simple matrix manipulation will prove the equivalencies of expressions (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 $1 \le n \le N$, and, we also attach the subscript N to these data matrices so as to explicitly recognize there 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

$$\chi_N = \begin{matrix} k_2=1 \\ \updownarrow \\ k_2=p+1 \end{matrix} \begin{bmatrix} x(q+2) & \cdots & x(q-p+2) \\ \vdots & & \\ x(t+q+1) & \cdots & x(t+q-p+1) \\ \vdots & & \vdots \\ x(N) & \cdots & x(N-p) \\ & \ddots & \\ & & x(N) \end{bmatrix} \qquad y_N = \begin{bmatrix} x(1) & & \\ & \ddots & \\ & & \\ x(t) & \cdots & x(1) \\ \vdots & & \vdots \\ x(N-q-1) & \cdots & x(N-q-t) \\ \vdots & & \vdots \\ x(N+p-q-1) & \cdots & x(N+p-q-t) \end{bmatrix} \begin{matrix} k_1=1 \\ \uparrow \\ \text{prewindowing} \\ \downarrow \\ k_1=t \end{matrix} \qquad (10.5)$$

Upon examination of expressions (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 $\mathcal{X}_N$ (or $\mathcal{Y}_N$) composed of its k_1^{st} through $(N-q-2+k_2)^{st}$ 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 kernal matrix $\mathcal{Y}_N$ are there due to the implicit prewindow assumption that $x(n) = 0$ for $2-t \leq n \leq 0$. 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 $k_1 \geq t$ will avoid any data prewindowing. On the other hand, a degree of prewindowing is incorporated whenever k_1 is such that $1 \leq k_1 \leq t-1$. 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 $\mathcal{X}_N$ are there due to the implicit postwindow assumption that $x(n) = 0$ for $N+1 \leq n \leq N+p$. This contrived assumption on an unobserved segment of the time series is also to be avoided. *Upon examination of the kernel matrix* $_N$, it is apparent that postwindowing may be avoided by selecting $k_2 \leq 1$. 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 10.1 in which it is noted that each involves combinations of maximum windowing and no windowing. The <u>covariance</u> 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 <u>largest</u> 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 <u>covariance</u> method typically provides the best modeling performance from the ARMA adaptive class of autocorrelation estimators.

The three other methods listed in Table 10.1 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 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 10.1 will provide comparable modeling performance. This is a consequence of the fact that the windowed portions of the data matrices $\mathcal{X}_N$ and $\mathcal{Y}_N$ play a proprtionately 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.

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

Table 10.1: Four ARMA Adaptive Autocorrelation Estimator Methods

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 expression (10.2) into the basic ARMA modeling equation (6.6). The resultant autoregressive parameter vector is then obtained by solving the normal equations

$$\mathcal{X}_N^{\dagger}\mathcal{Y}_N\mathcal{Y}_N^{\dagger}\mathcal{X}_N\,\underline{a}_N = \alpha\,\underline{e}_1 \qquad (10.6)$$

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 $\underline{a}_N$ is

one. From this expression it is apparent that the data matrix product $X_N^* X_N Y_N^* X_N$ completely identifies the ARMA model's autoregressive parameter vector. In order to compute $\underline{a}_N$, it will be then necessary to compute the data matrix product's inverse at each value of N where the autoregressive parameter vector is required. This can be a particularly imposing computational task if real time signal processing is to be archieved.

Adaptive Algorithm: $k_2 = 1$

When the autoregressive 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 expression (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., $1 \leq k_1 \leq t$). Namely, upon examination of expression (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

$$X_{N+1}^* X_{N+1} = X_N^* X_N + \underline{y}_N^* \underline{x}_N \qquad N \geq t \qquad (10.7a)$$

where $\underline{x}_N$ and $\underline{y}_N$ are the above mentioned 1x(p+1) and 1xt row vectors, that are appended to X_N and Y_N, respectively. These vectors are specified by

$$\underline{x}_N = [x(N+1),\ x(N),\ \ldots,\ x(N+1-p)] \qquad (10.7b)$$

$$\underline{y}_N = [x(N-q),\ x(N-q-1),\ \ldots,\ x(N+1-q-t)] \qquad (10.7c)$$

It is to be noted that recursive expression (10.7a) holds only for time indices $N \geq t$ 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 purturbation matrix $\underline{y}_N^* \underline{x}_N$ 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 expression (10.7a) will be different for various choices of k_1.

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

$$X_{N+1}^* Y_{N+1} Y_{N+1}^* X_{N+1} = X_N^* Y_N Y_N^* X_N + \underline{x}_N^* \underline{x}_N + \underline{z}_N^* \underline{z}_N + (\underline{y}_N \underline{y}_N^*)\, \underline{z}_N^* \underline{z}_N \qquad N \geq t \qquad (10.8a)$$

where $\underline{z}_N$ is the 1x(p+1) vector given by

$$\underline{z}_N = \underline{y}_N\, Y_N^* X_N \qquad (10.8b)$$

An examination of this recursive expression indicates that $t(p+1)$ operations are required to compute $\underline{z}_N$ 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 $\underline{Y}_N^*\underline{Y}_N$ and $\underline{X}_N^*\underline{Y}_N\underline{Y}_N^*\underline{X}_N$ are available.

When updates of the autoregressive parameter vectors $\underline{a}_N$ are not required at each time index N, we could then use recursions (10.7) and (10.8) to compute the data matrix products $\underline{Y}_N^*\underline{Y}_N$ and $\underline{X}_N^*\underline{Y}_N\underline{Y}_N^*\underline{X}_N$ in a computationally efficient manner. At those time instants at which the evaluation of $\underline{a}_N$ 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 autoregressive parameter vector at each (or nearly each) 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 product $[\underline{X}_N^*\underline{Y}_N\underline{Y}_N^*\underline{X}_N]^{-1}$. To effect this recursion, we note from relationship (10.8) that the matrix products $\underline{X}_N^*\underline{Y}_N\underline{Y}_N^*\underline{X}_N$ at two contiguous time indices (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 10.2. 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 $\underline{X}_N^*\underline{Y}_N\underline{Y}_N^*\underline{X}_N$ 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 term 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 10.2 is for the particular nonpostwindowing selection $k_2=1$ wherein the parameter k_1 will be typically selected to satisfy $1 \leq k_1 \leq t$. 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 the 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 $o(p\ np)$ for each new time series observation.

STEP 0	The input to commence the algorithm at $N = q+p+k_1+1$ is $X_N^* Y_N$ and $[X_N^* Y_N Y_N^* X_N]^{-1}$
STEP 1	$N = q+p+k_1+1$
STEP 2	Compute $X_{N+1}^* Y_{N+1}$ using expression (10.7)
STEP 3	$\underline{z}_N = \underline{y}_N Y_N^* X_N$
STEP 4	$\underline{u}_1 = \underline{z}_N$, $Y_1 = \underline{z}_N$, $A_1^{-1} = [X_N^* Y_N Y_N^* X_N]^{-1}$ Compute $[A_1 + \underline{u}_1^* Y_1]^{-1}$ using Lemma 9.1
STEP 5	$\underline{u}_2 = \underline{z}_N^*$, $Y_2 = \underline{z}_N$, $A_2^{-1} = [A_1 + \underline{u}_1^* Y_1]^{-1}$ Compute $[A_2 + \underline{u}_2^* Y_2]^{-1}$ using Lemma 9.1
STEP 6	$\underline{u}_3 = (Y_N \underline{y}_N^*)\, \underline{z}_N$, $Y_3 = \underline{z}_N$, $A_3^{-1} = [A_2 + \underline{u}_2^* Y_2]^{-1}$ Compute $[A_3 + \underline{u}_3^* Y_3]^{-1} = [X_{N+1}^* Y_{N+1} Y_{N+1}^* X_{N+1}]^{-1}$ using Lemma 9.1
STEP 7	$\underline{c} = [X_{N+1}^* Y_{N+1} Y_{N+1}^* X_{N+1}]^{-1} \underline{e}_1$ $\underline{z}_{N+1} = c(1)^{-1} \underline{c}$ where $c(1)$ is the first component of $\underline{c}$
STEP 8	Let $N = N+1$, GO TO STEP 2

TABLE 10.2: Adaptive Algorithm for Computing $\underline{z}_{N+1}$

XI. Acknowledgments

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

XII. 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 often exceed that of existing techniques. Studies are currently under way to more effectively use this SVD adaption for achieving yet further performance improvements.

XIII. References

1. H. Akaike, 'Fitting autoregressive models for prediction,' Ann. Inst. Statist. Math., Vol. 21, pp. 243-247, 1969.

2. H. Akaike, 'Power spectrum estimation through autoregressive model fitting,' Ann. Inst. Statist. Math, Vol. 21, pp. 407-419, 1969.

3. H. Akaike, Maximum likelihood identification of Gaussian autoregressive moving-average models, Biometrica, vol. 60, no. 2, pp. 255-265, Aug. 1973.

4. H. Akaike, 'A new look at the statistical model identificant,' IEEE Trans. Autom. Control, Vol. AC-19, pp. 716-723, December 1974.

5. N. O. Andersen, 'On the calculation of filter coefficients for maximum entropy spectral analysis,' Geophys., Vol. 39, pp. 69-72, February 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, 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, Dover Publications, Inc., New York, 1958.

9. G. Box and G. Jenkins, TIME SERIES ANALYSIS: FORECASTING AND CONTROL (REVISED EDITION), San Francisco, Holden Day, 1976.

10. T. P. Bronez and J. A. Cadzow, An algebraic approach to superresolution array processing, 1981 ICASSP Proceedings, pp. 302-305, Atlanta, Ga., 1981 and accepted for publication in the IEEE Trans. on Aerospace and Electronics Systems.

11. S. Bruzzone and M. Kaveh, On some suboptimum ARMA spectral estimators, IEEE Trans. Acousts., Speech, Signal Processing, Vol. ASSP-28, pp. 753-734, Dec. 1980.

12. K. J. Bry and J. LeRoux, Comparison of some algorithms for identifying autoregressive signals in the presence of observation noise, 1982 ICASSP, Paris, France, pp. 224-227, May 1982.

13. J. P. Burg, 'Maximum entropy spectral analysis,' Proc. 37th meeting Society of Exploration Geophysicists, Oklahoma City, OK, October 31, 1967.

14. J. P. Burg, '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,' Proc. of the RADC spectrum estimation workshop, pp. 81-97, October 1979.

16. J. A. Cadzow, 'High performance spectral estimation-a new ARMA method,' IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-28, pp. 524-529, October 1980.

17. J. A. Cadzow, 'Autoregressive-moving average spectral estimation: a model equation error procedure,' IEEE Trans. Geoscience and Remote Sensing, Vol. GE-19, pp. 24-28, January 1981.

18. J. A. Cadzow and Koji Ogino, Two-dimensional spectral estimation, IEEE Trans. Acoustics, Speech, and 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,' First ASSP Workshop on Spectral Estimation, McMaster Univ., Hamilton, Ont., Canada, August 1981.

20. J. A. Cadzow, 'ARMA modeling of time series,' IEEE Trans. Patt. Analy. and Mach. Intel., special issue on digital signal and waveform analysis, vol. PAMI-4, no. 2, pp. 124-128, March 1982.

21. J. A. Cadzow, 'Rational time series modeling: an effective method,' IEEE Trans. Aerosp. Elect. Syste., Vol. AES- , pp. , August 1982.

22. J. A. Cadzow and B. Baseghi, 'Data adaptive ARMA modeling of time series,' 1982 ICASSP, Paris, France, pp. 256-261, May 1982.

23. D. G. Childers (Editor), 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 inversin formulas for matrices classified in terms of their distance from Toeplitz matrices,' Linear Algebra and its applications, Vol. 27, pp. 31-60, 1979.

26. W. Gersh, Estimation of the autoregressive parameters of mixed autoregressive moving-average time series, IEEE Trans. on Automatic Control, 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. Automatic Control, 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. Automatic Control, vol. AC-19, no. 6, pp. 774-783, Dec. 1974.

30. P. R. Gutowski, E. A. Robinson, and, S. Treitel, Spectral estimation: fact or fiction?, Vol. GE-16, pp. 80-84, April 1978.

31. S. S. Haykin (Editor), NONLINEAR METHODS OF SPECTRAL ANALYSIS, New York: Springer-Verlag, 1979.

32. T. L. Henderson, Geometric methods for determining system poles from transient response, IEEE Trans. Acoustics, Speech and 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. Automa. Contr., Vol. AC-25, pp. 996-998, October 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. Acoustics, 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, November 1981.

38. J. F. Kinkel, J. Perl, L. Scharf, and A. Stubberud, '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, April 1979.

39. V. C. Klema and A. J. Laub, 'The singular value decomposition: its computation and some applications,' IEEE Trans. Autom. Control, Vol. AC-25, pp. 164-176, April 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. Acoustics, 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,' First ASSP Workshop on Spectral Estimtaion, McMasters Univ., Hamilton, Ont., Canada, August 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, October 1977.

45. R. K. Mehra, On-line identification of linear dynamic systems with applications to Kalman filtering, IEEE Trans. Automatic Control, Vol. AC-16, no. 1, pp. 12-22, Feb. 1971.

46. M. Morf, T. Kailath, and L. Ljung, 'Fast algorithms for recursive identification,' Proc. 1976 IEEE Conf. Decision and Control, Clearwater, FL., pp. 916-921, December 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, October 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 Polytechnic Institute, Blacksburg, VA, 1981.

50. A. V. Oppenheim and R. W. Schafer, DIGITAL SIGNAL PROCESSING, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1975.

51. Y. Pao and D. T. Lee, Performance characteristics of the Cadzow modified direction ARMA method for spectrum estimation, 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. E. Parzen, 'Statistical spectral analysis (single channel case) in 1968,' Dept. Statistics, Stanford University, Stanford, CA., Tech. Rept. 11, June 1968.

54. E. Parzen, 'Some recent advances in time series modeling,' IEEE Trans. Automat. Control, Vol. AC-19, pp. 723-730, December 1974.

55. V. F. Pisarenko, The retrieval of harmonics from a covariance function, Geophys. J. Royal Astr. Soc., pp. 347-366, 1973.

56. R. Prony, Essai experimentale et analytique, Paris J. l'Ecole Polytechnique, vol. 1, pp. 24-76, 1975.

57. L. R. Rabiner and B. Gold, THEORY AND APPICATION OF DIGITAL SIGNAL PROCESSING, Prentice-Hall, Inc., Englewood Cliffs, N. J., 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, March 1898.

59. J. L. Shanks, Recursion filters for digital processing, Geophysics, vol. 32, pp. 33-51, Feb. 1967.

60. S. A. Tretter and K. Steiglitz, Power spectrum identifications in terms of rational models, IEEE Trans. Automatic Control, vol. AC-12, pp. 185-188, April 1967.

61. D. W. Tufts and R. Kumaresan, Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood, a paper appearing in this special issue.

62. T. J. Ulrych and R. W. Clayton, Time series modelling and maximum entropy, 'Phys. Earth Planetary Interiors, Vol. 12, pp. 188-200, August 1976.

63. G. Walker, 'On periodicity in series of related terms,' Proc. Roy. Soc. London, Series 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, December 1975.

66. G. U. Yule, 'On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers,' Philosophical Trans. Royal Soc. London, Series A., Vol. 226, pp. 276-298, July 1927.