AD 710390

NOLTR 70-109

THE SPECTRAL ANALYSIS OF DISCRETE TIME SERIES IN TERMS OF LINEAR REGRESSIVE MODELS

By
Edward C. Whitman

23 JUNE 1970

NOL

UNITED STATES NAVAL ORDNANCE LABORATORY, WHITE OAK, MARYLAND

NOLTR 70-109

ATTENTION

This document has been approved for public release and sale, its distribution is unlimited.

THE SPECTRAL ANALYSIS OF DISCRETE TIME SERIES IN TERMS OF LINEAR REGRESSIVE MODELS

Prepared by:

Edward C. Whitman

ABSTRACT: This paper considers several methods for the spectral analysis of discrete time series modeled as linear regressive processes. The most general of these models is the so-called "mixed-type" process in which the present value of the series is given as a weighted sum of both its own past values and those of an uncorrelated random sequence. Special cases of this class are the autoregression and the moving average. In the present approach, the determination of the weighting coefficients of the model is equivalent to specifying the sampled power spectrum of the process, and the central problem treated here is that of estimating such a set of parameters on the basis of a sample sequence of the series.

After a development of the necessary mathematical background, the estimation problem is formulated and solved from several alternative points of view, with particular attention to statistical stability, sample size requirements, and possible approximation errors. For each method, the results are demonstrated by a computational analysis of the spectra of a set of computer generated examples. A general purpose spectral analysis algorithm for digital computation is proposed and discussed.

Methods for extending an existing spectral estimate to take into account newly available data are briefly explained, and suggestions for further research are presented.

PUBLISHED 23 JUNE 1970

U. S. NAVAL ORDNANCE LABORATORY
WHITE OAK, MARYLAND

NOLTR 70-109 23 June 1970

The Spectral Analysis of Discrete Time Series in Terms of Linear Regressive Models

This report is concerned with the spectral analysis of discrete time series modeled as linear regressive processes and derives several methods for estimating spectral representations in those terms. The theoretical results are confirmed by a series of computational examples, and the application of these techniques to several areas of interest is discussed. The work was partially funded under WEPTASK MAT 03L-000/ZF17-312-001, Prob 009. The report will be of interest to those concerned with the theory of discrete stochastic processes, digital spectral analysis, adaptive processing and control, and sampled data systems.

The author wishes to acknowledge his indebtedness to personnel of the Computer Applications Division (Code 332) for their aid in preparing the computer programming that underlies many of the results.

In its original form, this paper was submitted to the Graduate Faculty of the University of Maryland in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering.

GEORGE G. BALL
Captain, USN

E. H. Beach
E. H. BEACH
By direction

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure | Page

BLANK PAGE

CHAPTER I

INTRODUCTION

The subject of the research reported here is the power spectral analysis of discrete time series in terms of linear regressive models. As such, it lies within a large body of techniques and methods which has grown up over the years for the interpretation and use of the power spectrum in signal description and processing. The great utility of spectral analysis has become such a truism in so many scientific and technical areas that little point is served in discussing it at any length here. Of most interest to the practicing electrical engineer is its role as a powerful tool in the design and evaluation of circuits and systems intended for use with random or information bearing signals, but not to be overlooked is its value in tracing cause-effect relationships in other fields of physics, engineering, and even económic analysis.

As digital computation has become more widespread, and as more and more physical data are collected in sampled or digital form, spectral measurement techniques specifically intended for use with discrete sequences of data have assumed increased importance. In addition, a substantial body of theory has grown up for treating control systems operating on sampled data and for developing adaptive processing techniques specifically utilizing the advantages of the digital computer. The present research has been aimed at extending the techniques available for the spectral estimation of discrete time series and is largely based on well known statistical results in the study of regressive series. This approach is somewhat different from traditional methods which rely on the transform properties of mean lagged products from the process of interest, or more recent developments in the area of Fast Fourier Transforms (FFT's). Theoretical treatments of the former are found in Hannan[1], Bartlett[2], and Grenander and Rosenblatt[3], with the best known expositions for practical use found in Blackman and Tukey[4] and Bendat and Piersol[5]. An enormous quantity of literature on the latter method has appeared since its introduction by Cooley and Tukey[6]. The author has found Cochran et al.[7], Welch[8], and Cooley, Lewis and Welch[9] particularly valuable.

In the present work, the discrete process of interest is modeled as the solution of a linear stochastic difference equation driven by a stationary random sequence. Under this assumption, specification of the coefficients of the equation and the variance of the driving sequence is equivalent to specifying the spectrum of the process. In many applications, among them adaptive processing and control, this representation is very convenient, since the number of parameters to be determined is often quite small, and because the coefficients of such a modeling equation can be used directly to implement digital filters for whitening and compensation. Also, this formulation is exact for any process that results from passing white noise through a digital filter with a rational transfer function, and it is thus appropriate to a wide variety of physical situations. The fundamental problem considered here, therefore, is that of modeling a process as a linear regression and then operating on an available sample sequence to derive estimates of the regression coefficients and the variance of the driving noise. These are then used to prepare the required estimate of the power spectrum. This is quite similar to the familiar "system identification problem" in which sample sequences of the input and output of a digital system are made the basis for an estimate of the parameters of the system itself[10-16]. In the present case, however, only the output is assumed available for study, and the input is considered to be a stationary random sequence of unknown variance.

A good deal of study has been devoted to the problem of estimating the coefficients of a stochastic difference equation by observing a sample sequence of the solution. The most tractable form in this regard is a difference equation that produces an autoregression, a process in which the present value can be expressed as a weighted sum of past values with the addition of an independent noise perturbation. The problem of estimating the coefficients of an autoregression was first treated by Mann and Wald[17], who used a maximum likeihood approach to derive the estimation equations and the statistical properties of the estimates in the limit of large sample size. This result has been concisely discussed in Hannan[1], and has been applied to spectral analysis by Steiglitz[18]. It can be extended to approximate the spectra of non-autoregressive processes on the ground (shown by Wold[19]) that a large class of discrete series can be represented as infinite order autoregressions. This approach, which yields so-called "all-pole" estimates, has been utilized by Whittle[20] and Tretter and Steiglitz[21], among others.

Another discrete random process of considerable importance is the moving average, in which the present value is given as the weighted sum of past values of an uncorrelated sequence. The most general of those processes having a rational spectral density function is a combination of the moving average and the autoregression: its present value is given by a weighted sum of both its own past values and those of an independent random sequence. It will be referred to here as a "process of mixed type". A general solution for the large sample estimation of the parameters of a mixed type process, based on a linear least squares approach, has been provided by Whittle[22,23], but its towering computational difficulties have precluded wide acceptance (For a discussion of these difficulties, see Hannan[1] and Walker[24].). Similarly, the estimation equations derived on the basis of maximum likelihood are highly non-linear and cannot be solved in closed form. Thus, previous workers have been led to a number of numerical, iterative methods, including those proposed by Tretter and Steiglitz[21] and Zetterberg[25].

Fortunately, there is a straightforward procedure for separately estimating the autoregressive coefficients of a mixed type process, and with these in hand, the original sample sequence can be filtered to yield a signal consisting only of the moving average part of the process (or at least an approximation to it). The remainder of the spectral analysis then consists of estimating the coefficients of this moving average, and this problem has also received attention in the past. The most fundamental approach is that of Wold[19], who provides a method for determining the coefficients of a moving average whose theoretical correlation sequence matches an observed sequence of mean lagged products (when such a moving average exists). This can be computationally awkward, however, and frequently the experimental data are such that no solution can be found. An iterative method is presented by Walker[24] and has been extended to deal with the general mixed type process[26], but in general, spectral estimation for moving average processes has been somewhat neglected in the literature.

The present work has sought to apply the statistical results described above specifically to the estimation of sampled power spectra, and has concentrated on non-iterative methods and the more straightforward spectral approximations. The several methods and procedures can be combined to yield a number of alternative approaches to the spectral analysis as a whole. These were analyzed in terms of their error performance, sample size requirements and applicability to typical situations. An extensive computational program was then carried out in which time series with known spectra were artificially generated and analyzed using the methods considered here. A direct comparison of the alternative treatments has thus been possible for a fairly broad class of spectral examples.

The next chapter is a summary of mathematical definitions and principles required in the derivations that follow. It is drawn largely from digital system theory, statistics, and early work on regressive series. It also contains an original matrix method for computing the autocorrelation sequence of a process of mixed type from a knowledge of the set of regression coefficients that define the process. The purpose of this chapter of mathematical preliminaries is basically to make the work as a whole reasonably self contained.

Chapter III is a study of spectral analysis for mixed type processes using all-pole approximations. A new criterion of optimality is proposed for deriving the well known all-pole estimates, and the distinction between approximation and statistical errors is studied in some detail.

The following chapter considers the problem of separately estimating the autoregressive coefficients of a mixed type process, with particular emphasis on the statistical errors that arise from using a finite sample size. The effect of these errors on the autoregressive part of the estimated spectrum is also considered, and the use of the estimated autoregressive polynomial for filtering out the moving average part of the process is analyzed to determine the magnitude of the errors that result from that operation.

Chapter V provides means for completing the spectral analysis by covering methods for determining the spectrum of the moving average part. The first of these is a simple all-pole approximation, the second Wold's approach, and the third, an approximating procedure that provides asymptotically exact estimates as an intermediate model order is increased without limit. Again, the error performance of the alternative treatments is described.

The final section of each of these chapters consists of a series of computational examples intended to demonstrate the methods derived. The latter are applied to computer generated series corresponding to five standard spectral examples chosen for their generality and described in Appendix A. The final plots of the separate spectral estimates provide a convenient means for comparing the performance of the methods studied here.

Chapter VI proposes two methods for extrapolating spectral estimates when new data becomes available, so that one can avoid repeating the entire estimation procedure.

The final chapter is a general discussion of the findings of the study with recommendations for using one or the other of the spectral measurement techniques, depending upon the particular situation. It attempts to highlight the strengths and weaknesses of the various methods and contains suggestions for further research in areas uncovered by the present work.

CHAPTER II

MATHEMATICAL PRELIMINARIES

The present chapter provides a selection of mathematical and statistical results that will be useful in the development that follows. The treatments here are necessarily abbreviated, but references are provided for more detailed and rigorous approaches.

A. z-Transform Theory

One of the most efficient ways of treating discrete time series is in terms of the z-transform calculus widely used in the analysis of sampled data control systems. Good elementary developments are found in Tou[27] and Ragazzini and Franklin[28], and this section will provide only those salient points required below.

For any number sequence $\{f_j\}_{j=-\infty}^{\infty}$, the z-transform is defined as

$$F(z) = \sum_{j=-\infty}^{\infty} f_j z^{-j} \qquad (2\text{-}1)$$

where z is complex[1]. If the sequence f_j is a set of consecutive samples of a continuous time function f(t), taken at

---

1. Strictly speaking, this expression is known in the literature as the double-sided z-transform, appropriate to number sequences defined for both positive and negative argument. Somewhat more common is the single-sided z-transform used when $f_j = 0$ for $j<0$ and identical with (2-1) except that the lower summation limit is 0.

sampling interval T, it can be shown that F(z), evaluated for $z = e^{sT}$, represents the Laplace transform of the impulse sampled version of the continuous signal.

z-transform techniques are of great value in dealing with linear sampled-data systems. The behavior of such a system in operating on an input number sequence to produce an output sequence is completely specified by its unit sample response, h_j, defined as the output yielded by the device in response to the input δ_j, where

$$\delta_j = 1 \ , \ j = 0$$
$$\delta_j = 0 \ , \ j \neq 0 \qquad (2\text{-}2)$$

If h_j is zero for $j<0$, the system is said to be realizable, in the sense that its output does not anticipate its input. The response, x_j, to any input sequence, f_j, can be written in terms of a convolution sum as follows:

$$x_j = \sum_{k=-\infty}^{\infty} f_k h_{j-k} \qquad (2\text{-}3)$$

This is directly analogous to the convolution integral relating the input and output of a continuous linear system through the impulse response. This relationship can be expressed very conveniently in the z-transform domain. If X(z) is the z-transform of the sequence x_j, and F(z) is that of the sequence f_j, then it can be shown that

$$X(z) = F(z)H(z) \tag{2-4}$$

where H(z) is the z-transform of the unit sample response:

$$H(z) \equiv \sum_{j=0}^{\infty} h_j z^{-j} \tag{2-5}$$

The digital system function H(z) thus also completely characterizes the system, which is frequently portrayed as

Figure 2-1. Digital system characterized by H(z).

A system is said to be stable if its output in response to a bounded input is also bounded. For a digital system characterized by its unit sample response h_j, a sufficient condition for stability is that

$$\sum_{j=0}^{\infty} |h_j| < \infty \tag{2-6}$$

It can be shown that an equivalent condition on the sampled system function H(z) is that it have no poles or other singularities on or outside of the unit circle of the z-plane.

Perhaps the most common digital system encountered in practice is one in which the input and output are related by a linear difference equation of finite order. This can be expressed in the form

$$\sum_{k=0}^{K} a_k f_{j-k} = \sum_{k=0}^{L} b_k x_{j-k} \tag{2-7}$$

where the elements of $\{a_k\}$ and $\{b_k\}$ are real, and j ranges over the integers. Normally $b_0 = 1$, and the output can be written as

$$x_j = \sum_{k=0}^{K} a_k f_{j-k} - \sum_{k=1}^{L} b_k x_{j-k} \tag{2-8}$$

This relationship can be expressed in the z-transform domain as

$$X(z) = \frac{A(z)}{B(z)} F(z) = H(z)F(z) \tag{2-9}$$

where

$$A(z) = \sum_{j=0}^{K} a_j z^{-j} \tag{2-10a}$$

$$B(z) = \sum_{j=0}^{L} b_j z^{-j} \quad , \; b_0 = 1 \tag{2-10b}$$

and

$$H(z) = \frac{A(z)}{B(z)} \tag{2-11}$$

The condition for stability of this system is that the polynomial B(z) have all its roots inside the unit circle.

Several methods exist for inverting a z-transform expression to find the corresponding sample sequence. The most basic is in the form of a contour integral in the z-plane:

$$f_j = \frac{1}{2\pi i} \oint_{|z|=1} z^{j-1} F(z)\, dz \tag{2-12}$$

where $i = \sqrt{-1}$. The derivation of this result and an exhaustive compendium of other methods is found in Tou[27]. Later in this chapter, an original matrix method will be presented for inverting a z-transform of the form of equation (2-11).

B. Discrete Time Stochastic Processes

The present study is concerned entirely with discrete time stochastic processes, of which examples might be the sequence of equally spaced samples from a continuous noise waveform, or an economic or demographic time series. It will be assumed that the processes of interest are stationary, meaning that their statistical properties do not change with time. They will further be assumed to be ergodic, this allowing one to interchange time and ensemble averages. (For detailed treatments of the implications of stationarity and ergodicity, see Lee[29], Ragazzini and Franklin[28], Tou[27], and Parzen[30].)

An important concept in the treatment of any stochastic process x_k is that of the autocorrelation function, defined as

$$\phi_{xx}(j) = E\left[x_k \; x_{k+j}\right], \quad j = 0, \pm 1, \pm 2, \ldots . \quad (2\text{-}13)$$

where the operator $E[\cdot]$ denotes statistical expectation, and k may range over all the integers. The requirement of stationarity insures that the value of $\phi_{xx}(j)$ does not depend on a particular choice of the index k. Ergodicity implies that $\phi_{xx}(j)$ may also be defined as

$$\phi_{xx}(j) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} x_k \, x_{k+j} \qquad (2\text{-}14)$$

A related concept is that of the cross-correlation between the sequences x_j and y_j, defined as

$$\phi_{xy}(j) = E\left[x_k \, y_{k+j}\right] , \quad j = 0, \pm 1, \pm 2, \ldots\ldots \qquad (2\text{-}15)$$

It should be noted that $\phi_{xx}(j)$ is an even function of its argument, whereas $\phi_{xy}(j)$ generally is not. In fact, by inspection,

$$\phi_{xy}(-j) = \phi_{yx}(j) \qquad (2\text{-}16)$$

The sampled power spectrum of the process x_j is defined as the double-sided z-transform of the autocorrelation function for x_j:

$$S_{xx}(z) = \sum_{j=-\infty}^{\infty} \phi_{xx}(j) z^{-j} \qquad (2\text{-}17)$$

The same inversion formula presented as equation (2-12) can be used to gain the autocorrelation sequence from the sampled power spectrum:

$$\phi_{xx}(j) = \frac{1}{2\pi i} \oint_{|z|=1} S_{xx}(z) z^{j-1} dz, \ j = 0, \pm 1, \pm 2, \ldots. \qquad (2\text{-}18)$$

In particular, the mean square value of x_j is given by $\phi_{xx}(0)$:

$$E[x_j^2] = \phi_{xx}(0) = \frac{1}{2\pi i} \oint_{|z|=1} S_{xx}(z)\frac{dz}{z} \tag{2-19}$$

A frequently recurring operation is that of passing a discrete time random process through a digital system specified by the system function H(z):

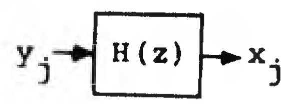


Figure 2-2. Digital filtering of the sequence y_j.

It can be shown (as in Tou[27] or Ragazzini and Franklin[28]) that the sampled power spectrum for x_j is given by

$$S_{xx}(z) = H(z)H(z^{-1})S_{yy}(z) \tag{2-20}$$

where $S_{yy}(z)$ is the sampled spectrum for the sequence y_j.

We are normally most interested in evaluating the sampled power spectrum on the z-plane unit circle where z can be expressed

$$z = e^{i\theta} \tag{2-21}$$

with θ measured in radians from the positive real axis. In dealing with a sampled data sequence, it is particularly convenient to determine the behavior of $S_{xx}(z)$ for $z = e^{i\omega T}$, where ω is a real radian frequency, and T is the sampling interval in seconds. On the unit circle, $S_{xx}(z)$ is always real, and as a function of ω will be even and periodic every $2\pi/T$ radians. Another property of some usefulness is that on the unit circle, equation (2-20) can be written as

$$S_{xx}(z) = |H(z)|^2 S_{yy}(z) \qquad (2-22)$$

where $z = e^{i\omega T}$.

C. Linear Regressive Series

One of the most widely studied classes of discrete random processes is that known as linear regressive series. These are generated by difference equations of the form

$$\begin{aligned} x_t = n_t + a_1 n_{t-1} + a_2 n_{t-2} + \cdots + a_K n_{t-K} \\ - b_1 x_{t-1} - b_2 x_{t-2} - \cdots - b_L x_{t-L} \end{aligned} \qquad (2-23)$$

or

$$\sum_{j=0}^{L} b_j x_{t-j} = \sum_{j=0}^{K} a_j n_{t-j} \quad , a_0 = b_0 = 1 \qquad (2-24)$$

where t ranges over the integers. Here, x_t is the process of interest, the $\{a_j\}$ and $\{b_j\}$ are real, and n_t is a sequence of zero-mean, uncorrelated, identically distributed random variables with common variance σ^2. No generality is lost by assuming that a_0 and b_0 are unity, since the effect of other values can be absorbed into the variance of n_t. In general, K or L or both can be infinite, but if both are finite, the process x_t is called a finite linear regression. The sequence n_t can be considered to be the driving function for the system of equation (2-24), and it can be shown that since n_t is zero mean, x_t must be also.

From previous considerations, it should be apparent that

x_t can be pictured as the output of the digital system

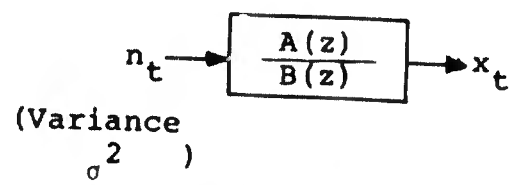


Figure 2-3. Digital system model for the generation of x_t.

By assumption,

$$\phi_{nn}(j) = \sigma^2, \; j=0$$
$$\phi_{nn}(j) = 0 \; , \; j \neq 0 \qquad (2\text{-}25)$$

which can be expressed more compactly using the δ-function defined in equation (2-2) as

$$\phi_{nn}(j) = \sigma^2 \delta_j \qquad (2\text{-}26)$$

Using equation (2-17), the sampled power spectrum for n_t becomes simply

$$S_{nn}(z) = \sigma^2 \qquad (2\text{-}27)$$

and thus by equation (2-20), the sampled power spectrum for x_t becomes

$$S_{xx}(z) = \sigma^2 \frac{A(z)A(z^{-1})}{B(z)B(z^{-1})} \qquad (2\text{-}28)$$

In the most general case, the polynomials A(z) and B(z) are each of first order or greater, and the process is said to be "of mixed type". Two important special cases emerge immediately. If all the b_j, save b_0, vanish, then

$$x_t = \sum_{j=0}^{K} a_j n_{t-j} \ , \ a_0 = 1 \tag{2-29}$$

and the process is known as a moving average with

$$S_{xx}(z) = \sigma^2 A(z) A(z^{-1}) \tag{2-30}$$

If all the a_j are zero, except for a_0, such that

$$x_t = n_t - \sum_{j=1}^{L} b_j x_{t-j} \ , \tag{2-31}$$

the process is called an autoregression with

$$S_{xx}(z) = \frac{\sigma^2}{B(z)B(z^{-1})} \tag{2-32}$$

Both moving averages and autoregressions can be of finite or infinite order. For convenience, the $\{a_j\}$ and $\{b_j\}$ will be referred to as the moving average and autoregressive coefficients respectively.

In the systems model for each of these process types, the driving function is the uncorrelated random sequence n_t, which has been defined to be stationary and ergodic. The stationarity of the processes themselves depends upon the coefficient sets $\{a_j\}$ and $\{b_j\}$ and is closely bound up with the stability of the associated digital system. Wold[19] treats this question in great detail and provides the following criteria for the stationarity of the three main process types:

1.) Every moving average of finite order is stable and hence stationary if all the coefficients a_j are finite. An infinite order moving average will be stationary if the sum $\sum_{j=0}^{\infty} a_j^2$ is convergent.

2.) An autoregression of finite order will be stationary if all the zeros of the polynomial B(z) are found within the unit circle of the z-plane. This criterion can also be applied to infinite order autoregressions if a closed form expression for B(z) can be found such that the zeros can be determined. (See Whittle[20].)

3.) In an obvious extension, a process of mixed type will be stationary if Σa_j^2 is convergent and if B(z) has all of its zeros within the unit circle.

D. Wold's Decomposition

A theoretical result of exceptional importance, now known as Wold's Decomposition, is set forth by Wold[19]. Perhaps the most concise exposition, however, is found in Whittle[20] along with a discussion of its physical and mathematical implications. Wold's theorem states, essentially, that any stationary discrete parameter process with finite variance can be represented as the sum of two mutually uncorrelated processes:

$$x_t = y_t + z_t \tag{2-33}$$

where y_t is deterministic, and z_t can be expressed as a one-sided moving average over a stationary uncorrelated sequence n_t:

$$z_t = \sum_{j=0}^{\infty} d_j n_{t-j} \tag{2-34}$$

In all of our considerations, the deterministic part will be zero, and thus each process can be represented as a moving average of generally infinite order, although it may have been initially defined as an autoregression or a mixed-type process.

The converse result is also treated by Wold and Whittle, and it is shown that every stationary moving average has an autoregressive representation of generally infinite order. This, in turn, can be extended to show that every well-behaved process of mixed type can be written in an autoregressive representation that converges in mean square to the original process.

For processes of finite order, both of these results are easily demonstrable. A representation of the form

$$A(z) = \sum_{j=0}^{K} a_j z^{-j} \tag{2-35}$$

can be interpreted as the Laurent expansion about the origin for the function A(z). This expansion contains only negative powers of z and will be valid on the entire unit circle. If A(z) has no zeros on the unit circle, then the reciprocal function

$$D(z) = \frac{1}{A(z)} \tag{2-36}$$

will also possess a Laurent expansion about the origin which includes the unit circle in its zone of convergence. We know that this expansion is unique. The ratio $1/A(z)$ can be expanded by long division in an infinite series in negative powers of z. Thus there exists a unique polynomial

$$D(z) = \sum_{j=0}^{\infty} d_j z^{-j} \tag{2-37}$$

which satisfies equation (2-36) on the unit circle.

E. Correlation Properties of Linear Regressive Series

Much of the analysis in following chapters will be based on the autocorrelation properties of the three canonical processes defined above, and this section will present a development of the correlation relationships that will be needed later. The process of mixed type will be treated first, and the moving average and autoregression handled as special cases.

The basic equation defining a process of mixed type is given as

$$\sum_{j=0}^{L} b_j x_{t-j} = \sum_{j=0}^{K} a_j n_{t-j}, \quad a_0 = b_0 = 1 \tag{2-38}$$

As a first step in obtaining $\phi_{xx}(j)$ as defined in equation (2-13) both sides of this equation are multiplied by x_{t-k} and the ensemble average taken:

$$E\left[\sum_{j=0}^{L} b_j x_{t-j} x_{t-k}\right] = E\left[\sum_{j=0}^{K} a_j n_{t-j} x_{t-k}\right]$$

or (2-39)

$$\sum_{j=0}^{L} b_j \, E\left[x_{t-j}x_{t-k}\right] = \sum_{j=0}^{K} a_j \, E\left[n_{t-j}x_{t-k}\right]$$

Using the evenness of the autocorrelation function and equations (2-13) and (2-15) this becomes

$$\sum_{j=0}^{L} b_j \, \phi_{xx}(k-j) = \sum_{j=0}^{K} a_j \, \phi_{nx}(j-k) \text{ , all integer } k \tag{2-40}$$

If the $\{b_j\}$ and the right hand side are known, the possibility exists of generating a soluble system for a portion of the autocorrelation sequence by varying the index k. In particular, the system of M+1 equations obtained with

$$k = 0,1,2,3,\ldots,L,\ldots,M \quad , \; M \geq L \tag{2-41}$$

will involve only M+1 different values of $\phi_{xx}(j)$. Thus, providing that the right hand side can be evaluated, the system

$$\sum_{j=0}^{L} b_j \, \phi_{xx}(k-j) = \sum_{j=0}^{K} a_j \, \phi_{nx}(j-k) \text{ , } k=0,1,2,\ldots,M \tag{2-42}$$

can be solved for the first M+1 values of $\phi_{xx}(j)$.

To evaluate the right hand side of equation (2-42), it is necessary to deal with the cross-correlation sequence $\phi_{nx}(j)$. Equation (2-38) is multiplied by n_{t-k} and again expectations are taken. The result is

$$\sum_{j=0}^{L} b_j \, \phi_{nx}(k-j) = \sum_{j=0}^{K} a_j \, \phi_{nn}(k-j) \text{ , all integer } k \tag{2-43}$$

Since n_t is an uncorrelated sequence,

$$\phi_{nn}(j) = \sigma^2 \delta_j \tag{2-44}$$

and hence

$$\sum_{j=0}^{L} b_j \phi_{nx}(k-j) = \sigma^2 a_k \quad , \; k=0,1,2,\ldots,K$$

$$\sum_{j=0}^{L} b_j \phi_{nx}(k-j) = 0 \quad , \; k>K \tag{2-45}$$

It is now instructive to digress and consider the unit sample response of the system $H(z) = A(z)/B(z)$. Following equation (2-3), the system output can be expressed as the convolution sum

$$x_j = \sum_{k=0}^{\infty} h_k n_{j-k} \tag{2-46}$$

and multiplying both sides of equation (2-46) by n_{j-i} and averaging yields

$$\phi_{nx}(i) = \sum_{j=0}^{\infty} h_j \phi_{nn}(i-j) \tag{2-47}$$

But using equation (2-44),

$$\phi_{nx}(i) = \sigma^2 h_i \quad , \text{ all integer } i \tag{2-48}$$

Thus, if the system is driven by white noise, the cross-correlation sequence between the input and output is identical with the unit sample response of the system, scaled by the noise variance σ^2. If the system is realizable, therefore, the input-output cross-correlation sequence will be zero for arguments less than zero. Heuristically, this is readily seen by noting that $\phi_{nx}(j)$ for negative j is a measure of the correlation between the present output value and an input value some time in the future. If the system is realizable, and successive input values uncorrelated, this must vanish.

Returning to the system of equation (2-45), the fact that A(z)/B(z) represents a realizable system and hence that $\phi_{nx}(j)$ is zero for negative j greatly simplifies the equations. Beginning with k = 0 and writing out successive equations, the system becomes

$$
\begin{array}{lllll}
b_o\phi_{nx}(0) & & & & = \sigma^2 a_o \\
b_o\phi_{nx}(1) & + b_1\phi_{nx}(0) & & & = \sigma^2 a_1 \\
\vdots & \vdots & & & \vdots \\
b_o\phi_{nx}(K) & + b_1\phi_{nx}(K-1) & + \cdots & + b_K\phi_{nx}(0) & = \sigma^2 a_K \\
b_o\phi_{nx}(K+1) & + b_1\phi_{nx}(K) & + \cdots & + b_{K+1}\phi_{nx}(0) & = 0 \\
\vdots & \vdots & & \vdots & \vdots \\
b_o\phi_{nx}(L) & + b_1\phi_{nx}(L-1) & + \cdots & + b_L\phi_{nx}(0) & = 0 \\
\vdots & \vdots & & \vdots & \vdots \\
b_o\phi_{nx}(M) & + b_1\phi_{nx}(M-1) & + \cdots & + b_L\phi_{nx}(M-L) & = 0
\end{array}
\qquad (2\text{-}49)
$$

Since the $\{a_j\}$ and $\{b_j\}$ are known, it is apparent that the system can be solved recursively for the sequence $\phi_{nx}(j)$.

Since it is possible to compute successive values of the sequence $\phi_{nx}(j)$ out to any argument, it is clear that the system of equation (2-42) can be solved by evaluating the right hand side for as many equations as desired. Since $\phi_{nx}(j)$ is zero for all negative j, however, the right hand sum vanishes whenever k>K. Thus, the system can be written

$$\sum_{j=0}^{L} b_j \phi_{xx}(k-j) = \sum_{j=0}^{K} a_j \phi_{nx}(j-k) \quad , \; k = 0,1,2,...K$$

$$\sum_{j=0}^{L} b_j \phi_{xx}(k-j) = 0 \quad , \; k>K \tag{2-50}$$

It is now convenient to use matrix notation. Since it is necessary to have at least L+1 equations of the form of equation (2-50) to find a solution, the dimension of the pertinent matrices must be L+1 or larger. If M is the maximum argument for which the values of $\phi_{nx}(j)$ and $\phi_{xx}(j)$ are desired, and $M \geq L$, then the dimensionality of the entire system will be M+1. Defining

$$\phi_{nx} = \begin{bmatrix} \phi_{nx}(0) \\ \phi_{nx}(1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \phi_{nx}(M) \end{bmatrix} \quad , \quad a = \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_K \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{2-51a}$$

and

$$B = \begin{bmatrix} b_0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ b_1 & b_o & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ b_2 & b_1 & b_0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & & & & & & & & & \cdot \\ \cdot & \cdot & \cdot & & & & & & & & & \cdot \\ \cdot & \cdot & \cdot & & & & & & & & & \cdot \\ \cdot & \cdot & \cdot & & & & & & & & & \cdot \\ b_L & b_{L-1} & \cdot & \cdot & \cdot & \cdot & \cdot & b_0 & & 0 & \cdots & 0 \\ 0 & b_L & b_{L-1} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & b_0 & \cdots 0 \cdots & 0 \\ & & & & & & & & & & & 0 \\ 0 \cdots\cdots 0 & b_L & b_{L-1} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & b_1 & b_0 \end{bmatrix} \qquad (2\text{-}51b)$$

where ϕ_{nx} and a are of dimension M+1 and B is a square matrix of order M+1, the system of equation (2-45) can be written as

$$B\,\phi_{nx} = \sigma^2 a \qquad \text{or} \qquad \phi_{nx} = \sigma^2 B^{-1} a \qquad (2\text{-}52)$$

Now letting c be a column vector of dimension M+1 representing the right hand side of equation (2-50) and defining

$$\phi_{xx} = \begin{bmatrix} \phi_{xx}(0) \\ \phi_{xx}(1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \phi_{xx}(M) \end{bmatrix} \qquad (2\text{-}53a)$$

and

$$
\beta = \begin{bmatrix}
b_0 & b_1 & b_2 & \cdot & \cdot & \cdot & \cdot & \cdot & b_L & 0 & \cdot & \cdot & \cdot & 0 \\
b_1 & (b_0+b_2) & b_3 & b_4 & \cdot & \cdot & b_L & 0 & 0 & \cdot & \cdot & \cdot & & 0 \\
b_2 & (b_1+b_3) & (b_0+b_4) & b_5 & \cdots & & & & & & & & & \vdots \\
\vdots & \vdots & \vdots & \vdots & & & & & & & & & & \vdots \\
b_L & b_{L-1} & \cdots & \cdots & b_1 & b_0 & & & & & & & & \vdots \\
0 & b_L & b_{L-1} & \cdots & \cdots & b_1 & b_0 & & & & & & & \vdots \\
\vdots & & & & & & b_1 & b_0 & & & & & & 0 \\
0 & 0 & \cdots & b_L & b_{L-1} & \cdots & \cdots & b_1 & \cdots & & & & & b_o
\end{bmatrix} \qquad (2\text{-}53b)
$$

equation (2-44) can be written

$$\beta\phi_{xx} = c \qquad \text{or} \qquad \phi_{xx} = \beta^{-1}c \qquad (2\text{-}54)$$

Since the form of the matrix β is rather complicated, the following rule is offered for constructing its elements <u>when the rows and columns are indexed from 0 to M</u>:

a.) The elements of the zeroth column are given by

$$
\begin{aligned}
\beta_{i0} &= b_i \quad , \; i = 0,1,\ldots\ldots,L \\
\beta_{i0} &= 0 \quad , \; i = L+1,\; L+2,\ldots,M
\end{aligned} \qquad (2\text{-}55)
$$

b.) All other elements of the matrix are given by

$$\beta_{ij} = b_{i-j} + b_{i+j} \quad , \quad j \neq 0 \qquad (2\text{-}56)$$

where b_j is understood to be zero for $j<0$ or $j>L$. Now constructing a square matrix A of order M+1,

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \cdots & a_K & 0 & 0 & \cdots & 0 \\ a_1 & a_2 & a_3 & \cdots & a_K & 0 & 0 & \cdots & & 0 \\ \vdots & \vdots & & & & & & & & \vdots \\ a_{K-1} & a_K & 0 & 0 & \cdots & & & & & 0 \\ a_K & 0 & 0 & \cdots & & & & & & 0 \\ 0 & 0 & \cdots & & & & & & & 0 \\ \vdots & \vdots & & & & & & & & \vdots \\ 0 & 0 & \cdots & & & & & & & 0 \end{bmatrix} \qquad (2\text{-}57)$$

the column of right hand sides of the system (2-50) is expressed

$$c = A\phi_{nx} \qquad (2\text{-}58)$$

Now combining equations (2-52), (2-54), and (2-58), the final result is

$$\phi_{xx} = \sigma^2 \beta^{-1} A B^{-1} a \qquad (2\text{-}59)$$

and can be solved for any dimension greater than the order of the autoregressive part of the process. It should be noted that since the right hand side of equation (2-50) vanishes for $k>K$, the system becomes a recursion once the first L+1 values of $\phi_{xx}(j)$ are known. Thus, the matrix method need only be used to find these L+1 values, and thereafter,

$$\phi_{xx}(k) = -\sum_{j=1}^{L} b_j \phi_{xx}(k-j), \quad k>L \tag{2-60}$$

In a moving average, where all of the autoregressive coefficients save b_0 are zero, equation (2-45) shows directly that

$$\begin{aligned} \phi_{nx}(j) &= \sigma^2 a_j \quad , j \geq 0 \\ \phi_{nx}(j) &= 0 \quad\quad , j<0 \end{aligned} \tag{2-61}$$

The autocorrelation can be written as

$$\phi_{xx}(j) = \sigma^2 \sum_{k=0}^{K-|j|} a_k a_{k+|j|} = \sigma^2 \sum_{k=|j|}^{K} a_{k-|j|} a_k \tag{2-62}$$

from which it follows that for a K^{th} order moving average

$$\phi_{xx}(j) = 0 \ , \ |j|>K \tag{2-63}$$

In a matrix equation, the moving average autocorrelation becomes

$$\phi_{xx} = \sigma^2 Aa \tag{2-64}$$

which is found from equation (2-59) by noting that for a moving average, β and B are identity matrices.

For a pure autoregression, equation (2-45) becomes

$$\sum_{j=0}^{L} b_j \phi_{nx}(k-j) = \sigma^2 \delta_k \ , \ k=0,1,2,...M \tag{2-65}$$

and since the matrix a is

$$a = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} , \tag{2-66}$$

the matrix form is still

$$\phi_{nx} = \sigma^2 B^{-1} a \tag{2-67}$$

The equations for the autocorrelation become

$$\sum_{j=0}^{L} b_j \, \phi_{xx}(k-j) = \sigma^2 \delta_k, \; k = 0,1,2,\ldots M \tag{2-68}$$

and in matrix form, the solution is

$$\phi_{xx} = \sigma^2 \beta^{-1} a \tag{2-69}$$

where again a is given by equation (2-66).

The matrix method derived above for determining the auto- and cross-correlation sequences of a regressive process, given the regression coefficients, is particularly well suited to digital computation providing that matrix manipulation subroutines are available. In addition, by setting $\sigma^2=1$, equation (2-52) can be used to determine the sample sequence corresponding to any digital system function or z-transform expression of the form A(z)/B(z), because of the identity of equation (2-48).

F. Sampling Properties of the Mean Lagged Products of Regressive Series

All of the estimation schemes to be treated here involve the set of mean lagged products computed from a sample sequence of the process as an estimate of its correlation function. It is thus necessary to know certain statistical properties of the mean lagged products and their relationship to the statistical correlation function. A short study of these elementary properties is presented here.

For present purposes, the mean lagged product over N samples for the process x_t and lag j will be defined as

$$\hat{\phi}_{xx}(j) = \frac{1}{N} \sum_{k=1}^{N} x_k x_{k+j} \qquad (2\text{-}70)^{1}$$

The expected value of $\hat{\phi}_{xx}(j)$ is the theoretical correlation $\phi_{xx}(j)$:

$$E[\hat{\phi}_{xx}(j)] = \frac{1}{N} \sum_{k=1}^{N} E[x_k x_{k+j}] = \frac{1}{N} \sum_{k=1}^{N} \phi_{xx}(j) \qquad (2\text{-}71)$$

$$= \phi_{xx}(j)$$

Thus, $\hat{\phi}_{xx}(j)$ is an unbiased estimate of $\phi_{xx}(j)$.

The covariance between $\hat{\phi}_{xx}(i)$ and $\hat{\phi}_{xx}(j)$ is

$$\text{Cov}[\hat{\phi}_{xx}(i), \hat{\phi}_{xx}(j)] = E[\hat{\phi}_{xx}(i)\hat{\phi}_{xx}(j)] - \phi_{xx}(i)\phi_{xx}(j) \qquad (2\text{-}72)$$

From equation (2-70),

$$\hat{\phi}_{xx}(i)\ \hat{\phi}_{xx}(j) = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} x_k x_{k+i} x_\ell x_{\ell+j} \qquad (2\text{-}73)$$

---

1. In the present work, a symbol bearing a carat denotes an estimate of the quantity represented by the symbol itself. Here, for example, the mean lagged product $\hat{\phi}_{xx}(j)$ is an estimate of the theoretical correlation $\phi_{xx}(j)$.

and thus

$$E[\hat{\phi}_{xx}(i)\hat{\phi}_{xx}(j)] = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N} E[x_k x_{k+i} x_\ell x_{\ell+j}] \qquad (2\text{-}74)$$

To proceed further with the 4th order expected value in the summation, it will be assumed that x_t is Gaussian. A brief treatment of the non-Gaussian case is included in Hannan[1], where it is seen that the results approach those derived here as $N\to\infty$. If x_t is Gaussian, then

$$E[\hat{\phi}_{xx}(i)\hat{\phi}_{xx}(j)] = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\begin{bmatrix}\phi_{xx}(i)\phi_{xx}(j)\\ +\phi_{xx}(k-\ell-j)\phi_{xx}(k-\ell+i)\\ +\phi_{xx}(k-\ell)\phi_{xx}(k-\ell+i-j)\end{bmatrix} \qquad (2\text{-}75)$$

and from equation (2-72),

$$\mathrm{Cov}[\hat{\phi}_{xx}(i),\hat{\phi}_{xx}(j)] = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\begin{bmatrix}\phi_{xx}(k-\ell-j)\phi_{xx}(k-\ell+i)\\ +\phi_{xx}(k-\ell)\phi_{xx}(k-\ell+i-j)\end{bmatrix} \qquad (2\text{-}76)$$

Since k and ℓ always appear together as the difference $(k-\ell)$, this is somewhat more conveniently expressed as

$$\mathrm{Cov}[\hat{\phi}_{xx}(i),\hat{\phi}_{xx}(j)] = \frac{1}{N^2}\sum_{k=1-N}^{N-1}(N-|k|)\begin{bmatrix}\phi_{xx}(k-j)\phi_{xx}(k+i)\\ +\phi_{xx}(k)\phi_{xx}(k+i-j)\end{bmatrix} \qquad (2\text{-}77)$$

An expression for the variance of $\hat{\phi}_{xx}(j)$ follows directly from equation (2-77) by setting $i = j$:

$$\mathrm{Var}[\hat{\phi}_{xx}(j)] = \frac{1}{N^2} \sum_{k=1-N}^{N-1} (N-|k|) \left[\phi_{xx}^{2}(k) + \phi_{xx}(k+j)\,\phi_{xx}(k-j) \right] \qquad (2\text{-}78)$$

In order for the $\hat{\phi}_{xx}(j)$ to be a consistent estimate of the $\phi_{xx}(j)$, $\mathrm{Var}[\hat{\phi}_{xx}(j)]$ must approach 0 as $N \to \infty$. Since

$$\mathrm{Var}[\hat{\phi}_{xx}(j)] < \frac{1}{N} \sum_{k=1-N}^{N-1} [\phi_{xx}^{2}(k) + \phi_{xx}(k+j)\,\phi_{xx}(k-j)] \qquad (2\text{-}79)$$

it is apparent that if

$$\lim_{N \to \infty} \left[\frac{1}{N} \sum_{k=1-N}^{N-1} \phi_{xx}^{2}(k) \right] = 0 \qquad (2\text{-}80)$$

the mean lagged products will be consistent estimates of the theoretical correlation function. Hannan[1] demonstrates that this will be true for any process that can be represented as an infinite order moving average in the manner of Wold's Decomposition. He further shows that the mean lagged products are consistent estimates of the $\phi_{xx}(j)$ even when x_t is not Gaussian.

A great deal has been written on the asymptotic normality of the probability distributions of the mean lagged products as the sample size N increases without limit. Among other workers, Walker[31] shows that this property holds for the serial correlation coefficients of a finite order autoregression, and states that his result can be extended to the case of the infinite order moving average

$$x_t = \sum_{j=0}^{\infty} a_j n_{t-j} \qquad (2\text{-}81)$$

so long as $\sum_{j=0}^{\infty} a_j$ and $\sum_{j=0}^{\infty} ja_j$ are absolutely convergent. A lengthy discussion of asymptotic normality for the mean lagged products is found in Hannan[1], and an interesting comparison of the actual and approximating normal distributions is found in Anderson[32]. On this basis, the mean lagged products will be assumed to be normally distributed for large sample size in the following sections.

CHAPTER III

ALL-POLE ESTIMATION FOR PROCESSES OF MIXED TYPE

In this section, the spectral analysis of mixed-type processes is treated in terms of autoregressive models from a new point of view. The effects of approximation and statistical error are studied in detail, and a series of computational examples is presented.

A. All-pole Approximation by Spectral Whitening

As defined in the preceding chapter, a process of mixed type is a linear regressive series which might be described as a moving average and an autoregression in combination. The basic difference equation is given by

$$\sum_{j=0}^{L} b_j x_{t-j} = \sum_{j=0}^{K} a_j n_{t-j}\ , \quad a_0 = b_0 = 1 \qquad (3\text{-}1)$$

where t ranges over the integers. As before, n_t represents a sequence of uncorrelated random numbers with zero mean and variance σ^2. It was shown above that associated with this process is a transform domain system function

$$H(z) = \frac{A(z)}{B(z)} \qquad (3\text{-}2)$$

where

$$A(z) = \sum_{j=0}^{K} a_j z^{-j} \quad , \; a_0 = 1 \qquad (3\text{-}3)$$

and

$$B(z) = \sum_{j=0}^{L} b_j z^{-j} \quad , \; b_0 = 1 \qquad (3\text{-}4)$$

For stability, it will be assumed that all of the zeros of B(z) lie within the unit circle, and for reasons that will become apparent later, this same requirement will be imposed on A(z). The moving average and autoregression are special cases of the process defined by equation (3-1).

It was noted above that the sampled power spectrum for the mixed-type process is given by

$$S_{xx}(z) = \sigma^2 \frac{A(z)A(z^{-1})}{B(z)B(z^{-1})} \qquad (3\text{-}5)$$

The central problem that is treated in this study is that of devising and comparing methods for operating on a sample of the process x_t to derive an estimate of its sampled power spectrum in terms of a convenient model. One of the earliest approaches to this problem, treated by Steiglitz[18], Whittle[20], and Hannan[1], is the formation of so-called <u>all-pole</u> estimates, which model the process as an autoregression of sufficiently high order to represent the spectrum with acceptable accuracy. In the past, the estimation equations have been derived either on the basis of a least squares minimization or by assuming that the process is Gaussian and proceeding by maximum likelihood arguments. If the order of the autoregressive model is designated M, the result of these techniques is an estimate of the M autoregressive coefficients $\{d_j\}$ and the input noise variance σ^2 which "best" fit the data. The resulting spectral estimate

$$S_{xx}(z) = \sigma^2 \frac{1}{D(z)D(z^{-1})} \tag{3-6}$$

with

$$D(z) = \sum_{j=0}^{M} d_j z^{-j} \quad , \; d_0 = 1 \tag{3-7}$$

has no zeros; hence the designation all-pole.

A disadvantage of the maximum likelihood approach is that it strictly applies only to Gaussian processes, although Tretter[33] and Whittle[20] have provided justifications for more general applicability. A linear least squares derivation of the all-pole estimation equations is presented in Appendix B, but here an alternative approach will be followed that is independent of the probabilistic structure of the processes x_t and n_t and leads, somewhat more intuitively, to the same estimation equations.

Essentially, the procedure to be described here seeks to find a suitably optimum moving average filter which will convert the process x_t into white noise of unit variance. Such a system is portrayed here:

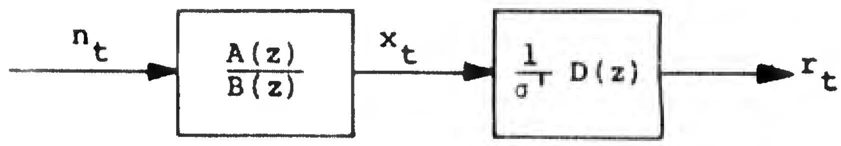


Figure 3-1. The whitening and normalizing of x_t by a moving average filter.

where D(z) is of the form of equation (3-7). The sampled power spectrum for r_t is given straight-forwardly as

$$S_{rr}(z) = \frac{\sigma^2}{\sigma'^2} \frac{A(z)A(z^{-1})}{B(z)B(z^{-1})} D(z)D(z^{-1}) \tag{3-8}$$

If the coefficients of D(z) and the constant σ' are chosen such that $S_{rr}(z) \approx 1$, then

$$S'_{xx}(z) = \frac{\sigma'^2}{D(z)D(z^{-1})} \approx \sigma^2 \frac{A(z)A(z^{-1})}{B(z)B(z^{-1})} = S_{xx}(z) \tag{3-9}$$

and the digital system 1/D(z), driven by an uncorrelated sequence of variance σ'^2, constitutes the desired finite order autoregressive model for the process.

Because of the understanding it provides of the statistical estimation problem, the following "deterministic whitening problem" will be studied in detail:

> Given a process of mixed type characterized by the system function A(z)/B(z) and input noise variance σ^2, how does one best choose the coefficients of an M^{th} order moving average filter and the gain constant σ' such that the original process is whitened and normalized to unit variance?

This preliminary study of the deterministic situation has the advantage of separating the roles of the two types of error that arise in all-pole spectral estimations: that due to the finite order of the approximation, and that which arises from the statistical spread of the data.

The first step in the solution of the problem is that of determining a suitable criterion of optimality. Since the aim of the exercise is to find a system that will "undo" the effect of the filter $A(z)/B(z)$ and scale the output to unit variance, intuition suggests that the whitening filter function must be near to the inverse $\frac{1}{\sigma}B(z)/A(z)$ and should yield the output sequence n_t/σ when driven by x_t. Since it is hoped that ultimately

$$r_t \approx \frac{n_t}{\sigma} \qquad (3\text{-}10)$$

the following approximation error will be defined at time t:

$$\varepsilon_t \equiv r_t - \frac{n_t}{\sigma} \qquad (3\text{-}11)$$

A very natural criterion of optimality, and the one that will be used here, is that the "best" system will be that which minimizes the mean square value of the error signal ε_t. Thus, the optimum system parameters must obey the equations

$$\frac{\partial \overline{\varepsilon_t{}^2}}{\partial d_j} = 0 \quad , \; j = 1,2,\ldots,M$$

$$\frac{\partial \overline{\varepsilon_t{}^2}}{\partial \sigma'} = 0. \qquad (3\text{-}12)$$

It should be recalled that by assumption, $d_0 = 1$.

It is most convenient to work with the moving average representation of the process whose existence is guaranteed by Wold's decomposition. If the infinite order moving average system function is denoted H(z), then

$$H(z) = \frac{A(z)}{B(z)} = \sum_{j=0}^{\infty} h_j z^{-j} \qquad (3\text{-}13)$$

The weighting sequence $\{h_j\}$ can be considered to be the unit sample response for the system A(z)/B(z) and can be found using the method developed in the last chapter. This leads to the system

$$\sum_{j=0}^{L} b_j h_{n-j} = a_n \quad , \ n \leq K$$

$$\sum_{j=0}^{L} b_j h_{n-j} = 0 \quad , \ n > K \qquad (3\text{-}14)$$

$$h_j = 0 \quad , \ j < 0$$

By an extension of the previously derived correlation properties, the autocorrelation sequence for x_t can now be written in the form

$$\phi_{xx}(j) = \sigma^2 \sum_{k=0}^{\infty} h_k h_{k+j} = \sigma^2 \sum_{k=j}^{\infty} h_k h_{k-j} \ , \ j \geq 0 \qquad (3\text{-}15)$$

The error signal ϵ_t can be generated by the following digital system:

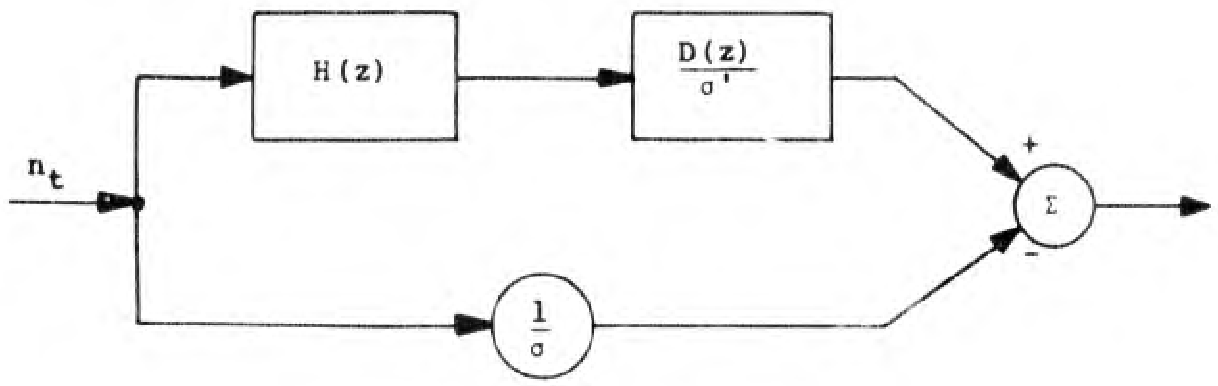


Figure 3-2. Digital system representation of the generation of ϵ_t.

whose transfer function is

$$T(z) = \frac{H(z)D(z)}{\sigma'} - \frac{1}{\sigma} \tag{3-16}$$

Now defining

$$C(z) \equiv H(z)D(z) , \tag{3-17}$$

the system transfer function of equation (3-16) can be written

$$T(z) = \frac{C(z)}{\sigma'} - \frac{1}{\sigma} = \frac{1}{\sigma'} \sum_{j=0}^{\infty} c_j z^{-j} - \frac{1}{\sigma} \tag{3-18}$$

or

$$T(z) = \left(\frac{1}{\sigma'} - \frac{1}{\sigma}\right) + \frac{1}{\sigma'} \sum_{j=1}^{\infty} c_j z^{-j} \tag{3-19}$$

since $c_0 = 1$. Therefore,

$$\varepsilon_t^2 = \phi_{\varepsilon\varepsilon}(0) = \sigma^2\left[\left(\frac{1}{\sigma'} - \frac{1}{\sigma}\right)^2 + \frac{1}{\sigma'^2}\sum_{j=1}^{\infty} c_j^2\right]$$

$$= \frac{\sigma^2}{\sigma'^2}\sum_{j=0}^{\infty} c_j^2 + 1 - \frac{2\sigma}{\sigma'} \tag{3-20}$$

This is the function that will be minimized to find the optimum whitening filter. To find the desired set $\{d_j\}$, this expression is used in the first group of equation (3-12). The result is the system

$$\sum_{j=1}^{\infty} c_j \frac{\partial c_j}{\partial d_k} = 0 \quad , \; k = 1,2,\ldots,M \tag{3-21}$$

where $\{c_j\}$ is the convolution of $\{h_j\}$ and $\{d_j\}$ given by

$$c_j = 0, \quad j<0$$

$$c_j = \sum_{k=0}^{j} d_k h_{j-k} \quad , \; j = 0,1,2,\ldots,M \tag{3-22}$$

$$c_j = \sum_{k=0}^{M} d_k h_{j-k} \quad , \; j>M$$

The required derivatives can now be written as follows:

For $j = 0,1,2,\ldots,M$

$$\frac{\partial c_j}{\partial d_k} = h_{j-k} \quad , \; k \leq j$$

$$= 0 \quad , \; k>j \tag{3-23a}$$

For $j>M$,

$$\frac{\partial c_j}{\partial d_k} = h_{j-k} \;, \quad k \leq M$$
$$= 0 \;, \quad k>M \qquad (3\text{-}23b)$$

Now equation (3-21) is written as

$$\sum_{j=1}^{K} c_j \frac{\partial c_j}{\partial d_k} + \sum_{j=M+1}^{\infty} c_j \frac{\partial c_j}{\partial d_k} = 0, \; k = 1,2,\ldots,M \qquad (3\text{-}24)$$

and using equations (3-23a) and (3-23b) yields

$$\sum_{j=k}^{\infty} c_j h_{j-k} = 0 \;, \; k = 1,2,\ldots,M \qquad (3\text{-}25)$$

Now if the $\{c_j\}$ are evaluated using the expressions of equation (3-22), the result is

$$\sum_{j=k}^{M} \sum_{m=0}^{j} d_m h_{j-k} h_{j-m} + \sum_{j=M+1}^{\infty} \sum_{m=0}^{M} d_m h_{j-m} h_{j-k} = 0 \;, \qquad (3\text{-}26)$$
$$k = 1,2,\ldots,M$$

Reversing the order of summation gives

$$\sum_{m=0}^{k} \sum_{j=0}^{\infty} d_m h_j h_{j-m+k} + \sum_{m=k+1}^{M} \sum_{j=0}^{\infty} d_m h_j h_{j-k+m} = 0 \;, \qquad (3\text{-}27)$$
$$k = 1,2,\ldots,M$$

and using equation (3-15), this becomes

$$\sum_{j=0}^{k} d_j \; \phi_{xx}(k-j) + \sum_{j=k+1}^{M} d_j \; \phi_{xx}(j-k) = 0 \;, \qquad (3\text{-}28)$$
$$k = 1,2,\ldots,M$$

or finally

$$\sum_{j=0}^{M} d_j \, \phi_{xx}(k-j) = 0 \; , \; k = 1,2,\ldots,M \; ; \; d_0 = 1 \qquad (3\text{-}29)$$

which is a system of M simultaneous linear equations in M unknowns. Defining the column vectors

$$d = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_M \end{bmatrix} \quad \text{and} \quad \phi_{xx} = \begin{bmatrix} \phi_{xx}(1) \\ \phi_{xx}(2) \\ \cdot \\ \cdot \\ \cdot \\ \phi_{xx}(M) \end{bmatrix} \qquad (3\text{-}30)$$

and the process correlation matrix Φ_{xx},

$$\Phi_{xx} = \begin{bmatrix} \phi_{xx}(0) & \phi_{xx}(1) & \phi_{xx}(2) & \cdots & \phi_{xx}(M-1) \\ \phi_{xx}(1) & \phi_{xx}(0) & \phi_{xx}(1) & \cdots & \phi_{xx}(M-2) \\ \phi_{xx}(2) & \phi_{xx}(1) & \phi_{xx}(0) & \cdots & \phi_{xx}(M-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \cdots \; \phi_{xx}(0) & \phi_{xx}(1) \\ \phi_{xx}(M-1) & \phi_{xx}(M-2) & \cdots & \phi_{xx}(1) & \phi_{xx}(0) \end{bmatrix} \qquad (3\text{-}31)$$

the system of equation (3-29) can be expressed as

$$\Phi_{xx} d = - \phi_{xx} \tag{3-32}$$

If the matrix Φ_{xx} is non-singular, the unknown coefficients $\{d_j\}$ can now be written as

$$d = - \Phi_{xx}^{-1} \phi_{xx} \tag{3-33}$$

For the processes considered here, the correlation matrix[1] Φ_{xx} is positive-definite (See Whittle[20]), and hence equation (3-33) will always have a unique solution. Furthermore, it has been remarked (by Whittle) that the roots of the corresponding polynomial D(z) lie within the z-plane unit circle.

Again using the expression for $\phi_{\epsilon\epsilon}(0)$ given by equation (3-20), the optimum constant σ' is found to be

$$\sigma' = \sigma \sum_{j=0}^{\infty} c_j^2 \tag{3-34}$$

This can be worked into a more convenient form by using the expression for c_j in terms of the $\{h_j\}$ and $\{d_j\}$. One first obtains

$$\sigma' = \sigma \sum_{j=0}^{M} \sum_{k=0}^{j} \sum_{m=0}^{j} d_k d_m h_{j-k} h_{j-m} + \sigma \sum_{j=M+1}^{\infty} \sum_{k=0}^{M} \sum_{m=0}^{M} d_k d_m h_{j-k} h_{j-m} \tag{3-35}$$

1. In the literature, this form is most often denoted the covariance matrix. Since x_t is a zero mean process, however, the covariance and correlation functions are identical. Both designations will be used here.

in which it is necessary to reverse the order of summation so that the autocorrelation of equation (3-15) can be recognized. Ultimately, equation (3-35) becomes

$$\sigma' = \frac{1}{\sigma} \sum_{j=0}^{M} d_j \sum_{k=0}^{M} d_k \phi_{xx}(j-k) \tag{3-36}$$

But by equation (3-29), the inner summation vanishes except when $j = 0$. Thus, finally

$$\sigma' = \frac{1}{\sigma} \sum_{j=0}^{M} d_k \; \phi_{xx}(k) \tag{3-37}$$

Using the matrix forms defined above, σ' can also be expressed as

$$\sigma' = \frac{1}{\sigma} \left[\phi_{xx}(0) + d^T \phi_{xx} \right] \tag{3-38}$$

where d^T denotes the transpose of the d-vector.

Equations (3-33) and (3-37) have been derived by setting the derivatives of $\overline{\epsilon_t^2}$ equal to zero. To show that the present solution is indeed a minimum, and not a maximum or a stationary point, it is necessary to demonstrate that the second derivatives of $\overline{\epsilon_t^2}$ at this point are all positive. Using equations (3-20) and (3-23),

$$\begin{aligned} \frac{\partial^2 \overline{\epsilon_t^2}}{\partial d_k^2} &= \frac{2\sigma^2}{\sigma'^2} \sum_{j=1}^{\infty} \left[c_j \frac{\partial^2 c_j}{\partial d_k^2} + \left(\frac{\partial c_j}{\partial d_k} \right)^2 \right] \\ &= \frac{2\sigma^2}{\sigma'^2} \sum_{j=1}^{\infty} \left(\frac{\partial c_j}{\partial d_k} \right)^2 , \quad k = 1,2,\ldots,M \end{aligned} \tag{3-39}$$

and thus the second derivatives with respect to the $\{d_j\}$ are everywhere positive. Differentiating equation (3-20) twice with respect to σ' yields

$$\frac{\partial^2 \overline{\epsilon_t^2}}{\partial \sigma'^2} = -\frac{\sigma}{\sigma'^3} \left[6 \frac{\sigma}{\sigma'} \sum_{j=0}^{\infty} c_j^2 - 4 \right] \tag{3-40}$$

and now employing equation (3-34),

$$\frac{\partial^2 \overline{\epsilon_t^2}}{\partial \sigma'^2} = \frac{\sigma}{\sigma'^3} (6-4) = \frac{2\sigma}{\sigma'^3} > 0 \tag{3-41}$$

The solution is therefore truly a minimum.

It is of interest to know the value of the minimum mean square error that results from the use of the optimum system. This is also found from equations (3-20) and (3-37):

$$\overline{\epsilon^2_{t_{min}}} = \frac{1}{\sigma'^2} \sum_{j=0}^{M} d_j \phi_{xx}(j) - \frac{2\sigma}{\sigma'} + 1 \tag{3-42}$$

$$= 1 - \frac{\sigma}{\sigma'}$$

From this expression, it is clear that as the mean square error approaches zero, σ' will approach σ more and more closely.

In summary, the coefficient set $\{d_j\}$ found from equation (3-33) and the constant σ' of equation (3-37) represent the parameters of the Mth order moving average filter which will optimally whiten and normalize the process x_t. In the present formulation, only the autocorrelation function x_t is required to find the $\{d_j\}$, whereas the input noise variance is also needed to compute σ'.

Another interpretation of this result, useful in a following chapter, can be gained by considering the minimized quantity $\phi_{\varepsilon\varepsilon}(0)$ in its role as the inverse z-transform of the sampled power spectrum for the error:

$$\phi_{\varepsilon\varepsilon}(0) = \frac{\sigma^2}{2\pi i} \oint_{|z|=1} \left| \frac{H(z)D(z)}{\sigma'} - \frac{1}{\sigma} \right|^2 \frac{dz}{z} \tag{3-43}$$

Since the path of integration is the unit circle, this can be written as

$$\phi_{\varepsilon\varepsilon}(0) = \frac{\sigma^2}{2\pi} \int_0^{2\pi} \left| \frac{H(z)D(z)}{\sigma'} - \frac{1}{\sigma} \right|^2 d\theta, \quad z=e^{i\theta} \tag{3-44}$$

which can be considered to be the squared magnitude of the difference between $\sigma H(z)D(z)/\sigma'$ and 1, averaged around the unit circle of the z-plane. If this quantity is sufficiently minimized, then

$$\sigma \frac{A(z)}{B(z)} \approx \frac{\sigma'}{D(z)} \quad , \quad \forall z \ni |z| = 1 \tag{3-45}$$

and thus

$$\sigma^2 \frac{A(z)A(z^{-1})}{B(z)B(z^{-1})} \approx \frac{\sigma'^2}{D(z)D(z^{-1})} \quad . \tag{3-46}$$

Armed with a priori knowledge of the correlation structure of x_t and the input noise variance, the M^{th} order deterministic whitening problem for the process can be solved and the results used to formulate an approximate M^{th} order autoregressive model defined as

$$x_t' = n_t' - \sum_{j=1}^{M} d_j x_{t-j}' \qquad (3\text{-}47)$$

where n_t' is an uncorrelated random sequence with variance σ'^2. In the special case where x_t is an autoregression of order M, there is no approximation involved, $\sigma' = \sigma$, $\overline{\varepsilon_t^2} = 0$, and the set $\{d_j\}$ is identically equal to the set $\{b_j\}$. This can be deduced immediately from the fact that in the case of an autoregression, the system of equation (3-29) is identical with that of equation (2-68) if the k = 0 case is removed from the latter.

The error signal ε_t whose mean square was minimized to find the optimum "whitening" system will henceforth be known as the "whitening" error. For purposes of approximating the spectrum of the original process by that of the derived auto-regression, a more satisfying error measure can be visualized with reference to the following system:

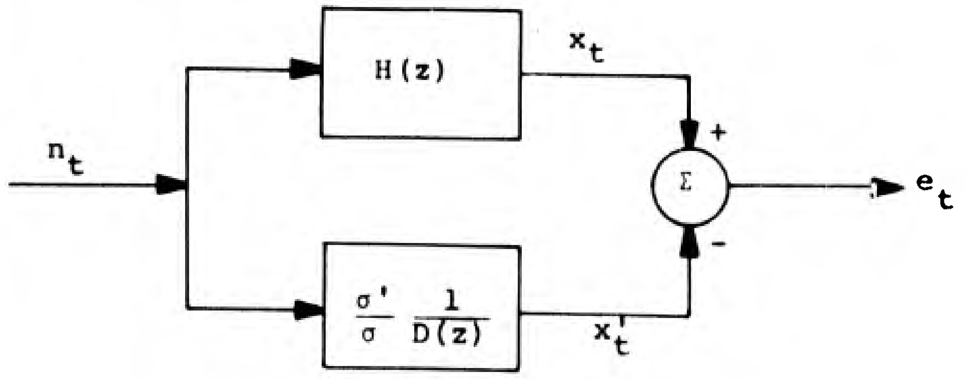


Figure 3-3. Digital system for generating the "approximation" error e_t.

Here, e_t is the difference between x_t and its approximation x_t'. The system transfer function between n_t and e_t is

$$T'(z) = \frac{H(z)D(z) - \sigma'/\sigma}{D(z)} \qquad (3\text{-}48)$$

Comparing this with the transfer function between n_t and ε_t found in equation (3-16), it becomes apparent that e_t and ε_t are related by the autoregressive relation

$$e_t = \sigma'\varepsilon_t - \sum_{j=1}^{M} d_j e_{t-j} \qquad (3\text{-}49)$$

or in a system diagram:

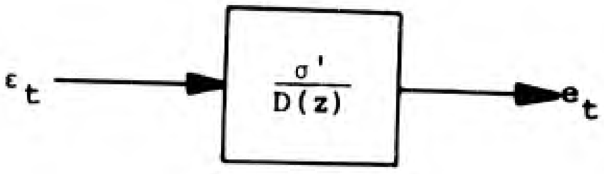


Figure 3-4. Digital system showing the relationship between ε_t and e_t.

Thus, the "approximation" error e_t can be generated by passing the "whitening" error ε_t through the filter $\sigma'/D(z)$. As $\overline{\varepsilon_t^2}$ approaches zero, $\overline{e_t^2}$ must vanish also. For reasons of mathematical tractability, our attention will remain with the whitening error ε_t.

B. All-pole Estimation Based on Mean Lagged Products

It was shown in the last section that given <u>a priori</u> knowledge of the autocorrelation sequence and the input noise variance for a mixed type process, a finite order autoregressive approximation could be derived, optimum in a certain sense. This representation can be used in turn to provide an approximation of the spectrum. In the statistical estimation problem of prime concern here, however, no prior information is available, and one must deal only with a sample of the process itself to derive the spectral representation desired. Fortunately, the equation system for the set $\{d_j\}$ involves only the correlation sequence of the process, which can be estimated from the data sample as the set of mean lagged products. Unfortunately, using the "whitening" approach, σ is needed to compute σ', and without a modification, the present estimation scheme is not complete. A solution of this difficulty is found by noting that as the order of the autoregressive model increases, σ' approaches σ. This is seen from Wold's demonstration[19] that the limit of a sequence of autoregressive approximations of increasing order is the unique infinite order autoregressive representation of the process. This implies directly that

$$\lim_{M\to\infty} \overline{\varepsilon_t^2} = 0 \tag{3-50}$$

and since equation (3-42) shows that

$$\lim_{\overline{\varepsilon_t^2}\to 0} \sigma' = \sigma \tag{3-51}$$

then

$$\lim_{M \to \infty} \sigma' = \sigma \tag{3-52}$$

Using this result and equation (3-37),

$$\lim_{M \to \infty} \left[\sum_{j=0}^{M} d_j \, \phi_{xx}(j) \right] = \lim_{M \to \infty} \sigma\sigma' = \sigma^2 \tag{3-53}$$

Now defining

$$\hat{\sigma}^2 = \sum_{j=0}^{M} d_j \, \phi_{xx}(j) \tag{3-54}$$

$$\lim_{M \to \infty} \hat{\sigma}^2 = \sigma^2 \tag{3-55}$$

and for sufficiently high values of M, $\hat{\sigma}^2$ will serve as an estimate for σ^2 that does not itself involve the latter. The price that is paid for this convenience, however, is increased whitening error, since the system using $\hat{\sigma}$ instead of σ' is sub-optimum. It can be shown using equation (3-42) that this new error is given by

$$\overline{\epsilon_t^2}' = 2(1-\sigma/\hat{\sigma}) \doteq 2\overline{\epsilon_t^2} \tag{3-56}$$

and also approaches zero as M increases without limit. In the following, $\hat{\sigma}^2$ of equation (3-54) will be used as the estimate for the variance of the input noise sequence.

The error that has figured in the discussions above is that associated with the approximation of a general regression by a finite order autoregression. This indeed is the error that would be obtained if the correlation function used in the calculations could be measured from the available data without error. As we have seen in the discussion of the sampling properties of the mean lagged products, however, this is not possible since the lagged products will always display a certain variance. The effects of this statistical spread on the estimation accuracy will now be treated to provide some guidance in choosing the required sample size. The following development is based largely upon that of Mann and Wald[17] as interpreted in Hannan[1]. First the variance in the estimates of the $\{d_j\}$ caused by a finite sample size will be studied, and then the variance for $\hat{\sigma}^2$.

Let the set of noiseless coefficient estimates, i.e. those obtained by a priori knowledge of the $\{\phi_{xx}(j)\}$, be denoted as before $\{d_j\}$. The set derived on the basis of the mean lagged products will be designated $\{\hat{d}_j\}$. These two sets must satisfy the respective systems

$$\sum_{j=0}^{M} d_j \, \phi_{xx}(k-j) = 0 \; , \; k = 1,2,\ldots,M \qquad (3\text{-}57a)$$

$$\sum_{j=0}^{M} \hat{d}_j \hat{\phi}_{xx}(k-j) = 0 \; , \; k = 1,2,\ldots,M \qquad (3\text{-}57b)$$

Now define the set $\{u_k\}$:

$$u_k \equiv \sum_{j=0}^{M} d_j \, \hat{\phi}_{xx}(k-j) \; , \; k = 1,2,\ldots,M \qquad (3\text{-}58)$$

Combining these with equations (3-57a) and (3-57b) yields

$$\sum_{j=0}^{M} (\hat{d}_j - d_j)\, \hat{\phi}_{xx}(k-j) = -u_k \quad , \; k = 1,2,\ldots,M \qquad (3\text{-}59)$$

Since $\hat{d}_0 = d_0 = 1$, this is

$$\sum_{j=1}^{M} (\hat{d}_j - d_j)\, \hat{\phi}_{xx}(k-j) = -u_k \quad , \; k = 1,2,\ldots,M \qquad (3\text{-}60)$$

Now defining an M by M matrix of mean lagged products, $\hat{\phi}_{xx}$, in analogy with equation (3-31) and also

$$e = \begin{bmatrix} \hat{d}_1 - d_1 \\ \hat{d}_2 - d_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{d}_M - d_M \end{bmatrix} \quad \text{and} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_M \end{bmatrix} \qquad (3\text{-}61)$$

equation (3-60) can be expressed as

$$\hat{\phi}_{xx}\, e = -u \qquad (3\text{-}62)$$

Using the definition of a mean lagged product, equation (3-58) is now written as

$$u_k = \frac{1}{N} \sum_{j=0}^{M} \sum_{m=1}^{N} d_j x_m x_{m+k-j} \quad , \; k = 1,2,\ldots,M \qquad (3\text{-}63)$$

where it will be recalled that N is the sample size. Interchanging the order of summation gives

$$u_k = \frac{1}{N} \sum_{m=1}^{N} x_m \sum_{j=0}^{M} d_j x_{m+k-j} \quad , \quad k = 1,2,\ldots,M \tag{3-64}$$

Referring to Figure 3-1 above it can be seen that the inner summation is $\sigma' r_{m+k}$, where it will be recalled that r_t is the whitened and normalized version of x_t. Thus,

$$u_k = \frac{\sigma'}{N} \sum_{j=1}^{N} x_j r_{j+k} \quad , \quad k = 1,2,\ldots,M \tag{3-65}$$

Now it is possible to consider the statistical properties of the set $\{u_j\}$.

$$E[u_k] = \frac{\sigma'}{N} \sum_{j=1}^{N} E[x_j r_{j+k}] = \sigma' \phi_{xr}(k) \quad , \quad k = 1,2,\ldots,M \tag{3-66}$$

It can be shown that

$$\sigma' \phi_{xr}(k) = \sum_{j=0}^{M} d_j \, \phi_{xx}(k-j) \tag{3-67}$$

and therefore, by equation (3-57a),

$$\phi_{xr}(k) = 0 \quad , \quad k = 1,2,\ldots,M \tag{3-68}$$

Hence

$$E[u_k] = 0 \quad , \quad k = 1,2,\ldots,M \tag{3-69}$$

Since the mean of each u_k is zero, the covariance between u_j and u_k is given by

$$\mathrm{Cov}[u_j, u_k] = E[u_j u_k] = \frac{\sigma'^2}{N^2} \sum_{\ell=1}^{N} \sum_{m=1}^{N} E[x_\ell r_{\ell+j} x_m r_{m+k}] \qquad (3\text{-}70)$$

To obtain further results in compact form, it is necessary at this point to make an approximation. In using the present, "whitening" technique, one is in essence formulating a system that will make

$$r_t \approx n_t/\sigma$$

or (3-71)

$$\sigma' r_t \approx n_t$$

If the approximation is made reasonably close by increasing the model order M, r_t will be very nearly a scaled version of n_t and thus approaches an uncorrelated sequence. This is to say that

$$\lim_{M \to \infty} \sigma^2 \phi_{rr}(j) = \phi_{nn}(j) = \sigma^2 \delta(j) \qquad (3\text{-}72)$$

Using this limit expression, equation (3-70) becomes

$$E[u_j u_k] = \frac{\sigma^2}{N^2} \sum_{\ell=1}^{N} \sum_{m=1}^{N} E[x_\ell x_m] \; E[r_{\ell+j} r_{m+k}]$$

$$= \frac{\sigma^2}{N^2} \sum_{\ell=1}^{N} \sum_{m=1}^{N} \phi_{xx}(\ell-m) \; \delta(\ell+k-m-j) \qquad (3\text{-}73)$$

$$= \frac{\sigma^2}{N^2} \sum_{\ell=1}^{N} \phi_{xx}(j-k) = \frac{\sigma^2}{N} \phi_{xx}(j-k)$$

Therefore, when the model order is sufficiently high to yield a good spectral representation,

$$\mathrm{Cov}[u_j, u_k] \approx \frac{\sigma^2}{N} \phi_{xx}(k-j) \tag{3-74}$$

It should be pointed out that in the case where x_t is a pure autoregression, this approximation becomes exact when the model order is equal to that of the process. If this approximating step is not taken, progress can only be made with equation (3-70) by assuming that x_t is Gaussian and expanding the fourth moment expression to yield

$$\mathrm{Cov}[u_j, u_k] = \frac{\sigma^2}{N^2} \sum_{\ell=1}^{N} \sum_{m=1}^{N} \left\{ \phi_{xx}(\ell-m)\phi_{rr}(\ell+k-m-j) + \phi_{xr}(j)\phi_{xr}(k) + \phi_{xr}(m+j-\ell)\phi_{xr}(\ell+k-m) \right\} \tag{3-75}$$

which can only be related to the measured data with difficulty. The approximation proposed above discards the second two terms within the brackets and gives sufficiently accurate results for the analysis of variability when the model order is high enough to provide a good spectral representation. Accepting a small error, the approximate covariance matrix for the $\{u_j\}$ can be written

$$\phi_{uu} = \frac{\sigma^2}{N} \phi_{xx} \tag{3-76}$$

As the number of samples N increases without limit, the $\hat{\phi}_{xx}(j)$ become asymptotically jointly normal. Since by definition the $\{u_j\}$ are linear combinations of the mean lagged products, they must become jointly normal also. Furthermore, as $N \to \infty$,

$\hat{\Phi}_{xx}$ converges in probability to Φ_{xx}. From equation (3-62),

$$e = -\hat{\Phi}_{xx}^{-1} u \tag{3-77}$$

and thus

$$\underset{N \to \infty}{\text{plim}}\ e = -\Phi_{xx}^{-1} u \tag{3-78}$$

Since, as we have seen above, the $\{u_j\}$ are all zero mean, in the limit,

$$E[e_j] = 0 \ , \ j = 1,2,\ldots,M \tag{3-79}$$

Using a well-known property of systems of linear equations in jointly distributed random variables (see Cramer[34]), in the limit of large sample size

$$\Phi_{ee} \simeq \Phi_{xx}^{-1} \Phi_{uu} (\Phi_{xx}^{-1})^T \tag{3-80}$$

Because of equation (3-76) and the symmetry of Φ_{xx}, however,

$$\Phi_{ee} = \frac{\sigma^2}{N} \Phi_{xx}^{-1} \tag{3-81}$$

Thus, the error representing the difference between the noiseless coefficients and those derived from the mean lagged products will be asymptotically normal with zero mean and a covariance matrix given by equation (3-81). It is apparent that the latter are unbiased and consistent estimates of the former.

The modified estimate for the ıput noise variance is

$$\hat{\sigma}^2 = \sum_{j=0}^{M} \hat{d}_j \hat{\phi}_{xx}(j) \qquad (3\text{-}82)$$

Using the mean lagged product definition, this becomes

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=1}^{N} x_j \sum_{k=0}^{M} \hat{d}_k x_{j-k} \qquad (3\text{-}83)$$

Again it is necessary to use the approximation that $r_t \doteq n_t/\sigma$, or

$$\sum_{j=0}^{M} \hat{d}_j x_{k-j} \doteq n_k \qquad (3\text{-}84)$$

This leads to

$$\hat{\sigma}^2 \doteq \frac{1}{N} \sum_{j=1}^{N} x_j n_j \qquad (3\text{-}85)$$

Taking the expected value yields

$$E\left[\frac{1}{N} \sum_{j=1}^{N} x_j n_j\right] = \phi_{nx}(0) = \sigma^2 \qquad (3\text{-}86)$$

Thus,

$$\lim_{M \to \infty} E[\hat{\sigma}^2] = \sigma^2 \qquad (3\text{-}87)$$

and for large M, the estimator becomes nearly unbiased. If x_t is Gaussian, the variance of the estimate can be approached by considering the expectation of the square of the expression on the right of equation (3-85) and using the Gaussian fourth moment expansion. The result is that

$$\mathrm{Var}[\sigma^2] \simeq \frac{\sigma^2 \phi_{xx}(0) + \sigma^4}{N} \qquad (3\text{-}88)$$

with the approximation becoming better for increasing model order. Furthermore, since $\phi_{xx}(0) \geq \sigma^2$,

$$\frac{\sigma^2 \phi_{xx}(0) + \sigma^4}{N} \geq \frac{2\sigma^4}{N} \qquad (3\text{-}89)$$

showing that the standard deviation of $\hat{\sigma}^2$ is on the order of $\sigma^2\sqrt{\frac{2}{N}}$. This is the result derived by Steiglitz (reference 18) using a somewhat different approach involving the asymptotic properties of the χ^2 distribution. Although these final results apply strictly only to Gaussian processes, they probably retain sufficient accuracy in the general case for use in a variability analysis.

In summary, the foregoing has shown that an all-pole approximation for a mixed type process, satisfying a certain optimality criterion, can be formulated on the basis of the observed mean lagged products for a finite sample sequence. The coefficients of this representation are given by the matrix expression

$$d = -\hat{\Phi}_{xx}^{-1} \hat{\phi}_{xx} \qquad (3\text{-}90)$$

with the input noise variance estimated by

$$\hat{\sigma}^2 = \sum_{j=0}^{M} \hat{d}_j \hat{\phi}_{xx}(j) \qquad (3\text{-}91)$$

The coefficients thus computed have been shown to be unbiased and consistent estimates of those that would be obtained under noiseless conditions. The variance estimate becomes asymptotically unbiased and consistent as the model order is increased. The estimation formulae (3-90) and (3-91) are identical to those formulated for purely autoregressive processes by Steiglitz[18] on the basis of maximum likeihood and by Hannan[1] and Whittle[20] from a linear least squares approach. The present treatment provides still another interpretation of these fairly well-known results and in the case of the mixed type process makes it possible to distinguish the two types of error that arise in the analysis.

C. Computational Examples of All-pole Estimation

In practical applications of the present technique, when no a priori knowledge of the process is available, the first step is to measure a sequence of mean lagged products over some sample size N hopefully chosen large enough to yield acceptable statistical error in the final result. The maximum lag required in these measurements is equal to the largest model order M that is anticipated. Of course at the outset, one has few grounds for knowing what M or N should be. After an initial model order is chosen, the matrix of mean lagged products is formed, and the autoregressive coefficient and input variance estimates are calculated using equations (3-90) and (3-91). As the model order M is made successively larger, the quality of the all-pole approximation will increase, and $\hat{\sigma}^2$ will approach the limiting value of σ^2. In fact, the best indication available of the convergence of the approximation to the actual spectrum is the leveling off of the curve of $\hat{\sigma}^2$ plotted against M. It is not necessary to construct and invert the matrix of mean lagged products at each step, since a recursive relationship exists which directly yields the set $\{d_j\}$ for the $(M+1)^{th}$ order from the corresponding set for the M^{th} order. This result is quoted by Whittle[20] and Bartlett[2], and a derivation is provided here in Appendix C. At any rate, when it is felt that the approximation order is sufficient for the purposes at hand, the statistical error inherent in the estimate can be analyzed by using equations (3-81) and (3-88) to yield the standard deviations of the estimates for the $\{d_j\}$ and σ^2. As derived, these results call for the theoretical covariance matrix and its inverse, but in the practical situation, the matrix of mean lagged products constitutes a covariance estimate of sufficient quality to be used for the variability analysis. Thus, having previously computed $\hat{\Phi}_{xx}^{-1}$ for finding the coefficient estimates, one can use the result in equation (3-81) to determine the magnitude of the statistical error. Should this be unacceptable, the sample size is increased and the estimate re-computed. The final all-pole spectral estimate is found by using the $\{\hat{d}_j\}$ and $\hat{\sigma}^2$ in the expression

$$\hat{S}_{xx}(z) = \frac{\hat{\sigma}^2}{\hat{D}(z)\hat{D}(z^{-1})} \qquad (3\text{-}92)$$

and can be evaluated as a function of the radian frequency ω by means of the substitution $z = e^{i\omega T}$.

The five standard examples described in Appendix A will be used to illustrate the all-pole procedure and the characteristics of the all-pole results. For the purposes of the present study, the maximum model order was taken to be eight, and approximations of all orders up to that maximum were computed and plotted so that the improvement of estimate quality with model order could be graphically demonstrated. An exception was made for the first example, itself a fourth order all-pole function, where attention was directed particularly to the behavior of the estimate as a function of samp.. size. Although each example was analyzed for several values of N, in all but the first example, the results of only one of these will be presented. As a basis for comparison, the result of using the theoretical correlation sequence (as found in Appendix A) will be presented for the case where M = 8 and used to evaluate the statistical errors which arise from the use of the mean lagged products.

1. Example # 1

As described in Appendix A, the first example is a 4th order autoregression with system function

$$H(z) = \frac{1}{B(z)} = \frac{1}{1 - 2.036z^{-1} + 1.812z^{-2} - 0.770z^{-3} + 0.129z^{-4}}$$

and input noise sequence variance $\sigma^2 = 1.0$. With all of the examples treated here, the input sequence is Gaussian.

Since there is no approximation error involved in the all-pole estimation of an autoregression, there is little of interest in the behavior of the spectral estimate for increasing model order. Hence, this opportunity will be taken to study the effect of the sample size on estimate accuracy for a given value of M - in this case, 4. In the absence of statistical error, the actual autoregressive coefficients would be returned, and an exact reproduction of the theoretical spectrum produced. The actual state of affairs is summed up in Table 3-1 and Figures 3-5 and 3-6. In the first half of the table are listed the coefficient and variance estimates derived on the basis of 100, 200, 500, 1,000, and 2,000 samples. These should be compared with the actual values at the top. Figure 3-5 displays the spectra corresponding to these parameter estimates, in comparison with the actual spectrum. The improvement with increasing sample size is evident from both table and graphs. Figure 3-6 is an enlarged presentation of the estimate for N = 2,000.

The second half of Table 3-1 presents the standard deviations of the parameter estimates as a function of sample size. Although in a practical situation with no knowledge of the process beforehand, these calculations would be based on the matrix of mean lagged products, they are here computed from the theoretical covariance matrix and hence represent the true standard deviations for the estimates. Actual computation in a number of examples has shown that there is only a negligible

Table 3-1

All-pole Estimation Coefficients and Their Standard Deviations for Example # 1 as a Function of the Sample Size N.

N	Parameter Estimates $\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{\sigma}^2$
100	-1.69826	1.07022	-0.00480	-0.19648	1.41050
200	-1.91213	1.58757	-0.54649	0.02121	1.11354
500	-2.03121	1.79247	-0.73369	0.11027	1.05251
1,000	-2.00864	1.74421	-0.69524	0.09590	1.03578
2,000	-1.99592	1.73154	-0.72503	0.11155	1.04435
Actual Values	-2.036	1.812	-0.770	0.129	1.0
N	**Theoretical Estimate Standard Deviations** $\sigma(\hat{b}_1)$	$\sigma(\hat{b}_2)$	$\sigma(\hat{b}_3)$	$\sigma(\hat{b}_4)$	$\sigma(\sigma^2)$*
100	0.0992	0.2130	0.2130	0.0992	0.406
200	0.0700	0.1510	0.1510	0.0700	0.287
500	0.0442	0.0950	0.0950	0.0442	0.181
1,000	0.0314	0.0673	0.0673	0.0314	0.129
2,000	0.0222	0.0477	0.0477	0.0222	0.090

*Computed from Equation (3-88)

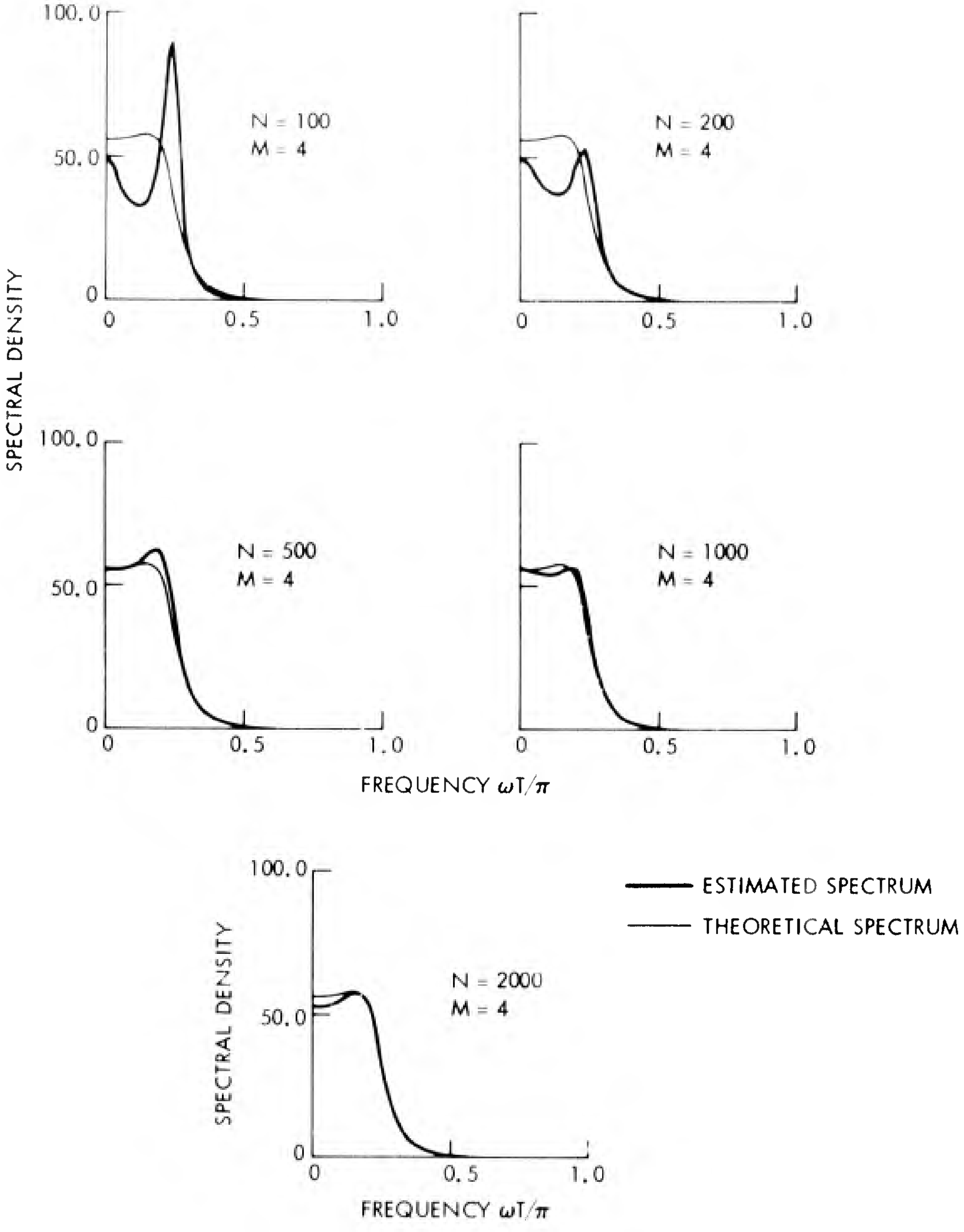


FIG. 3-5 4TH ORDER SPECTRAL ESTIMATES FOR EXAMPLE #1 COMPUTED OVER FIVE SAMPLE SIZES

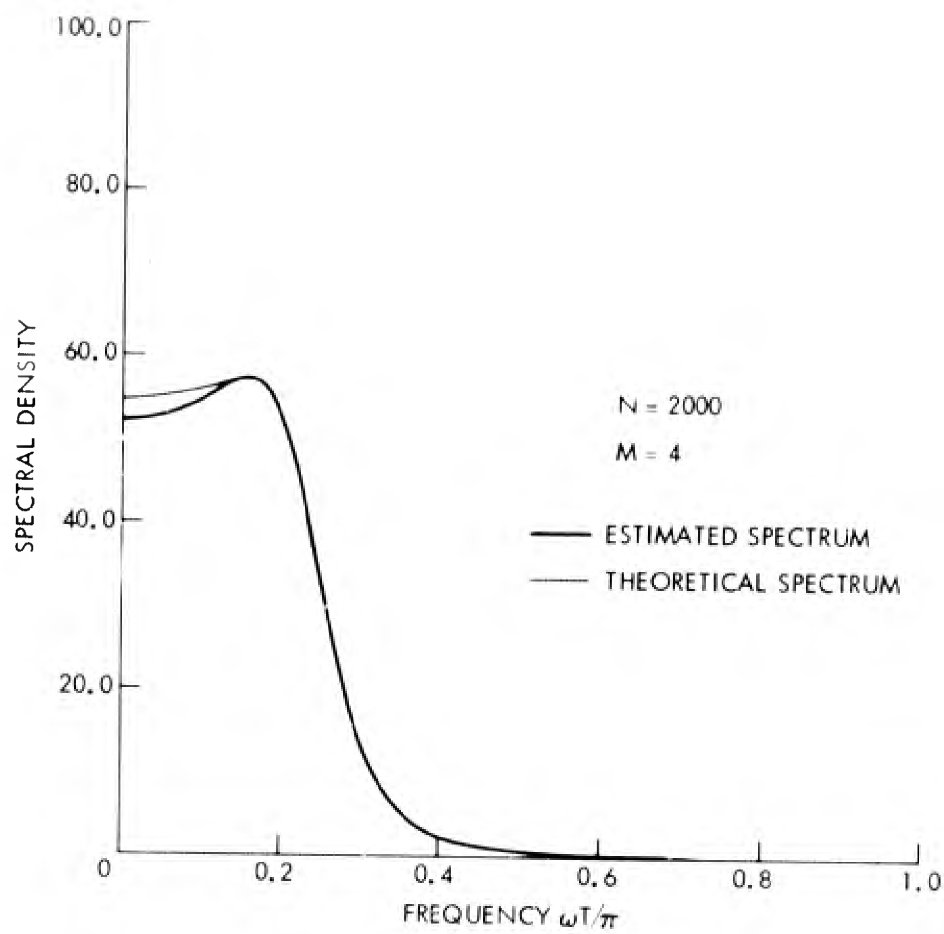


FIG. 3-6 4TH ORDER SPECTRAL ESTIMATE FOR EXAMPLE #1 COMPUTED OVER 2000 SAMPLES difference in the results for reasonable sample sizes. To give one numerical example of this type of calculation, the case where N = 2,000 will be used. From equation (3-81), the covariance matrix for the estimate errors is given by

$$\Phi_{ee} = \frac{\sigma^2}{N} \Phi_{xx}^{-1}$$

Using the theoretical autocorrelation sequence presented in Appendix A, the 4th order covariance matrix has the inverse

$$\Phi_{xx}^{-1} = \begin{bmatrix} 0.98280 & -1.93531 & 1.57691 & -0.50682 \\ -1.93531 & 4.53242 & -4.22733 & 1.57691 \\ 1.57691 & -4.22733 & 4.53242 & -1.93531 \\ -0.50682 & 1.57691 & -1.93531 & 0.98280 \end{bmatrix}$$

If N = 2,000 and σ^2 = 1.0, then

$$\Phi_{ee} = \begin{bmatrix} 0.000491 & -0.000968 & 0.000788 & -0.000253 \\ -0.000968 & 0.002265 & -0.002115 & 0.000788 \\ 0.000788 & -0.002115 & 0.002265 & -0.000968 \\ -0.000253 & 0.000788 & -0.000968 & 0.000491 \end{bmatrix}$$

As can be verified by reference to Table 3-1, the estimate standard deviations are the square roots of the diagonal terms. If required, the correlation coefficients for pairs of parameter errors can be also obtained from Φ_{ee}. For the errors in $\hat{b}_1$ and $\hat{b}_2$, for example

$$\rho_{12} = \frac{\mathrm{Cov}[e_1, e_2]}{\sigma_{e_1} \sigma_{e_2}}$$

$$= \frac{-0.000968}{(.0222)(.0477)} = -0.915$$

and it is apparent that the errors can be rather strongly correlated. From equation (3-88), with N = 2,000,

$$\mathrm{Var}[\hat{\sigma}^2] \; \frac{15.2437 + 1.0}{2,000} = 0.00813$$

Thus, the standard deviation of the variance estimate is 0.09 as indicated. Table 3-1 shows that the variability analysis well predicts the magnitude of the parameter errors, although to be sure, many of the estimates lie beyond one standard deviation of the mean. This is not too surprising, because the strong correlation between the separate errors tends to produce whole sets of out-liers. At any rate, the quality of the error analysis improves markedly for the larger sample sizes, where the assumptions made in the theoretical approach are better justified.

The present example can also be used to provide a clear demonstration of the behavior of the variance estimate $\hat{\sigma}^2$ as the autoregressive approximation becomes nearly exact. If the present process is modeled as an autoregression of increasing order, beginning with M = 1, the successive estimates for σ^2, based on 2,000 samples, are found in the following table:

Table 3-2. Variance Estimate Sequence for Example #1 based on 2,000 samples.

M	1	2	3	4	5	6	7	8
$\hat{\sigma}^2$	3.8477	1.3904	1.0575	1.0444	1.0434	1.0434	1.0432	1.0429

Note that after the correct order (4) is reached, the variance estimate decreases only negligibly, since except for the effects of statistical error, the 4^{th} order model should represent the process exactly. This is also in accord with the findings of Appendix C. Another manifestation of the convergence of the approximation is that the values of the first four coefficient estimates scarcely vary for successive model orders greater than 4, and that the higher order coefficients become very small by comparison.

2. Example # 2

Example # 2 is a fourth order moving average with system function

$$H(z) = A(z) = 1 + 2.036z^{-1} + 1.812z^{-2} + 0.770z^{-3} + 0.129z^{-4}$$

and input noise variance σ^2 again 1.0. All-pole spectral estimates for orders 1 through 8, all based on 1,000 samples, are portrayed in Figure 3-7 with the 8^{th} order estimate enlarged for clarity in Figure 3-8. The corresponding sequence of variance estimates was found to be as shown in the table:

Table 3-3. Variance Estimate Sequence for Example # 2 based on 1,000 samples.

M	1	2	3	4	5	6	7	8
$\hat{\sigma}$	3.2863	1.8566	1.3434	1.1367	1.0511	1.0435	1.0427	1.0211

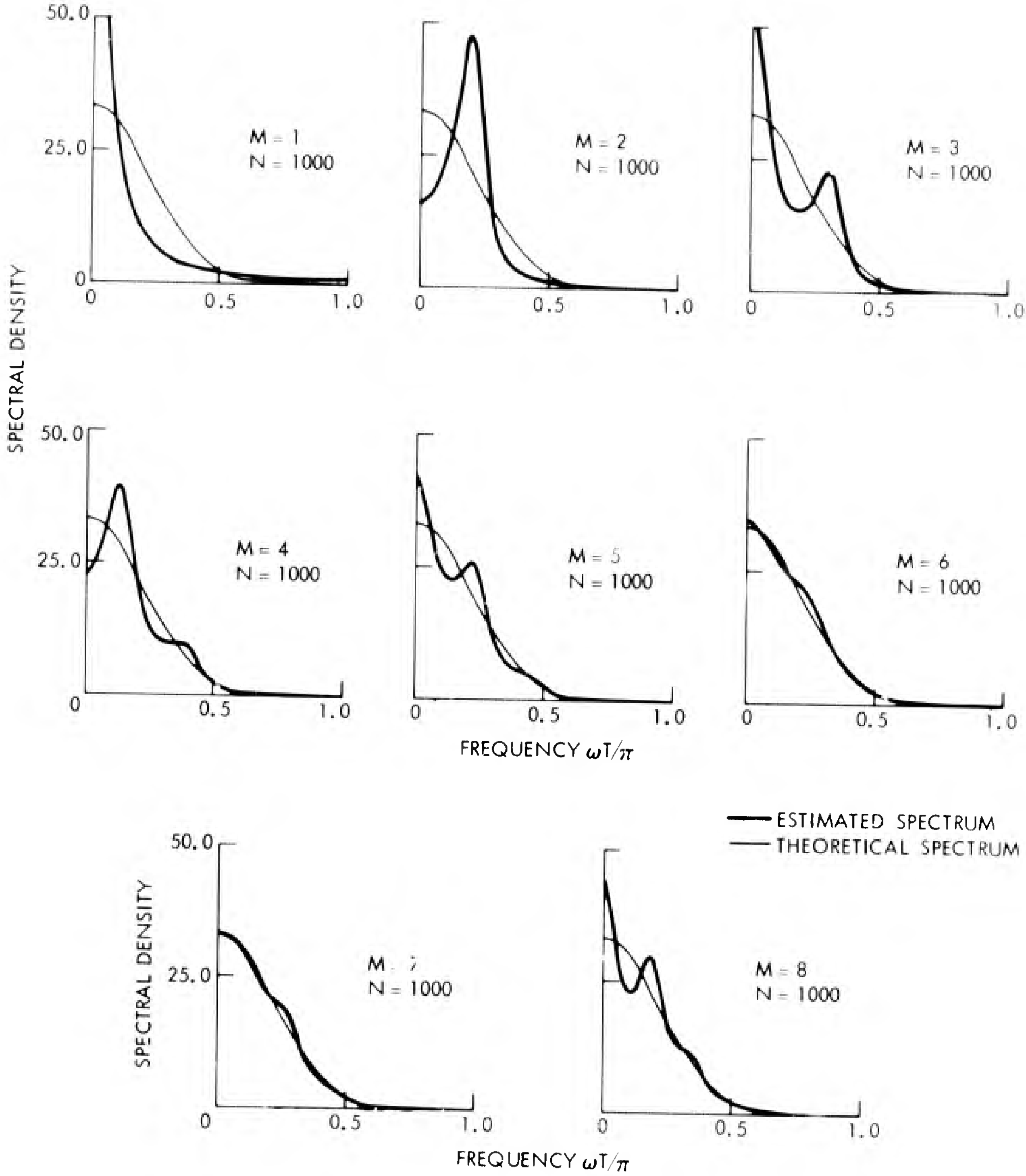


FIG. 3-7 SEQUENCE OF ALL-POLE ESTIMATES FOR EXAMPLE #2 BASED ON N = 1000 SAMPLES

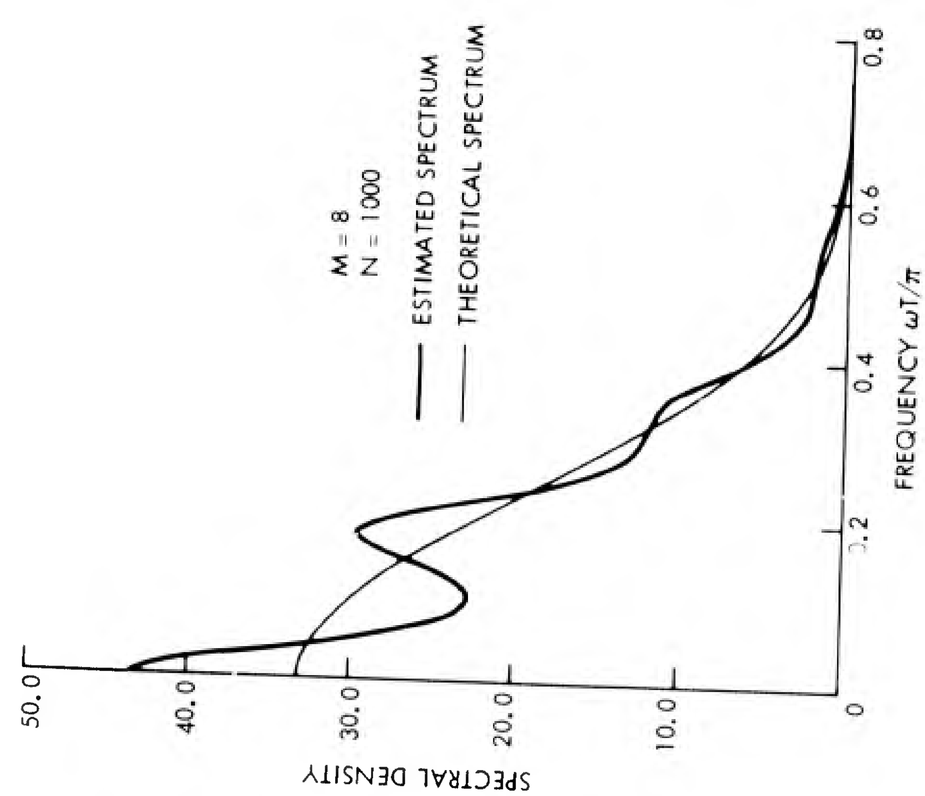


FIG. 3-8 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #2 BASED ON N = 1,000 SAMPLES

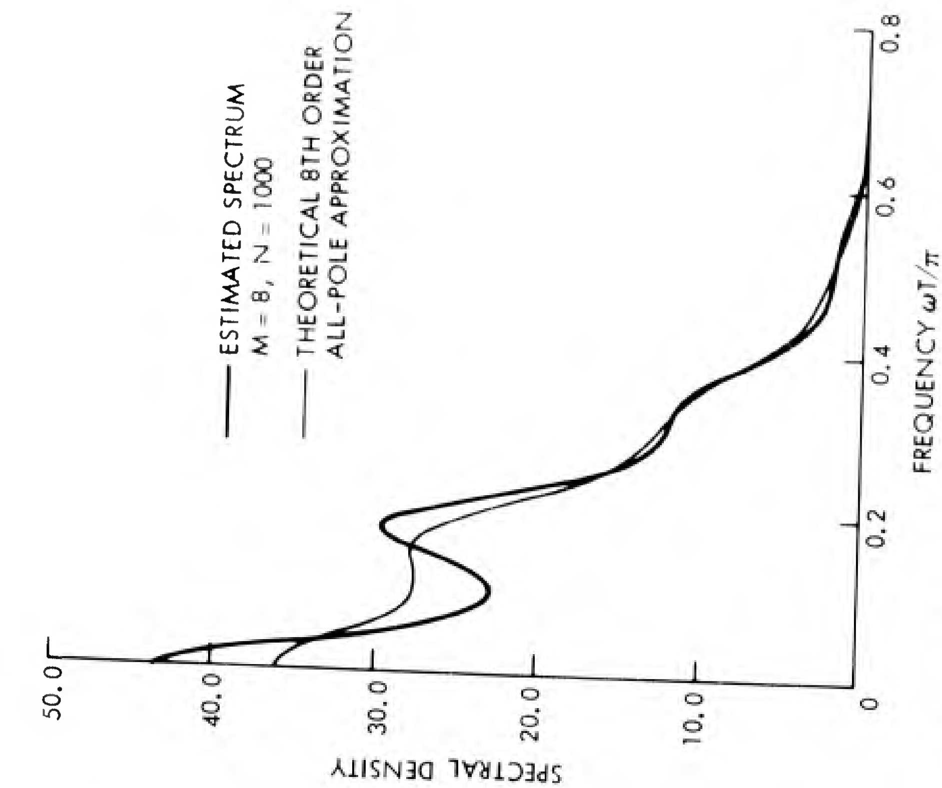


FIG. 3-9 A COMPARISON OF THE 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #2, BASED ON 1,000 SAMPLES; AND THE THEORETICAL 8TH ORDER ALL-POLE APPROXIMATION

The slow decrease of $\hat{\sigma}^2$ for M>6 indicates that the quality of the approximation has become reasonably good. This is confirmed by the spectral comparisons.

The statistical error inherent in the estimates will be studied by concentrating upon the 8^{th} order model and comparing the result based on 1,000 samples with that obtained from using the theoretical autocorrelation sequence (essentially the solution of the deterministic whitening problem). Table 3-4 presents both sets of parameters, their differences, and the theoretical standard deviations of the latter. It can be seen that the first five estimates are within one standard deviation of the theoretical values, and the last three are within two. According to equation (3-88), the standard deviation for $\hat{\sigma}^2$ is 0.1014.

When using the theoretical autocorrelation sequence as the basis for the calculations, the successive values of the constant σ', from equation (3-37), are found in Table 3-5:

Table 3-5. Sequence of σ' for Example #2 based on the theoretical correlation function.

M	1	2	3	4	5	6	7	8
σ'	3.2709	1.8527	1.3135	1.0925	1.0225	1.0144	1.0144	1.0071

Thus σ' does indeed approach the input standard deviation as the model order increases. When M = 8, the mean square whitening error $\overline{\epsilon_{t_{min}}^2}$ becomes 0.00709, implying that the RMS whitening error is 8.4% of the intended RMS value of the

Table 3-4

A Comparison of Theoretical and Experimental
8^{th} Order All-pole Estimation Coefficients
for Example # 2. N = 1,000 Samples.

j	$\hat{d}_j$	d_j	$e_j = \hat{d}_j - d_j$	$\sigma(e_j)$
1	-2.00950	-2.02738	+0.01788	0.0314
2	2.27206	2.31140	-0.03934	0.0711
3	-1.75831	-1.80971	+0.05140	0.1016
4	0.89373	0.95217	-0.05844	0.1163
5	-0.11857	-0.21475	+0.09618	0.1163
6	-0.29836	-0.15606	-0.14229	0.1016
7	0.31628	0.19060	+0.12568	0.0711
8	-0.14374	-0.07891	-0.06482	0.0314

whitened and normalized signal r_t (which is unity). This in turn indicates that the assumption of equation (3-71) in the variability analysis is not as well satisfied as one might hope, leading to possible errors in the estimate standard deviations.

Finally, the spectrum corresponding to the 8^{th} order theoretical model is compared to that generated from the mean lagged products over 1,000 samples in Figure 3-9. The difference is solely the result of statistical error in the latter.

3. Example # 3

The third example has a relatively simple system function:

$$H(z) = \frac{A(z)}{B(z)} = \frac{1 - 0.859z^{-1}}{1 + 0.853z^{-2}}$$

with the variance of n_t again unity. A comparison of all-pole spectral estimates for orders 1 through 8 and based on 1,000 samples is presented in Figure 3-10, and the corresponding sequence of variance estimates is given in Table 3-6.

Table 3-6. Variance Estimate Sequence for Example #3 based on 1,000 Samples.

M	1	2	3	4	5	6	7	8
$\hat{\sigma}^2$	6.5219	1.6320	1.2308	1.1399	1.0698	1.0252	1.0079	0.9939

Although the enlarged plot of the 8^{th} order spectral model in Figure 3-11 displays fairly good agreement with the actual spectrum, it is apparent from the behavior of the vari-

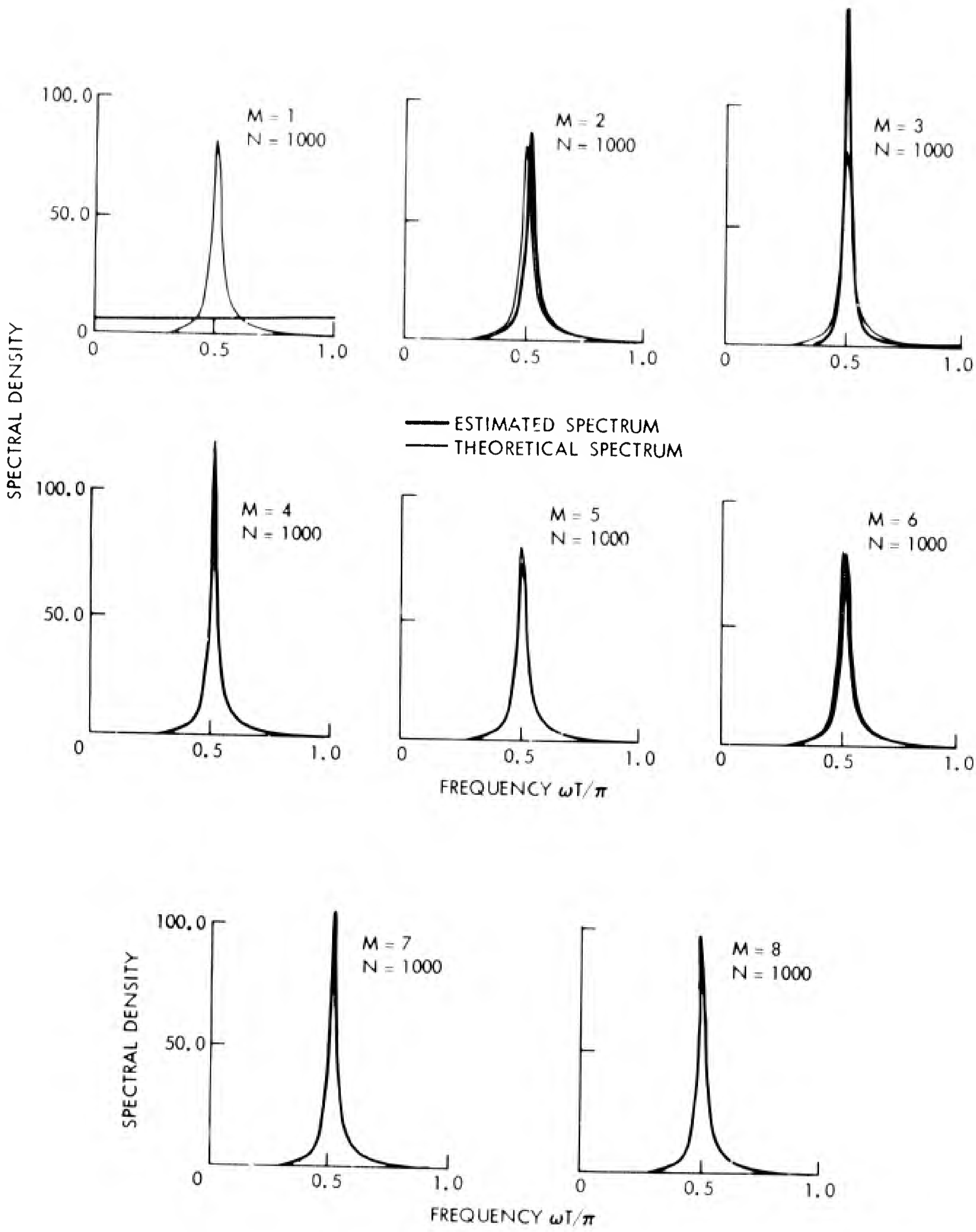


FIG. 3-10 SEQUENCE OF ALL-POLE ESTIMATES FOR EXAMPLE #3 BASED ON N = 1,000 SAMPLES

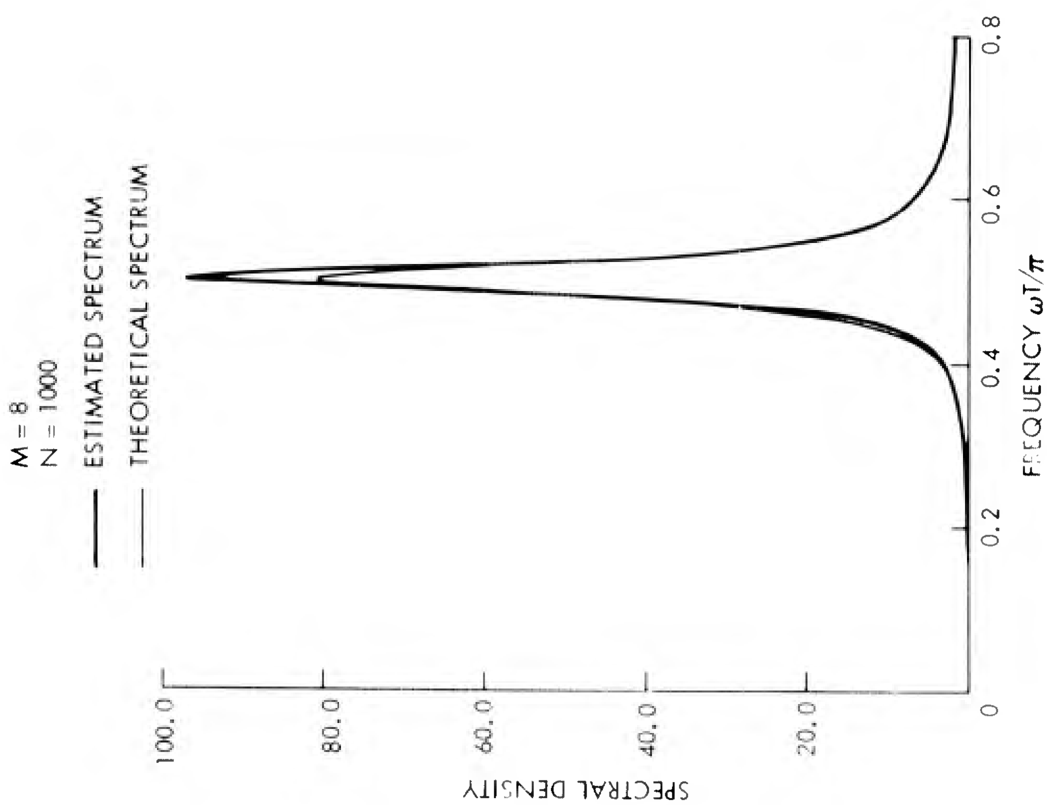


FIG. 3-11 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #3 BASED ON N = 1,000 SAMPLES

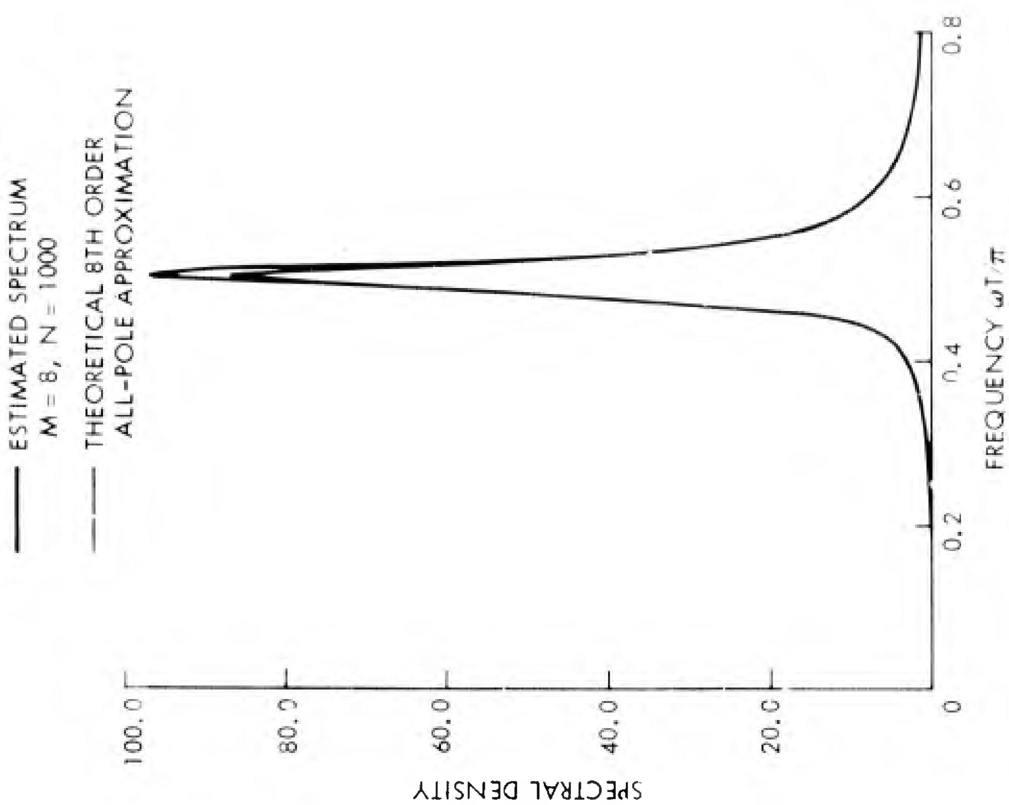


FIG. 3-12 A COMPARISON OF THE 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #3, BASED ON 1,000 SAMPLES, AND THE THEORETICAL 8TH ORDER ALL-POLE APPROXIMATION

ance sequence that further improvement can be expected. Restricting attention to the 8^{th} order model, as before, a comparison of the approximation coefficients based on theoretical autocorrelations with those based on 1,000 sample mean lagged products is presented in Table 3-7. Also tabulated are the statistical coefficient errors and the theoretical standard deviation of the latter. For the most part, the errors lie within one standard deviation of their mean of zero. The standard deviation of the variance estimate, by equation (3-88), is 0.086 for 1,000 samples.

It is interesting to note that the value of σ', computed for this process in the deterministic whitening problem, is 1.03363 and thus yields a mean square whitening error of 0.0325. This in turn shows that the RMS error in approximating n_t/σ by r_t is 0.18 or 18% of the nominal value. Despite this substantial error, the variability analysis seems to have maintained sufficient accuracy.

The final figure for the present example (Figure 3-12) compares, as in the preceding section, the 8^{th} order spectra based on theoretical correlations and the mean lagged products over 1,000 samples.

4. Example # 4

The double peaked spectrum of the fourth example arises from the system function

$$H(z) = \frac{A(z)}{B(z)} = \frac{1 - 1.7835z^{-1} + 0.793z^{-2}}{1 - 1.337z^{-1} + 1.632z^{-2} - 0.987z^{-3} + 0.660z^{-4}}$$

Table 3-7

A Comparison of Theoretical and Experimental 8^{th} Order All-pole Estimation Coefficients for Example # 3. N = 1,000 Samples.

j	$\hat{d}_j$	d_j	$e = \hat{d}_j - d_j$	$\sigma(e_j)$
1	0.87301	0.81987	+0.05314	0.0309
2	1.53848	1.51173	+0.02675	0.0394
3	1.30769	1.21220	+0.09549	0.0596
4	0.97733	0.94076	+0.03656	0.0672
5	0.74136	0.69114	+0.05022	0.0672
6	0.48911	0.45754	+0.03157	0.0596
7	0.23093	0.23457	-0.00364	0.0394
8	0.11776	0.11593	+0.00182	0.0309

with $\sigma^2 = 1.0$. Successive all-pole estimates, based on 5,000 samples, are displayed in Figure 3-13, with an enlarged presentation of the 8^{th} order spectrum found in Figure 3-14. It is apparent from these plots that the present spectrum is sufficiently complex that even an 8^{th} order approximation serves only poorly to represent it. This is borne out by the sequence of variance estimates:

Table 3-8. Variance Estimate Sequence for Example #4 based on 5,000 Samples.

M	1	2	3	4	5	6	7	8
$\hat{\sigma}^2$	5.1791	2.8188	2.7930	2.4218	2.0110	1.7346	1.5457	1.4203

It is of interest to note that substantial decrease can still be found in this sequence when M = 16, indicating that a model order of that order would be necessary to adequately represent the process. Even so, the present 8^{th} order estimate has unequivocally indicated the double-peaked nature of the spectrum and has located the peaks with reasonable accuracy. Such a result might well suffice for certain applications. Table 3-9 is analogous to Tables 3-4 and 3-7 discussed above and highlights the statistical error present in this case when M = 8 and N = 5,000. Again the theoretical standard deviations well predict the magnitude of the error, despite the fact that the solution of the deterministic whitening problem reveals a 54.6% RMS error in approximating n_t/σ by r_t. It is somewhat surprising that the

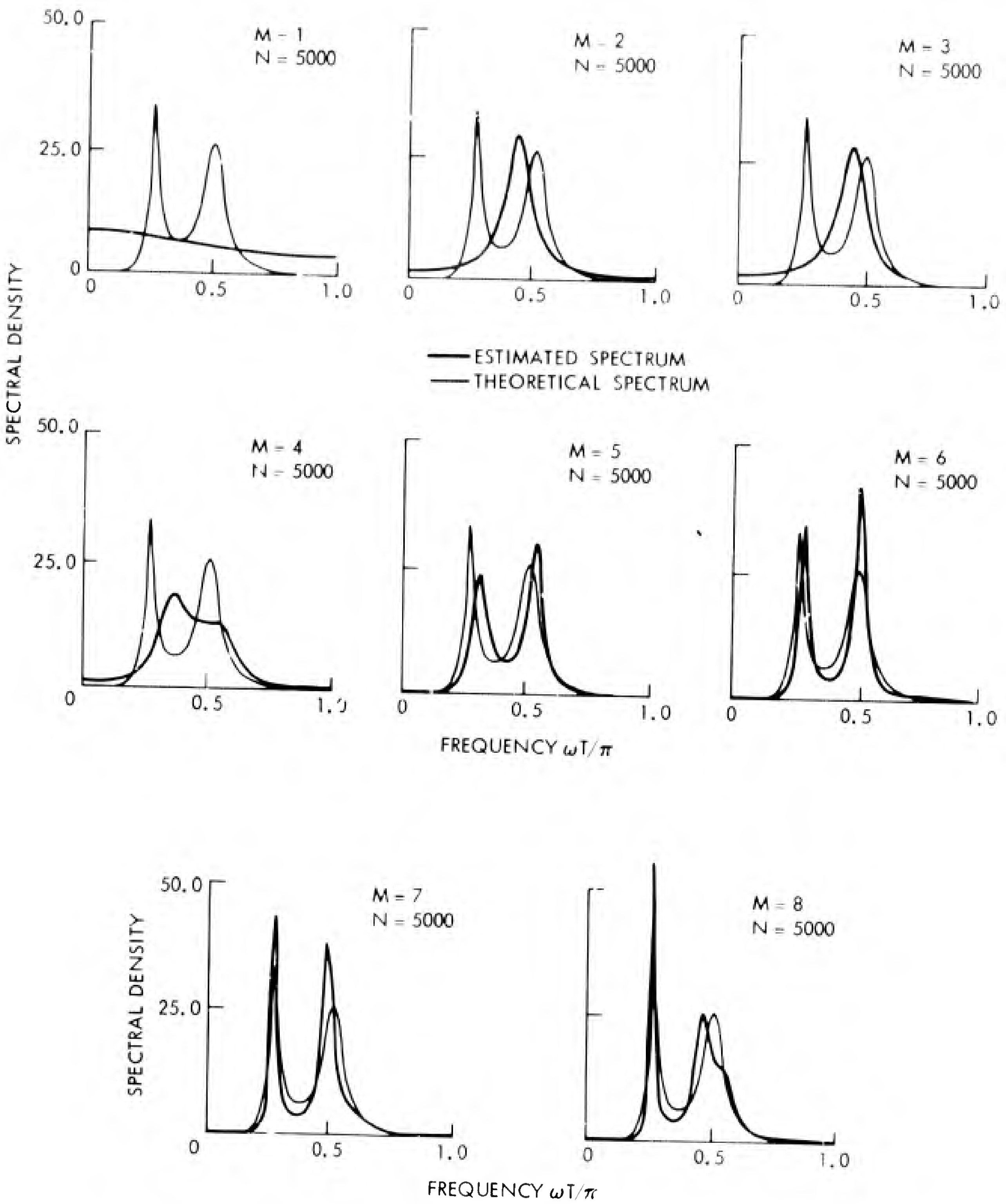


FIG. 3-13 SEQUENCE OF ALL-POLE ESTIMATES FOR EXAMPLE #4 BASED ON N = 5,000 SAMPLES

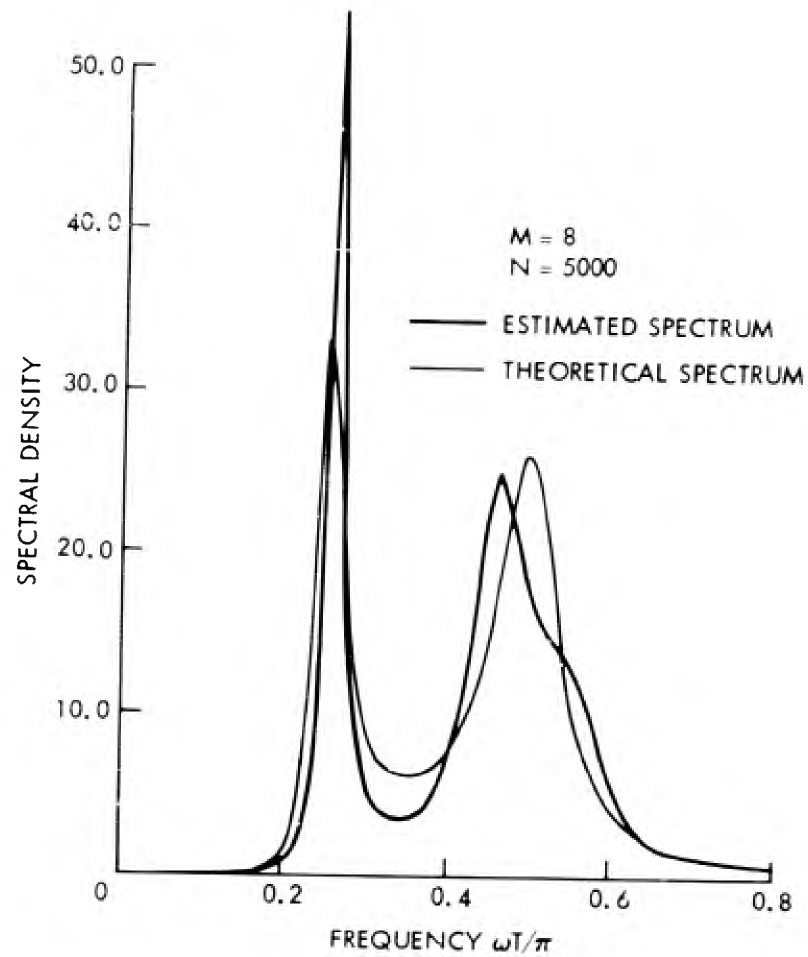


FIG. 3-14 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #4 BASED ON N = 5,000 SAMPLES

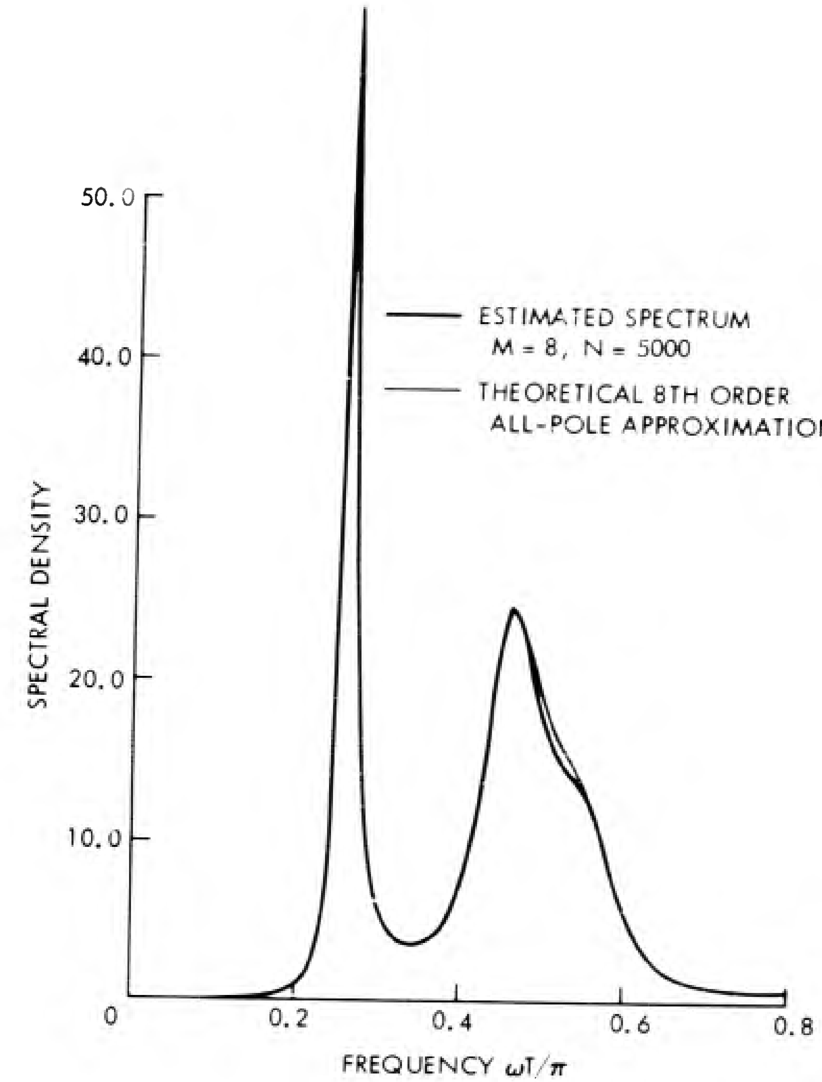


FIG. 3-15 A COMPARISON OF THE 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #4, BASED ON 5,000 SAMPLES, AND THE THEORETICAL 8TH ORDER ALL-POLE APPROXIMATION

Table 3-9

A Comparison of Theoretical and Experimental 8^{th} Order All-pole Estimation Coefficients for Example # 4. N = 5,000 Samples.

j	$\hat{d}_j$	d_j	$e = \hat{d}_j - d_j$	$\sigma(e_j)$
1	0.08503	0.06005	+0.02497	0.0114
2	1.12870	1.12994	-0.00124	0.0108
3	0.55629	0.53368	+0.02262	0.0156
4	0.74134	0.73542	+0.00592	0.0146
5	0.72377	0.71181	+0.01197	0.0146
6	0.62164	0.61830	+0.00333	0.0156
7	0.32798	0.31595	+0.01203	0.0108
8	0.28425	0.28138	+0.00287	0.0114

variability analysis holds up so well in the face of this violation of one of the key assumptions.

Figure 3-15 compares the theoretical and experimental all-pole estimates for the present case, when the latter is based on 5,000 samples.

5. Example # 5

The band pass character of the fifth spectral example is produced by the system function

$$H(z) = \frac{A(z)}{B(z)} = \frac{1 - 1.74z^{-1} + 0.81z^{-2}}{1 - 1.352z^{-1} + 1.338z^{-2} - 0.662z^{-3} + 0.240z^{-4}}$$

and an input noise variance $\sigma^2 = 1.0$. The sequence of successive all-pole estimates, based on 5,000 samples, is shown in Figure 3-16, with the final estimate for M = 8 enlarged in Figure 3-17. As in the preceding example, it is evident that even the 8th order approximation leaves a good deal to be desired. The sequence of variance estimates supports this observation:

Table 3-10. Variance Estimate Sequence for Example #5 Based on 5,000 Samples.

M	1	2	3	4	5	6	7	8
$\hat{\sigma}^2$	2.8910	1.6877	1.5575	1.4240	1.3098	1.2214	1.1538	1.1079

It is found, in fact, that the variance estimate levels off for M = 15 or so.

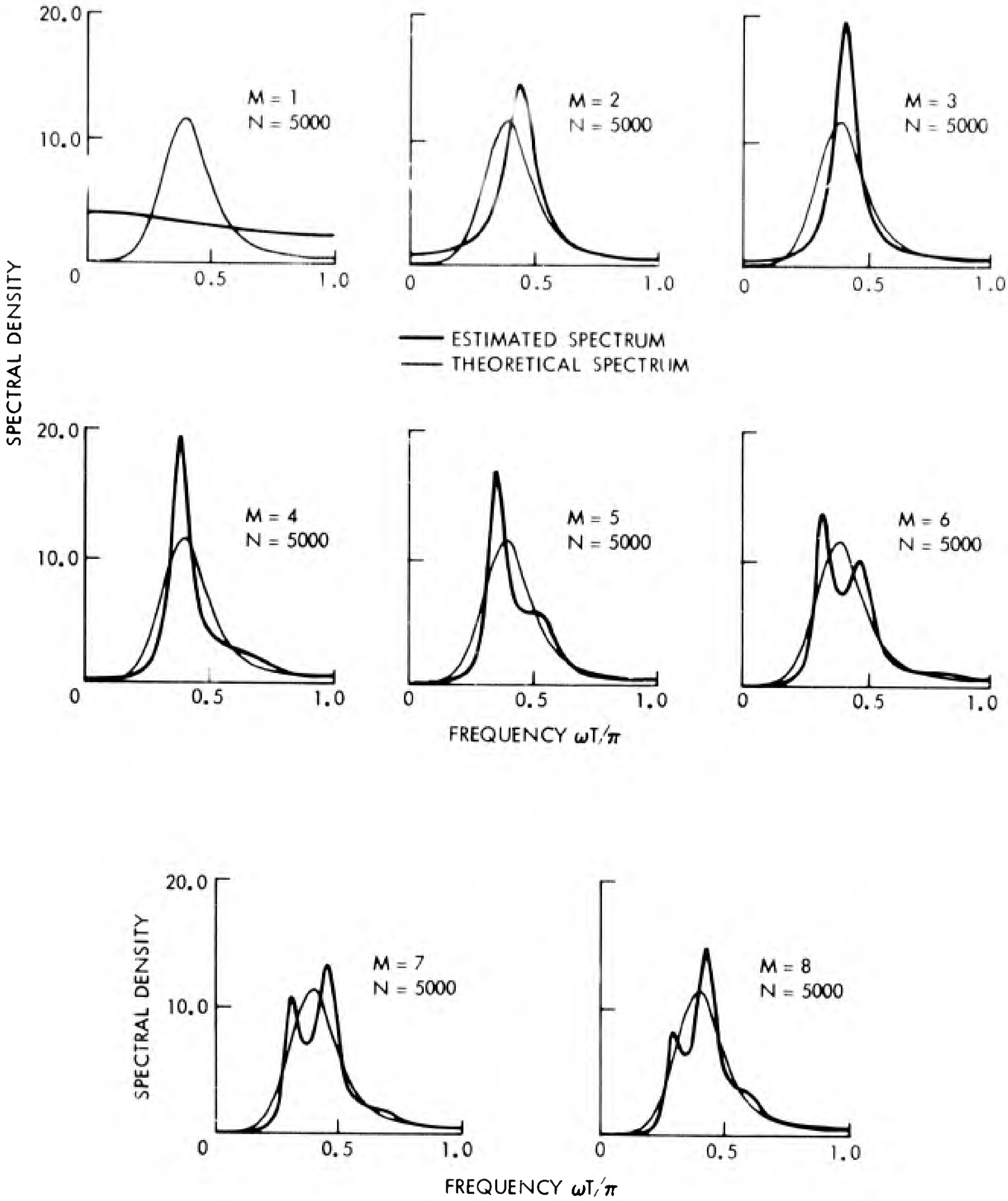


FIG. 3-16 SEQUENCE OF ALL-POLE ESTIMATES FOR EXAMPLE #5 BASED ON N = 5,000 SAMPLES

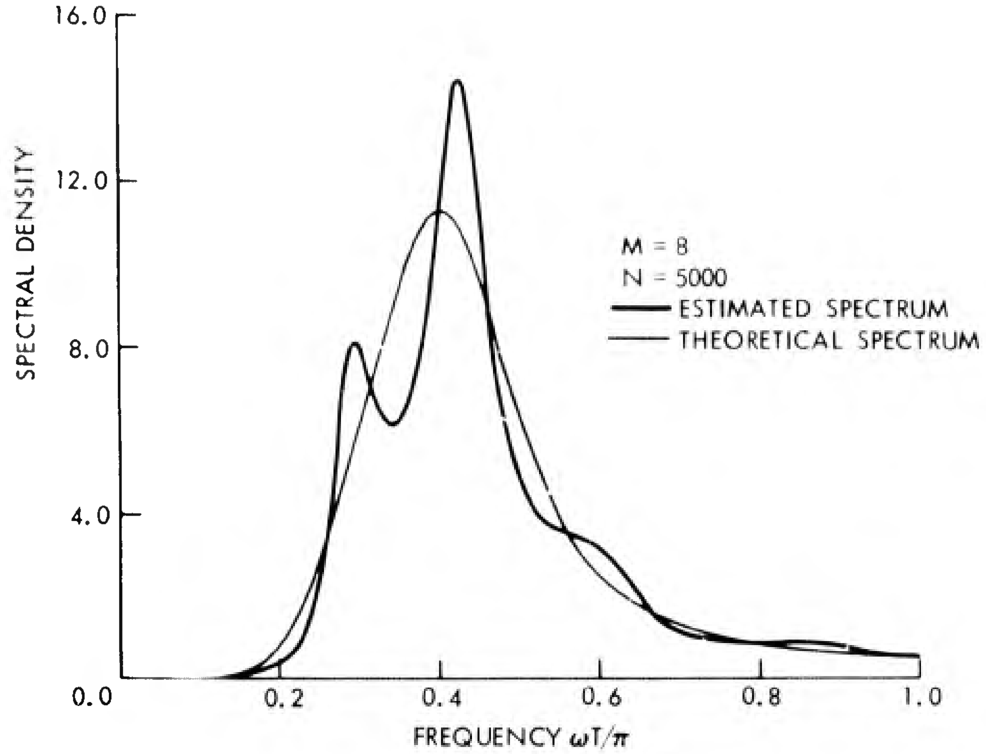


FIG. 3-17 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #5 BASED ON N = 5,000 SAMPLES

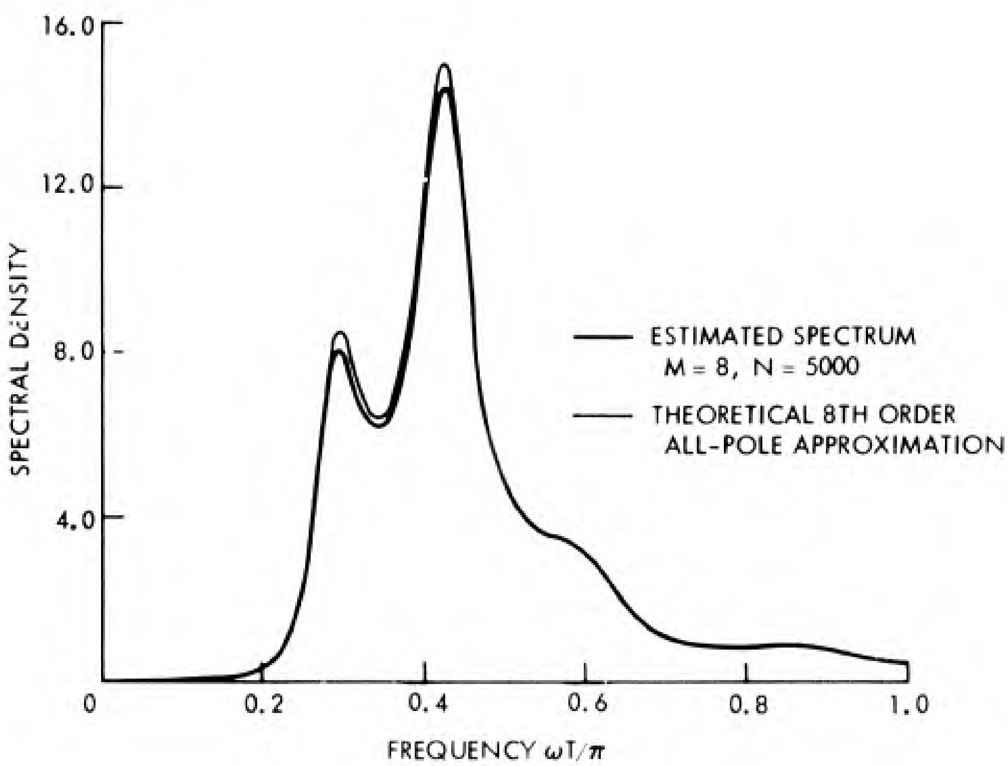


FIG. 3-18 A COMPARISON OF THE 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #5, BASED ON 5,000 SAMPLES, AND THE THEORETICAL 8TH ORDER ALL-POLE APPROXIMATION

The statistical properties for the parameter estimates when M = 8 are outlined in Table 3-11, which shows that the error magnitudes, with only one exception, are less than one standard deviation. The standard deviation of the variance estimate is 0.063 for 5,000 samples, and the theoretical "whitening" approach reveals an RMS error of 31% in the approximation of equation (3-71).

Finally, Figure 3-18 shows the spectrum corresponding to the 8^{th} order model based on theoretical autocorrelations and that based on the mean lagged products computed over 5,000 samples.

The set of all-pole estimates described above provides a fairly complete demonstration of the strengths and weaknesses of the technique. These will be discussed extensively in the Conclusions, but it should be apparent here that for many processes typical of those that might be met in actual practice, and exemplified by the fourth and fifth examples, only a large order model will suffice to represent the process. These spectra are characterized by sharp peaks and the presence of zeros near the z-plane unit circle. If a computational limit is placed on the approximation order - as has artificially been done here - the method may be unacceptable. On the other hand, the technique is ideal for all-pole spectra (such as Example # 1), and adequate results can be obtained for most slowly varying low pass functions (Example # 2). Perhaps the greatest advantage of the all-pole technique is that no a priori knowledge whatever is required of

Table 3-11

A Comparison of Theoretical and Experimental 8^{th} Order All-pole Estimation Coefficients for Example # 5. N = 5,000 Samples.

j	$\hat{d}_j$	d_j	$e_j = \hat{d}_j - d_j$	$\sigma(e_j)$
1	0.29183	0.26959	+0.02224	0.0132
2	1.03637	1.03400	+0.00237	0.0132
3	0.82327	0.80848	+0.01478	0.0180
4	0.78602	0.78057	+0.00544	0.0190
5	0.66759	0.66275	+0.00484	0.0190
6	0.49754	0.50058	-0.00305	0.0180
7	0.28422	0.28096	+0.00326	0.0132
8	0.19940	0.20874	-0.00934	0.0132

the process, and that one has at least theoretical assurance that a good representation can eventually be found by going to a high enough model order.

CHAPTER IV

ESTIMATION OF AUTOREGRESSIVE COEFFICIENTS IN A MIXED-TYPE PROCESS

For many applications, all-pole spectral estimates are inappropriate, either because the process of interest requires an autoregressive model of excessive order, or because there is some interest in separately identifying the autoregressive and moving average parts of a mixed-type process. The present chapter considers the problem of separately estimating the autoregressive coefficients of a process of mixed type so that such a decomposition is possible.

A. Statistical Estimation of the Autoregressive Coefficients

The digital system model for a mixed-type process can be portrayed in a form that clearly separates the autoregressive and moving average parts:

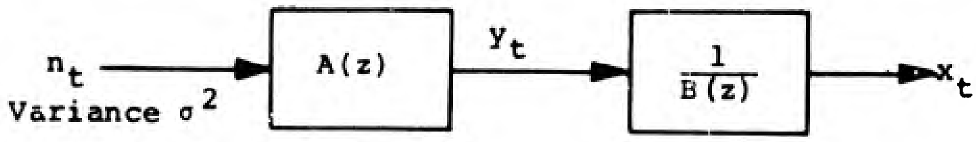


Figure 4-1. Decomposition of the system model for a process of mixed type.

Defining

$$S_{yy}(z) = \sigma^2 A(z)A(z^{-1}) \tag{4-1}$$

and

$$S_{rr}(z) = \frac{1}{B(z)B(z^{-1})} \tag{4-2}$$

the sampled spectrum for x_t can be written

$$S_{xx}(z) = S_{yy}(z)S_{rr}(z) \tag{4-3}$$

where $S_{yy}(z)$ and $S_{rr}(z)$ are the spectra of a moving average and an autoregression, respectively.

If the orders of A(z) and B(z) are known a priori, the separate identification of the two factors of the spectrum of x_t is possible due to a scheme proposed by Hsia and Landgrebe[35]. This method, fundamentally different from the iterative approaches of Tretter and Steiglitz[21] and Zetterberg[25], yields estimates of the set $\{b_j\}$ based only on the mean lagged products of x_t. The present section gives an independent derivation of Hsia and Landgrebe's result and includes an analysis of the statistical errors inherent in the estimates, a problem which is not considered in the earlier work.

The present approach can most easily be developed as a logical consequence of the correlation properties of a process of mixed type as derived in Chapter II. Equation (2-50) of that section was

$$\sum_{j=0}^{L} b_j \phi_{xx}(k-j) = \sum_{j=0}^{K} a_j \phi_{nx}(j-k), \quad k = 0,1,2,\ldots,K$$
$$\sum_{j=0}^{L} b_j \phi_{xx}(k-j) = 0 \quad , k>K \tag{4-4}$$

Of most interest in the present context is the second of these, which involves only the autoregressive coefficients and the correlation function $\phi_{xx}(j)$. In particular, consider the L-equation system

$$\sum_{j=0}^{L} b_j \phi_{xx}(k-j) = 0, \; k = K+1, \; K+2, \; \ldots\ldots, \; K+L \tag{4-5}$$

or

$$\sum_{j=1}^{L} b_j \phi_{xx}(k-j) = -\phi_{xx}(k) \quad , k = K+1, \; K+2, \; \ldots, \; K+L \tag{4-6}$$

since $b_0 = 1$. If the correlation sequence $\phi_{xx}(j)$ is known for arguments through $j = K+L$, these equations can be solved for the L unknown $\{b_j\}$. If

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_L \end{bmatrix} \qquad \phi_{xx} = \begin{bmatrix} \phi_{xx}(K+1) \\ \phi_{xx}(K+2) \\ \cdot \\ \cdot \\ \cdot \\ \phi_{xx}(K+L) \end{bmatrix} \tag{4-7a}$$

and

$$\Phi'_{xx} = \begin{bmatrix} \phi_{xx}(K) & \phi_{xx}(K-1) & \phi_{xx}(K-2) & \cdots & \phi_{xx}(K-L+1) \\ \phi_{xx}(K+1) & \phi_{xx}(K) & \phi_{xx}(K-1) & \cdots & \phi_{xx}(K-L+2) \\ \phi_{xx}(K+2) & \phi_{xx}(K+1) & \phi_{xx}(K) & \cdots & \vdots \\ \vdots & & & & \vdots \\ \phi_{xx}(K+L-2) & \cdots & \phi_{xx}(K+1) & \phi_{xx}(K) & \phi_{xx}(K-1) \\ \phi_{xx}(K+L-1) & \phi_{xx}(K+L-2) & \cdots & \phi_{xx}(K+1) & \phi_{xx}(K) \end{bmatrix}$$

then

$$\Phi'_{xx} b = -\phi'_{xx}$$

$$\text{or} \qquad (4\text{-}8)$$

$$b = -\Phi'^{-1}_{xx} \phi'_{xx}$$

which is identical in form to the equation for all-pole estimation. In an actual spectral measurement, the correlation sequence is replaced by the appropriate set of mean lagged products, and equation (4-8) becomes

$$\hat{b} = -\hat{\Phi}'^{-1}_{xx} \hat{\phi}'_{xx} \qquad (4\text{-}9)$$

The analysis of the statistical error inherent in the estimation of the autoregressive coefficients roughly parallels that of the last chapter for all-pole estimation. The actual and estimated coefficients obey the equations

$$\sum_{j=0}^{L} b_j \, \phi_{xx}(k-j) = 0 \qquad (4\text{-}10a)$$

$$k = K+1, K+2, \ldots, K+L$$

$$\sum_{j=0}^{L} \hat{b}_j \, \hat{\phi}_{xx}(k-j) = 0 \qquad (4\text{-}10b)$$

Defining

$$u_k \equiv \sum_{j=0}^{L} b_j \, \hat{\phi}_{xx}(k-j) \;, \; k = K+1, K+2, \ldots, K+L \qquad (4\text{-}11)$$

the systems of equation (4-10) can be combined to yield

$$\sum_{j=1}^{L} (\hat{b}_j - b_j) \, \hat{\phi}_{xx}(k-j) = -u_k \;, \; k = K+1, K+2, \ldots, K+L \qquad (4\text{-}12)$$

Now defining the error vector and a vector of the $\{u_j\}$

$$e = \begin{bmatrix} \hat{b}_1 - b_1 \\ \hat{b}_2 - b_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{b}_L - b_L \end{bmatrix} \; ; \quad u = \begin{bmatrix} u_{K+1} \\ u_{K+2} \\ \cdot \\ \cdot \\ \cdot \\ u_{K+L} \end{bmatrix} \qquad (4\text{-}13)$$

the system of equation (4-12) can be written as

$$\hat{\phi}'_{xx} \, e = -u \qquad (4\text{-}14)$$

Now using the mean lagged product definition, equation (4-11) yields

$$u_k = \frac{1}{N} \sum_{j=0}^{L} b_j \sum_{m=1}^{N} x_m x_{m+k-j} \quad , \; k = K+1,\; K+2,\; \ldots,\; K+L \tag{4-15}$$

By manipulating the indices and reversing the summation order, this becomes

$$u_k = \frac{1}{N} \sum_{m=k+1}^{N+k} x_{m-k} \sum_{j=0}^{L} b_j x_{m-j} \quad , \; k = K+1,\; K+2, \ldots, K+L \tag{4-16}$$

Recalling the equation that defines the mixed-type process,

$$\sum_{j=0}^{L} b_j x_{k-j} = \sum_{j=0}^{K} a_j n_{k-j} \tag{4-17}$$

the inner summation of equation (4-16) can be replaced by the right hand side of (4-17) so that

$$u_k = \frac{1}{N} \sum_{m=k+1}^{N+k} x_{m-k} \sum_{j=0}^{K} a_j n_{m-j} \quad , \; k = K+1,\; K+2, \ldots, K+L \tag{4-18}$$

From Figure 4-1, however, it is apparent that

$$\sum_{j=0}^{K} a_j n_{m-j} = y_m \tag{4-19}$$

and hence

$$u_k = \frac{1}{N} \sum_{m=k+1}^{N+k} x_{m-k} y_m \quad , \; k = K+1,\; K+2, \ldots, K+L \tag{4-20}$$

Taking the expected value,

$$E[u_k] = \phi_{xy}(k) \, , \; k = K+1,\; K+2, \ldots, K+L \tag{4-21}$$

The process x_t can be written in autoregressive form as

$$x_t = y_t - \sum_{j=1}^{L} b_j x_{t-j} \tag{4-22}$$

and by Wold's Decomposition can be considered to be a moving average over the past of y_t. Although y_t is not an uncorrelated sequence, the fact that it is itself a K^{th} order moving average over n_t implies that it will not be correlated beyond K lags. Since x_t is computed from the past of y_t, it will thus be uncorrelated with future values of y_t more than K lags distant. That is

$$\phi_{xy}(k) = 0 \quad , \; k>K \tag{4-23}$$

and thus

$$E[u_k] = 0, \; k = K+1, K+2, \ldots, K+L \tag{4-24}$$

Now since the $\{u_k\}$ are zero mean, the covariance between u_j and u_k can be written from equation (4-20) as

$$\mathrm{Cov}[u_j, u_k] = \frac{1}{N^2} \sum_{\ell=j+1}^{N+j} \sum_{m=k+1}^{N+k} E[x_{\ell-j} y_\ell x_{m-k} y_m] \qquad j,k = K+1, \ldots, K+L \tag{4-25}$$

At this point, it is necessary to assume that the processes in question are Gaussian so that the expected value can be expanded as

$$E[x_{\ell-j} y_\ell x_{m-k} y_m] = \phi_{xy}(j)\phi_{xy}(k) + \phi_{xx}(\ell-j-m+k)\phi_{yy}(\ell-m) + \phi_{xy}(m-\ell+j)\phi_{xy}(\ell-m+k) \tag{4-26}$$

Given the ranges of j and k involved, previous considerations show that the first product must be zero. Therefore,

$$\mathrm{Cov}[u_j,u_k] = \frac{1}{N^2}\sum_{\ell=j+1}^{N+j}\sum_{m=k+1}^{N+k}\phi_{xx}(\ell-j-m+k)\,\phi_{yy}(\ell-m)$$

$$+\frac{1}{N^2}\sum_{\ell=j+1}^{N+j}\sum_{m=k+1}^{N+k}\phi_{xy}(m-\ell+j)\,\phi_{xy}(\ell-m+k) \qquad (4\text{-}27)$$

Concentrating on the second on these summations, it is noted that a given summation term is non-zero if simultaneously

$$(m-\ell+j) \le K \quad \text{and} \quad (\ell-m+k) \le K \qquad (4\text{-}28)$$

This requires that

$$k-K \le m-\ell \le K-j \qquad (4\text{-}29)$$

The maximum value of the right hand side of equation (4-29) occurs when j attains its minimum, which is K + 1. Similarly, the left hand side is a minimum when k is minimum, or K + 1. For a given term in the second sum of (4-27) to be non-zero, the indices must obey the inequality

$$1 \le m-\ell \le -1 \qquad (4\text{-}30)$$

which is a contradiction. Thus, the entire second sum must vanish. The covariance can now be expressed as

$$\mathrm{Cov}[u_j,u_k] = \frac{1}{N^2}\sum_{\ell=j+1}^{N+j}\sum_{m=k+1}^{N+k}\phi_{xx}(\ell-j-m+k)\,\phi_{yy}(\ell-m) \qquad (4\text{-}31)$$

Since the indices ℓ and m appear only as the difference (ℓ - m) and because $\phi_{yy}(j)$ is identically zero for j>K, this can be written

$$\mathrm{Cov}[u_j, u_k] = \frac{1}{N^2} \sum_{m=-K}^{K} (N - |m|)\,\phi_{xx}(m-j+k)\,\phi_{yy}(m) \qquad (4\text{-}32)$$

$$j,k = K+1,\ K+2,\ \ldots,\ K+L$$

Since the number of samples can be expected to be much larger than the order of the moving average part, i.e. N>>K, a very close approximation is

$$\mathrm{Cov}[u_j, u_k] \doteq \frac{1}{N} \sum_{m=-K}^{K} \phi_{xx}(m-j+k)\,\phi_{yy}(m) \qquad (4\text{-}33)$$

Returning now to the main thrust of the argument, equation (4-14) shows that the error vector is given by

$$e = -\hat{\phi}'^{-1}_{xx} u \qquad (4\text{-}34)$$

Again, as the number of samples grows large, the mean lagged products approach the theoretical correlations in probability:

$$\operatorname*{plim}_{N \to \infty} \hat{\phi}'_{xx} = \phi'_{xx} \qquad (4\text{-}35)$$

and hence

$$\operatorname*{plim}_{N \to \infty} e = -\phi'^{-1}_{xx} u \qquad (4\text{-}36)$$

Since the $\{u_j\}$ have a zero mean, in the limit the errors will have zero mean also, and the $\{\hat{b}_j\}$ become unbiased estimates of the actual autoregressive coefficients. Furthermore, the covariance matrix for the error becomes

$$\Phi_{ee} = \Phi_{xx}'^{-1} \Phi_{uu} (\Phi_{xx}'^{-1})^T \tag{4-38}$$

Since Φ_{xx}' is non-symmetrical, no further simplification can be performed. The fact that every element of Φ_{uu} is multiplied by the factor 1/N shows that the $\{\hat{b}_j\}$ are also consistent estimates for the $\{b_j\}$ since the error variances tend to zero as the number of samples increases.

Strictly speaking, the variability analysis carried through above applies only to the estimation of Gaussian processes. In practical situations, however, it will usually be true that the results apply with sufficient accuracy for non-Gaussian processes as well.

In an actual estimation problem, one might choose, more or less arbitrarily, a given number of samples for an initial analysis. Having estimated the $\{b_j\}$ on the basis of mean lagged products, the coefficients can be used to formulate a digital filter to yield the process y_t. $\phi_{yy}(j)$ can then be estimated on the basis of mean lagged products of the filtered process and combined with $\phi_{xx}(j)$, using equation (4-33), to provide the elements of the matrix Φ_{uu}. The standard deviations of the estimates can then be determined as a means for judging whether the number of samples chosen was sufficient for the required accuracy.

B. Spectral Error Due to Coefficient Estimation Errors

Depending upon the use to which the analysis will be put, the accuracy criteria which determine the sample size may refer to the coefficient estimates $\{\hat{b}_j\}$ or to the spectral estimates prepared on the basis of these. For the latter alternative, it is necessary to find a relationship between the coefficient errors treated above and the resulting errors in the spectral estimate for the process.

In terms of the coefficient errors and the actual coefficient values, the estimates are given by

$$\hat{b}_j = b_j + e_j \quad , \; j = 0, 1, 2, \ldots, L \tag{4-39}$$

The $\{e_j\}$ are zero mean, approximately Gaussian random variables with covariance structure described by equation (4-38)[1]. The autoregressive spectrum corresponding to the estimates is given by

$$\hat{S}_{rr}(z) = \frac{1}{\hat{B}(z)\hat{B}(z^{-1})} \tag{4-40}$$

where

$$\hat{B}(z) = \sum_{j=0}^{L} \hat{b}_j z^{-j} \tag{4-41}$$

Using equations (4-39) and (4-41), equation (4-40) becomes

$$\hat{S}_{rr}(z) = \frac{1}{\sum_{j=0}^{L} \sum_{k=0}^{L} (b_j + e_j)(b_k + e_k) z^{j-k}} \tag{4-42}$$

---

1. $\hat{b}_o = b_0 = 1$, and $e_0 = 0$, but the zeroth term is included for notational convenience.

Multiplying out the factors within the summation then yields

$$\hat{S}_{rr}(z) = \frac{1}{B(z)B(z^{-1}) + \sum_{j=0}^{L} \sum_{k=0}^{L} (b_j e_k + b_k e_j + e_j e_k) z^{j-k}}$$

(4-43)

$$= S_{rr}(z) \frac{1}{1 + S_{rr}(z) \sum_{j=0}^{L} \sum_{k=0}^{L} (b_j e_k + b_k e_j + e_j e_k) z^{j-k}}$$

If the coefficient errors are sufficiently small, $\hat{S}_{rr}(z)$ can be well represented by a first order approximation:

$$\hat{S}_{rr}(z) \approx S_{rr}(z) \left[1 - S_{rr}(z) \sum_{j=0}^{L} \sum_{k=0}^{L} (b_j e_k + b_k e_j + e_j e_k) z^{j-k}\right]$$

(4-44)

$$\approx S_{rr}(z) - S_{rr}^2(z) \sum_{j=0}^{L} \sum_{k=0}^{L} (b_j e_k + b_k e_j + e_j e_k) z^{j-k}$$

The second term of equation (4-44) is the approximate spectral error as a function of z. Since the $\{e_j\}$ are zero mean, the expected value of $\hat{S}_{rr}(z)$ is given by

$$E[\hat{S}_{rr}(z)] \approx S_{rr}(z) - S_{rr}^2(z) \sum_{j=0}^{L} \sum_{k=0}^{L} \mathrm{Cov}(e_j, e_k) z^{j-k} \qquad (4\text{-}45)$$

with the second term representing a bias which vanishes in the limit of large sample size when $\mathrm{Cov}(e_j, e_k)$ approaches zero. On the unit circle, when $z = e^{i\omega T}$, equation (4-45) becomes

$$E\left[\hat{S}_{rr}(e^{i\omega T})\right] \doteq S_{rr}(e^{i\omega T}) - S_{rr}^2(e^{i\omega T})\left[\sum_{j=0}^{L} Var(e_j) + 2\sum_{k=1}^{L}\sum_{j=0}^{L-k} Cov(e_j, e_{j+k})\cos k\omega T\right] \quad (4-46)$$

To investigate the consistency of the spectral estimate, it is necessary to derive an expression for the variance of $\hat{S}_{rr}(z)$. To be sure, there is a problem in defining the variance of a complex-valued random variable, but since $\hat{S}_{rr}(z)$ is real on the unit circle - which is really the only place where it is of interest - a formal derivation using the definition appropriate to real variables will yield a meaningful answer on that locus. Following the first order argument presented above, the spectral variance is approximated by computing the variance of the approximate spectral error:

$$Var[\hat{S}_{rr}(z)] \doteq E\left[S_{rr}^4(z)\left\{\sum_{j=0}^{L}\sum_{k=0}^{L}(b_je_k+b_ke_j+e_je_k)z^{j-k}\right\}^2\right] - S_{rr}^4(z)\left[\sum_{j=0}^{L}\sum_{k=0}^{L} Cov(e_j,e_k)z^{j-k}\right]^2 \quad (4-47)$$

After a good deal of algebra and recourse to the fourth moment properties of Gaussian random variables, the final result is

$$\mathrm{Var}[\hat{S}_{rr}(z)] \doteq E^2[\hat{S}_{rr}(z)] - S^2_{rr}(z)$$

$$+ S^4_{rr}(z) \left| \sum_{j=0}^{L} \sum_{k=0}^{L} \mathrm{Cov}(e_j, e_k) z^{j+k} \right|^2 \tag{4-48}$$

$$+ 2S^4_{rr}(z) \ \mathrm{Re}\left[B^2(z) \sum_{j=0}^{L} \sum_{k=0}^{L} \mathrm{Cov}(e_j, e_k) z^{j+k} \right]$$

Although this expression is quite difficult to evaluate, its limiting behavior is clear. As the number of samples increases, $E[\hat{S}_{rr}(z)]$ approaches $S_{rr}(z)$, and the coefficient error covariances approach zero. Thus, the expression of equation (4-48) approaches zero also, and at the same time, the first order approximation becomes more and more exact. The spectral estimate is therefore asymptotically unbiased and consistent as $N \to \infty$.

C. Separation of the Moving Average Part by Digital Filtering

It was mentioned above that a set of estimates for the $\{b_j\}$ could be used to formulate a digital filter by which the moving average process y_t appearing in the decomposition of Figure 4-1 could be generated from the original process x_t. This moving average can then be analyzed in isolation. The procedure is demonstrated in the following system diagram:

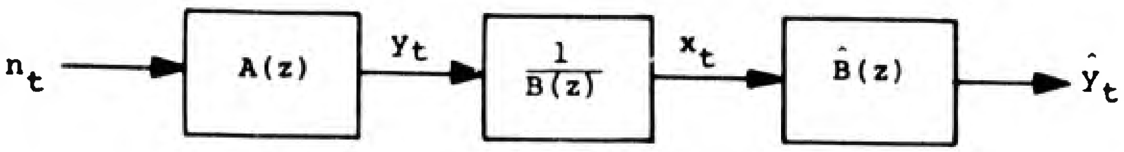


Figure 4-2. Use of the polynomial $\hat{B}(z)$ to generate an approximation to the moving average y_t.

Here $\hat{B}(z)$ is a moving average system function corresponding to the estimated set $\{\hat{b}_j\}$;

$$\hat{B}(z) = \sum_{j=0}^{L} \hat{b}_j z^{-j} \tag{4-49}$$

If the estimates are error free so that $\hat{B}(z) = B(z)$, then $\hat{y}_t = y_t$, and the resulting output can be analyzed to secure a representation or an estimate for A(z) or for $S_{yy}(z)$. In the case, however, where the coefficient estimates are contaminated by statistical error, the output $\hat{y}_t$ will only approximate y_t, and the apparent system function $\hat{A}(z)$ and spectrum will only approximate the actual functions associated with the original process. Since this type of filtering will be used repeatedly in what follows to isolate the moving average y_t in a general process of mixed type, it is of considerable importance to know the effect of the errors in the $\{\hat{b}_j\}$ on the apparent spectrum of y_t. Although it is quite difficult to obtain exact expressions, the formulation of rough criteria of acceptability for the estimation errors, as seen in their effect on $\hat{A}(z)$, is fairly straightforward.

Following equation (4-39), the system function $\hat{B}(z)$ can be expressed as

$$\hat{B}(z) = \sum_{j=0}^{L} (b_j + e_j) z^{-j}$$

$$\text{or} \tag{4-50}$$

$$\hat{B}(z) = B(z) + E(z)$$

where

$$E(z) = \sum_{j=1}^{L} e_j z^{-j} \tag{4-51}$$

since $e_0 = 0$. Thus, the system diagram for y_t can be portrayed:

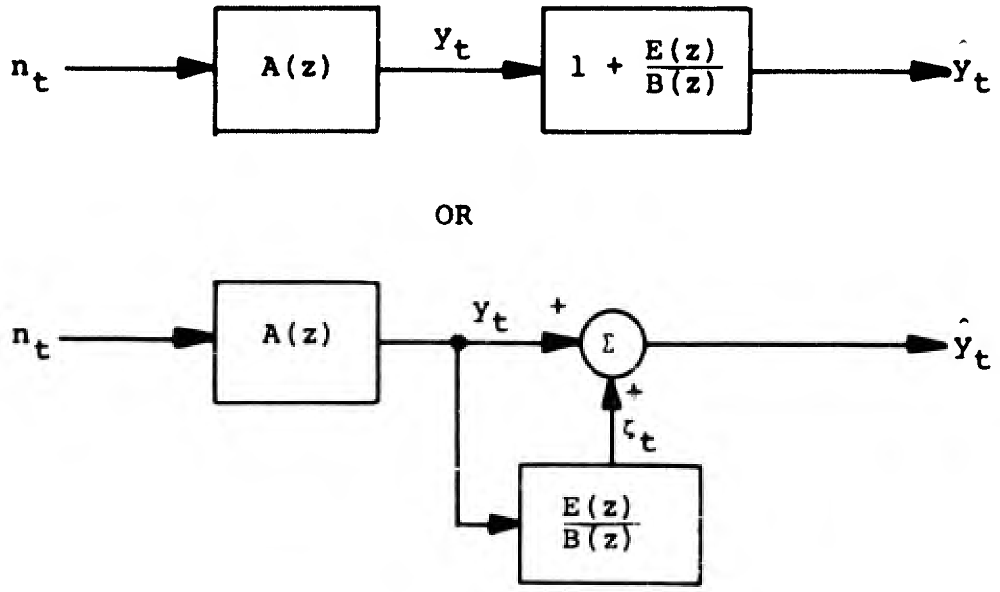


Figure 4-3. Reformulations of the system diagram for $\hat{y}_t$.

In this way, the "estimated" sequence can be interpreted as the sum of the actual sequence y_t and a sequence of errors

ζ_t, with the latter arising from the system shown here:

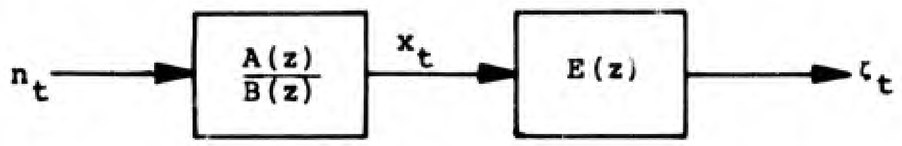


Figure 4-4. System diagram for the generation of ζ_t.

A comparison of the mean square error with the mean square value of y_t is a convenient criterion for the quality of the approximation of y_t by $\hat{y}_t$. By Figure 4-4, the error can be expressed as

$$\zeta_t = \sum_{j=1}^{L} e_j x_{t-j} \tag{4-52}$$

and the mean square follows as

$$\overline{\zeta_t^2} = \sum_{j=1}^{L} \sum_{k=1}^{L} E[e_j e_k x_{t-j} x_{t-k}] \tag{4-53}$$

It is reasonable to assume that the coefficient errors are nearly statistically independent of the individual x_t. If a different realization of the process x_t is filtered to produce $\hat{y}_t$, using a previously estimated set $\{\hat{b}_j\}$, this will certainly bo true. Assuming this independence,

$$\overline{\zeta_t^2} = \sum_{j=1}^{L} \sum_{k=1}^{L} E[e_j e_k] E[x_{t-j} x_{t-k}] \tag{4-54}$$

which by virtue of the zero mean of the $\{e_j\}$ and the definition of autocorrelation becomes

$$\overline{\zeta_t^{\,2}} = \sum_{j=1}^{L} \sum_{k=1}^{L} \mathrm{Cov}(e_j, e_k)\ \phi_{xx}(j-k) \qquad (4\text{-}55)$$

The covariance function for the $\{e_j\}$ has been treated above, and the present quantity can be computed from this information as an indication of the accuracy of the filtering.

A somewhat rougher, but computationally simple, estimate of the quality of the approximation can be gained by noting from the system diagram for $\hat{y}_t$ that it is desirable that $\left|\frac{E(z)}{B(z)}\right|^2$ or its average around the unit circle be much less than unity. Although this is fairly inconvenient to compute, a closely related figure of merit is the ratio

$$R = \frac{\sum_{j=1}^{L} \mathrm{Var}(e_j)}{\sum_{j=0}^{L} b_j^{\,2}} \qquad (4\text{-}56)$$

which can be interpreted as a comparison of the mean square properties of E(z) and B(z) as moving average filters. If R is small, $\hat{y}_t$ will closely approximate y_t.

The development of the present chapter has dealt with the estimation of the autoregressive coefficients of a mixed-type process when the numerator and denominator orders of the associated system are known a priori. It has also considered the factors involved in selecting the accuracy requirements and their relationship to the sample size over which the analysis is computed. The particular criterion selected for determining the sample size will depend upon the use to which the results will be put. Fortunately, good performance in one of these areas will carry over to the others.

D. Computational Examples of the Estimation of the Autoregressive Coefficients

Only the last three of the five standard examples are processes of mixed type for which the methods of the present chapter are applicable. With a knowledge of the orders of the moving average and autoregressive parts of the process (K and L, respectively) and after measurement of a suitable set of mean lagged products, equation (4-9) can be solved directly for the desired estimates.

1. Example # 3

As seen in Appendix A or in the computational examples of the last chapter, the autoregressive polynomial for the 3rd example is

$$B(z) = 1 + 0.853\ z^{-2}$$

so that $b_1 = 0.0$ and $b_2 = 0.853$. Furthermore, $K = 1$ and $L = 2$. Using the same set of mean lagged products (over 1,000 samples) upon which the all-pole estimates of the last chapter were based, the matrix $\hat{\phi}'_{xx}$, defined in equation (4-7b) becomes

$$\phi'_{xx} = \begin{bmatrix} \hat{\phi}_{xx}(1) & \hat{\phi}_{xx}(0) \\ \hat{\phi}_{xx}(2) & \hat{\phi}_{xx}(1) \end{bmatrix} = \begin{bmatrix} -0.52455 & 6.56377 \\ -5.60531 & -0.52455 \end{bmatrix}$$

and the column vector $\hat{\phi}'_{xx}$ is

$$\hat{\phi}'_{xx} = \begin{bmatrix} \hat{\phi}_{xx}(2) \\ \hat{\phi}_{xx}(3) \end{bmatrix} = \begin{bmatrix} -5.60531 \\ 0.48098 \end{bmatrix}$$

Thus, the coefficient estimates are

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = -\hat{\phi}'^{-1}_{xx} \, \hat{\phi}'_{xx} = \begin{bmatrix} 0.00585 \\ 0.85445 \end{bmatrix}$$

as compared to the actual values of 0.0 and 0.853.

For the error calculations, the theoretical process correlation function will be used in lieu of its estimate (the sequence of mean lagged products). As in the previous computational examples, the final result will thus represent the actual standard deviations of the estimates considered as random variables. Ultimately, equation (4-38) must be evaluated for the present case. ϕ'^{-1}_{xx} and $(\phi'^{-1}_{xx})^T$ are easily obtained, but Φ_{uu} requires somewhat more work. The elements of this matrix are found from equation (4-33), where $\phi_{yy}(m)$ represents the theoretical autocorrelation sequence associated with the process y_t. This is a moving average with input noise variance unity and system function

$$A(z) = 1 - 0.859z^{-1}$$

The autocorrelation sequence is found to be

$$\left\{\phi_{yy}(k)\right\}_{k=0}^{\infty} = 1.73788, -0.85900, 0, 0, 0, \ldots.$$

Using this in equation (4-33) yields the matrix

$$\Phi_{uu} = \frac{1}{N}\begin{bmatrix} 11.88430 & -1.61127 \\ -1.61127 & 11.88430 \end{bmatrix}$$

And now from equation (4-38),

$$\Phi_{ee} = \Phi_{xx}'^{-1}\,\Phi_{uu}(\Phi_{xx}'^{-1})^T = \frac{1}{N}\begin{bmatrix} 0.39062 & 0.04979 \\ 0.04979 & 0.29713 \end{bmatrix}$$

The standard deviations for the errors in the present case follow directly when N is set equal to 1,000. The results are summed up in the following table:

Table 4-1

Autoregressive Coefficient Estimates for Example # 3

j	$\hat{b}_j$	b_j	$e_j=\hat{b}_j-b_j$	$\sigma(e_j)$
1	0.00585	0.000	+0.00585	0.01975
2	0.85445	0.853	+0.00145	0.01717

It is seen that the estimates are well within one standard deviation of the actual values.

It is of interest to compare the spectra corresponding to the actual and estimated autoregressive polynomials. Forming

$$S_{rr}(z) = \frac{1}{B(z)B(z^{-1})}$$

and

$$\hat{S}_{rr}(z) = \frac{1}{\hat{B}(z)\hat{B}(z^{-1})} \quad ,$$

a comparison for $z = e^{i\omega T}$ is shown in Figure 4-5. As predicted by the error analysis for the $\{\hat{b}_j\}$, the agreement is very good.

As described in Section C, the estimated polynomial $\hat{B}(z)$ can be used as a digital filter to obtain a process associated solely with the moving average polynomial A(z). From Figure 4-2, we see that the system function relating n_t and $\hat{y}_t$ is given by $\frac{A(z)\hat{B}(z)}{B(z)}$. In the present example, this can be evaluated as

$$\frac{A(z)\hat{B}(z)}{B(z)} = \frac{(1 - 0.859z^{-1})(1 + 0.00585z^{-1} + 0.85445z^{-2})}{1 + 0.853z^{-2}}$$

$$= \frac{1 - 0.85300z^{-1} + 0.84952z^{-2} - 0.73397z^{-3}}{1 + 0.853z^{-2}}$$

The spectrum associated with the filter output $\hat{y}_t$ should now

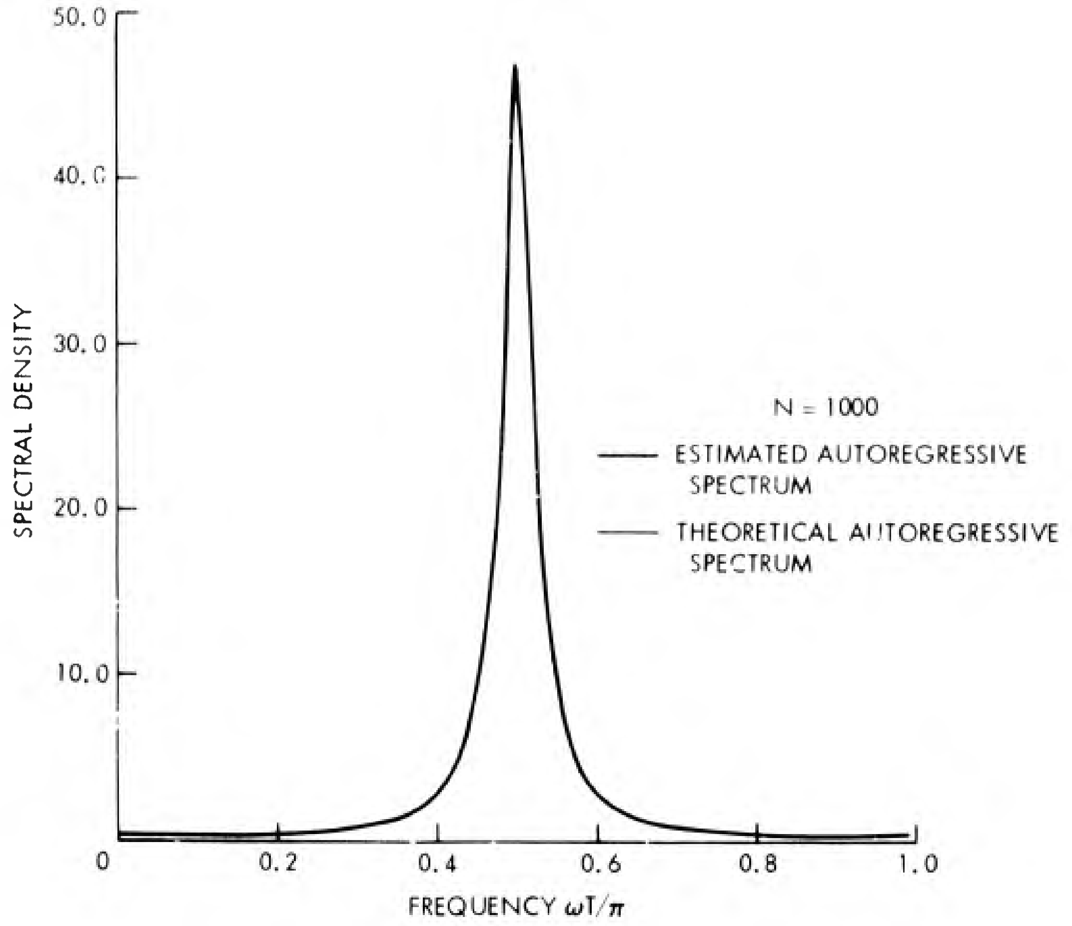


FIG. 4-5 ESTIMATE OF THE AUTOREGRESSIVE SPECTRUM OF EXAMPLE #3 BASED ON 1,000 SAMPLES

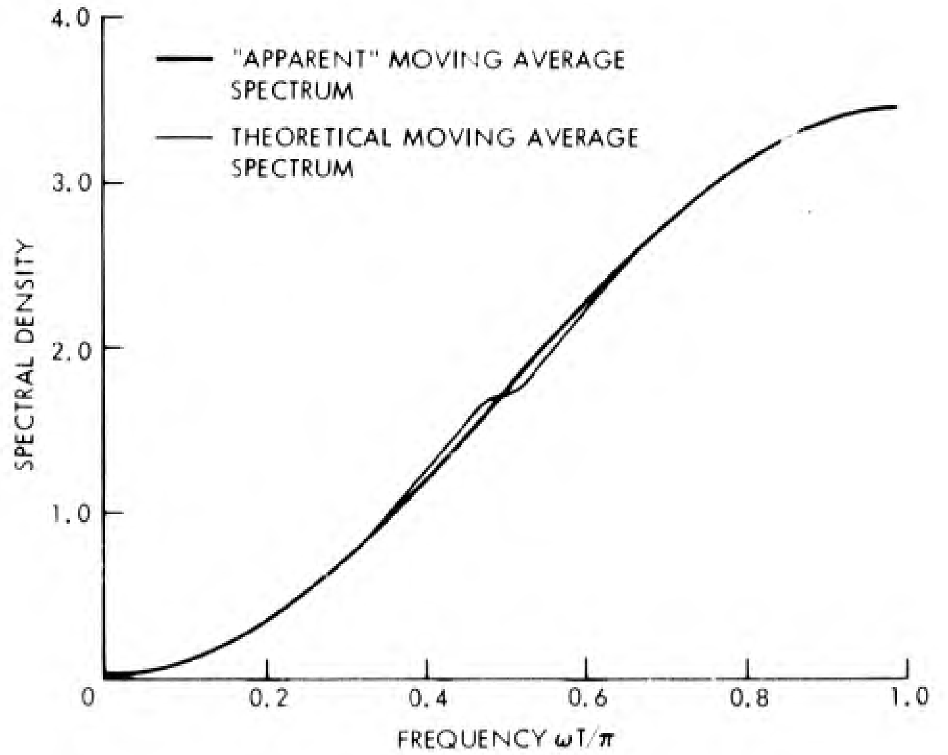


FIG. 4-6 A COMPARISON OF THE "APPARENT" AND THEORETICAL MOVING AVERAGE SPECTRA FOR EXAMPLE #3

approximate $\sigma^2 A(z)A(z^{-1})$, that associated with the moving average part of the process. A comparison is drawn in Figure 4-6, and again, the agreement is very good. As shown above, the filtered output $\hat{y}_t$ can be interpreted as the sum of the true moving average process y_t and an error ζ_t. The mean square value of this error signal is given by equation (4-55) and can be evaluated by knowing the covariance structure of the process x_t and that of the coefficient estimate errors. Using the numerical matrix Φ_{ee} presented above,

$$\overline{\zeta_t^{\,2}} = 0.00439$$

for the estimates based on 1,000 samples. This is very small compared to the nominal mean square value of y_t,

$$\phi_{yy}(0) = 1.73788$$

The other figure of merit derived in Section C (equation (4-56)) emerges here as

$$R = 0.00040$$

2. Example # 4

In the next example, the 4^{th} order autoregressive polynomial is

$$B(z) = 1 - 1.337z^{-1} + 1.632z^{-2} - 0.987z^{-3} + 0.660z^{-4}$$

with K = 2 and L = 4. Using the procedure detailed above, the estimate vector, based on the same mean lagged products used in the previous chapter (over 5,000 samples) becomes

$$\hat{b} = \begin{bmatrix} -1.33564 \\ 1.62015 \\ -0.97754 \\ 0.65017 \end{bmatrix}$$

with covariance matrix

$$\Phi_{ee} = \frac{1}{N} \begin{bmatrix} 4.21371 & -8.30061 & 4.95810 & -3.76785 \\ -8.30061 & 25.01910 & -13.43440 & 13.69301 \\ 4.95810 & -13.43440 & 7.85709 & -7.24302 \\ -3.76785 & 13.69301 & -7.24302 & 8.16127 \end{bmatrix}$$

When Φ_{ee} is evaluated with N = 5,000, the error analysis can be summarized in the following table:

Table 4-2

Autoregressive Coefficient Estimates for Example # 4

j	$\hat{b}_j$	b_j	$e_j = \hat{b}_j - b_j$	$\sigma(e_j)$
1	-1.33564	-1.33700	+0.00136	0.02901
2	1.62015	1.63200	-0.01185	0.07072
3	-0.97754	-0.98700	+0.00946	0.03963
4	0.65017	0.66000	-0.00983	0.04043

Again the estimates are well within one standard deviation of the actual values of the coefficients, and accordingly, a comparison of the spectrum $\frac{1}{B(z)B(z^{-1})}$ with its estimate, as in Figure 4-7, reveals only negligible difference.

When the polynomial $\hat{B}(z)$ is used as a filter to yield an approximation to the moving average process y_t, the system function associated with $\hat{y}_t$ becomes

$$\frac{A(z)\hat{B}(z)}{B(z)} = \frac{\left[\begin{array}{c} 1 - 3.11914z^{-1} + 4.79526z^{-2} - 4.92624z^{-3} + 3.67839z^{-4} \\ -1.93476z^{-5} + 0.51558z^{-6} \end{array}\right]}{1 - 1.337z^{-1} + 1.632z^{-2} - 0.987z^{-3} + 0.660z^{-4}}$$

and the corresponding spectrum is compared with $A(z)A(z^{-1})$ in Figure 4-8. The mean square value of the error signal ζ_t defined above can be found from equation (4-55) to be

$$\overline{\zeta_t^{\,2}} = 0.0141$$

which is small compared to the nominal mean square value of y_t, which is 4.80972. In the present case, the ratio R becomes 0.0132.

3. Example # 5

The last example again has K = 2 and L = 4 with

$$B(z) = 1 - 1.352z^{-1} + 1.338z^{-2} - 0.662z^{-3} + 0.240z^{-4}$$

With N = 5,000 samples (again, the same sample sequence as that used previously), the $\{\hat{b}_j\}$ become

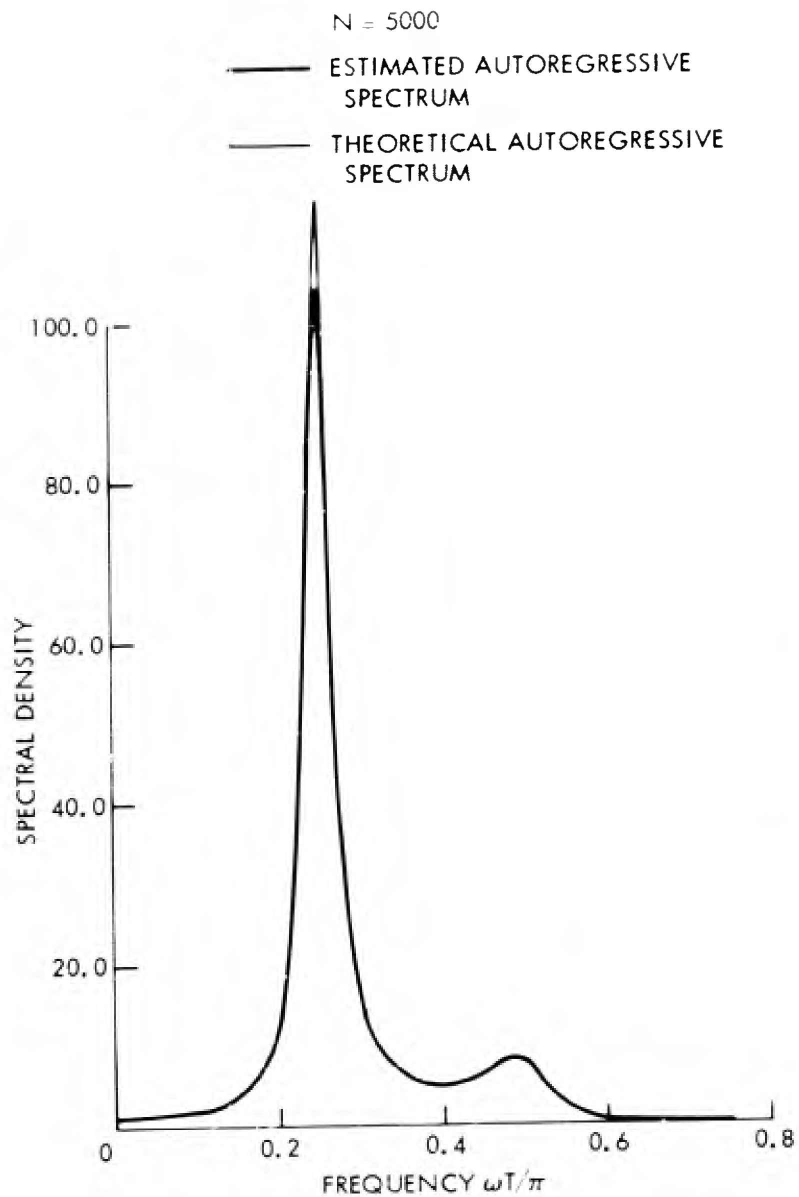


FIG. 4-7 ESTIMATE OF THE AUTOREGRESSIVE SPECTRUM OF EXAMPLE #4 BASED ON 5,000 SAMPLES

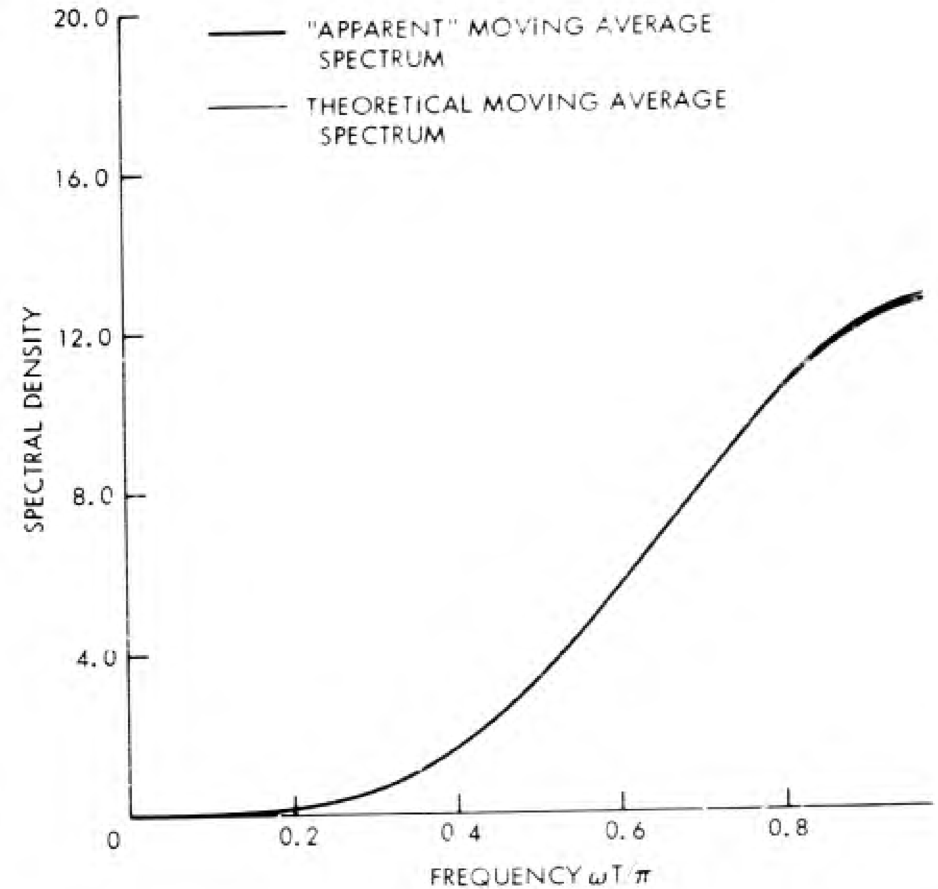


FIG. 4-8 A COMPARISON OF THE "APPARENT" AND THEORETICAL MOVING AVERAGE SPECTRA FOR EXAMPLE #4

$$\hat{b} = \begin{bmatrix} -1.49862 \\ 1.45736 \\ -0.76938 \\ 0.27201 \end{bmatrix}$$

with theoretical covariance matrix

$$\Phi_{ee} = \frac{1}{N} \begin{bmatrix} 2810.35 & -2729.27 & 2263.67 & -850.11 \\ -2729.27 & 2659.97 & -2202.27 & 833.49 \\ 2263.67 & -2202.27 & 1825.74 & -688.21 \\ -850.11 & 833.49 & -688.21 & 264.29 \end{bmatrix}$$

Immediately, the size of the entries in the matrix indicates that a very large sample size will be needed to provide the estimates with substantial reliability. In the present case with N = 5,000, the results are as follows:

Table 4-3

Autoregressive Coefficient Estimates for Example # 5

j	$\hat{b}_j$	b_j	$e_j = \hat{b}_j - b_j$	$\sigma(e_j)$
1	-1.49862	-1.35200	-0.14662	0.74901
2	1.45736	1.33800	+0.11936	0.72933
3	-0.76938	-0.66200	-0.10738	0.60411
4	0.27201	0.24000	+0.03201	0.23041

Although the estimates are again within one standard deviation of the actual coefficients, the standard deviations are so large that the estimates would certainly be suspect in a practical situation. From the elements of Φ_{ee}, it can be seen that a sample size on the order of 300,000 is required to reduce the standard deviation of the first two estimates to a reasonable 0.1.

At any rate, the spectral estimate $\hat{S}_{rr}(z)$ based on 5,000 samples is compared with the actual autoregressive spectrum in Figure 4-9. The effect of the substantial coefficient errors is readily apparent. If the present set of estimates is used in a digital filter to obtain y_t, the resulting spectrum of the approximation is compared with the actual spectrum in Figure 4-10. The sizable divergence mirrors the imprecision of the coefficient estimates, as does the mean square value of the error signal ζ_t: 1.671, in comparison with the nominal mean square value of y_t, 4.6837.

It is interesting that there exists so large a difference in the standard deviations of the $\{\hat{b}_j\}$ in Examples 4 and 5. Both processes are of the same general type with only relatively small differences in the respective regression coefficients. Moreover, the variability of their all-pole estimates is about the same. This seems to indicate a far greater sensitivity to the detailed correlation structure of the process in the present procedure than in all-pole estimation. Of course, only a far more detailed analysis and a larger set of examples will confirm such a conjecture.

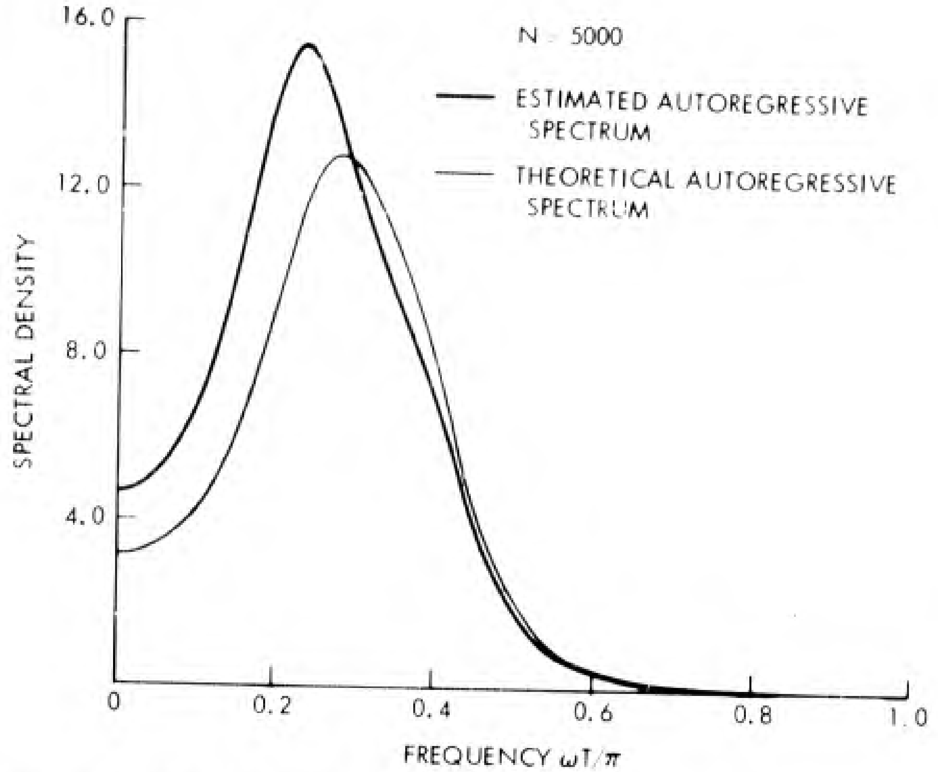


FIG. 4-9 ESTIMATE OF THE AUTOREGRESSIVE SPECTRUM OF EXAMPLE #5, BASED ON 5,000 SAMPLES

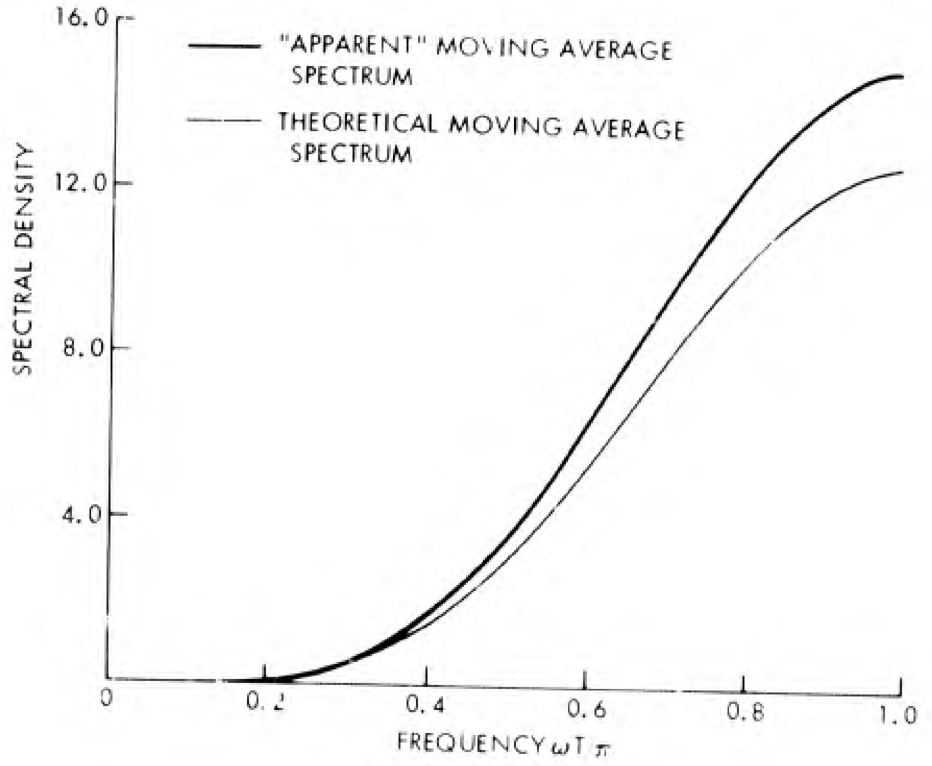


FIG. 4-10 A COMPARISON OF THE "APPARENT" AND THEORETICAL MOVING AVERAGE SPECTRA FOR EXAMPLE #5

CHAPTER V

SPECTRAL ESTIMATION FOR MOVING AVERAGE PROCESSES

By using the methods described in the last chapter, it is possible to estimate the autoregressive coefficients of any mixed-type process for which the orders of the polynomials A(z) and B(z) are known beforehand. Using these coefficients, a digital filter can be devised for processing x_t to yield the moving average y_t (or at least an approximation to it) generated by the following system:

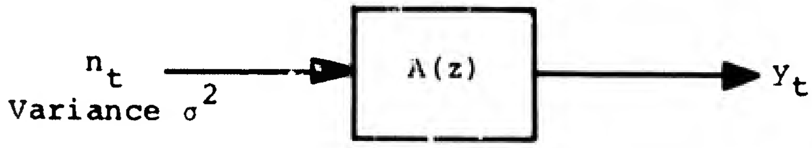


Figure 5-1. Generation of the moving average y_t.

In accordance with the decomposition of the original process described in the preceding chapter, the spectral analysis of x_t can now be completed by using a sample of y_t to determine either the set $\{a_j\}$ or the associated spectrum $\sigma^2 A(z) A(z^{-1})$.

We are thus led to study in general the procedures available for analyzing spectra of processes of the moving average type. The problem is considerably more vexing than that of the autoregressive spectrum, for which a straightforward estimation scheme is available. As noted in the Introduction, moving average estimates have not been extensively treated in the literature, and in fact no single entirely satisfactory method exists. Here, several approaches will be described and compared.

A. All-Pole Estimation for the Moving Average Part

In view of Chapter III, an obvious way of dealing with the moving average process y_t is to form an all-pole representation that can be used in combination with the previously estimated autoregressive coefficients to yield an all-pole estimate for the mixed-type spectrum as a whole. Considering first the noiseless case where B(z) is known without error, the process y_t is modeled by the system shown here:

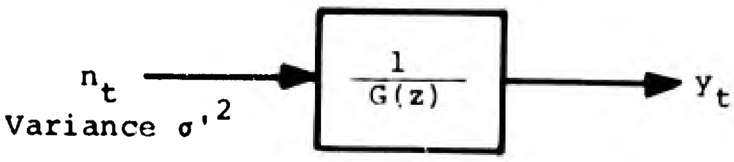


Figure 5-2. All-pole approximation for y_t.

The polynomial G(z) and the constant σ' are derived from a p^{th} order all-pole approximation for the process y_t based on the theoretical autocorrelation function. The spectrum for x_t is then approximated as

$$S'_{xx}(z) = \frac{\sigma'^2}{B(z)B(z^{-1})G(z)G(z^{-1})} \tag{5-1}$$

which is an M^{th} order all-pole function, where

$$M = P + L \tag{5-2}$$

since L is the order of the autoregressive part. Hence, $S'_{xx}(z)$ can be written as

$$S'_{xx}(z) = \frac{\sigma'^2}{F(z)F(z^{-1})} \tag{5-3}$$

with

$$F(z) = B(z)G(z) = \sum_{j=0}^{M} f_j z^{-j} \;, \; f_0 = 1 \tag{5-4}$$

This M^{th} order autoregressive model for x_t will in general not be the same as the M^{th} order representation derived directly from the autocorrelation sequence $\phi_{xx}(j)$. Thus, it will be sub-optimum in terms of the criterion described in Chapter III for formulating such approximations, and unless there is some external reason for identifying the autoregressive part separately, one is better advised, for a given model order, to compute an all-pole estimate of that order directly from x_t. Actually, in situations where the computational facilities at hand are limited in the sense that an upper bound is placed on the order of the systems to be solved, there may be an advantage to the present procedure.

The separate parts of such a calculation can be combined into a model whose order exceeds the computational maximum and which provides greater accuracy overall than a direct all-pole approximation of the highest possible order.

When the set of mean lagged products of x_t and y_t are used as a basis for the estimates, the result is expressed as

$$\hat{S}_{xx}(z) = \frac{\hat{\sigma}^2}{\hat{B}(z)\hat{B}(z^{-1})\hat{G}(z)\hat{G}(z^{-1})} \qquad (5\text{-}5)$$

and the errors in the estimates of the $\{g_j\}$ can be determined using the formulas derived in the latter part of Chapter III. Their effect on the estimate of the spectrum as a whole can then be approached by using the results of Section B of the last chapter. Because of the inherent inferiority of the present procedure to direct all-pole estimation of x_t, it will not be treated further here. Computational examples, however, will be presented later in this chapter.

B. Wold's Method

In the all-pole estimation of a moving average, no attempt is made to identify the individual moving average coefficients, and the ultimate spectral representation is an approximation quite different in functional form from the actual spectrum. If it is desired to represent the spectrum in its actual form, it is necessary to procure estimates for the $\{a_j\}$ themselves. The following method, which provides separate moving average coefficient estimates, is due to Wold[19], and has been reproduced by Hsia and Landgrebe[35].

The spectrum of the moving average y_t is given by

$$S_{yy}(z) = \sigma^2 A(z)A(z^{-1}) \tag{5-6}$$

where

$$A(z) = \sum_{j=0}^{K} a_j z^{-j}, \quad a_0 = 1 \tag{5-7}$$

A(z), in turn, is written in terms of its roots as a product of factors:

$$A(z) = \prod_{j=1}^{K} (1 - u_j/z) \tag{5-8}$$

where the $\{u_j\}$ are the roots of the polynomial A(z). By inspection, it is seen that if u_j is a root of A(z), then $1/u_j$ must be a root of $A(z^{-1})$. Equation (5-6) therefore implies immediately that for every root u_j of the sampled power spectrum, there exists another root given by $1/u_j$. One of these will lie inside the unit circle, and the other will be outside.

The spectrum can also be expressed in terms of its correlation transform as

$$S_{yy}(z) = \sum_{j=-K}^{K} \phi_{yy}(j) z^{-j} \tag{5-9}$$

Since $\phi_{yy}(j)$ is an even function of its argument, this can be re-written as

$$S_{yy}(z) = \phi_{yy}(0) + \sum_{j=1}^{K} \phi_{yy}(j)(z^{j} + z^{-j}) \tag{5-10}$$

As an aid to factoring this expression, the auxiliary variable r is introduced:

$$r \equiv z + z^{-1} \tag{5-11}$$

Now the following relationships can be demonstrated:

$$\begin{aligned} z + z^{-1} &= r \\ z^{2} + z^{-2} &= r^{2} - 2 \\ z^{3} + z^{-3} &= r^{3} - 3r \\ z^{4} + z^{-4} &= r^{4} - 4r^{2} + 2 \\ &\text{etc.} \end{aligned} \tag{5-12}$$

These are used to write the spectrum as

$$V(r) = \sum_{j=0}^{K} v_j r^{j} \tag{5-13}$$

where the $\{v_j\}$ are functions of the correlation sequence $\phi_{yy}(j)$. As a concrete example, when K = 4,

$$\begin{aligned} v_0 &= \phi_{yy}(0) - 2\,\phi_{yy}(2) + 2\,\phi_{yy}(4) \\ v_1 &= \phi_{yy}(1) - 3\,\phi_{yy}(3) \\ v_2 &= \phi_{yy}(2) - 4\,\phi_{yy}(4) \\ v_3 &= \phi_{yy}(3) \\ v_4 &= \phi_{yy}(4) \end{aligned} \tag{5-14}$$

For each of the K roots of $V(r)$, there will be two roots of $S_{yy}(z)$, since by equation (5-11),

$$z_{j1},\ z_{j2} = \frac{r_j}{2} \pm \frac{\sqrt{r_j^2 - 4}}{2} \tag{5-15}$$

It is readily seen that

$$z_{j1} z_{j2} = 1 \tag{5-16}$$

and thus that

$$z_{j1} = 1/z_{j2} \tag{5-17}$$

which indicates as before that the roots of the spectrum are paired as reciprocals and can be identified with the $\{u_j\}$ that emerge when $A(z)$ is factored. The identification of the spectral roots is the heart of Wold's method. From the correlation sequence of the process, the polynomial $V(r)$ is formed and factored to yield the roots of the spectrum. These are then related to the system polynomial $A(z)$ through equation (5-8).

At this point, however, a difficulty appears. Without _a priori_ knowledge of the root locations of $A(z)$ with respect to the unit circle, it is impossible to determine which of the paired spectral roots should be identified with $A(z)$ and which with $A(z^{-1})$: it is completely arbitrary whether one considers u_j to be z_{j1} or z_{j2}. For every root of $V(r)$, two possibilities are generated for a root of $A(z)$, and hence there are 2^K possible polynomials $A(z)$ which will be associated with the correlation sequence $\phi_{yy}(j)$ and the spectrum $S_{yy}(z)$. Since, however, one of each of the paired roots is inside the unit circle and the other outside, only one of these 2^K polynomials will have all of its roots within the unit circle (i.e. that for which the inner root is chosen for each r_j). If it is known at the outset that A(z) has no roots outside of the unit circle, this choice is the obvious solution and yields what Wold has called the regular moving average. This process is the only one with the given spectrum and autocorrelation which can be passed through a stable autoregressive filter to recover the original noise sequence n_t.

This study has throughout assumed that A(z) has no zeros outside the unit circle, and hence the regular solution provides the correct polynomial. In practice, the mean lagged products, as computed from a sample of the process, are substituted into equation (5-14) to evaluate the coefficients of the polynomial V(r). The roots of this polynomial are successively substituted in equation (5-15), and the z_j that falls inside the unit circle is identified with u_j. The relationship between the $\{u_j\}$ and the sequence $\{a_j\}$ is found from the equation

$$\sum_{j=0}^{K} a_j z^{-j} = \prod_{j=1}^{K} (1 - u_j/z) \qquad (5\text{-}18)$$

and as an example, when K = 4,

$$
\begin{aligned}
a_1 &= -(u_1 + u_2 + u_3 + u_4) \\
a_2 &= u_1u_2 + u_1u_3 + u_1u_4 + u_2u_3 + u_2u_4 + u_3u_4 \\
a_3 &= -(u_1u_2u_3 + u_1u_2u_4 + u_1u_3u_4 + u_2u_3u_4) \\
a_4 &= u_1u_2u_3u_4
\end{aligned}
\qquad (5\text{-}19)
$$

It should be realized that the present technique involves no approximations and leads directly from a measured correlation sequence to that unique regular process corresponding to the sequence, <u>when such a process exists.</u> Unfortunately, this last condition is not always fulfilled and leads to the most severe practical difficulties that arise in applying the procedure. The problem is most concisely described by presenting the following theorem due to Wold[19]:

> A necessary and sufficient condition that there exists a moving average of order K with an autocorrelation sequence equal to $\{\phi_{yy}(j)\}_{j=-K}^{K}$ is that the auxiliary polynomial V(r) defined by equation (5-13) have no zero r_j of odd multiplicity in the real interval $-2<r_j<2$.

The proof will not be presented here, but the significance of this requirement can be grasped by noting that by equation (5-15), a value of r_j in the range $-2<r_j<2$ yields two distinct complex spectral roots, both <u>on</u> the unit circle, with no clear-cut choice between them. As Wold points out, and as will readily be appreciated by the reader should he attempt the practical application of the Wold method, the positions of the roots of the polynomial V(r) are a very sensitive functi of the correlation sequence $\phi_{yy}(j)$. Furthermore, when A(z) has roots close to the unit circle, the corresponding r_j are found to lie very close to the "forbidden region", $-2<r_j<2$. Thus, when attempting to analyze processes with influential zeros, only a very small amount of statistical error in the mean lagged products of y_t can suffice to place a root of V(r) in the region where no K^{th} order moving average exists corresponding to the apparent correlation function. In practice, this happens quite frequently, particularly whe. the sequence y_t has been derived from a mixed type process by filtering based on estimates of the autoregressive coefficients. The unavoidable errors in the latter cause the supposedly pure moving average y_t to contain a weak autoregressive part, and this aggravates the problem. Wold suggests one possible solution to this dilemma when it arises, particularly if the offending root is near one end of the excluded region. This is to replace the actual r_j by the nearer of +2 or -2, thus yielding double real spectral roots at z = -1 or +1, respectively. Although this is intuitively unappealing and places a root of A(z) on the unit circle, contrary to the original assumptions, the resulting spectrum is often a surprisingly good representation of the actual spectral function. This will be seen in the examples.

In the event that no problem of this kind arises and a set of estimates for the $\{a_j\}$ are prepared, the estimator for the input noise variance σ^2 emerges directly from the expression for the mean square value of a moving average process as

$$\hat{\sigma}^2 = \frac{\hat{\phi}_{yy}(0)}{1 + \sum_{j=1}^{K} \hat{a}_j^2} \tag{5-20}$$

For practical work, it would be desirable at this point to derive the relationships between the means and variances of the estimates $\{\hat{a}_j\}$ and the sample size N. Unfortunately, due to the mathematical complexity of the estimation procedure, its abundant non-linearity, and the lack of a general expression for the roots of V(r) when its order exceeds four, this problem has proved insurmountable, There is, however, an alternative approach which provides a measure of the quality of the spectral estimate as a whole.

The present attack is based on the premise that the Wold method - when it works - yields that spectral function uniquely associated with the measured sequence of mean lagged products by the z-transformation expression

$$\hat{S}_{yy}(z) = \sum_{j=-K}^{K} \hat{\phi}_{yy}(j) z^{-j} \tag{5-21}$$

Taking the expected value of both sides yields

$$E\left[\hat{S}_{yy}(z)\right] = \sum_{j=-K}^{K} E\left[\hat{\phi}_{yy}(j)\right] z^{-j} \tag{5-22}$$

and by equation (2-71), this becomes

$$\begin{aligned} E\left[\hat{S}_{yy}(z)\right] &= \sum_{j=-K}^{K} \phi_{yy}(j) z^{-j} \\ &= S_{yy}(z) \end{aligned} \tag{5-23}$$

Thus, Wold's technique yields an unbiased estimate of the spectrum for all z. (This assumes, of course, that $\hat{y}_t$ is exactly y_t, which can be assured by choosing a large enough sample size in estimating the $\{b_j\}$.)

With the proviso that the result is only meaningful on the unit circle, where it is real, the variance of the spectral estimate can also be derived:

$$\begin{aligned}
\text{Var}\left[\hat{S}_{yy}(z)\right] &= E\left[\hat{S}_{yy}^2(z)\right] - S_{yy}^2(z) \\
&= \sum_{j=-K}^{K} \sum_{k=-K}^{K} E\left[\hat{\phi}_{yy}(j)\hat{\phi}_{yy}(k)\right] z^{-(j+k)} - S_{yy}^2(z) \\
&= \sum_{j=-K}^{K} \sum_{k=-K}^{K} \left\{\text{Cov}\left[\hat{\phi}_{yy}(j),\hat{\phi}_{yy}(k)\right] + \phi_{yy}(j)\phi_{yy}(k)\right\} z^{-(j+k)} \\
&\quad - S_{yy}^2(z)
\end{aligned} \tag{5-24}$$

Finally,

$$\text{Var}\left[\hat{S}_{yy}(z)\right] = \sum_{j=-K}^{K} \sum_{k=-K}^{K} \text{Cov}\left[\hat{\phi}_{yy}(j),\hat{\phi}_{yy}(k)\right] z^{-(j+k)} \tag{5-25}$$

where the covariance is given by equation (2-78). To obtain a single figure of merit, the variance can be averaged around the unit circle to yield

$$\begin{aligned}
\text{Var}_{avg} &= \frac{1}{2\pi i} \oint_{|z| = 1} \text{Var}\left[\hat{S}_{yy}(z)\right] \frac{dz}{z} \\
&= \sum_{j=-K}^{K} \text{Cov}\left[\hat{\phi}_{yy}(j),\ \hat{\phi}_{yy}(-j)\right] \\
&= \sum_{j=-K}^{K} \text{Var}\left[\hat{\phi}_{yy}(j)\right]
\end{aligned} \tag{5-26}$$

Since the considerations of Chapter II have indicated that the covariance function for the mean lagged products goes to zero for increasingly large sample sizes, the variance $\hat{S}_{yy}(z)$ must also approach zero, and hence $\hat{S}_{yy}(z)$ is a consistent estimate of the spectrum. Using equation (2-80), the average variance can be shown to obey the inequality

$$\mathrm{Var}_{avg} < \frac{2}{N} \sum_{j=-K}^{K} \sum_{k=-K}^{K} \phi_{yy}^{2}(k) = \frac{2(2K+1)}{N} \sum_{k=-K}^{K} \phi_{yy}^{2}(k)$$

or

$$\mathrm{Var}_{avg} < \left(\frac{4K+2}{N}\right) \sum_{j=-K}^{K} \phi_{yy}^{2}(k) \qquad (5\text{-}27)$$

The average spectral standard deviation can now be compared with the spectral magnitude in general to judge the quality of the estimate.

To summarize, Wold's method holds great promise in theory but often displays substantial difficulty in practice. Even when the polynomial V(r) can be factored, the method is extremely vulnerable to statistical error that can render it insoluble. The treatment of Hsia and Landgrebe[35], based on this approach, is regrettably marred by several seriously misleading errors, makes no mention of the difficulties described here, and includes no error analysis for the estimate.

C. Moving Average Estimates by Spectral Inversion

As shown above, statistical errors in the mean lagged products often cause Wold's method to fail. The technique to be described now, which does not have this shortcoming, was developed as an outgrowth of the "whitening" approach to all-pole estimation found in Chapter III and yields an approximate solution to the problem of estimating the moving average coefficients. In the noiseless case, where the calculations are based on the theoretical correlation sequence of the moving average, it can be shown that the estimates converge to the actual values as the approximation order increases. Because of the important role of an auxiliary process whose spectrum is the multiplicative inverse of the spectrum of the original moving average, this method has been called the "spectral inversion technique".

Essentially, the procedure consists of deriving an all-pole representation of arbitrarily high order for the process and noting that the resulting autoregressive polynomial is approximately the multiplicative inverse of A(z). If this polynomial is used as the system function for the generation of an auxiliary moving average process, a K^{th} order all-pole representation of the latter will provide a K^{th} order polynomial which approximates the reciprocal of the autoregressive polynomial derived as an all-pole representation for the original moving average. Hence, it will approximate A(z). It was later found that the resulting estimation equations are practically identical with those derived by Durbin[36] who bases a maximum likeihood estimate for the $\{a_j\}$ on the coefficients of an autoregressive model for the original process.

Following the results of Chapter III, an M^{th} order all-pole approximation for y_t yields the polynomial D(z), with zeros within the unit circle, which best approximates the reciprocal function 1/A(z) according to the criterion defined in the development of the all-pole estimates above. In the present application, and assuming that the autocorrelation function for y_t is known without error, the estimation equations become

$$d = -\Phi_{yy}^{-1}\phi_{yy} \qquad (5\text{-}28)$$

by equation (3-33). Since Φ_{yy} is usually non-singular, a solution to equation (5-28) exists, and the result is used to form the polynomial

$$D(z) = \sum_{j=0}^{M} d_j z^{-j} \; , \; d_0 = 1 \qquad (5\text{-}29)$$

According to equation (3-45), the criterion used in deriving equation (5-28) insures that at least on the unit circle

$$D(z) \approx \frac{1}{A(z)} \qquad (5\text{-}30)$$

with the quality of the approximation increasing with the order M.

Using this same M^{th} order polynomial D(z), consider the following moving average system:

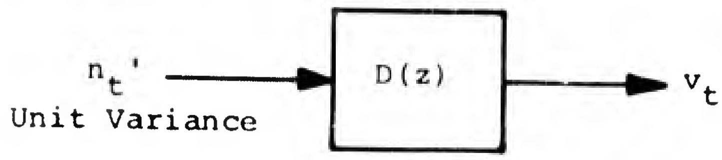


Figure 5-3. Generation of the auxiliary moving average v_t.

The spectrum of v_t is

$$S_{vv}(z) = D(z)D(z^{-1}) \approx \frac{1}{A(z)A(z^{-1})} \qquad (5\text{-}31)$$

which, except for a factor of σ^2, is approximately the inverse of the true moving average spectrum $S_{yy}(z)$. A further consequence of equation (5-30) is that an approximate autoregressive representation for v_t can be written as

$$v_t \approx n_t' - \sum_{j=1}^{K} a_j v_{t-j} \qquad (5\text{-}32)$$

and again as $M \to \infty$, this becomes exact. At this point, the technique of Chapter III is used for a second time to derive an optimum K^{th} order autoregressive representation for v_t on the basis of its correlation function. Since the coefficients of the polynomial D(z) are known, the correlation function for v_t can be expressed directly as

$$\phi_{vv}(j) = \phi_{vv}(-j) = \sum_{k=0}^{M-j} d_k d_{k+j} \qquad (5\text{-}33)$$

and the new K^{th} order estimation system becomes

$$a' = -\Phi_{vv}^{-1}\phi_{vv} \qquad (5\text{-}34)$$

This is always soluble to yield a K^{th} order polynomial $A'(z)$ with zeros within the unit circle. Because of equation (5-32), however, it is apparent that

$$a_j' \simeq a_j \; , \; j = 1, 2, \ldots, K \qquad (5\text{-}35)$$

and that $A'(z) \simeq A(z)$. The quality of the approximation becomes better and better as the representation of equation (5-32) becomes more accurate, i.e. as $M \to \infty$ and $D(z) \to 1/A(z)$.

The required expression for the input noise variance can be obtained directly from the initial autoregressive model for the moving average, or can be found using the same formula presented for the Wold method (equation (5-20)) once estimates for the $\{a_j\}$ are known.

To analyze the error inherent in the present scheme, it is again necessary to distinguish two sources of inaccuracy. In the first place, there is the "approximation" error due to the finite order of the model for the inverse spectrum and which appears in isolation in the noiseless case (i.e. when the theoretical correlation function is used). Secondly, there is the statistical error arising from the use of mean lagged products in the correlation matrices and which will be present regardless of the intermediate model order. For clarity, these are best treated separately, although in practice, the two are inextricably bound up together. Accordingly, it is convenient to define and use three separate coefficient sets:

1.) The $\{a_j\}$ - The actual moving average coefficients whose estimates are sought.

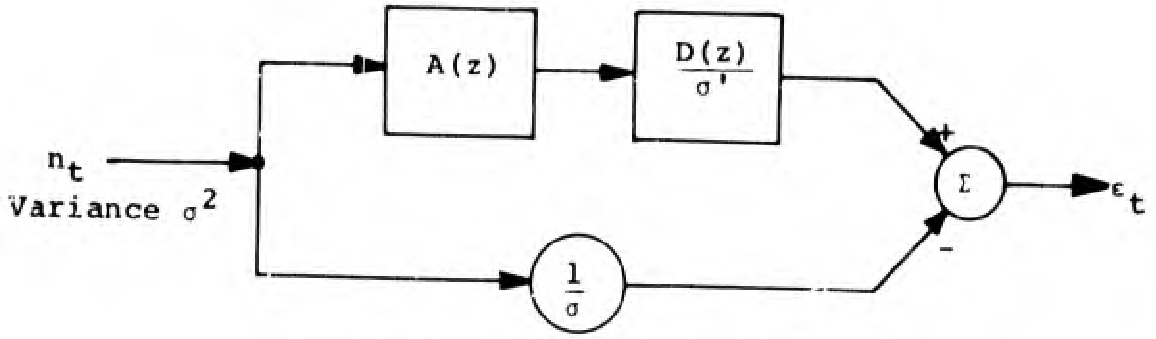
2.) The $\{a_j'\}$ - The coefficient values obtained using a noiseless M^{th} order spectral inversion.

3. The $\{\hat{a}_j\}$ - Coefficient estimates obtained using an M^{th} order spectral inversion based on mean lagged products.

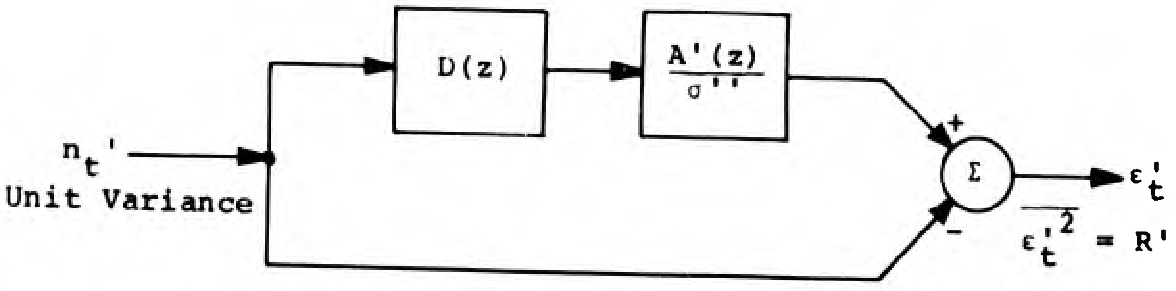
The relationship between the $\{a_j'\}$ and the $\{a_j\}$ will be treated first.

Inherent in the noiseless all-pole approximation corresponding to equation (5-28) is the system of Figure 5-4A, where A(z) is of K^{th} order, and D(z) is of M^{th} order. Knowledge of the correlation properties of y_t and of the input noise variance makes possible a choice of D(z) and σ' such that $\overline{\varepsilon_t^2}$ is a minimum. Let the resulting minimum value of $\overline{\varepsilon_t^2}$ be denoted R. With the coefficients $\{d_j\}$ obtained in this first step, the second part of the spectral inversion technique can be represented by the system of Figure 5-4B, where D(z) is as before, and A'(z) is K^{th} order. Here A'(z) and σ'' are chosen such that $\overline{\varepsilon_t'^2}$ is a minimum - and this will be denoted R'. Now it is noted that the system of Figure 5-4A can be re-drawn in the form of Figure 5-4C. There, $\sigma A(z)/\sigma'$ is a bona fide K^{th} order system and thus can yield no better error performance than the optimum K^{th} order system $A'(z)/\sigma''$ appearing in Figure 5-4B. This implies that

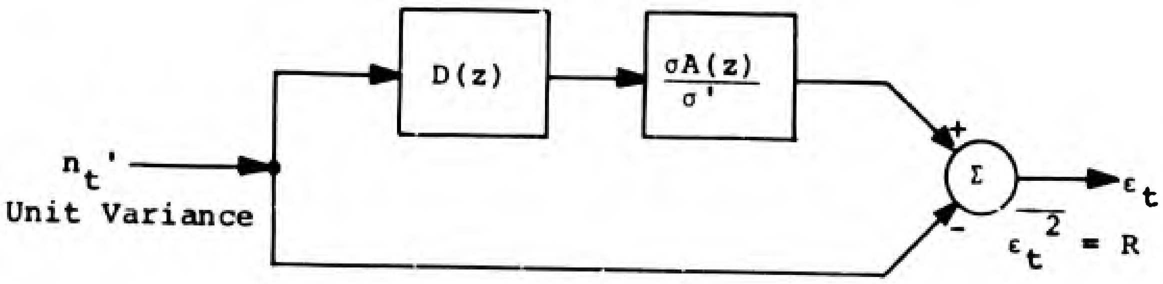
$$R' \leq R \tag{5-36}$$



A. System diagram for the first step of the spectral inversion technique.



B. System diagram for the second step of the spectral inversion technique.



C. The system diagram of figure 5-4A redrawn.

Figure 5-4. System diagrams pertaining to the analysis of the approximation error for the spectral inversion technique.

This fact is the key to bounding the approximation error in $A'(z)$.

From equation (3-43), the minimum mean square errors can be written as

$$R = \frac{1}{2\pi i} \oint \left| \frac{\sigma}{\sigma'} A(z)D(z) - 1 \right|^2 \frac{dz}{z} \tag{5-37}$$

and

$$R' = \frac{1}{2\pi i} \oint \left| \frac{1}{\sigma''} A'(z)D(z) - 1 \right|^2 \frac{dz}{z} \tag{5-38}$$

where the contour of integration is the z-plane unit circle. Now we note that

$$\left[\frac{\sigma}{\sigma'} A(z)D(z) - 1\right] - \left[\frac{1}{\sigma''} A'(z)D(z) - 1\right] = D(z)\left[\frac{\sigma}{\sigma'} A(z) - \frac{1}{\sigma''} A'(z)\right] \tag{5-39}$$

Using the triangle inequality,

$$\left| D(z)\left[\frac{\sigma}{\sigma'} A(z) - \frac{1}{\sigma''} A'(z)\right]\right| \leq \left|\frac{\sigma}{\sigma'} A(z)D(z) - 1\right| + \left|\frac{1}{\sigma''} A'(z)D(z) - 1\right| \tag{5-40}$$

Squaring both sides yields

$$|D(z)|^2 \left|\frac{\sigma}{\sigma'} A(z) - \frac{1}{\sigma''} A'(z)\right|^2 \leq \left|\frac{\sigma}{\sigma'} A(z)D(z) - 1\right|^2 + \left|\frac{1}{\sigma''} A'(z)D(z) - 1\right|^2$$

$$+ 2\left|\frac{\sigma}{\sigma'} A(z)D(z) - 1\right| \left|\frac{1}{\sigma'} A'(z)D(z) - 1\right| \tag{5-41}$$

Now performing a contour integration of both sides of equation (5-41) around the unit circle yields

$$\frac{1}{2\pi i}\oint\left|D(z)\right|^2\left|\frac{\sigma}{\sigma'}A(z) - \frac{1}{\sigma''}A'(z)\right|^2\frac{dz}{z} \tag{5-42}$$

$$\leq R + R' + \frac{2}{2\pi i}\oint\left|\frac{\sigma}{\sigma'}A(z)D(z)-1\right|\left|\frac{1}{\sigma''}A'(z)D(z)-1\right|\frac{dz}{z}$$

Now consider the integral

$$\frac{1}{2\pi i}\oint\left\{\left|\frac{\sigma}{\sigma'}A(z)D(z)-1\right| - \left|\frac{1}{\sigma''}A'(z)D(z)-1\right|\right\}^2\frac{dz}{z} \geq 0 \tag{5-43}$$

This indicates that

$$\frac{1}{2\pi i}\oint\left|\frac{\sigma}{\sigma'}A(z)D(z)-1\right|^2\frac{dz}{z} + \frac{1}{2\pi i}\oint\left|\frac{1}{\sigma''}A'(z)D(z)-1\right|^2\frac{dz}{z} \tag{5-44}$$

$$\geq \frac{2}{2\pi i}\oint\left|\frac{\sigma}{\sigma'}A(z)D(z)-1\right|\left|\frac{1}{\sigma''}A'(z)D(z)-1\right|\frac{dz}{z}$$

and hence, equation (5-42) becomes

$$\frac{1}{2\pi i}\oint\left|D(z)\right|^2\left|\frac{\sigma}{\sigma'}A(z) - \frac{1}{\sigma''}A'(z)\right|^2\frac{dz}{z} \leq 2(R + R') \tag{5-45}$$

By equation (5-36),

$$\frac{1}{2\pi i}\oint_{|z|=1}\left|D(z)\right|^2\left|\frac{\sigma}{\sigma'}A(z) - \frac{1}{\sigma''}A'(z)\right|^2\frac{dz}{z} \leq 4R \tag{5-46}$$

The integral on the left hand side, which goes to zero as

$\frac{\sigma}{\sigma'}A(z)$ approaches $\frac{1}{\sigma''}A'(z)$, is a good measure of the approximation of A(z) by A'(z) on the unit circle. It is less than four times the mean square "whitening" error of the first step of the spectral inversion method. The significance of this finding is greatly clarified by considering the limiting behavior of equation (5-46) as the approximation order M increases without limit. It is apparent from the foregoing that

$$\lim_{M \to \infty} \sigma' = \sigma \tag{5-47a}$$

$$\lim_{M \to \infty} \sigma'' = 1 \tag{5-47b}$$

$$\lim_{M \to \infty} D(z) = \frac{1}{A(z)} \tag{5-47c}$$

When the limits are substituted into equation (5-46), the following approximate result is found:

$$\frac{1}{2\pi i} \oint \frac{|A(z) - A'(z)|^2}{|A(z)|^2} \frac{dz}{z} \leq \frac{4}{2\pi i} \oint \frac{|A(z) - 1/D(z)|^2}{|A(z)|^2} \frac{dz}{z} \quad \text{as } M\to\infty \tag{5-48}$$

The left hand side can be interpreted as the average percentage error in the approximation of A(z) by A'(z). The right hand side can similarly be viewed as four times the average percentage error in the approximation of A(z) by 1/D(z) and serves as an upper bound. Thus, as M increases and R approaches zero, A'(z) will approach A(z) as closely as desired. Since the two percentage errors are of the same order of magnitude, an obvious implication is that if A(z) can be well represented by an M^{th} order all-pole approximation, the spectral inversion technique will yield good results with order M. As the examples will show, this becomes a two-edged sword. At least in the noiseless case, however, the coefficient estimates obtained by the spectral inversion technique approach the true values arbitrarily closely as $M \to \infty$.

The development thus far has been concerned with studying the relationship between a_j and a_j' in the noiseless case. In using mean lagged products as the basis for an M^{th} order spectral inversion one obtains the set $\{\hat{a}_j\}$, whose elements differ from those of $\{a_j'\}$ by amounts depending upon the statistical error present in the mean lagged products. We turn now to the study of this effect.

Suppose that the first step of the spectral inversion technique yields the M estimated autoregressive coefficients $\{\hat{d}_j\}$. These can be expressed as

$$\hat{d}_j = d_j + e_j \quad , \quad j = 1, 2, \ldots\ldots, M \tag{5-49}$$

where the $\{d_j\}$ represent the coefficients that would be found in the noiseless case, and the $\{e_j\}$ are statistical errors of the type treated in Chapter III. As the number of samples N grows large, these will become zero mean, jointly normal, and possess a covariance structure given by

$$\Phi_{ee} = \frac{1}{N} \Phi_{yy}^{-1} \tag{5-50}$$

For the next stage of the estimation procedure, the correlation function associated with the inverse spectrum must be computed using equation (5-33). When the analysis is based on mean lagged products, this becomes

$$\hat{\phi}_{vv}(j) = \sum_{k=0}^{M-j} \hat{d}_k \hat{d}_{k+j} \tag{5-51}$$

To compute the expected value of $\hat{\phi}_{vv}(j)$, one uses equation (5-49) to show that

$$\begin{aligned} E[\hat{\phi}_{vv}(j)] &= \sum_{k=0}^{M-j} [d_k d_{k+j} + Cov(e_k, e_{k+j})] \\ &= \phi_{vv}(j) + \sum_{k=0}^{M-j} Cov(e_k, e_{k+j}) \end{aligned} \tag{5-52}$$

where $\phi_{vv}(j)$ is the value obtained in the noiseless case. Equation (5-52) shows that $\hat{\phi}_{vv}(j)$ will in general be a biased estimator of $\phi_{vv}(j)$, becoming unbiased in the limit of large sample size N. To study the covariance properties of the $\hat{\phi}_{vv}(j)$, one notes first that

$$E[\hat{\phi}_{vv}(j)\hat{\phi}_{vv}(k)] = \sum_{\ell=0}^{M-j} \sum_{m=0}^{M-k} E[\hat{d}_\ell \hat{d}_{\ell+j} \hat{d}_m \hat{d}_{m+k}] \tag{5-53}$$

Now since the $\hat{d}_j$ become Gaussian as $N \to \infty$, appeal can be made to the Gaussian fourth moment properties to treat the expected value within the summation. The final result is

$$\mathrm{Cov}[\hat{\phi}_{vv}(j),\ \hat{\phi}_{vv}(k)] =$$

$$\sum_{\ell=0}^{M-j}\sum_{m=0}^{M-k}\left[\begin{array}{l} \mathrm{Cov}(e_\ell,e_m)\mathrm{Cov}(e_{\ell+j},e_{m+k}) + \mathrm{Cov}(e_\ell,e_{m+k})\mathrm{Cov}(e_{\ell+j},e_m) \\ +d_\ell d_{m+k}\mathrm{Cov}(e_{\ell+j},e_m) + d_{\ell+j}d_{m+k}\mathrm{Cov}(e_\ell,e_m) \\ +d_\ell d_m \mathrm{Cov}(e_{\ell+j},e_{m+k}) + d_{\ell+j}d_m\mathrm{Cov}(e_\ell,e_{m+k}) \end{array}\right] \tag{5-54}$$

This covariance approaches zero for all j and k as the sample size N increases without limit.

The second step expressions of equation (5-34) can be written out as

$$\sum_{j=0}^{K} a'_j\ \phi_{vv}(k-j) = 0\ ,\quad k = 1,\ 2,\ \ldots,\ K \tag{5-55}$$

in the noiseless case and as

$$\sum_{j=0}^{K} \hat{a}_j\ \hat{\phi}_{vv}(k-j) = 0\ ,\quad k = 1,\ 2,\ \ldots,\ K \tag{5-56}$$

when the estimates are based on the mean lagged products. Again, a set of auxiliary variables must be defined:

$$u_k \equiv \sum_{j=0}^{K} a_j'\ \hat{\phi}_{vv}(j-k)\ ,\quad k = 1,\ 2,\ \ldots,\ K \tag{5-57}$$

and in terms of these, the differences between a_j' and a_j can be related as follows:

$$\sum_{j=0}^{K} (\hat{a}_j - a_j')\hat{\phi}_{vv}(j-k) = -u_k \quad , \; k = 1, 2, \ldots, K \tag{5-58}$$

Now defining

$$\xi_k = \hat{a}_k - a_k' \quad , \; k = 1, 2, \ldots, K \tag{5-59}$$

and the column vectors

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \cdot \\ \cdot \\ \cdot \\ \xi_K \end{bmatrix} \quad \text{and} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_K \end{bmatrix} \tag{5-60}$$

equation (5-58) becomes

$$\hat{\Phi}_{vv}\xi = -u \quad \text{or} \quad \xi = -\hat{\Phi}_{vv}^{-1}u \tag{5-61}$$

Now it is necessary to study the statistical properties of the $\{u_j\}$. Using equations (5-57) and (5-52),

$$\begin{aligned} E[u_j] &= \sum_{k=0}^{K} a_k' \, E[\hat{\phi}_{vv}(k-j)] \\ &= \sum_{k=0}^{K} a_k'\phi_{vv}(k-j) + \sum_{k=0}^{K} a_k' \sum_{\ell=0}^{M-k+j} \text{Cov}(e_\ell, e_{\ell+k-j}) \end{aligned} \tag{5-62}$$

By equation (5-55), the first summation is zero, and hence

$$E[u_j] = \sum_{k=0}^{K} \sum_{\ell=0}^{M-k+j} a_k' \mathrm{Cov}(e_\ell, e_{\ell+k-j}) \quad , \; j = 1,2,..,K \qquad (5\text{-}63)$$

The covariance between u_j and u_k is found by forming

$$E[u_j u_k] = \sum_{\ell=0}^{K} \sum_{m=0}^{K} a_\ell' a_m' \; E[\hat{\phi}_{vv}(\ell-j)\hat{\phi}_{vv}(m-k)] \qquad (5\text{-}64)$$

and then using equations (5-54) and (5-63) to yield finally

$$\mathrm{Cov}[u_j, u_k] = \sum_{\ell=0}^{K} \sum_{m=0}^{K} a_\ell' a_m' \mathrm{Cov}[\hat{\phi}_{vv}(\ell-j), \hat{\phi}_{vv}(m-k)] \qquad (5\text{-}65)$$

where the covariance function under the summation can be evaluated using equation (5-54). It can be seen that in the limit as $N \to \infty$, both $E[u_j]$ and $\mathrm{Cov}[u_j, u_k]$ approach zero asymptotically for all j and k.

Now as the number of samples grows large, the behavior of the mean and covariance of the $\{\hat{\phi}_{vv}(j)\}$ indicates that

$$\underset{N \to \infty}{\mathrm{plim}} \; \hat{\Phi}_{vv} = \Phi_{vv} \qquad (5\text{-}66)$$

Thus, at the same time

$$\underset{N \to \infty}{\mathrm{plim}} \; \xi = -\Phi_{vv}^{-1} u \qquad (5\text{-}67)$$

Since the $\{u_j\}$ are asymptotically zero mean, it is clear from equation (5-67) that the $\{\xi_j\}$ must also have zero mean in the limit as $N \to \infty$. Hence, the $\{\hat{a}_j\}$ are asymptotically unbiased estimates of the $\{a_j'\}$. To check the consistency of the estimates and to determine their variance, it is necessary to investigate the covariance matrix for the $\{\xi_j\}$. Since the errors are found from the $\{u_j\}$ by a straightforward linear transformation, their covariance matrix - in the limit of large sample size - is related to that of the $\{u_j\}$ by the matrix equation

$$\Phi_{\xi\xi} = \Phi_{vv}^{-1}\Phi_{uu}(\Phi_{vv}^{-1})^T \qquad (5\text{-}68)$$

In practical situations, this result is of minimal usefulness due to the extreme complexity of the elements of Φ_{uu} (found by combining the results of equations (5-50), (5-54), and (5-65)). No simplifications have been found, and equation (5-68) really only serves to show that the estimate variances indeed approach zero for large sample size and thus that the $\{\hat{a}_j\}$ are consistent estimates of the $\{a_j'\}$.

In the work of Durbin[36] mentioned above, an approximate expression for $\Phi_{\xi\xi}$ is presented as

$$\Phi_{\xi\xi} \simeq \frac{1}{N}\Phi_{vv}^{-1} \qquad (5\text{-}69)$$

where Φ_{vv} is a K^{th} order covariance matrix whose elements are given by equation (5-33). This result is intuitively appealing since it is precisely what would be obtained if the K^{th} order all-pole estimate which forms the second step of the spectral inversion technique were based on actual mean lagged products from the process v_t. Another interesting point brought out by Durbin's approach is that the efficiency of the method approaches unity for large sample sizes as the intermediate model order M increases without limit.

In treating the variability of all-pole and autoregressive estimates, it was possible to depend upon the asymptotic joint normality of the auxiliary variables $\{u_j\}$ to infer joint normality for the estimates. This is by no means so clearly justified here. By equation (5-57), the $\{u_j\}$ are indeed defined as linear sums of estimated correlations as before, but in the present case, the latter are not mean lagged products, whose Gaussian character is little disputed, but rather finite sums of products of Gaussian random variables (the $\{\hat{d}_j\}$) as shown in equation (5-51). Thus, no claim will be made here for the asymptotic joint normality of the $\{\xi_j\}$, although the possibility is certainly likely.

As a means of avoiding the tedium of evaluating equation (5-68), a fairly good estimate of the magnitude of the statistical error inherent in the data - at least in the limiting case where the intermediate model order M approaches infinity - can be found using the same relationships developed in the treatment of Wold's method, culminating in equations (5-26) and (5-27).

It has been pointed out that the spectral inversion technique will always provide a solution to the problem of estimating the coefficients of a moving average on the basis of mean lagged products, whereas Wold's method will fail when no moving average exists which fits the data. The objection can be raised, however, that if no moving average exists corresponding to the observed sequence of mean lagged products, then spectral inversion supplies a meaningless result. To an extent, this is true. In such cases, the set of mean lagged products actually corresponds to a process of mixed type with a weak autoregressive part. The approximation of the spectral inversion procedure, however, subsumes this residual autoregressive part into a moving average estimate that best represents the process as a whole. Indeed, it is this fortuitous approximation that makes a solution possible. At any rate, it is clear that as both the original sample size and the model order grow large, the spectral inversion estimates converge to the true moving average coefficients.

D. Computational Examples of Moving Average Spectral Estimates

The first step in each of the methods considered here for the analysis of moving average spectra is the measurement of a set of mean lagged products from a sample sequence of the process. In the case of the moving average part of a process of mixed type, this sample sequence is obtained by passing the original process through a moving average filter whose system function is the estimate of the autoregressive polynomial for the process, as found from the method of the last chapter. Depending upon the quality of the autoregressive estimate, the filtered sequence will be a more or less accurate representation of the true moving average y_t. Since it is the spectrum of this approximation that is actually estimated, the final result will include the effects of errors in estimating the autoregressive polynomial $B(z)$. The particular block of samples of the original process x_t which is filtered to obtain the approximate moving average sample sequence need not be the same as that used in computing $\hat{B}(z)$, nor must the respective sample sizes be identical. For simplicity in the present examples, however, the same original sample sequence was used in both parts of the analysis, so that the spectral estimate as a whole arises from a single block of samples.

The first of the five standard examples is a pure autoregression and thus will not be treated in the present chapter. The second is a pure moving average, and the set of mean lagged products presented in Appendix A will be used directly for the estimates here. For the remaining three examples (processes of mixed type), Table 5-1 presents mean lagged products measured out to eight lags from the sample sequence obtained by filtering the original process with the autoregressive polynomial estimated in the last chapter from the same original sample block. The theoretical moving average autocorrelation sequence is also presented for comparison.

1. All-pole Estimation of the Moving Average Part

All-pole estimates for Example # 2 have been presented in Chapter III, and the present section will deal with autoregressive representations of the moving average parts of the last three examples and their combination with the computational results of the last chapter to yield so-called

Table 5-1

Theoretical Autocorrelation and Mean Lagged Product Sequences for the Moving Average Parts of Three Mixed-type Examples

Example # 3

N = 1,000

Lag j	Theoretical $\phi_{yy}(j)$	N=1,000 $\hat{\phi}_{yy}(j)$
0	1.73788	1.77029
1	-0.85900	-0.93736
2	0.0	0.11048
3	0.0	-0.10071
4	0.0	0.06791
5	0.0	0.06456
6	0.0	-0.05602
7	0.0	0.01811
8	0.0	0.03848

Example # 4

N = 5,000

Lag j	Theoretical $\phi_{yy}(j)$	N=5,000 $\hat{\phi}_{yy}(j)$
0	4.80972	4.93328
1	-3.19782	-3.29425
2	0.79300	0.83207
3	0.0	-0.01344
4	0.0	0.02549
5	0.0	-0.02613
6	0.0	0.02515
7	0.0	-0.05985
8	0.0	0.09768

Example # 5

N = 5,000

Lag j	Theoretical $\phi_{yy}(j)$	N=5,000 $\hat{\phi}_{yy}(j)$
0	4.68370	5.70871
1	-3.14940	-3.89929
2	0.81000	1.05421
3	0.0	-0.00399
4	0.0	0.01036
5	0.0	-0.02556
6	0.0	0.04537
7	0.0	-0.07700
8	0.0	0.09949

"two-step all-pole estimates".

a. Example # 3

With a sample size of 1,000, the polynomial B(z) was estimated in the last chapter as

$$\hat{B}(z) = 1 + 0.00585z^{-1} + 0.85445z^{-2}$$

Using the corresponding mean lagged products of Table 5-1, a 6^{th} order all-pole estimate for the moving average part is found to be

$$\frac{1}{\hat{G}(z)} = \frac{1}{\begin{bmatrix} 1 + 0.83782z^{-1} + 0.61919z^{-2} + 0.47358z^{-3} \\ +0.27431z^{-4} + 0.10492z^{-5} + 0.042272z^{-6} \end{bmatrix}}$$

with $\hat{\sigma}^2 = 1.02867$. The final 8^{th} order two-step all-pole estimate for the process system function is given as $1/\hat{F}(z)$, where

$$\hat{F}(z) = \hat{G}(z)\hat{B}(z)$$

The coefficients $\{\hat{f}_j\}$ are found in Table 5-2, along with the two-step 8^{th} order all-pole estimate based on the theoretical correlation sequence (coefficients $\{f_j\}$), and the theoretical 8^{th} order <u>direct</u> all-pole estimate computed in Chapter III. The corresponding input noise variance estimates are also included.

Table 5-2

Two-step All-pole Estimates for Example # 3.

j	$\hat{f}_j$	f_j	d_j
1	0.84366	0.81775	0.87301
2	1.47853	1.50742	1.53848
3	1.19307	1.20377	1.30769
4	0.80614	0.92799	0.97733
5	0.51118	0.67368	0.74136
6	0.27771	0.43496	0.48911
7	0.08990	0.20631	0.23093
8	0.03650	0.10197	0.11776
Variance Estimate	1.02867	1.03544	0.99392

The numerical differences between $\hat{f}_j$ and f_j are due to statistical error in both parts of the analysis, whereas that between f_j and d_j reflects the fact that the $\{f_j\}$ represent a sub-optimum 8^{th} order approximation, with the $\{d_j\}$ representing the best. The corresponding spectra are compared in Figures 5-5 and 5-6. The former compares the direct and two-step all-pole estimates based on the theoretical correlation structure with the actual spectrum, and the latter displays the spectral estimate based on 1,000 samples. The differences between the theoretical spectra are small but apparent, but

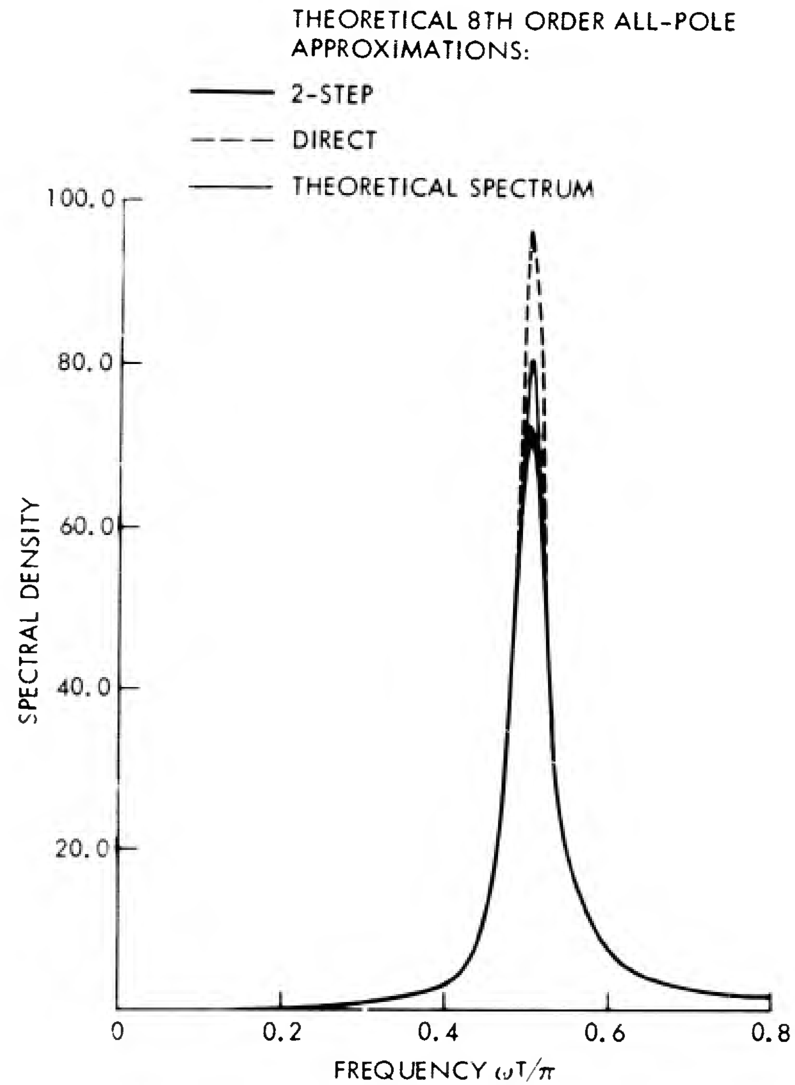


FIG. 5-5 A COMPARISON OF THE DIRECT AND 2-STEP 8TH ORDER ALL-POLE APPROXIMATIONS WITH THEORETICAL SPECTRUM FOR EXAMPLE #3

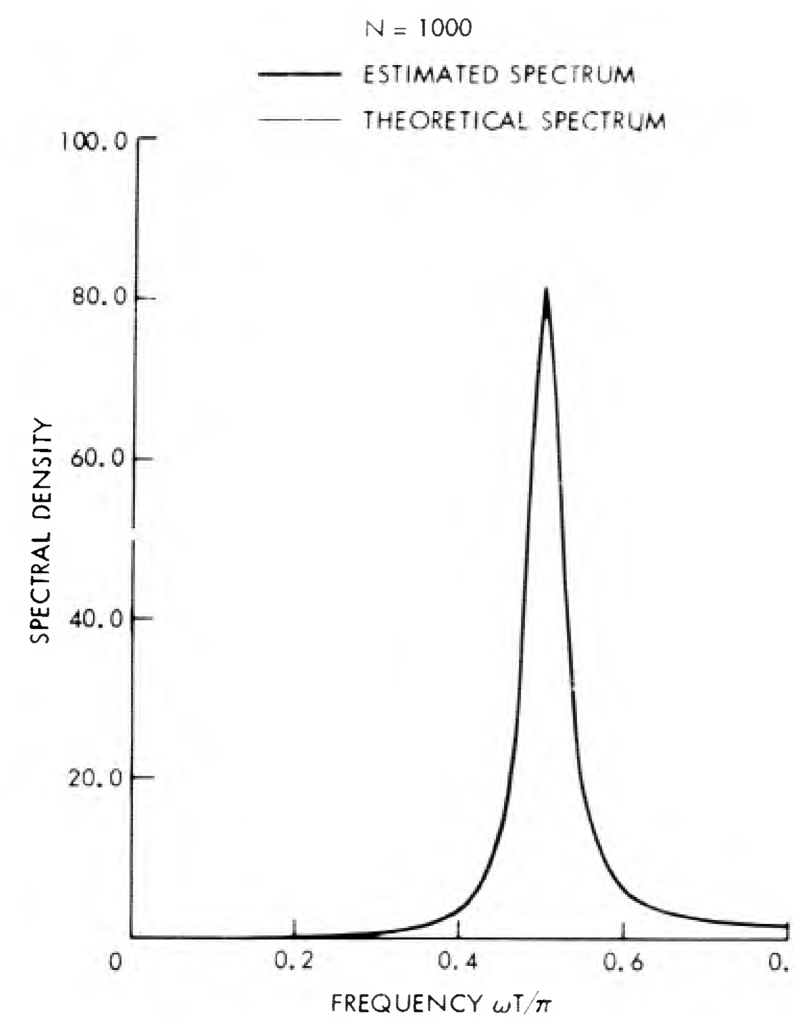


FIG. 5-6 THE 2-STEP 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #3 BASED ON N = 1,000 SAMPLES

as occasionally happens, the statistical error in the 1,000 sample estimate has fortuitously compensated the approximation error to provide the excellent agreement observed in Figure 5-6.

b. Example # 4

The estimate for B(z), based on 5,000 samples was

$$\hat{B}(z) = 1 - 1.33564z^{-1} + 1.62015z^{-2} - 0.97754z^{-3} + 0.65017z^{-4}$$

A 4^{th} order autoregressive approximation for A(z), stemming from the mean lagged products of Table 5-1 is

$$\frac{1}{\hat{G}(z)} = \frac{1}{1 + 1.33780z^{-1} + 1.20485z^{-2} + 0.80277z^{-3} + 0.33124z^{-4}}$$

with $\hat{\sigma}^2 = 1.52665$. Combining these gives the two-step 8^{th} order all-pole estimate whose coefficients are given in the following table, again with the theoretical direct and two-step approximations for comparison.

Table 5-3

Two-step All-pole Estimates for Example # 4.

j	$\hat{f}_j$	f_j	d_j
1	0.00216	0.01581	0.06005
2	1.03818	1.04760	1.12994
3	0.38342	0.37470	0.53368
4	0.55348	0.56195	0.73542
5	0.55019	0.54986	0.71181
6	0.53527	0.53915	0.61830
7	0.19813	0.19532	0.31595
8	0.21536	0.21623	0.28138
Variance Estimate	1.52665	1.52216	1.42528

The corresponding spectra are shown in Figures 5-7 and 5-8. These confirm the result found in the table - that there is substantial difference between the direct and two-step approximations, but relatively little between the latter and the actual estimate. This implies, to be sure, a small statistical error, but the approximation error is such that the estimate is fairly poor.

c. Example # 5

The autoregressive polynomial for the last example was estimated on the basis of 5,000 samples to be

$$\hat{B}(z) = 1 - 1.49862z^{-1} + 1.45736z^{-2} - 0.76938z^{-3} + 0.27201z^{-4}$$

and the 4^{th} order all-pole estimate for the moving average part, based on the same 5,000 samples is

$$\frac{1}{\hat{G}(z)} = \frac{1}{1 + 1.44121z^{-1} + 1.35811z^{-2} + 0.91920z^{-3} + 0.37625z^{-4}}$$

with $\hat{\sigma}^2 = 1.52097$. The composite two-step all-pole estimate has autoregressive coefficients $\{\hat{f}_j\}$ as given by the following table, along with the corresponding $\{f_j\}$ and $\{d_j\}$ as defined previously:

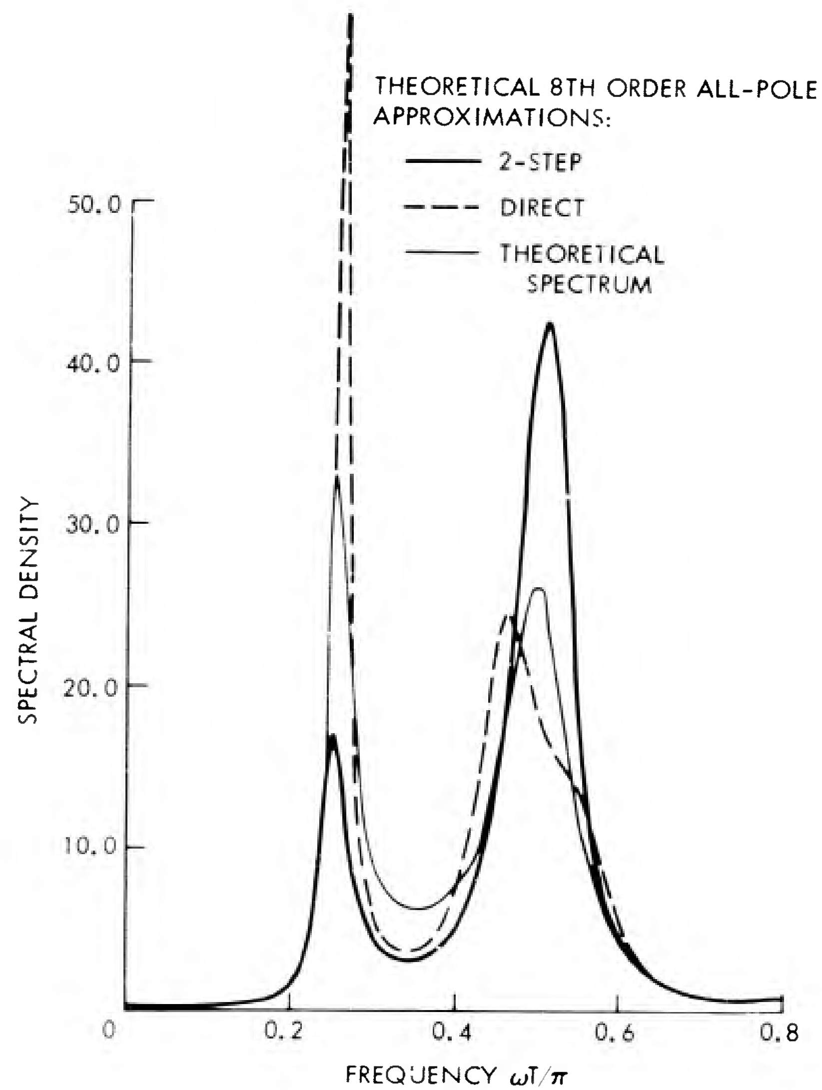


FIG. 5-7 A COMPARISON OF THE DIRECT AND 2-STEP 8TH ORDER ALL-POLE APPROXIMATIONS WITH THE THEORETICAL SPECTRUM FOR EXAMPLE #4

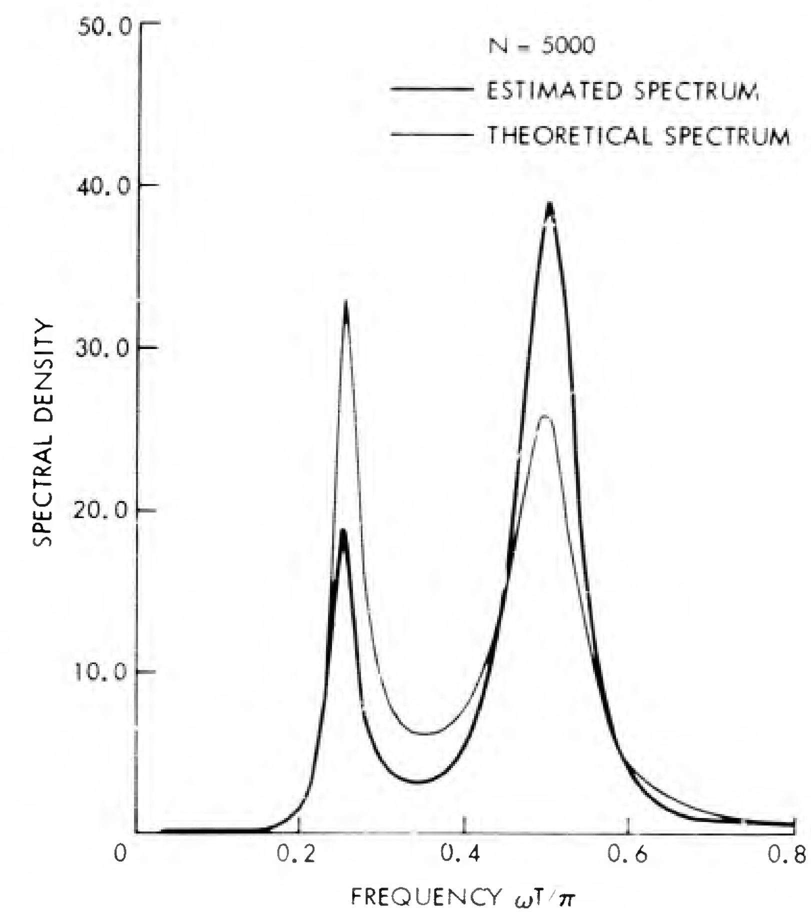


FIG. 5-8 THE 2-STEP 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #4 BASED ON N = 5,000 SAMPLES

Table 5-4

Two-step All-pole Estimates for Example # 5.

j	$\hat{f}_j$	f_j	d_j
1	-0.05741	0.01587	0.26959
2	0.65564	0.73803	1.03400
3	0.21489	0.31550	0.80848
4	0.14115	0.22164	0.78057
5	0.12287	0.15207	0.66275
6	0.21053	0.20958	0.50058
7	-0.03944	-0.02855	0.28096
8	0.10234	0.08313	0.20874
Variance Estimate	1.52097	1.38774	1.10627

The substantial differences between the $\{\hat{f}_j\}$ and the $\{f_j\}$ are due mainly to the large statistical error made in the estimation of the $\{b_j\}$ and the subsequent difference between the actual and apparent spectra for the moving average part (See Figure 4-10). The large disparity between the $\{f_j\}$ and the $\{d_j\}$ - and particularly between the corresponding values for $\hat{\sigma}^2$ - shows that the two-step estimate is markedly different from the optimum approximation. This is borne out by

Figures 5-9 and 5-10 where spectral comparisons are drawn.

The computational examples further indicate that if only the final spectral estimates are of interest, the direct all-pole estimate is preferable to the two-step version. Not only is the former mathematically simpler, but it produces substantially less error of both types.

2. Estimation by Wold's Method

The spectrum of Example # 2 and those of the moving average parts of the last three examples are here analyzed using the method of Wold.

a. Example # 2

Using the mean lagged products presented in Appendix A for N = 2,000, the equations (5-14) can be evaluated to generate the polynomial

$$V(r) = 0.0371495r^4 + 0.906883r^3 + 3.31884r^2 + 4.34652r + 2.02663$$

With recourse to a digital computer subroutine, the roots are found to be

$$r_1 = -2.05156$$

$$r_2 = 19.92020$$

$$r_3 = -1.07263 + i0.42936$$

$$r_4 = -1.07263 - i0.42936$$

Since none of these lies in the real interval $-2<r_j<2$, a 4th order moving average exists with a correlation sequence identical to the sequence of mean lagged products. Thus, the

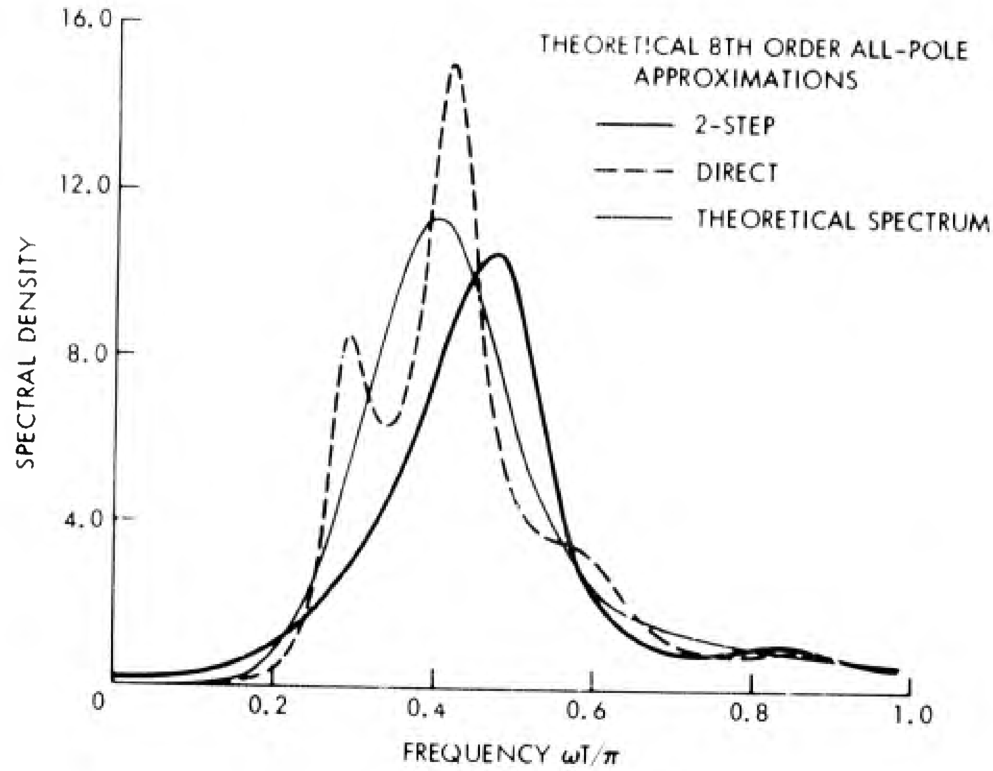


FIG. 5-9 A COMPARISON OF THE DIRECT AND 2-STEP 8TH ORDER ALL-POLE APPROXIMATIONS WITH THE THEORETICAL SPECTRUM FOR EXAMPLE #5

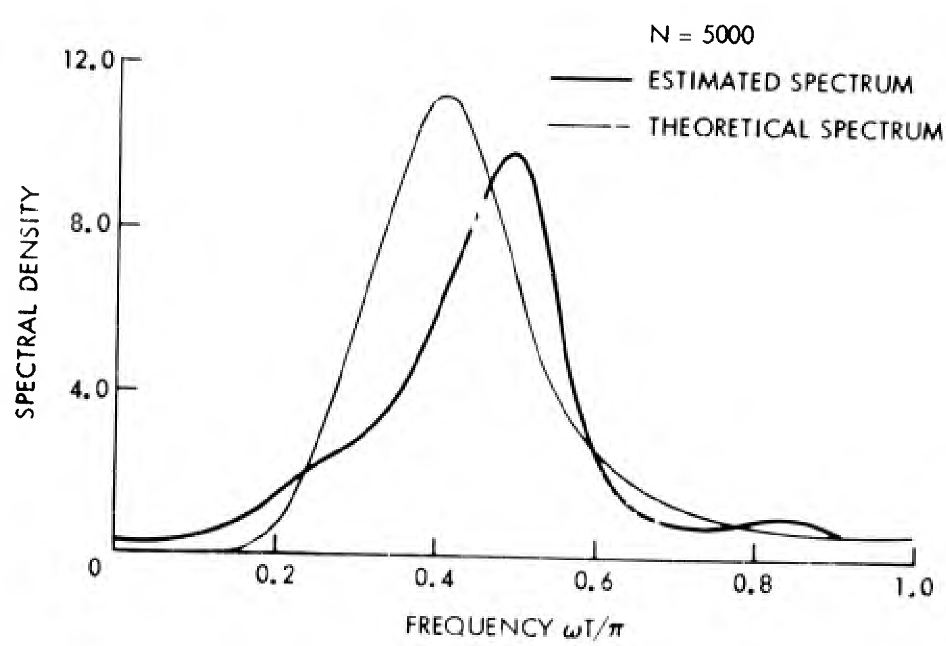


FIG. 5-10 THE 2-STEP 8TH ORDER ALL-POLE ESTIMATE FOR EXAMPLE #5 BASED ON N = 5,000 SAMPLES

solution may proceed by substituting each of the roots into equation (5-15) to find the corresponding z-plane zeros. The result is

$$\text{From } r_1: \quad z_{11} = -0.79725$$

$$z_{12} = -1.25431$$

$$\text{From } r_2: \quad z_{21} = -0.05033$$

$$z_{22} = -19.86987$$

$$\text{From } r_3: \quad z_{31} = -0.40557 - i0.66597$$

$$z_{32} = -0.66706 + i1.09533$$

$$\text{From } r_4: \quad z_{41} = -0.40557 + i0.66597$$

$$z_{42} = -0.66706 - i1.09533$$

Now choosing the four z-plane roots that lie within the unit circle (z_{11}, z_{21}, z_{31}, z_{41}), equation (5-19) yields the coefficient estimates:

$$\hat{a}_1 = 1.65873 \qquad (a_1 = 2.036)$$

$$\hat{a}_2 = 1.33564 \qquad (a_2 = 1.812)$$

$$\hat{a}_3 = 0.54788 \qquad (a_3 = 0.770)$$

$$\hat{a}_4 = 0.02440 \qquad (a_4 = 0.129)$$

(The values in parentheses are the actual coefficients.) The input variance estimate is now computed from equation (5-20) to be

$$\hat{\sigma}^2 = 1.498$$

The sampled power spectrum corresponding to the present coefficient estimate is compared to the actual spectrum in Figure 5-11. The agreement is very good, despite the fact that there are substantial errors in both the coefficient and variance estimates. (The form of the estimator of σ^2 is such that the coefficient and variance errors tend to cancel each other out, in the sense that the mean square value of the estimated process is identical with the measured mean square value of the sample sequence.) Another interesting comparison is that shown in Figure 5-12 between the theoretical pole-zero pattern for the process system function and that constructed in the estimate. They are very different, despite the close agreement of the corresponding spectra. As a check, the theoretical covariance sequence for the process was used as the basis for a Wold estimate, and the proper coefficients and system pole-zero pattern were indeed returned. This shows how sensitive the present method is to even small changes in the observed mean lagged products.

One measure of the error in the final spectral estimate is the spectral variance averaged around the z-plane unit circle as derived above. Using equation (5-26) and the expression for $\mathrm{Var}[\hat{\phi}_{yy}(j)]$ derived in Chapter II, this becomes

$$\mathrm{Var}_{avg} = \frac{2471.04}{N}$$

and in the present case where $N = 2{,}000$,

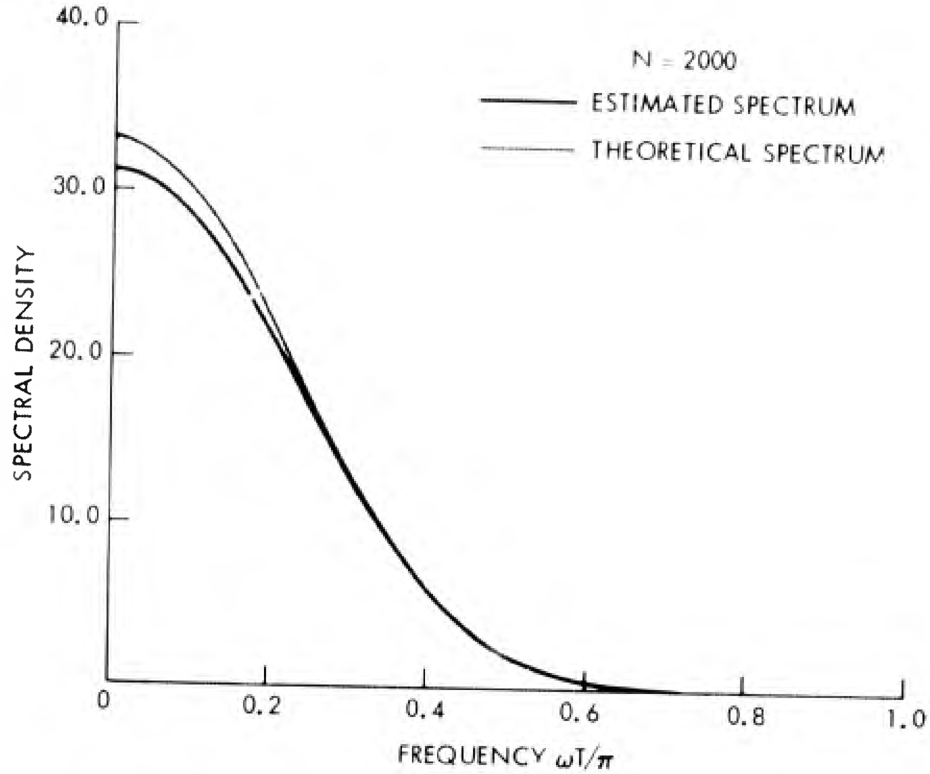


FIG. 5-11 AN ESTIMATE OF THE SPECTRUM OF EXAMPLE #2 FOUND FROM 2,000 SAMPLES BY WOLD'S METHOD

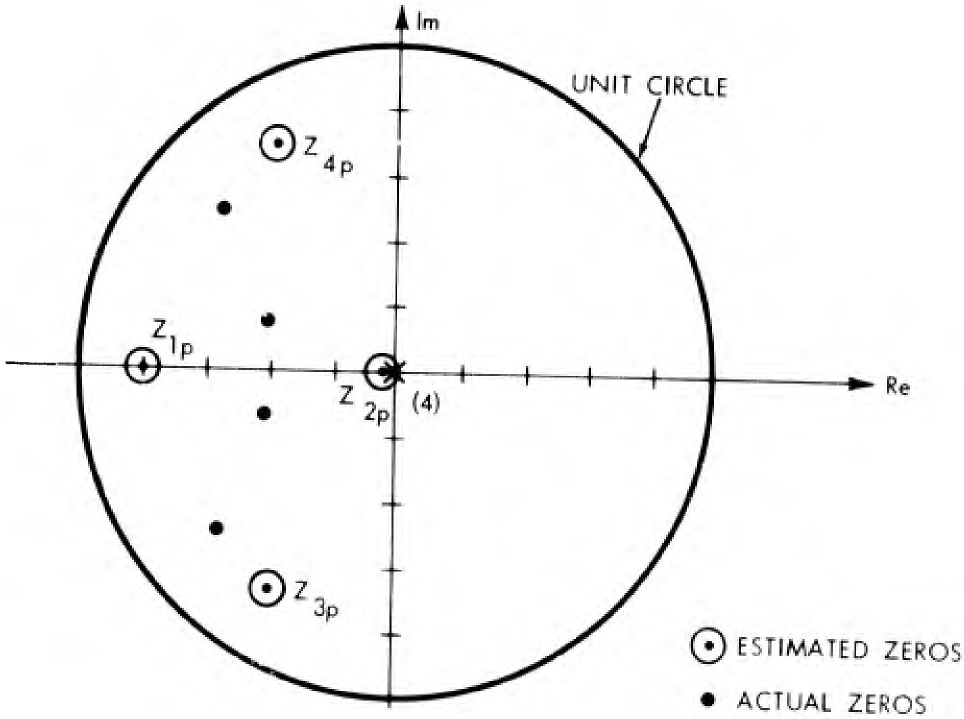


FIG. 5-12 EXAMPLE #2. COMPARISON OF THE POLE-ZERO PATTERN FOR A(z) FOUND FROM 2000 SAMPLES WITH WOLD'S METHOD AND THE ACTUAL POLE-ZERO PATTERN

$$Var_{avg} = 1.23$$

The average standard deviation is thus 1.11, which is not a large dispersion compared to the magnitude of the spectrum. The good spectral agreement is thus not surprising.

In using the mean lagged products measured over 1,000 samples for Example # 2, however, Wold's method fails to provide an exact solution. The polynomial V(r) becomes

$$V(r) = 0.017845r^4 + 0.964802r^3 + 3.51167r^2 + 4.31444r + 1.90646$$

with roots

$$r_1 = -1.90122$$

$$r_2 = -50.2434$$

$$r_3 = -0.960244 + i0.443087$$

$$r_4 = -0.960244 - i0.443087$$

Since $-2<r_1<2$, no 4^{th} order moving average exists with the measured correlation sequence. To continue, however, one can use Wold's suggestion of setting $r_1 = -2$ and plunging ahead. If this is done, the zeros of $\hat{A}(z)$ within the unit circle become

$$z_1 = -1.0$$

$$z_2 = -0.01991$$

$$z_3 = -0.36352 + i0.69068$$

$$z_4 = -0.36352 - i0.69068$$

as shown in Figure 5-13. It should be noted that the root at $z = -1$ arises from the approximation of r_1 by -2. Subsequent calculation reveals that

$$\hat{a}_1 = 1.74695$$

$$\hat{a}_2 = 1.37062$$

$$\hat{a}_3 = 0.63580$$

$$\hat{a}_4 = 0.01213$$

$$\hat{\sigma}^2 = 1.426$$

and the spectral estimate is plotted in Figure 5-14. The agreement with the theoretical spectrum is embarrassingly good, despite the somewhat arbitrary approximation.

b. Example # 3

From the sequence of mean lagged products measured over 1,000 samples for the moving average part of Example # 3, the polynomial V(r) is constructed to be

$$V(r) = 1.77029r - 0.93736$$

The single root is

$$r_1 = 1.89$$

which unfortunately lies within the "forbidden region", $-2<r<2$. Again, no exact solution exists. Approximating r_1 by +2.0, however, yields the z-plane root

$$z_1 = 1.0$$

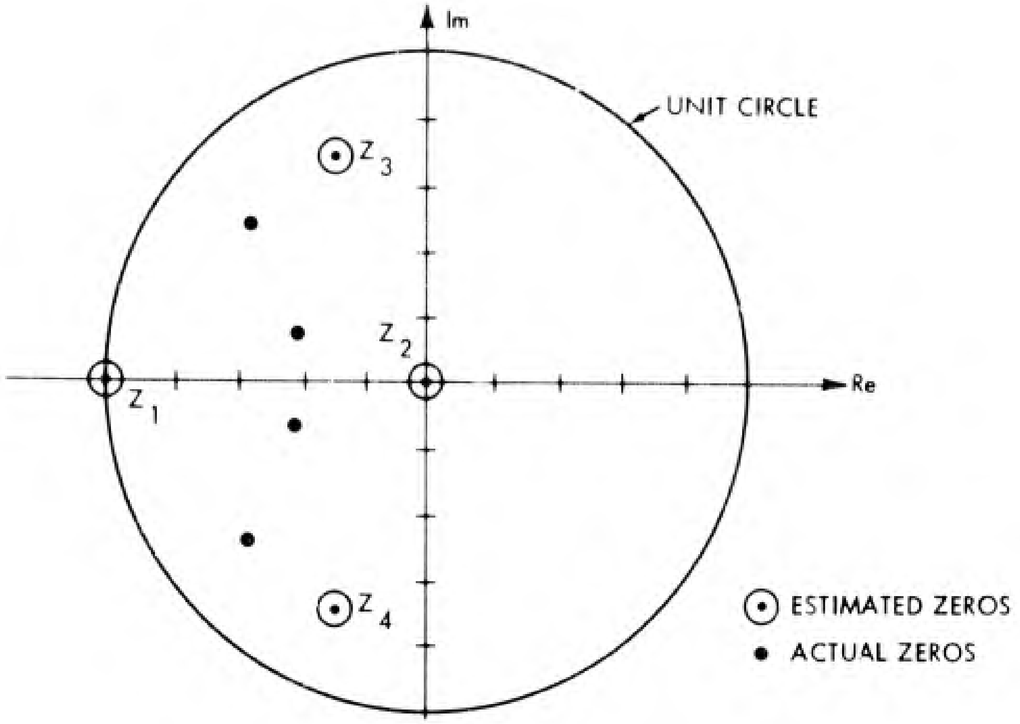


FIG. 5-13 EXAMPLE #2. COMPARISON OF THE POLE-ZERO PATTERN FOR A(z) FOUND FROM 1000 SAMPLES BY USING WOLD'S APPROXIMATE METHOD AND THE ACTUAL POLE-ZERO PATTERN. A 4TH ORDER POLE AT THE ORIGIN HAS BEEN OMITTED FOR CLARITY

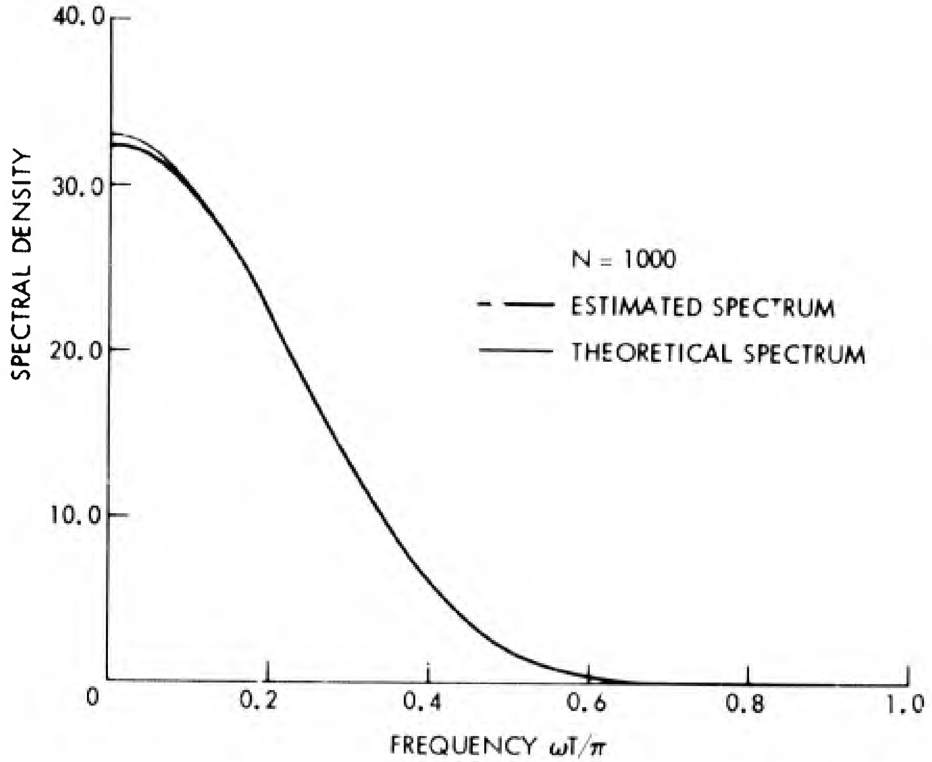


FIG. 5-14 AN ESTIMATE OF THE SPECTRUM OF EXAMPLE #2 BASED ON 1,000 SAMPLES AND USING WOLD'S APPROXIMATION

Now using equations (5-19) and (5-20),

$$\hat{a}_1 = -1.0 \qquad (a_1 = -0.859)$$

$$\hat{\sigma}^2 = 0.885 \qquad (\sigma^2 = 1.0)$$

which yields the spectrum of Figure 5-15, plotted for comparison with the actual moving average spectrum.

Using this estimate in conjunction with the autoregressive coefficients estimated in the previous chapter gives the total system function estimate for the original process x_t:

$$\frac{\hat{A}(z)}{\hat{B}(z)} = \frac{1 - z^{-1}}{1 + 0.00585z^{-1} + 0.85445z^{-2}}$$

The corresponding spectrum (computed with $\hat{\sigma}^2 = 0.885$) is compared with the actual spectrum in Figure 5-16. Again, the agreement is surprisingly good despite the use of the approximate root.

c. Example # 4

Using the 5,000 sample mean lagged products from Table 5-1, the moving average part of the 4th example gives the polynomial

$$V(r) = 0.832067r^2 - 3.29425r + 3.26915$$

with roots

$$r_1 = 1.97956 + i0.10149$$

$$r_2 = 1.97956 - i0.10149$$

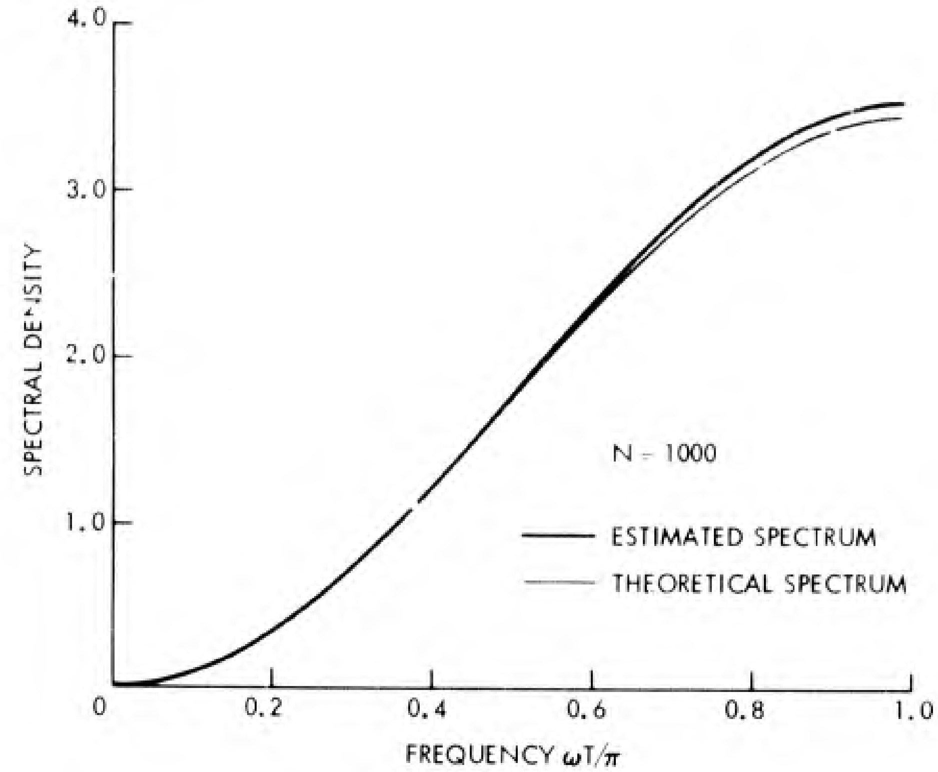


FIG. 5-15 AN ESTIMATE OF THE MOVING AVERAGE SPECTRUM OF EXAMPLE #3 FOUND USING WOLD'S METHOD ON 1,000 SAMPLES

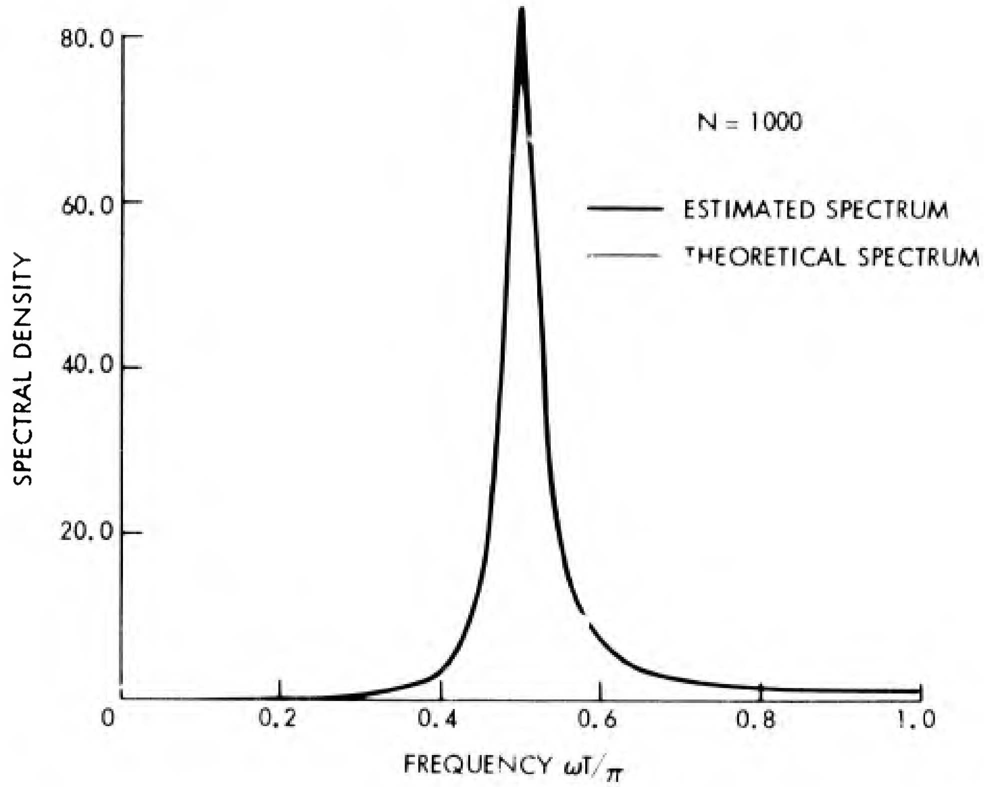


FIG. 5-16 AN ESTIMATE OF THE TOTAL SPECTRUM FOR EXAMPLE #3 BASED ON 1,000 SAMPLES AND USING WOLD'S APPROACH TO FIND THE MOVING AVERAGE PART

No real root lies between -2 and +2, so Wold's method can be applied exactly. The z-plane roots of $\hat{A}(z)$ which lie within the unit circle are

$$z_1 = 0.78963 + i0.20020$$

$$z_2 = 0.78963 - i0.20020$$

These are quite different from the actual process zeros, which are both positive real (See Figure A-11.). The coefficient and variance estimates become

$$\hat{a}_1 = -1.58080 \qquad (a_1 = -1.7835)$$

$$\hat{a}_2 = 0.66550 \qquad (a_2 = 0.7930)$$

$$\hat{\sigma}^2 = 1.254 \qquad (\sigma^2 = 1.0)$$

and the corresponding moving average spectral estimate is compared with the actual moving average spectrum in Figure 5-17, where the estimate is seen to be quite good. From equation (5-26), the average spectral variance is found to be

$$Var_{avg} = \frac{305.6}{N} = 0.00061$$

giving the average spectral standard deviation as 0.078. About half of the observed error can be laid to the spectral error inherent in the approximation of the process y_t (See Figure 4-8), and the remainder is consistent with the variance calculation.

When the present results are combined with the autore-

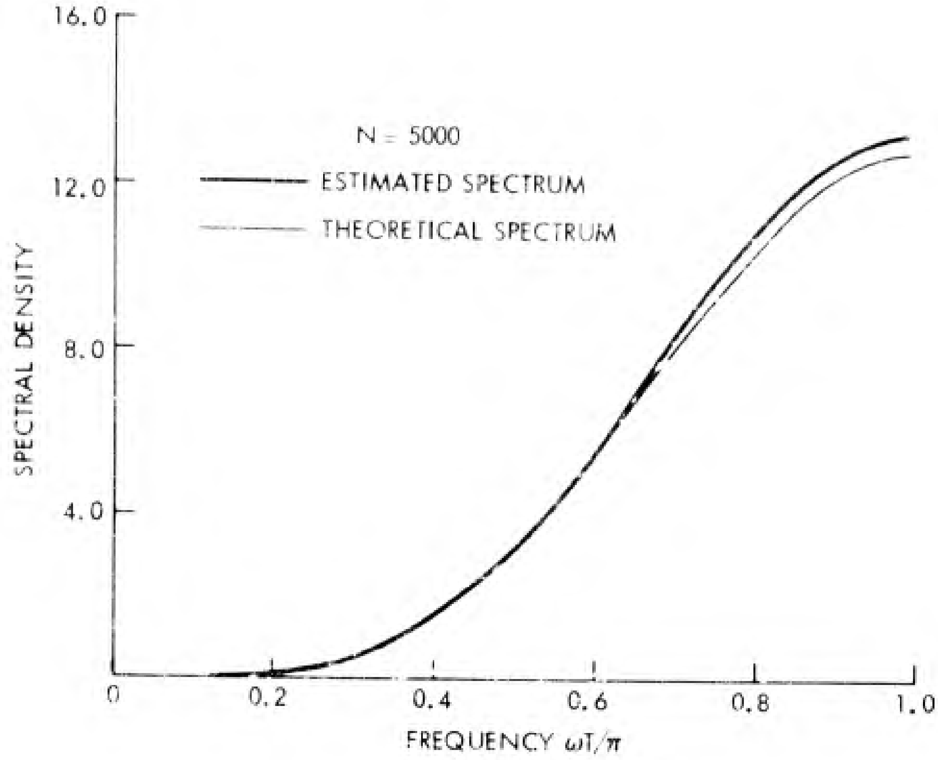


FIG. 5-17 AN ESTIMATE OF THE MOVING AVERAGE SPECTRUM OF EXAMPLE #4 FOUND USING WOLD'S METHOD ON 5 000 SAMPLES

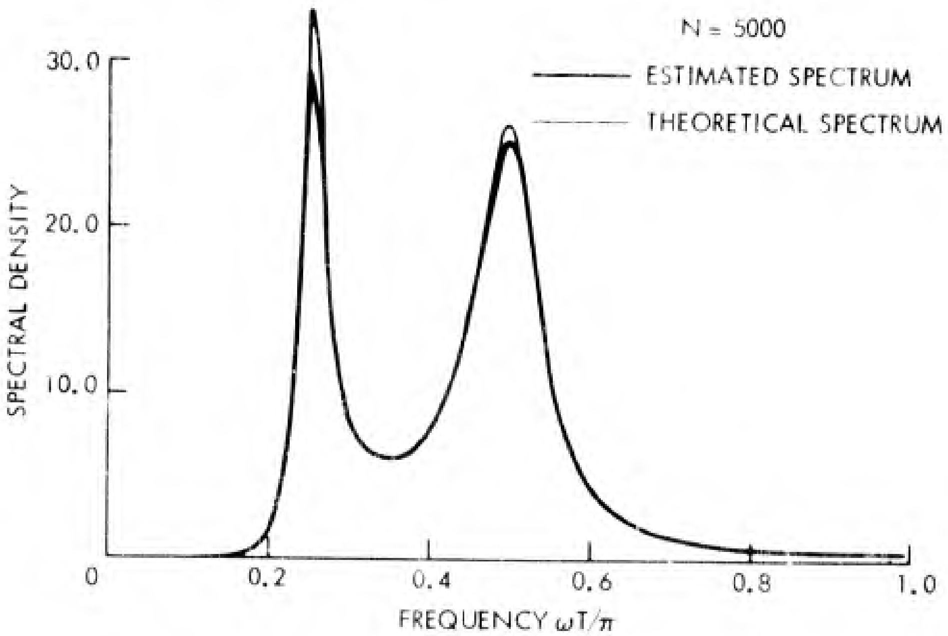


FIG. 5-18 AN ESTIMATE OF THE TOTAL SPECTRUM FOR EXAMPLE #4 BASED ON 5,000 SAMPLES AND USING WOLD'S APPROACH TO FIND THE MOVING AVERAGE PART

gressive estimates found previously, the final system estimate for the process x_t becomes

$$\frac{\hat{A}(z)}{\hat{B}(z)} = \frac{1 - 1.58080z^{-1} + 0.66550z^{-2}}{1 - 1.33564z^{-1} + 1.62015z^{-2} - 0.97754z^{-3} + 0.65017z^{-4}}$$

with $\hat{\sigma}^2 = 1.254$. The estimated and actual spectra are compared in Figure 5-18. The final result is quite good.

d. Example # 5

5,000 samples of the moving average part of the final example yield the polynomial

$$V(r) = 1.05420r^2 - 3.89929r + 3.60031$$

which unfortunately has both its roots between -2 and +2:

$$r_1 = 1.92083$$

$$r_2 = 1.77799$$

Again using the approximation that $r_1 \approx r_2 \approx 2.0$, $\hat{A}(z)$ displays a double zero at $z = +1$. Thus, using equations (5-19) and (5-20),

$$\hat{a}_1 = -2.0 \qquad (a_1 = -1.74)$$

$$\hat{a}_2 = 1.0 \qquad (a_2 = 0.81)$$

$$\hat{\sigma}^2 = 0.95 \qquad (\sigma^2 = 1.0)$$

The corresponding estimated moving average spectrum is shown with the actual spectrum in Figure 5-19. Most of the observ-

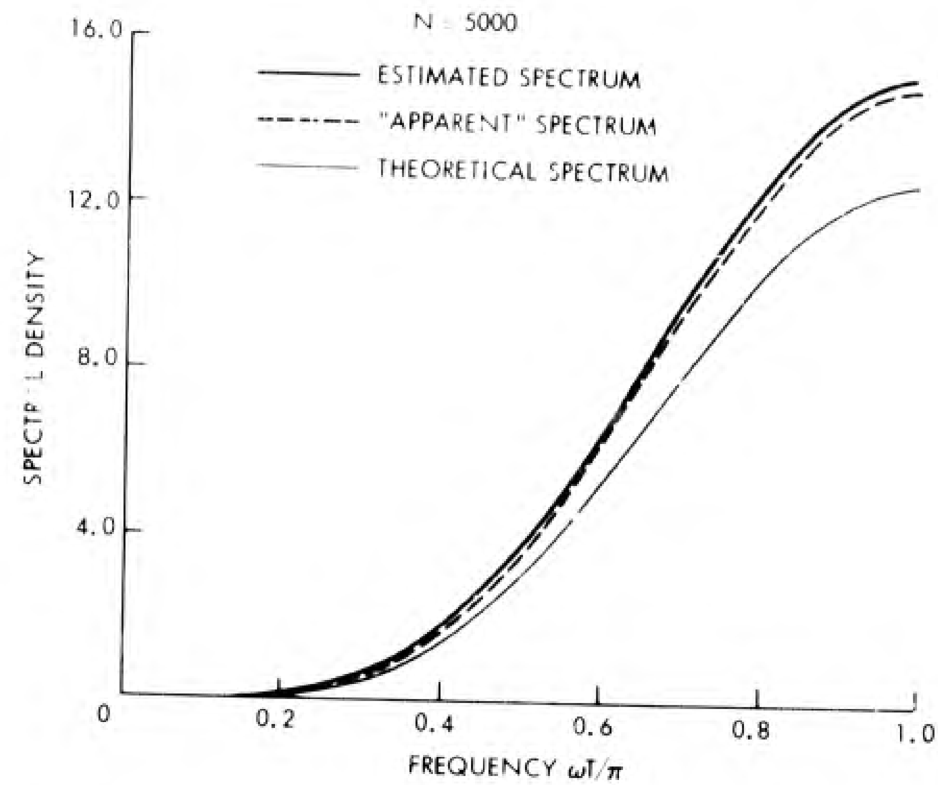


FIG. 5-19 AN ESTIMATE OF THE MOVING AVERAGE SPECTRUM OF EXAMPLE #5 FOUND WITH WOLD'S METHOD FROM 5,000 SAMPLES, IN COMPARISON WITH THE THEORETICAL AND "APPARENT" SPECTRA

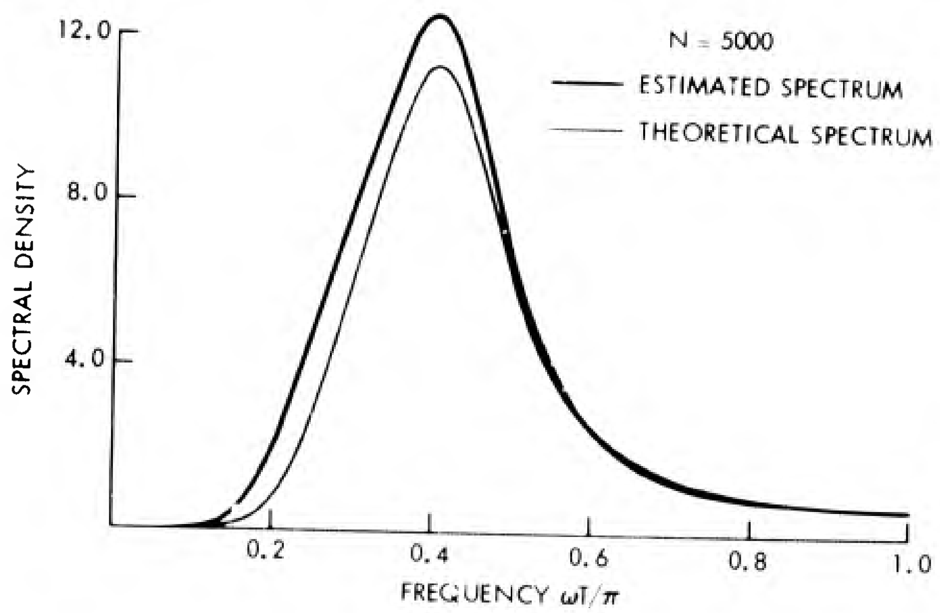


FIG. 5-20 AN ESTIMATE OF THE TOTAL SPECTRUM FOR EXAMPLE #5 BASED ON 5,000 SAMPLES AND USING WOLD'S APPROACH TO FIND THE MOVING AVERAGE PART

able difference can be attributed to the spectral error inherent in the approximation to y_t, which will be recalled as quite large due to the errors made in estimating the $\{b_j\}$. The apparent spectrum, originally presented in Figure 4-10, is also reproduced on Figure 5-19 to show the fairly close agreement with the result found with Wold's method. The remaining difference is due to statistical error in the moving average mean lagged products, and its average variance is given by

$$\mathrm{Var}_{avg} = \frac{295.6}{N} = 0.00059$$

The corresponding standard deviation is 0.077.

The total process system function becomes

$$\frac{\hat{A}(z)}{\hat{B}(z)} = \frac{1 - 2z^{-1} + z^{-2}}{1 - 1.49862z^{-1} + 1.45736z^{-2} - 0.76938z^{-3} + 0.27201z^{-4}}$$

with $\hat{\sigma}^2 = 0.95$. The final spectral result is presented in Figure 5-20, where it is apparent that the relatively poor results of the first part of the analysis (in Chapter IV) have carried over to reduce the quality of the estimate. Even so, the estimate is surprisingly good, considering the use of the approximations made in locating both zeros of $\hat{A}(z)$.

The foregoing calculations have demonstrated the difficulties that can arise in applying Wold's method when the mean lagged product sequence places a root of V(r) in the range $-2<r<2$. It is surprising, however, how little the final estimate suffers when the offending roots are replaced by acceptable approximations. This is one indication of a fortunate peculiarity of the Wold approach. Although the coefficient and variance estimates often display substantial error in comparison with the actual values, the final spectral result as a whole seems relatively insensitive to these deviations. Thus, even though the estimated system pole-zero pattern differs markedly from the actual pattern, the spectral estimate retains a reasonable quality. If accurate knowledge of the moving average coefficient values is not really required, and only the spectral estimate is of interest, Wold's method overcomes many of its more obvious shortcomings.

3. Moving Average Estimation by Spectral Inversion

In preparing computational examples of the treatment of moving average processes by the technique of spectral inversion, the intermediate model order (heretofore denoted M) was limited for convenience to 8, and this value was used in all of the final estimates presented here. It is apparent in several examples, however, that improved results could have been obtained by increasing the order of the intermediate autoregressive model beyond this limit, and the present results should not be interpreted as the best that can be done with the method in these cases. It is of considerable practical interest to determine the effect of some such limitation on the estimate, and this is possible to a large degree in the present findings.

a. Example # 2

The spectral inversion estimate for the second example, a pure moving average, is based on the sequence of mean lagged products measured over 1,000 samples and presented in Appendix A. The estimate formed by setting the intermediate model order M equal to 8 follows from the 8^{th} order all-pole estimate previously computed in Chapter III. The sequence of coefficients for that model was found to be

$$\{\hat{d}_j\}_{j=0}^{8} = 1, -2.00950, 2.27206, -1.75831, 0.89373, -0.11857, -0.29835, 0.31628, -0.14374$$

and the variance estimate was 1.02113. From equation (5-51) is computed the correlation sequence corresponding to the moving average process v_t defined in the derivation:

$$\{\hat{\phi}_{vv}(j)\}_{j=0}^{\infty} = 14.31450, -12.35211, 7.78321, -2.99934, -0.23047, 1.45233, -1.26050, 0.60512, -0.14374, 0, 0, \ldots$$

With these in hand, equation (5-34) becomes

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \end{bmatrix} = - \begin{bmatrix} 14.31450 & -12.35210 & 7.78321 & -2.99934 \\ -12.35210 & 14.31450 & -12.35210 & 7.78321 \\ 7.78321 & -12.35210 & 14.31450 & -12.35210 \\ -2.99934 & 7.78321 & -12.35210 & 14.31450 \end{bmatrix}^{-1} \begin{bmatrix} -12.35210 \\ 7.78321 \\ -2.99934 \\ -0.23047 \end{bmatrix}$$

$$= \begin{bmatrix} 2.00824 \\ 1.77101 \\ 0.75126 \\ 0.12221 \end{bmatrix}$$

The input noise variance estimate may be set equal to that found for the 8^{th} order autoregressive model (in this case 1.02113) or computed from equation (5-20), in which case the estimate becomes 1.035. When M is sufficiently large that the approximation error becomes negligible (i.e. when an autoregressive model of that order suffices to represent x_t), the two variance estimates become identical. When substantial approximation error is present, however, the latter is to be preferred since it is constructed to give the process model the same mean square value as that observed experimentally.

To separate the effects of approximation and statistical error in the coefficient estimates, a spectral inversion analysis was performed using the theoretical correlation sequence for the original process as the basis. Again, M was set equal to 8. A comparison of this result (which yields the coefficient set $\{a_j'\}$) with the experimental result (the $\{\hat{a}_j\}$) and the actual values is shown in the following table, along with the appropriate input variance estimate:

Table 5-5

Spectral Inversion Results for Example # 2

j	a_j	$\hat{a}_j$	a_j'	$\xi_j = \hat{a}_j - a_j'$	$\sigma(\xi_j)$
1	2.036	2.00824	2.03068	-0.02244	0.0313
2	1.812	1.77101	1.81735	-0.04634	0.0666
3	0.770	0.75126	0.80359	-0.05233	0.0666
4	0.129	0.12221	0.14115	-0.01895	0.0313
Variance Estimate	1.0	1.035	0.994	-----	-----

The difference between $\hat{a}_j$ and a_j', representative of the statistical error in the result, is also presented for comparison with the standard deviation of this error as computed from Durbin's approximate expression, given above as equation (5-69). The agreement is very good. The three corresponding spectra are displayed in Figure 5-21, where it is seen that the curves are practically indistinguishable. Thus, both statistical and approximation errors are very small, and the former is consistent with the small value of the average spectral standard deviation (computed for the Wold approach, but applicable here) of 1.11.

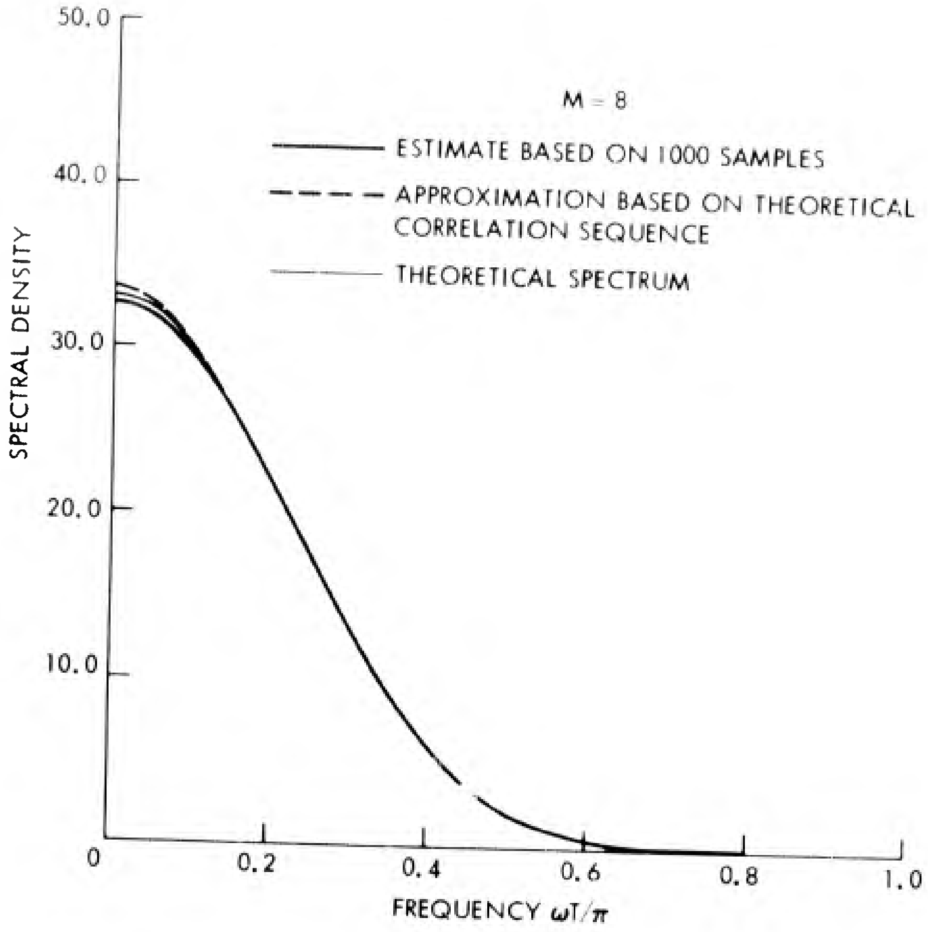


FIG. 5-21 SPECTRAL ESTIMATE FOR EXAMPLE #2 FOUND BY SPECTRAL INVERSION FROM 1,000 SAMPLES. THIS IS COMPARED WITH THE THEORETICAL SPECTRUM AND THE RESULT OBTAINED BY SPECTRAL INVERSION FROM THE THEORETICAL AUTOCORRELATION SEQUENCE.

The accuracy of the coefficient estimates in the present case indicates that an intermediate model order of 8 is sufficient to supply a good representation for the process. This has already been demonstrated by the behavior of the variance estimate for this example in Table 3-3. To demonstrate the effect of the intermediate model order on the spectral approximation, however, Figure 5-22 displays a sequence of estimates all based on the theoretical autocorrelation sequence, but with M ranging between unity and 8. The corresponding coefficient and variance estimates are found in the following table, the latter having been computed from equation (5-20):

Table 5-6

Spectral Inversion Estimates for Example # 2 as M Varies Between 1 and 8

M	a_1'	a_2'	a_3'	a_4'	$\hat{\sigma}^2$
1	0.74521	0.52819	0.33793	0.16479	4.070
2	1.28547	1.03794	0.57509	0.18193	2.210
3	1.66013	1.44487	0.74625	0.18838	1.341
4	1.89785	1.70763	0.83416	0.18041	1.062
5	2.00882	1.80283	0.83169	0.15962	0.995
6	2.03151	1.80545	0.81426	0.15236	0.999
7	2.02889	1.80946	0.81593	0.15055	0.996
8	2.03068	1.81735	0.80359	0.14115	0.994
Actual Values	2.036	1.812	0.770	0.129	1.0

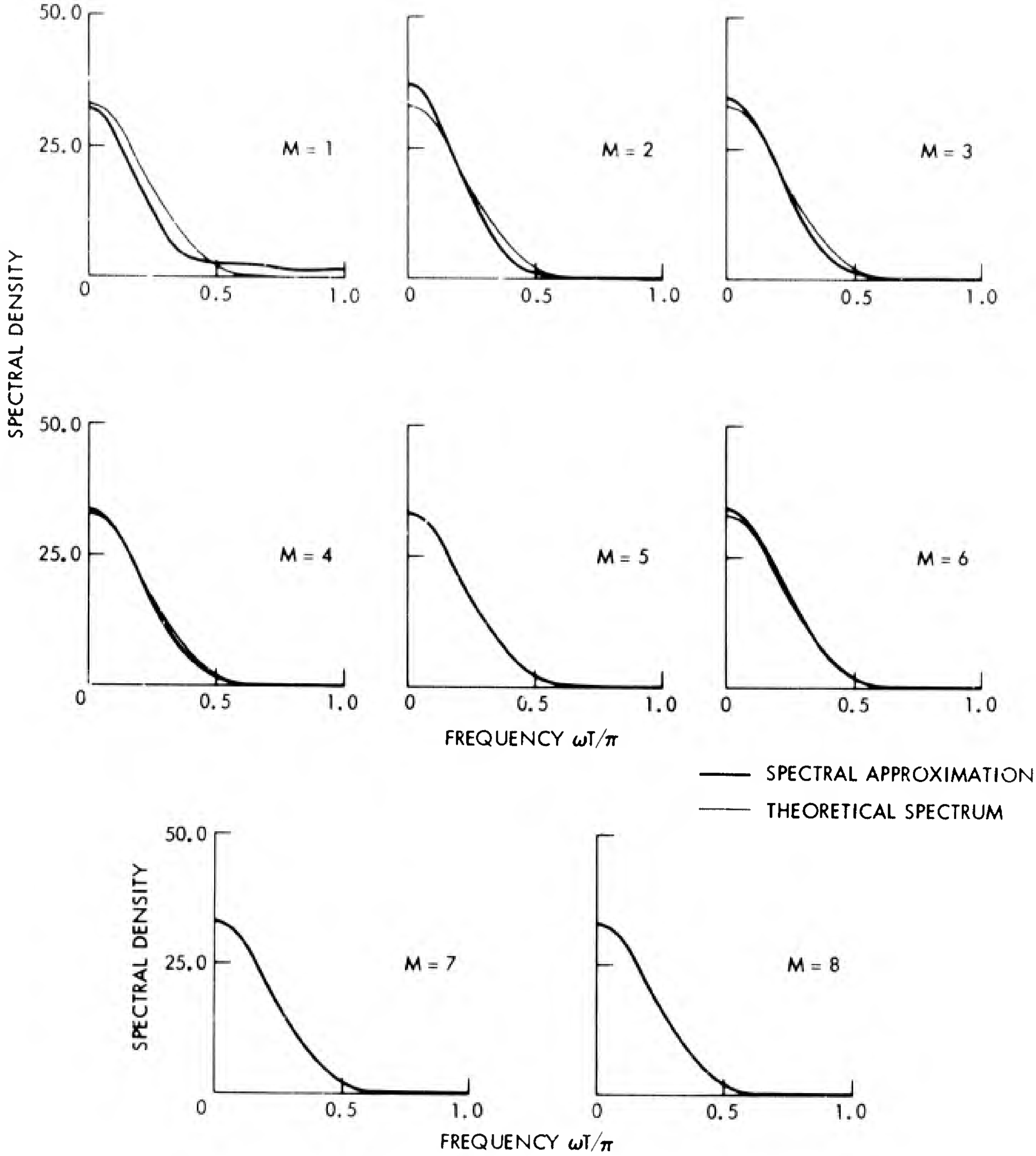


FIG. 5-22 SPECTRAL APPROXIMATIONS FOR EXAMPLE #2 FOR SUCCESSIVE INTERMEDIATE MODEL ORDERS. THE CALCULATIONS ARE BASED ON THE THEORETICAL AUTOCORRELATION SEQUENCE.

As can be seen from both table and graphs, the law of diminishing returns begins to set in when M = 5 or so, and the present spectrum can be more than adequately analyzed by the spectral inversion technique.

b. Example # 3

Using the 1,000 sample data for the moving average part of the third example (in Table 5-1), the spectral inversion estimate with M = 8 produces the values

$$\hat{a}_1 = -0.752994 \qquad (a_1 = -0.859)$$

and

$$\hat{\sigma}^2 = 1.13 \qquad (\sigma^2 = 1.0)$$

The latter estimate results from the application of equation (5-20), whereas the variance estimate for the 8^{th} order autoregressive model is 1.02593. The large difference indicates that larger values of M can be expected to improve the result substantially. Using the theoretical correlation sequence for the moving average part provides the numbers

$$a_1' = -0.80999$$
$$\hat{\sigma}^2 = 1.05$$

and the difference between a_1' and $\hat{a}_1$ is rather large to be explained in terms of the statistical error. (Durbin's result indicates that the statistical error standard deviation should be about 0.018). The disparity is probably caused by the errors inherent in the sequence repsesenting the moving average part, due to inaccuracy in estimating the $\{b_j\}$. This effect was absent in our treatment of the example just above. Nonetheless, a comparison of the estimated moving average spectrum with the actual function, as in Figure 5-23, shows very good agreement. When this result is combined with the estimates for the autoregressive coefficients, the total system function estimate for the third example becomes

$$\frac{\hat{A}(z)}{\hat{B}(z)} = \frac{1 - 0.75299z^{-1}}{1 - 0.00585z^{-1} + 0.85445z^{-2}}$$

with $\hat{\sigma} = 1.13$. The corresponding spectrum is compared with the nominal spectrum in Figure 5-24, and again, the estimate is a very good one.

c. Example # 4

The spectrum inversion treatment of the moving average part of the fourth example is based on the 5,000 sample mean lagged products of Table 5-1. Again, with M = 8, the technique provides the following table of estimates, errors, and error standard deviations:

Table 5-7

Spectral Inversion Results for Example # 4

j	a_j	$\hat{a}_j$	a_j'	$\xi_j = \hat{a}_j - a_j'$	$\sigma(\xi_j)$
1	-1.78350	-1.58333	-1.56936	-0.01397	0 0103
2	0.79300	0.69102	0.67772	+0.01330	0.0103
Variance Estimate	1.0	1.238	1.225	----	----

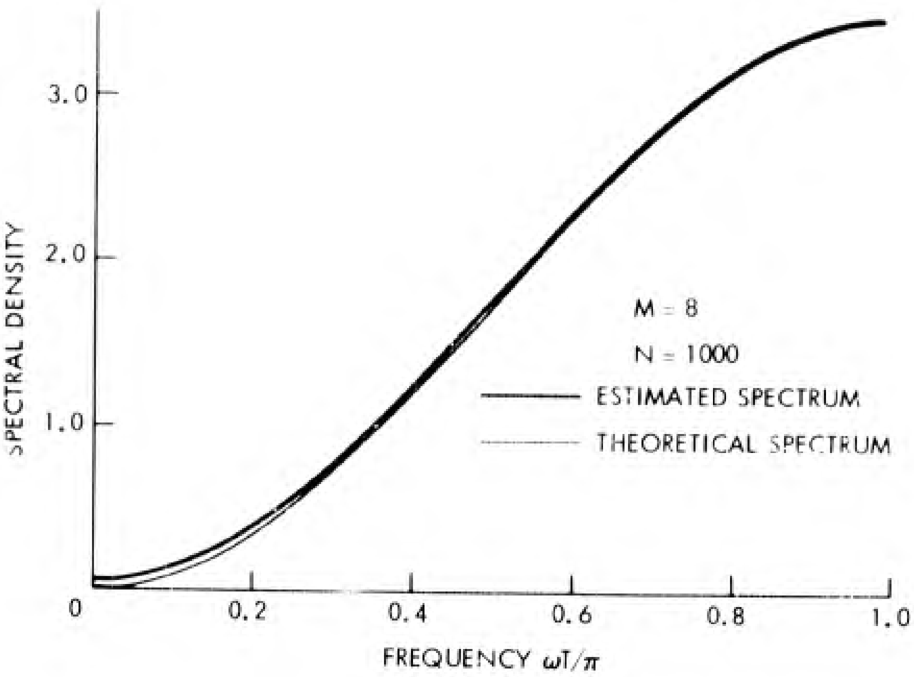


FIG. 5-23 AN ESTIMATE OF THE MOVING AVERAGE SPECTRUM OF EXAMPLE #3 FOUND WITH THE SPECTRAL INVERSION TECHNIQUE FROM 1,000 SAMPLES

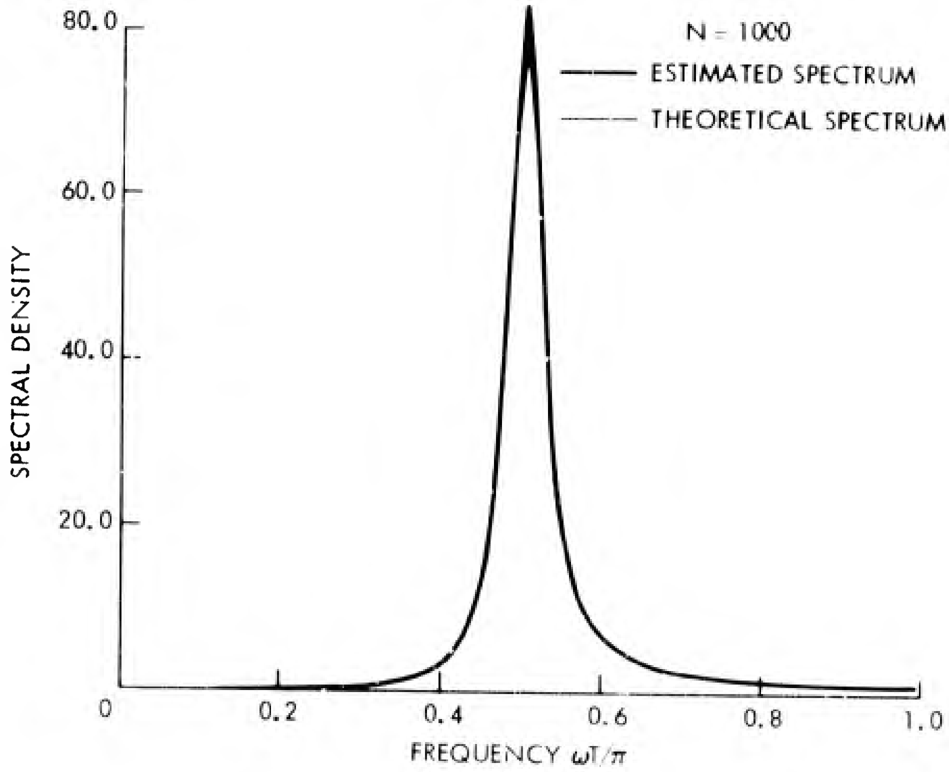


FIG. 5-24 TOTAL SPECTRAL ESTIMATE FOR EXAMPLE #3 BASED ON 1,000 SAMPLES WHEN THE MOVING AVERAGE PART HAS BEEN DETERMINED BY SPECTRAL INVERSION

Here, the $\{a_j'\}$ are the coefficients based on the theoretical correlation sequence, the variance estimates come from equation (5-20), and the standard deviations are computed from Durbin's formula. The 5,000 sample 8^{th} order all-pole variance estimate is 1.211, sufficiently different from the value presented above to indicate that a higher value of M would be desirable. The theoretical error standard deviations predict the magnitude of the statistical error quite well despite the error inherent in the moving average data due to the non-exact filtering operation. The estimated moving average spectrum is compared to the actual spectrum in Figure 5-25, where good agreement is found. The average spectral standard deviation was computed previously to be 0.078.

When the moving average results are combined with the autoregressive coefficient estimates of Chapter IV, the final system function estimate, based on 5,000 samples, becomes

$$\frac{\hat{A}(z)}{\hat{B}(z)} = \frac{1 - 1.58333z^{-1} + 0.69102z^{-2}}{1 - 1.33564z^{-1} + 1.62015z^{-2} - 0.97754z^{-3} + 0.65017z^{-4}}$$

with $\hat{\sigma}^2 = 1.238$. The corresponding spectrum is shown in Figure 5-26.

d. Example # 5

The results of the 5,000 sample spectral inversion analysis of the moving average part of the last example are presented in our usual comparative table:

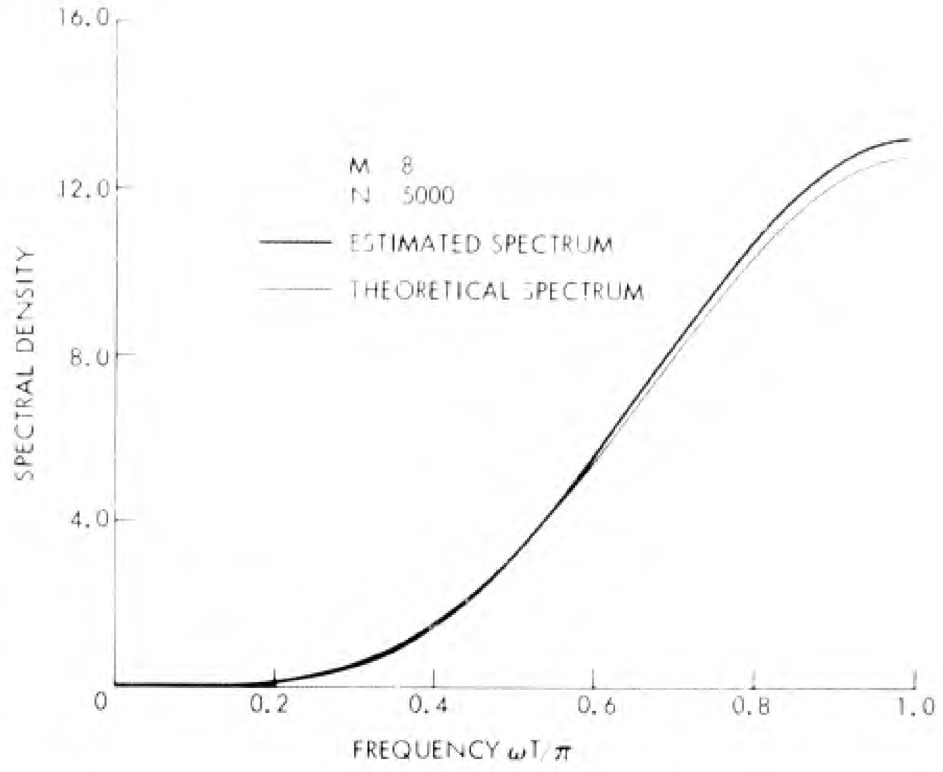


FIG. 5-25 AN ESTIMATE OF THE MOVING AVERAGE SPECTRUM OF EXAMPLE #4 FOUND WITH THE SPECTRAL INVERSION TECHNIQUE FROM 5,000 SAMPLES

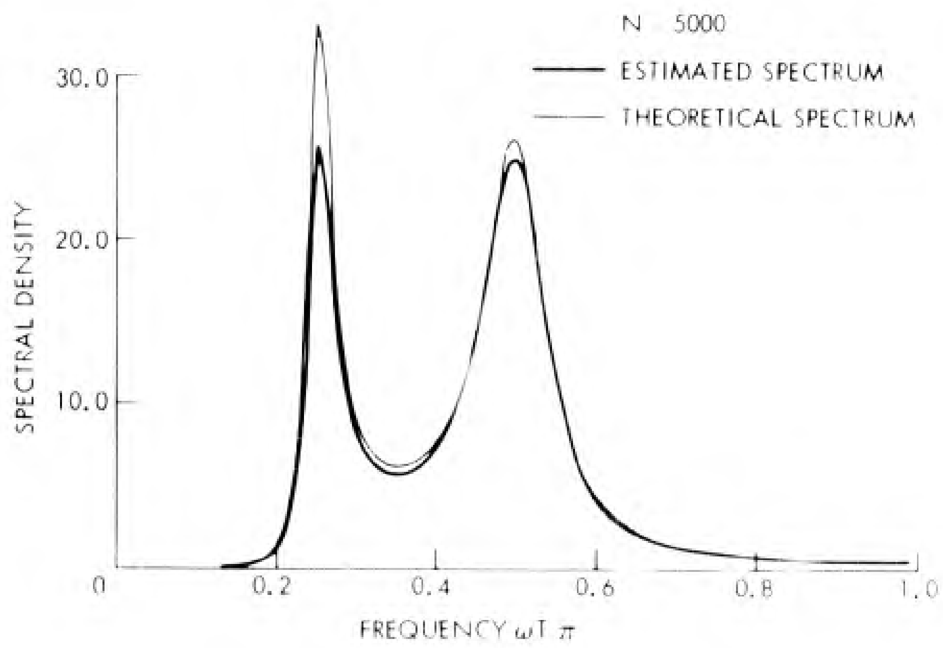


FIG. 5-26 TOTAL SPECTRAL ESTIMATE FOR EXAMPLE #4 BASED ON 5,000 SAMPLES WHEN THE MOVING AVERAGE PART HAS BEEN DETERMINED BY SPECTRAL INVERSION

Table 5-8

Spectral Inversion Results for Example # 5

j	a_j	$\hat{a}_j$	a_j'	$\xi_j = \hat{a}_j - a_j'$	$\sigma(\xi_j)$
1	-1.74	-1.68805	-1.61669	-0.07136	0.0096
2	0.81	0.79433	0.72432	+0.07000	0.0096
Variance Estimate	1.0	1.272	1.133	-----	----

As in the present treatment of the other examples, M = 8. The statistical error is far too large to be explainable in terms of Durbin's standard deviation, and must be due to the substantial errors made in estimating the autoregressive coefficients. It is also apparent that an increased intermediate model order would benefit the quality of the moving average estimates. The estimated spectrum for y_t is presented in Figure 5-27 in comparison with the actual spectrum and with the "apparent" spectrum (from Figure 4-10) with which the present analysis necessarily deals. Naturally, the estimate lies closer to the latter. For the final process system function, the previously made estimates for the $\{b_j\}$ are combined with the moving average results to yield

$$\frac{\hat{A}(z)}{\hat{B}(z)} = \frac{1 - 1.68805z^{-1} + 0.79433z^{-2}}{1 - 1.49862z^{-1} + 1.45736z^{-2} - 0.76938z^{-3} + 0.27201z^{-4}}$$

with $\hat{\sigma}^2 = 1.272$. Considering the magnitude of the errors in both the coefficients and the input variance estimate, the corresponding spectrum is a rather good representation of

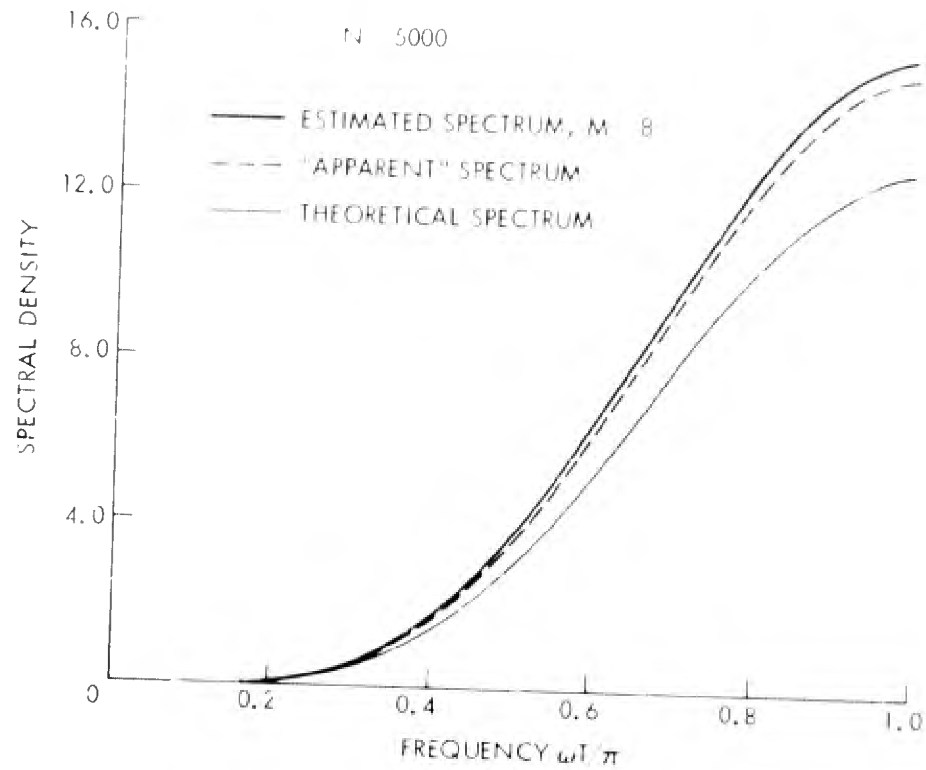


FIG. 5-27 AN ESTIMATE OF THE MOVING AVERAGE SPECTRUM OF EXAMPLE #5 FOUND WITH THE SPECTRAL INVERSION TECHNIQUE FROM 5,000 SAMPLES. THE THEORETICAL AND "APPARENT" SPECTRA ARE DISPLAYED FOR COMPARISON.

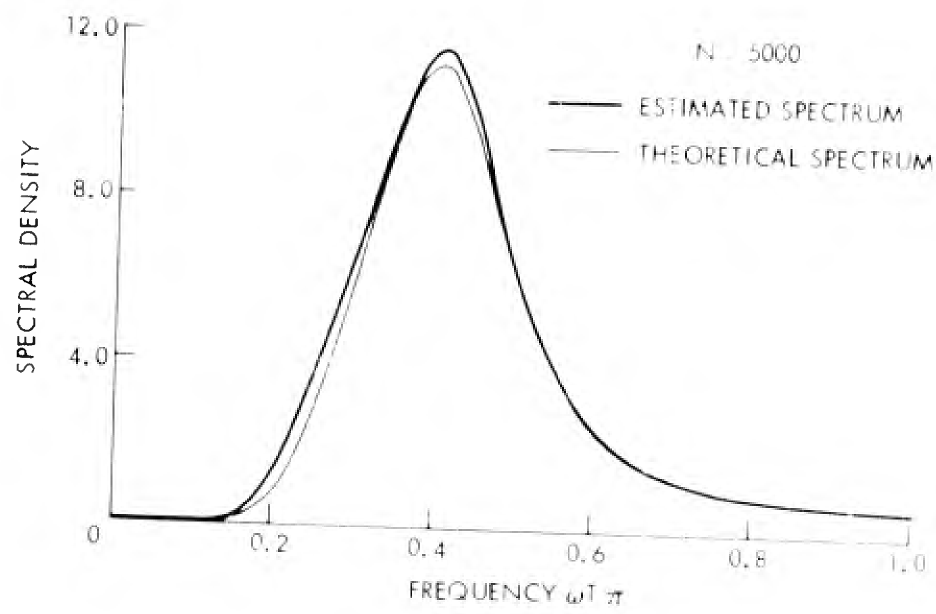


FIG. 5-28 TOTAL SPECTRAL ESTIMATE FOR EXAMPLE #5 BASED ON 5,000 SAMPLES WHEN THE MOVING AVERAGE PART HAS BEEN DETERMINED BY SPECTRAL INVERSION

the actual function, as shown in Figure 5-28. The spectral inversion method also seems to give acceptable spectral results in the face of significant parameter errors. The present example is one in which it has been apparent throughout that the original sample size is too small. Nonetheless, the final spectral results are quite good.

The present section has served to demonstrate that both Wold's method and the spectral inversion technique are superior to all-pole methods in dealing with spectral functions containing influential zeros. Of these two approaches, the Wold procedure seems to provide a slightly better total spectral estimate when no root approximations are required and the exact solution exists. This fails to happen much of the time, however, particularly when the data sequence in question is derived from filtering the original process with the estimated autoregressive polynomial. In these cases, the spectral inversion method comes into its own and has the added advantage of computational simplicity. It should be remembered, however, that both of these methods require a priori knowledge of the numerator and denominator orders of the original process, whereas the all-pole estimates can be carried through without this information.

CHAPTER VI

EXTRAPOLATION OF ESTIMATES BASED ON ADDITIONAL DATA

After a spectral estimate has been prepared on the basis of a certain number of data samples, the problem may arise of revising this estimate to include the additional information contained in further quantities of data which become available. It is apparent that the new data could simply be lumped with the old, mean lagged products re-computed, and the estimation procedure in use started from the beginning. Although this approach is straightforward, it is computationally inefficient, particularly since each of the estimation schemes described above requires at least one matrix inversion. The present section studies the means of avoiding the repeated inversion of the covariance matrix when the spectral calculations are extended over a new increment of data. In the first of these, suitable when a large number of additional samples is introduced, the resulting revision of the estimates is an approximation, valid under realistic conditions. In the other, applicable for a limited number of additional samples, an exact answer is obtained.

A. Extrapolation Based on a Block of Additional Samples

All-pole estimation will be used to demonstrate the two extrapolation methods treated here. The key equation is given in matrix form as

$$\hat{\phi}_{xx}\hat{d} = -\hat{\phi}_{xx} \qquad (6\text{-}1)$$

where the several matrices are defined as in Chapter III, and the superscript carats denote that the calculations are based on the mean lagged products of a sample sequence. Henceforth, the "xx" subscripts will be dropped for notational convenience. If the initial estimates are based on N_1 data samples, the mean lagged products are computed as

$$\hat{\phi}_1(j) = \frac{1}{N_1} \sum_{k=1}^{N_1} x_k x_{k+j} \qquad (6\text{-}2)$$

and the estimates themselves are found to be

$$\hat{d}_1 = -\hat{\phi}_1^{-1}\hat{\phi}_1 \qquad (6\text{-}3)$$

where the elements of $\hat{\phi}_1$ and $\hat{\phi}_1$ are computed from equation (6-2). Now suppose that N_2 additional samples are taken so that the total sample size is now

$$N = N_1 + N_2 \qquad (6\text{-}4)$$

The optimum estimates are found for the larger set using mean lagged products defined as

$$\hat{\phi}(j) = \frac{1}{N_1+N_2} \sum_{k=1}^{N_1+N_2} x_k x_{k+j} \qquad (6\text{-}5)$$

which can be written as

$$\hat{\phi}(j) = \frac{1}{N_1+N_2}\left[\sum_{k=1}^{N_1} x_k x_{k+j} + \sum_{k=N_1+1}^{N_1+N_2} x_k x_{k+j}\right] \tag{6-6}$$

$$= \frac{N_1}{N_1+N_2}\hat{\phi}_1(j) + \frac{N_2}{N_1+N_2}\hat{\phi}_2(j)$$

where

$$\hat{\phi}_2(j) \equiv \frac{1}{N_2}\sum_{k=N_1+1}^{N_1+N_2} x_k x_{k+j} \tag{6-7}$$

is the mean lagged product at lag j for the new increment of data. The new estimate, based on all of the data, must satisfy the equation

$$\left[\frac{N_1}{N_1+N_2}\hat{\phi}_1 + \frac{N_2}{N_1+N_2}\hat{\phi}_2\right]\hat{d} = -\frac{N_1}{N_1+N_2}\hat{\phi}_1 - \frac{N_2}{N_1+N_2}\hat{\phi}_2 \tag{6-8}$$

Now let

$$\hat{d} = \hat{d}_1 + e \tag{6-9}$$

so that e represents a column vector of corrections based on the new information. Since

$$\hat{\phi}_1 d_1 = -\hat{\phi}_1 \tag{6-10}$$

equations (6-8) and (6-9) combine to yield

$$(N_1\hat{\phi}_1 + N_2\hat{\phi}_2)e = -N_2(\hat{\phi}_2\hat{d}_1 + \hat{\phi}_2) \tag{6-11}$$

Dividing both sides by N_1 and defining the matrix Δ as

$$\Delta \equiv \frac{N_2}{N_1} \hat{\Phi}_2 \tag{6-12}$$

equation (6-11) becomes

$$(\hat{\Phi}_1 + \Delta)e = -(\Delta\hat{d}_1 + \frac{N_2}{N_1} \hat{\phi}_2) \tag{6-13}$$

Since

$$E[\hat{\Phi}_2] = E[\hat{\Phi}_1] = \Phi \tag{6-14}$$

and because usually $N_2 << N_1$, the matrix Δ can be considered to be a small perturbation of the matrix $\hat{\Phi}_1$. The solution for e becomes

$$e = -\left[\hat{\Phi}_1 + \Delta\right]^{-1} \left[\Delta d_1 + \frac{N_2}{N_1} \hat{\phi}_2\right] \tag{6-15}$$

and can be obtained efficiently if the inverse can be found in terms of $\hat{\Phi}_1^{-1}$, which has already been computed to obtain the first estimate.

Faddeeva[37] shows that if A is any square matrix and I is the corresponding identity matrix, then

$$\lim_{m \to \infty} \left[I + \sum_{j=1}^{m} A^j\right] = (I - A)^{-1} \tag{6-16}$$

if and only if $||A||<1$, where $||A||$ is an appropriately defined norm for the matrix A. Furthermore, if the series in equation (6-16) is truncated after the k^{th} term, then

$$||(I - A)^{-1} - (I + A + \ldots + A^k)|| \le \frac{||A||^{k+1}}{1 - ||A||} \qquad (6\text{-}17)$$

where the left hand side can be identified as the norm of the matrix representing the difference between $(I - A)^{-1}$ and the k^{th} order approximating series. Following Nash and Tuteur[38], the matrix $[\hat{\Phi}_1 + \Delta]$ is written as

$$[\hat{\Phi}_1 + \Delta] = \hat{\Phi}_1(I + \hat{\Phi}_1^{-1}\Delta) \qquad (6\text{-}18)$$

and thus

$$[\hat{\Phi}_1 + \Delta]^{-1} = (I + \hat{\Phi}_1^{-1}\Delta)^{-1}\hat{\Phi}_1^{-1} \qquad (6\text{-}19)$$

This is in the form where Faddeeva's result can be applied if the matrix $-\hat{\Phi}_1^{-1}\Delta$ is identified with A above, Faddeeva proposes several matrix norms, but the following will be used here:

$$||A|| = \max_j \sum_{k=1}^{M} |a_{jk}| \qquad (6\text{-}20)$$

If $||-\hat{\Phi}_1^{-1}\Delta|| < 1$, the expansion of equation (6-16) will converge and can be used to provide a first order approximation for $[\hat{\Phi}_1 + \Delta]^{-1}$. The norm here is a random variable, but since by equation (6-12),

$$\hat{\Phi}_1^{-1}\Delta = \frac{N_2}{N_1}\hat{\Phi}_1^{-1}\hat{\Phi}_2 \qquad (6\text{-}21)$$

and because of equation (6-14), it can be seen that $||-\hat{\Phi}_1^{-1}\Delta||$ will be on the order of N_2/N_1. Since normally, $N_2 << N_1$, the series of equation (6-16) will usually converge, and the inverse can be written

$$\left[\hat{\Phi}_1 + \Delta\right]^{-1} = \left[I + \sum_{j=1}^{\infty} (-\hat{\Phi}_1^{-1}\Delta)^j\right]\hat{\Phi}_1^{-1} \tag{6-22}$$

If only the linear term in the series is retained, then

$$\left[\hat{\Phi}_1 + \Delta\right]^{-1} \simeq \hat{\Phi}_1^{-1} - \hat{\Phi}_1^{-1}\Delta\,\hat{\Phi}_1^{-1} \tag{6-23}$$

By equation (6-17), the norm of the error matrix will be bounded by the expression

$$||E|| \leq \frac{||-\hat{\Phi}_1^{-1}\Delta||^2}{1 - ||-\hat{\Phi}_1^{-1}\Delta||}\,||\hat{\Phi}_1^{-1}|| \tag{6-24}$$

Thus, the percentage error in norm in approximating $[\hat{\Phi}_1 + \Delta]^{-1}$ by the expression of equation (6-23) is given by

$\dfrac{||-\hat{\Phi}_1^{-1}\Delta||^2}{1 - ||-\hat{\Phi}_1^{-1}\Delta||}$ and can easily be evaluated in a practical calculation since the product $\hat{\Phi}_1^{-1}\Delta$ is already required in computing the approximate inverse. For planning purposes, it should be noted that in accordance with considerations above, this percentage will be roughly on the order of $\dfrac{(N_2/N_1)^2}{1 - N_2/N_1}$.

Now if the approximate inverse is used in equation (6-15), the final result becomes

$$e \simeq -\left[\hat{\Phi}_1^{-1} - \hat{\Phi}_1^{-1}\Delta\,\hat{\Phi}_1^{-1}\right]\left[\Delta d_1 + \frac{N_2}{N_1}\hat{\Phi}_2\right] \tag{6-25}$$

which can be evaluated from quantities previously calculated and mean lagged products from the new data segment. In particular, no new matrix inversion is required. By studying the form of the estimation equations, it becomes clear that the percentage error in the elements of e caused by using the approximate inverse is also on the order of $\dfrac{\|-\hat{\Phi}_1^{-1}\Delta\|^2}{1 - \|-\hat{\Phi}_1^{-1}\Delta\|}$.

These results apply, with only notational modification, to the matrix equation for estimating autoregressive coefficients in a mixed-type process. Its use in both stages of the spectral inversion technique follows directly and will not be further discussed here.

As a computational example of this procedure, the extrapolation of an all-pole estimate for Example # 1 will be used. If 1,000 samples of the process are taken, the resulting sequence of mean lagged products is

$$\{\hat{\phi}_1(j)\}_{j=0}^{4} = 15.32552,\ 13.18281,\ 8.13402,\ 2.73565,\ -0.99685$$

The mean lagged products computed over the next 100 samples are

$$\{\hat{\phi}_1(j)\}_{j=0}^{4} = 11.59311,\ 9.76879,\ 5.71915,\ 1.73212,\ -0.72506$$

and for the entire block of 1,100 samples, the result is

$$\{\hat{\phi}(j)\}_{j=0}^{4} = 14.98624,\ 12.87241,\ 7.91449,\ 2.64442,\ -0.97214$$

In Chapter III, the estimate vector found from the block of $N_1 = 1{,}000$ samples was presented as

$$\hat{d}_1 = \begin{bmatrix} -2.00864 \\ 1.74420 \\ -0.69524 \\ 0.09590 \end{bmatrix}$$

If this same vector is estimated on the basis of the entire block of 1,100 samples by inverting the corresponding covariance matrix, the result is

$$\hat{d} = \begin{bmatrix} -1.98872 \\ 1.71170 \\ -0.67548 \\ 0.09201 \end{bmatrix}$$

It is this answer that we wish to reach using the present method on an additional 100 samples. Thus, the ratio $N_2/N_1 = 0.1$, and evaluating equation (6-25) with the data presented above, the correction vector is approximated as

$$e = \begin{bmatrix} 0.01921 \\ -0.03093 \\ 0.01837 \\ -0.00339 \end{bmatrix}$$

The actual correction vector is given by

$$e = \hat{d} - \hat{d}_1 = \begin{bmatrix} 0.01992 \\ -0.03250 \\ 0.01976 \\ -0.00389 \end{bmatrix}$$

and it can be seen that the agreement is very close. The respective percentage errors in the elements of the approximate correction vector are 3.56, 4.83, 7.05, 12.81 per cent. In the present case, the matrix $-\hat{\Phi}_1^{-1}\Delta$ becomes

$$-\hat{\Phi}_1^{-1}\Delta = -\begin{bmatrix} 0.06410 & -0.02938 & 0.00651 & 0.01811 \\ 0.03675 & 0.13087 & -0.03884 & -0.04389 \\ -0.04389 & -0.03884 & 0.13087 & 0.03675 \\ 0.01811 & 0.00651 & -0.02938 & 0.06410 \end{bmatrix}$$

and by equation (6-20)

$$||\Phi_1^{-1}\Delta|| = 0.25035$$

The predicted percentage error in the correction vector is thus on the order of

$$\frac{\left\| -\hat{\Phi}_1^{-1}\Delta \right\|^2}{1 - \left\| -\hat{\Phi}_1^{-1}\Delta \right\|} = 0.0834$$

or 8.34 per cent, and this is in agreement with the actual findings. On the other hand,

$$\frac{(N_2/N_1)^2}{1 - (N_2/N_1)} = 0.0111$$

or 1.11 per cent, and this is a fairly poor prediction of the actual error.

B. Extrapolation of Estimates for Several Additional Samples

Again, attention will be concentrated on all-pole estimates with the basic equation given by

$$\hat{d}_N = -\hat{\Phi}_N^{-1} \hat{\phi}_N \tag{6-26}$$

where the estimation order is considered to be M, and the subscript denotes that the estimates are based on a sample size of N elements. For the purposes of the present argument, the matrix of mean lagged products must be represented in the following form:

$$\hat{\Phi}_N = \frac{1}{N} \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_N \\ x_2 & x_3 & x_4 & & x_{N+1} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ x_M & x_{M+1} & \cdots\cdots & & x_{M+N+1} \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ x_2 & x_3 & \cdots & x_{M+1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ x_N & x_{N+1} & \cdot\cdot & x_{M+N+1} \end{bmatrix} \qquad (6\text{-}27)$$

If the right hand matrix, of dimension N x M, is denoted X_N, then

$$\hat{\Phi}_N = \frac{1}{N} X_N^T X_N \qquad (6\text{-}28)$$

It should be noted that this definition of the matrix of mean lagged products is somewhat different from that which has been used previously. This is best demonstrated with an example, in this case with M = 4 and N = 8:

$$\hat{\Phi}_8 = \frac{1}{8}\begin{bmatrix} x_1 & \cdots\cdots & x_8 \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ x_4 & \cdots\cdots & x_{11} \end{bmatrix} \begin{bmatrix} x_1 & \cdot\;\cdot\;\cdot & x_4 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_8 & \cdot\;\cdot\;\cdot & x_{11} \end{bmatrix}$$

$$= \frac{1}{8}\begin{bmatrix} \sum_1^8 x_j^2 & \sum_1^8 x_j x_{j+1} & \sum_1^8 x_j x_{j+2} & \sum_1^8 x_j x_{j+3} \\ \sum_1^8 x_j x_{j+1} & \sum_2^9 x_j^2 & \sum_2^9 x_j x_{j+1} & \sum_2^9 x_j x_{j+2} \\ \sum_1^8 x_j x_{j+2} & \sum_2^9 x_j x_{j+1} & \sum_3^{10} x_j^2 & \sum_3^{10} x_j x_{j+1} \\ \sum_1^8 x_j x_{j+3} & \sum_2^9 x_j x_{j+2} & \sum_3^{10} x_j x_{j+1} & \sum_4^{11} x_j^2 \end{bmatrix} \tag{6-29}$$

Evidently, $E[\hat{\Phi}_8] = \Phi$, but unlike the mean lagged product matrices used in practice, the entries for a given lag are not identically equal, in the sense, for example, that

$$\hat{\phi}_{12} = \sum_{j=1}^{8} x_j x_{j+1} \neq \hat{\phi}_{23} = \sum_{j=2}^{9} x_j x_{j+1} \tag{6-30}$$

since the summation limits are different. As N grows very large, however, this difference becomes negligible, and the form of equation (6-27) is, for all practical purposes, identical with the previous definition.

The effect on the matrix of mean lagged products of increasing the sample size by K elements is found by bordering the matrices X_N and X_N^T by the respective matrices X_K^T and X_K,

$$X_K = \begin{bmatrix} x_{N+1} & x_{N+2} & \cdots & x_{N+K} \\ \vdots & & & \vdots \\ x_{N+M} & \cdots & \cdots & x_{N+M+K-1} \end{bmatrix} \qquad (6\text{-}31)$$

which has dimension M x K. The result is

$$\hat{\Phi}_{N+K} = \frac{1}{N+K} \left[\begin{array}{c|ccc} X_N^T & x_{N+1} & \cdots & x_{N+K} \\ & x_{N+M} & \cdots & x_{N+M+K-1} \end{array} \right] \left[\begin{array}{ccc} & X_N & \\ \hline x_{N+1} & \cdots & x_{N+M} \\ \vdots & & \vdots \\ x_{N+K} & \cdots & x_{N+M+K-1} \end{array} \right] \qquad (6\text{-}32)$$

or

$$\hat{\Phi}_{N+K} = \frac{1}{N+K} \left[X_N^T \mid X_K \right] \left[\begin{array}{c} X_N \\ \hline X_K^T \end{array} \right] \qquad (6\text{-}33)$$

Therefore,

$$\hat{\Phi}_{N+K} = \frac{1}{N+K}(X_N^T X_N + X_K X_K^T)$$
$$= \frac{1}{N+K}(N\hat{\Phi}_N + X_K X_K^T) \qquad (6\text{-}34)$$

To find the estimates based on N + K samples, the inverse of $\hat{\Phi}_{N+K}$ must be computed. This can be done efficiently in terms of the matrix inversion lemma, as presented, for example, by Lee[39]. There, it is shown that if B is a matrix of dimension M x M and C is of dimension K x M, then

$$(B + C^T C)^{-1} = B^{-1} - B^{-1}C^T(CB^{-1}C^T + I)^{-1}CB^{-1} \qquad (6\text{-}35)$$

where I is a K x K identity matrix. In the present context, if $N\hat{\Phi}_N$ and X_K^T are associated with B and C respectively, then

$$\hat{\Phi}_{N+K}^{-1} = \frac{N+K}{N}\left[\hat{\Phi}_N^{-1} - \hat{\Phi}_N^{-1}X_K(X_K^T\hat{\Phi}_N^{-1}X_K + NI)^{-1}X_K^T\hat{\Phi}_N^{-1}\right] \qquad (6\text{-}36)$$

Defining

$$\Psi = X_K^T\hat{\Phi}_N^{-1}X_K + NI \qquad (6\text{-}37)$$

equation (6-36) becomes

$$\hat{\Phi}_{N+K}^{-1} = \frac{N+K}{N}\left[\hat{\Phi}_N^{-1} - \hat{\Phi}_N^{-1}X_K\Psi^{-1}X_K^T\hat{\Phi}_N^{-1}\right] \qquad (6\text{-}38)$$

The only new matrix inversion required in the evaluation of $\hat{\Phi}_{N+K}^{-1}$ is thus that of Ψ, which is of K^{th} order, where K is the number of additional samples. If $K<M$, equation (6-38) can be used to reduce the computational effort required in extending the estimation result over the additional data. In particular, when $K = 1$, Ψ^{-1} becomes the scalar constant

$$c = \frac{1}{X_1^T \hat{\Phi}_N^{-1} X_1 + N} \tag{6-39}$$

and only a few matrix multiplications are needed to produce the new inverse.

In the general case, the new estimates are given by the equation

$$\hat{d}_{N+K} = -\hat{\Phi}_{N+K}^{-1} \hat{\phi}_{N+K} \tag{6-40}$$

The elements of $\hat{\phi}_{N+K}$ can be expressed as

$$\begin{aligned} \hat{\phi}_{N+K}(j) &= \frac{1}{N+K} \sum_{k=1}^{N+K} x_k x_{k+j} \quad , \; j = 1, 2, \ldots, M \\ &= \frac{N}{N+K} \hat{\phi}_N(j) + \frac{1}{N+K} \sum_{k=N+1}^{N+K} x_k x_{k+j} \end{aligned} \tag{6-41}$$

Thus,

$$\hat{\phi}_{N+K} = \frac{N}{N+K} \hat{\phi}_N + \frac{1}{N+K} H_K \tag{6-42}$$

where the elements of H_K are given by

$$H(j) = \sum_{k=N+1}^{N+K} x_k x_{k+j} \tag{6-43}$$

Now substituting into equation (6-40),

$$\hat{d}_{N+K} = \frac{N+K}{N}\left[\hat{\Phi}_N^{-1} X_K \Psi^{-1} X_K^T \hat{\Phi}_N^{-1} - \hat{\Phi}_N^{-1}\right]\left[\frac{N}{N+K}\hat{\Phi}_N + \frac{H_K}{N+K}\right]$$

$$= \left[\hat{\Phi}_N^{-1} X_K \Psi^{-1} X_K^T \hat{\Phi}_N^{-1} - \hat{\Phi}_N^{-1}\right]\left[\hat{\Phi}_N + \frac{H_K}{N}\right] \qquad (6\text{-}44)$$

$$= \hat{d}_N + \left[\hat{\Phi}_N^{-1} X_K \Psi^{-1} X_K^T \hat{\Phi}_N^{-1}\right]\left[\hat{\Phi}_N + \frac{H_K}{N}\right] - \hat{\Phi}_N^{-1}\frac{H_K}{N}$$

The second two terms can be considered to be a correction to the original estimate, due to the K additional samples. If K exceeds the estimation order M, however, no computational effort is saved, and the present result is only of theoretical interest.

CHAPTER VII

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

This study has proposed and analyzed several methods for performing power spectral analysis on discrete time series modeled as linear regressions. Computational examples of these procedures have also been presented to provide some grounds for evaluating their effectiveness in practical situations. This final chapter attempts to give a comprehensive summary of the findings, their role in practical spectral analysis, and suggestions for further research along similar lines.

A. A Discussion of the Findings

In attempting a spectral analysis by regressive modeling, one is led naturally to the central problem of the mixed-type process. If the time series of interest can be considered such a process, the analysis can be directed to obtaining estimates of the relatively few parameters of the model, or to devising an alternative approximate representation, such as an all-pole estimate. In either approach, one avoids the difficulties associated with truncating and/or weighting the observed correlation sequence in the usual transform approach to the sampled power spectrum. In addition, the estimates are expressed in more useful terms, since the regression coefficients thus derived are directly applicable for use in digital filtering, adaptive processing, or signal identification.

For the spectral analysis of time series modeled as mixed-type processes (including the moving average and the autoregression as special cases), the first approach considered was that of all-pole estimation. Here, the process - itself a combination of moving average and autoregressive parts - is represented as an autoregression of finite order, and the coefficients of this autoregression are estimated from the mean lagged products of a sample sequence. In the present study, the estimation equations were derived by considering the properties of the optimum "whitening" filter for the process, but then turn out to be identical with those found by other workers on the basis of linear least squares or maximum likelihood.

In an all-pole model for a mixed-type process - and indeed in any approach involving the approximation of one functional form by another of a different type - there is a certain unavoidable approximation error which will appear in the final result no matter how accurately the correlation function for the process is specified. In the practical situation, the statistical error caused by the use of a finite sample size in computing mean lagged products is added to this approximation error to produce the total error of the spectral estimate. It has been shown above that the resulting coefficients are unbiased and consistent estimates of those that would be found in the noiseless case, and further that the approximation error tends to zero as the order of the autoregressive model increases. A fairly unambiguous indication that an adequate representation has been reached is the leveling off of the falling curve of estimated input variance against model order.

The great advantage of the all-pole approach is that no *a priori* knowledge of the process whatsoever is required to apply the method. In addition, one has theoretical assurance that by going to a high enough model order, a spectral representation of arbitrarily good accuracy can ultimately be obtained. An inherent disadvantage of this approach, however, is that in cases where the process of interest is truly a moving average or a mixed-type process, the final spectral estimate has a functional form entirely different from that which actually obtains, and thus no estimates for the actual process coefficients are found. Furthermore, in cases where the actual spectral function has zeros near the unit circle, only an excessively large model order will suffice to reduce the approximation error to an acceptable level. This has been adequately demonstrated by the computational examples. In practice, all-pole estimates can be useful in dealing with pure autoregressions (in which the results are true estimates of the autoregressive coefficients), processes with slowly varying low-pass spectra, and the simpler resonant processes. When these conditions can be assumed, the all-pole approach is sure, efficient, and easily analyzed from the standpoint of probable error.

If, for whatever reason, an all-pole estimate will not suffice for the purpose at hand and there exists sufficient knowledge beforehand for assuming the numerator and denominator orders of the system function generating the process, it is possible to carry through an analysis which yields separate estimates of the moving average and autoregressive factors of the spectrum. The first step of this procedure was outlined in Chapter IV, where separate estimates for the autoregressive coefficients were derived from the correlation relationships existing in a process of mixed type. These estimates are unbiased and consistent, and the effects of statistical error are readily predictable. As shown in particular by Example # 5, however, this statistical error can be substantial unless a large sample size is available. This is basically because the present calculations are largely based on mean lagged products of relatively long lag, and statistically these are much more variable than those for small lag.

Assuming that the process under examination is of mixed type, the spectral density can be factored as the product of the autoregressive spectrum associated with the coefficients estimated by the method of Chapter IV, and a moving average spectrum associated with the coefficients of the system function numerator polynomial. Following the argument of Chapter IV, a digital filter can be devised using the estimated autoregressive coefficients, which, when driven by the original process, will yield the moving average inherent in the generation of the former. To be sure, the filter output will only approximate the true moving average because of estimation errors in the autoregressive coefficients. The quality of the approximation, however, is directly related to the estimate covariances and can thus be taken into account in judging the quality of the final spectral estimate.

If sufficient confidence can be placed in the approximate moving average, a sample sequence can be used to obtain mean lagged products for analyzing the moving average factor of the original spectrum. Several methods for analyzing moving average processes were discussed in Chapter V. The first of these was simply all-pole estimation to yield an autoregressive model for the moving average part. When this is used in combination with the estimate for the autoregressive factor of the original spectrum to give a total estimate for the latter, the result is again an all-pole function, which unfortunately will be a suboptimum representation of the true spectrum in comparison with the approximation of the same order derived more simply by the method of Chapter III. Thus, there would appear to be no advantage to this approach in most cases.

Turning to methods which indeed supply individual estimates of the moving average coefficients, the first considered was Wold's. This approach begins with a sequence of mean lagged products and a known or assumed moving average order and supplies the coefficients of that unique moving average process - of the given order - whose theoretical correlation sequence matches the observed data. The procedure is analytic, and no approximations are involved. Intuitively, this is very appealing, and only in practice do certain shortcomings appear. As we have seen, the correlation sequence of a moving average must obey certain conditions, and if these are violated by the observed sequence of mean lagged products, Wold's procedure can naturally supply no solution unless certain rather arbitrary approximations are made. In practice, this happens quite frequently, particularly when the original process has spectral zeros near the unit circle. Another problem occurring with Wold's method is the computational difficulty associated with the factoring of the polynomial V(r) and the necessity of solving quadratic equations with complex coefficients. If the order of the moving averages to be studied exceeds 4, these make the application of the method forbiddingly inconvenient, especially if a general purpose computer algorithm is ultimately desired. Another peculiarity of Wold's approach is the extreme sensitivity of the coefficient estimates to even small changes in the mean lagged products. Thus, although the final spectral estimate is precisely what would be found by z-transformation of the mean lagged product sequence, the coefficient estimates themselves can be wide of the mark. This is shown, for instance, in the treatment of Example # 2, a pure moving average for which there is no question of approximation error. No attempt has been made to derive expressions for the statistical error present in the Wold coefficient estimates. The complexity and non-linearity of the estimation equations have prevented this, and one can only judge the quality of the spectral estimate as a whole by considering the statistical properties of the mean lagged products. If the goal of the analysis is really the set of moving average coefficients, Wold's approach provides no means for determining their statistical error.

The greatest advantage of the spectral inversion technique is that it always yields a solution. In addition, it is computationally simple, requiring essentially only the same operations used in all-pole estimation. The principal disadvantage of the method is the unavoidable approximation error that arises due to the first step of the procedure, where the process of interest is represented as a finite order autoregression. It has been shown above that the amount of the final approximation error is closely related to that found in this first step and manifests itself as a bias in the final coefficient estimates. This bias was actually observed by Durbin[36] in his experimental work on a first order moving average, and a theoretical prediction of its magnitude in the first order case is presented by Walker[24]. At any rate, as the intermediate model order is increased, the final approximation error (and the coefficient bias) will tend to zero. If the moving average of interest requires a high order all-pole model for sufficient accuracy, the order of the intermediate model required in the spectral inversion procedure will similarly be high. It could be said, therefore, that the present method suffers from the same limitations as all-pole estimation. This is not strictly true since the ultimate result is of limited order and represented in the correct functional form. The computational complexities of a high intermediate model order are, so to speak, hidden behind the scenes.

The amount of statistical error in the final moving average coefficient estimates is very difficult to compute exactly, and again one is moved to consider spectral error as a whole. Durbin's approximate result, however, seems to provide an adequate estimate of the statistical error, and it can be used with confidence. In a practical application, the required intermediate model order can be determined by the leveling off of the all-pole variance estimate in the first step of the procedure, and the statistical error can then be estimated using Durbin's expression with the covariance matrix of the second step.

On the whole, it appears that the spectral inversion technique is the best suited for use in a general purpose estimation algorithm for mixed-type processes. In combination with the method of Chapter IV for determining the autoregressive coefficients, it can easily be implemented on any digital computer with means for manipulating standard matrix expressions. Such an algorithm is shown diagrammatically in Figure 7-1, where it is assumed that the intermediate model order for the spectral inversion is found by successively increasing the order until the variance estimate levels off.

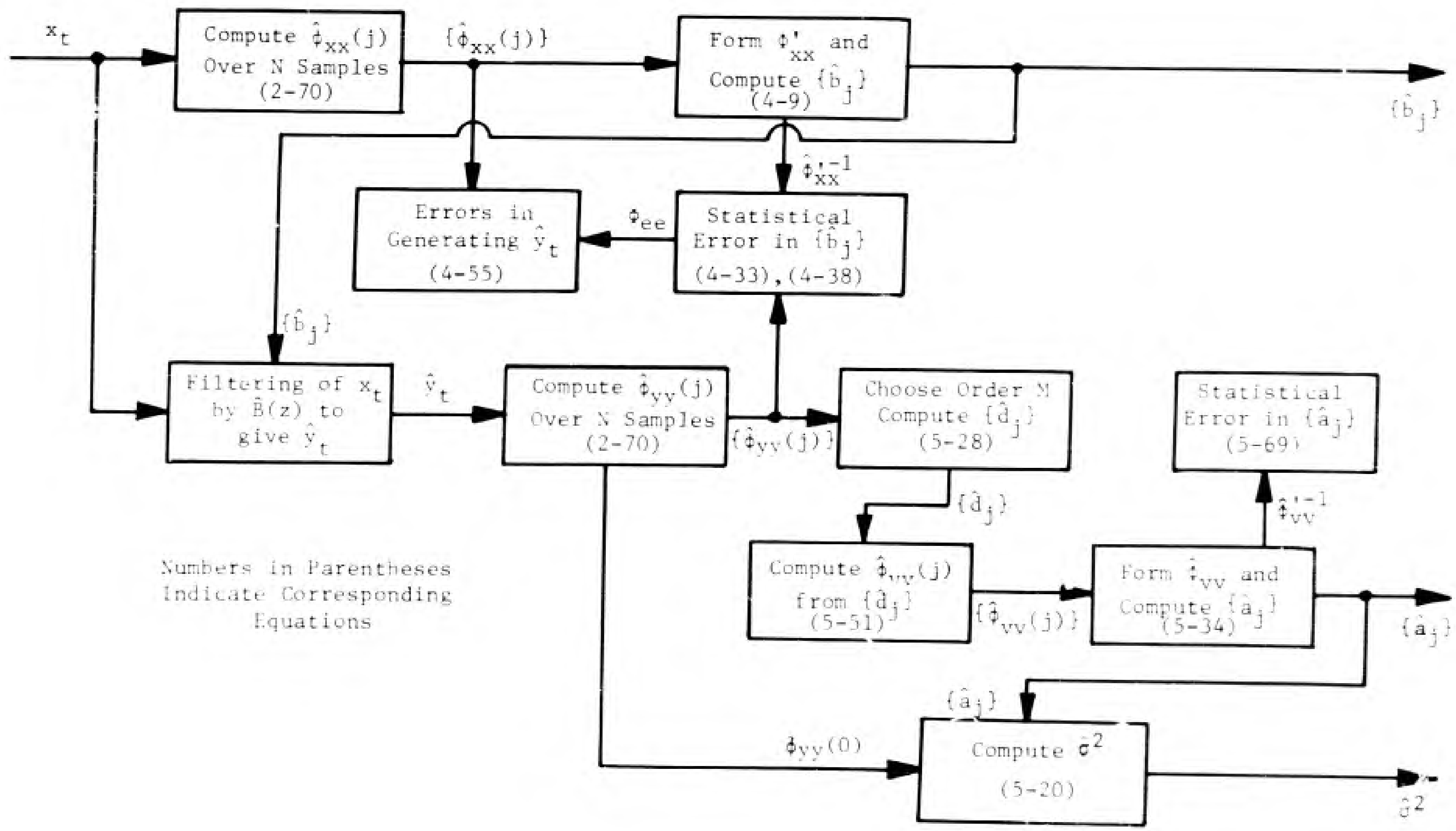


FIGURE 7-1. DIAGRAMMATIC REPRESENTATION OF A COMPLETE SPECTRAL ANALYSIS ALGORITHM USING THE SPECTRAL INVERSION TECHNIQUE.

If the intermediate model order can be pre-set to some standard sufficiently high value, even this iteration is unnecessary. The statistical error for both sets of coefficients can be readily determined from matrix inverses computed during the course of the estimation itself and sample sizes adjusted accordingly. It should be remembered, however, that the present procedure can only be applied when the orders of the numerator or denominator polynomials are assumed (or known) beforehand. When this information is lacking, an all-pole estimate for the entire process may be more appropriate.

As noted in Appendix A, sets of mean lagged products were also computed for processes generated by driving the standard system functions used for the Gaussian examples with a sequence of uniformly distributed random numbers, normalized to unit variance. These were used to provide computational examples of the application of the several estimation procedures to non-Gaussian processes. In every case, the results obtained were so similar to those found in the corresponding Gaussian example that no point is served in presenting the detailed findings here. Both the sets of coefficient estimates and the corresponding spectral functions displayed errors of the same magnitude found in the Gaussian cases, and these in turn were well predicted by the appropriate error analyses. Although the latter were derived for the most part by assuming Gaussian statistics, their predictions were found to be perfectly adequate here. Although the amount of experimentation performed upon non-Gaussian cases was limited, it seems safe to conclude that the results of the study apply adequately to such processes in practical situations. It should be recalled, in fact, that the Gaussian assumption was only used above in analyzing error performance and not in deriving the estimation equations themselves.

The results developed in Chapter VI can be used to further reduce the computational effort of a spectral analysis when data is continually arriving or when statistical considerations require the extension of the estimate over larger sample sizes. The approximate method derived on the basis of Faddeeva's theorem is best suited for extrapolating the estimates over large blocks of additional data - on the order of 1/10 to 1/2 the original sample size. The exact result, based on the matrix inversion lemma, is most useful for demonstrating theoretically the role played by each individual sample in determining the final estimate values. It is quite possible that this result could be extended to provide a practical recursive estimation scheme, and this stands as one of the most promising areas of future research.

B. Suggestions for Further Research

The following topics, suggested by the findings of the study, are considered to be logical extensions of the present work:

With the exception of direct all-pole estimation, the methods considered here have all required an *a priori* assumption of the moving average and autoregressive orders of a mixed-type process. Perhaps the most interesting area of further study is that of finding some means for satisfying this requirement on the basis of observations of the process itself. This would seem to have two parts. The first is the search for methods to determine a reasonable estimate of the required polynomial orders on the basis of either the observed correlation structure of the process or a rudimentary spectral estimate, such as an all-pole approximation. The second is a theoretical and experimental study of the effects on the spectral estimate of hypothesizing the wrong orders and proceeding on that basis. As we have seen in the computational examples, it is possible for two system functions of totally different form to yield very similar spectra, and thus such an incorrect working hypothesis might nonetheless provide a good spectral estimate. A closely related problem is raised in considering the use of the estimated coefficients in adaptive filtering and control. If the filters used are based on an erroneous assumption of the functional form of the filtered process, do practical difficulties arise even though the spectral estimate itself is very good?

In dealing with sample sequences of actual data, it is straightforward to fit a moving average or autoregressive model by successively increasing the model order until further additions are statistically insignificant. With no knowledge beforehand, however, the fitting of a mixed-type model is not so obvious, and here is where the investigations proposed above may be of great aid. Another approach might be the formulation of an iterative strategy in which both numerator and denominator orders could be manipulated to achieve an optimum fit to the data. It is not clear, though, just what criterion of optimality could be devised in the face of a complete lack of a priori knowledge of the form of the process, but studies along these lines would be of great interest to those working in practical spectral analysis.

The estimates presented here have been shown to be at least asymptotically unbiased and consistent. It would be desirable to extend these results to show to what extent they are efficient in the narrow statistical sense. As noted above, this has already been done by Durbin[36] for his version of the spectral inversion technique, and much can be inferred for all-pole estimates since they can be shown to have maximum likeihood in the Gaussian case, but for Wold's method and the autoregressive coefficient estimates of Chapter IV, no similar results seem to be available. The whole question of estimate optimality should probably be re-examined for this application from a more fundamental point of view. A good beginning is provided by Durbin's concept of "best unbiased linear estimates"[40], and several other areas of recent statistical estimation theory would be useful here.

It has already been mentioned that the results of Chapter VI hold promise of being further developed along the lines of sample-by-sample recursive estimation. Many of the ideas presented by Lee[39] are applicable here, and the principal difficulties seem to lie in formulating adequate representations for the estimation schemes in recursive terms. Once this is done, it should be possible to devise efficient means for providing spectral parameter estimates based on a sliding data "window" with up-dating at the arrival of each new sample. This would have broad practical application.

Finally, there is the investigation of possible applications for regressive models in adaptive filtering, processing, and control. Following the original work of Steiglitz[18], the use of an autoregressive model for the adaptive reconstruction of a continuous signal from its samples was treated by Tretter and Steiglitz[41], and an adaptive matched filter was proposed by Steiglitz and Thomas[42]. The additional flexibility and representational accuracy available from the mixed-type model should open further areas to this approach, and it is hoped in particular to use such a scheme to aid in the detection of a signal in a noise background of changing character. Success in these areas depends largely on solving the problems described above in connection with hypothesizing an appropriate model order, but if this is done, regressive models will be of great use in many such applications.

APPENDIX A

FIVE SPECTRAL EXAMPLES

To demonstrate the application of the estimation principles derived in the main body, five standard spectral examples have been generated and used. Sample sequences of processes corresponding to these spectra were produced on a digital computer by passing a sequence of uncorrelated Gaussian random numbers through a digital filter in the form of a recursive equation. The resulting number sequences were used for the calculation of the mean lagged products needed in all of the estimations. This Appendix indicates the rationale followed in selecting the examples, describes them in detail, and provides tables of the computed raw data.

The basic guidelines for the choice of the examples were as follows:

a.) It was felt necessary to include at least one of each of the three types of processes treated here, i.e. the moving average, the autoregression, and the process of mixed type.

b.) Insofar as possible, the examples were chosen to be typical of signals that might be met in practical problems. This required principally that the spectral magnitude tend to zero at the argument equivalent to half the sampling frequency, so that the resulting sequence could be taken to represent sampled continuous signals with negligible aliasing.

c.) For relative ease of computation, the maximum order of the numerator and denominator polynomials of the associated system function was taken to be 4. This also helped to insure that numerical complexity did not swamp the important conceptual points.

d.) In accordance with the assumptions made for the theoretical work, no poles or zeros of the system function were allowed outside of the z-plane unit circle.

e.) At least one example had to have several narrow spectral peaks so that the "resolving power" of the several estimation methods could be compared.

As noted in the main body, the sampled power spectra can be expressed as functions of the real radian frequency ω by making the substitution $z = e^{i\omega T}$, where T is the sampling interval in seconds. In this form, the spectral functions will be periodic in ω with period $2\pi/T$. Furthermore, the spectra are even functions of ω, and that portion lying between $\omega = \pi/T$ and $\omega = 2\pi/T$ will be the mirror image of the segment between $\omega = 0$ and $\omega = \pi/T$. For this reason, the spectral plots found in this Appendix and throughout the main body are presented only for the range $0 \leq \omega \leq \pi/T$. In addition, the spectra are plotted against the variable

$$v = \omega T/\pi$$

with corresponding range $0 \leq v \leq 1$. Thus $v = 1$ represents half the sampling frequency, or the nominal maximum frequency present in the corresponding continuous spectrum.

1. Standard Example # 1

The first example is a 4^{th} order autoregression with system function

$$H(z) = \frac{1}{B(z)} = \frac{1}{1 - 2.036z^{-1} + 1.812z^{-2} - 0.770z^{-3} + 0.129z^{-4}}$$

In all of the examples proposed here, the variance of the input noise sequence is unity. Thus, with n_t an uncorrelated sequence of unit variance, the difference equation representation becomes

$$x_t = n_t + 2.036x_{t-1} - 1.812x_{t-2} + 0.770x_{t-3} - 0.129x_{t-4}$$

and the spectral magnitude is plotted in Figure A-1. This example, which yields a relatively flat low-pass spectrum, was obtained by transforming the poles of a (continuous) 4^{th} order low-pass Butterworth filter into their z-plane equivalents using the relationship

$$z_p = e^{s_p T}$$

If the cut-off frequency of the filter is taken to be one eighth the sampling frequency, to limit the degree of aliasing, the four z-plane poles become

$$z_{P_1}, z_{P_2} = 0.553 \pm i0.491$$

$$z_{P_3}, z_{P_4} = 0.463 \pm i0.143$$

The system function is then written as a product of factors:

$$H(z) = \prod_{j=1}^{4} \frac{1}{(1 - z_{P_j}/z)}$$

to get the final form. The z-plane pole-zero pattern is portrayed in Figure A-2, where it should be noted that inherent in $H(z)$ is a 4^{th} order zero at the origin. For practical purposes, this can be ignored, since it disappears from the spectral expression $H(z)H(z^{-1})$.

Using the methods of Section E of Chapter II, the theoretical autocorrelation sequence for the process and the unit sample response of the associated system can be computed from the recursion coefficients. The first terms of these sequences are found to be

$$\{\phi_{xx}(j)\}_{j=0}^{\infty} = 15.24370,\ 13.15450,\ 8.22859,\ 2.95809,\ -0.72498,$$
$$-2.19703,\ -1.94326,\ -0.91528,\ 0.05949, \ldots\ldots$$

$$\{h_j\}_{j=0}^{\infty} = 1.0,\ 2.036,\ 2.33333,\ 1.83136,\ 0.93943,\ 0.12826,$$
$$-0.33197,\ -0.42117,\ -0.27841,\ \ldots\ldots$$

These are plotted in Figure A-3.

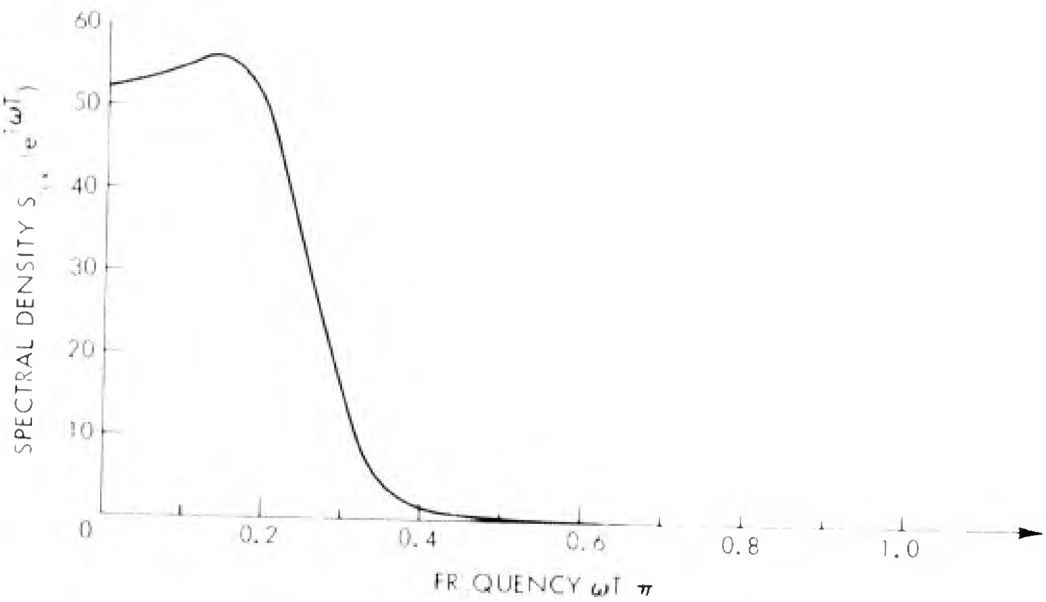


FIG. A-1 NOMINAL SAMPLED POWER SPECTRAL DENSITY FOR STANDARD EXAMPLE #1

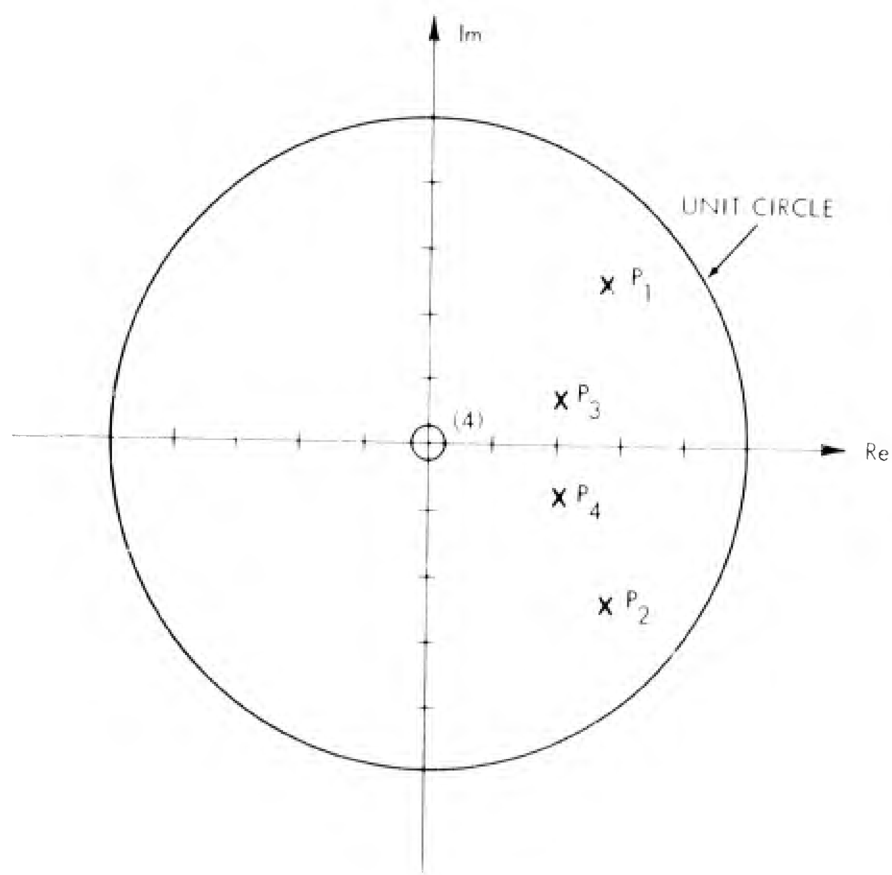


FIG. A-2 Z-PLANE POLE-ZERO PATTERN FOR THE SYSTEM FUNCTION OF EXAMPLE #1

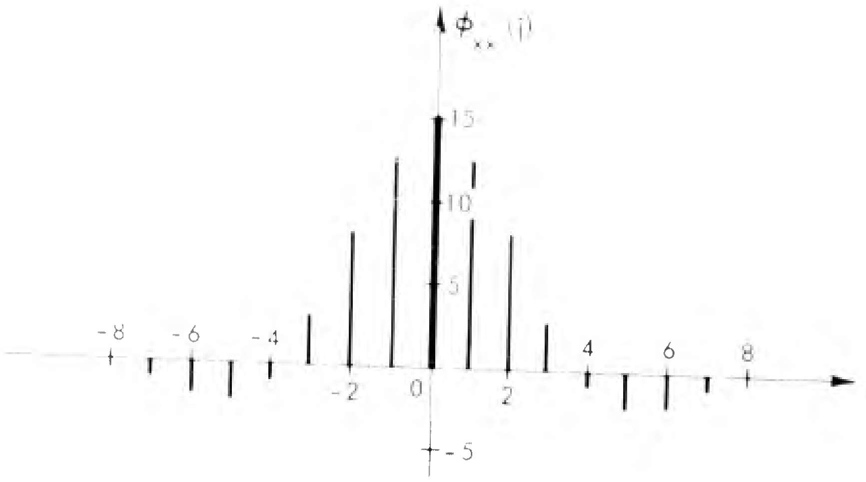


FIG. A-3A THEORETICAL AUTOCORRELATION SEQUENCE FOR EXAMPLE #1

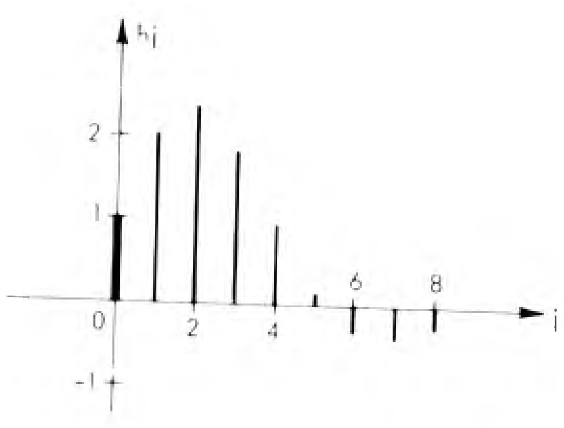


FIG. A-3B UNIT SAMPLE RESPONSE FOR THE SYSTEM ASSOCIATED WITH EXAMPLE #1

2. Standard Example # 2

The second example is a 4^{th} order moving average with system function

$$A(z) = 1 + 2.036z^{-1} + 1.812z^{-2} + 0.770z^{-3} + 0.129z^{-4}$$

and input variance $\sigma^2 = 1.0$. As a difference equation, the process is represented as

$$x_t = n_t + 2.036n_{t-1} + 1.812n_{t-2} + 0.770n_{t-3} + 0.129n_{t-4}$$

with spectral amplitude shown in Figure A-4. This low-pass spectrum was obtained by changing the poles of the first example into zeros and reflecting them about the z-plane imaginary axis. The four system function zeros become

$$z_{Z_1}, z_{Z_2} = -0.553 \pm i0.491$$

$$z_{Z_3}, z_{Z_4} = -0.463 \pm i0.143$$

as shown in the pole-zero diagram of Figure A-5. Again using the methods of Chapter II, the theoretical autocorrelation sequence for this example emerges as

$$\{\phi_{xx}(j)\}_{j=0}^{\infty} = 9.03818,\ 7.21980,\ 3.61347,\ 1.03264,\ 0.12900,\ 0,\ 0,\ 0,\ \ldots\ldots$$

and the unit sample response is

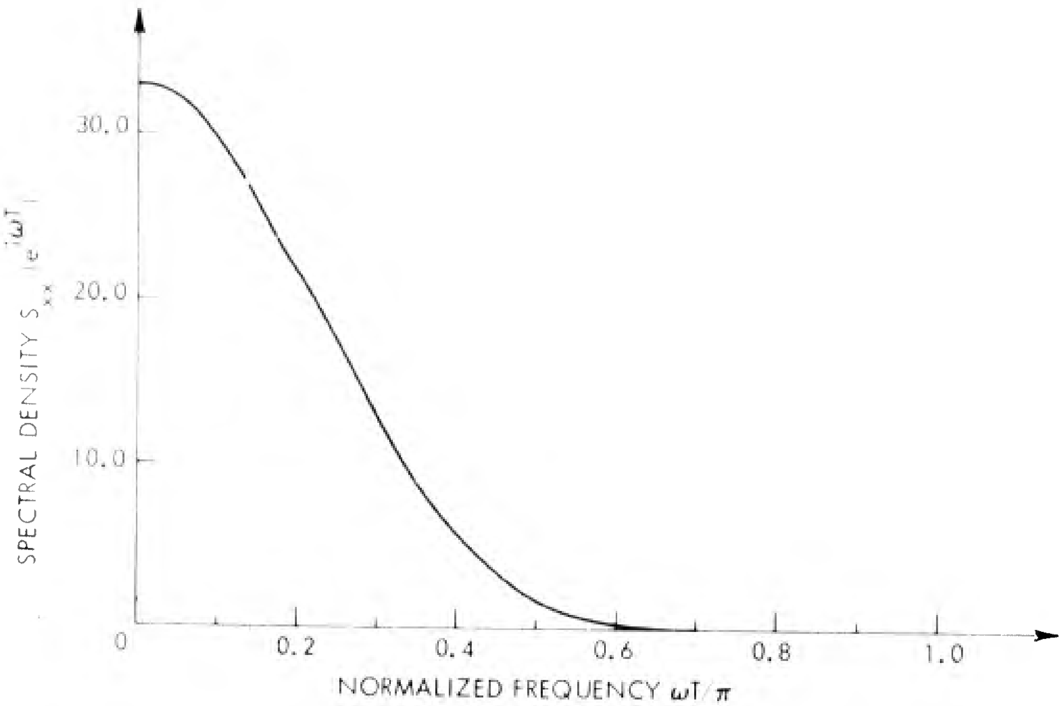


FIG. A-4 NOMINAL SAMPLED POWER SPECTRAL DENSITY FOR STANDARD EXAMPLE #2

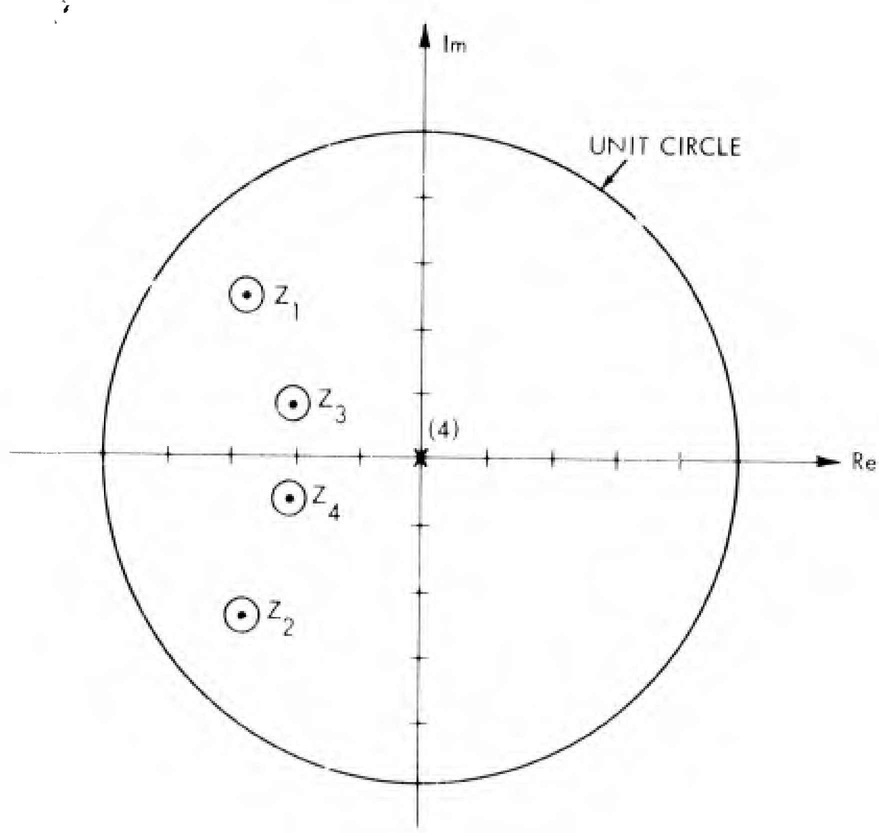


FIG. A-5 Z-PLANE POLE-ZERO PATTERN FOR THE SYSTEM FUNCTION OF EXAMPLE #2

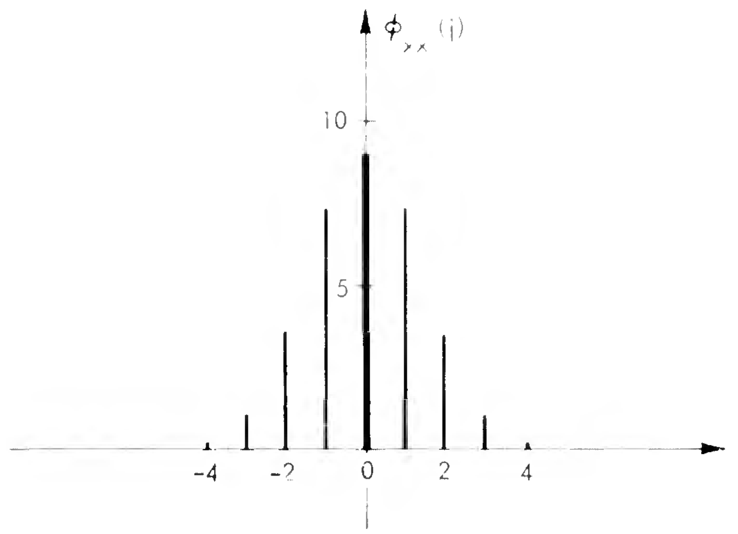


FIG. A-6A THEORETICAL AUTOCORRELATION SEQUENCE FOR EXAMPLE #2

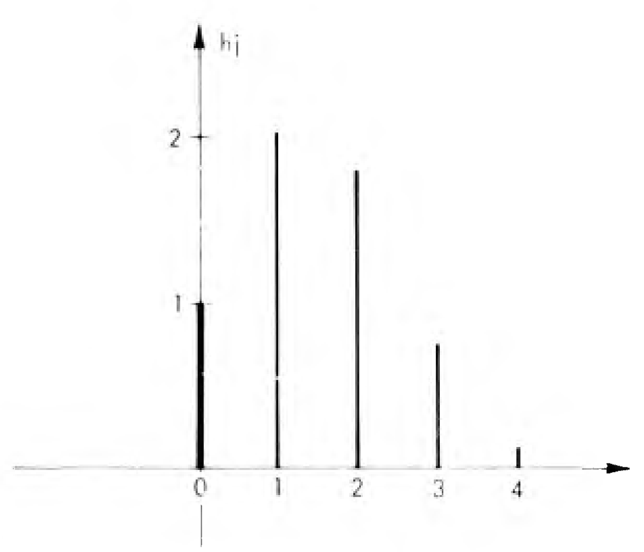


FIG. A-6B UNIT SAMPLE RESPONSE FOR THE SYSTEM ASSOCIATED WITH EXAMPLE #2

$$\{h_j\}_{j=0}^{\infty} = 1.0,\ 2.036,\ 1.812,\ 0.770,\ 0.129,\ 0,\ 0,\ 0,\ \ldots\ldots$$

These are both displayed in Figure A-6.

3. Standard Example # 3

The third example is a process of mixed type that provides a resonant spectrum roughly analogous to that observed in passing white noise through a series resonant RLC circuit. The system function and recursive representations are given by

$$H(z) = \frac{A(z)}{B(z)} = \frac{1 - 0.859z^{-1}}{1 + 0.853z^{-2}}$$

and

$$x_t = n_t - 0.859n_{t-1} - 0.853x_{t-2}$$

and the spectrum is plotted as Figure A-7. The resonant frequency occurs at one quarter of the sampling frequency - or $\nu = 0.5$ on the normalized scale - and the Q of the system is 10. The important z-plane poles and zeros are given by

$$z_{P_1}, z_{P_2} = \pm\, i0.9245$$

$$z_{Z_1} = 0.859$$

As seen in the pole-zero plot of Figure A-8, there is also a single zero at the origin. The initial portions of the theoretical autocorrelation and unit sample response sequences are given by

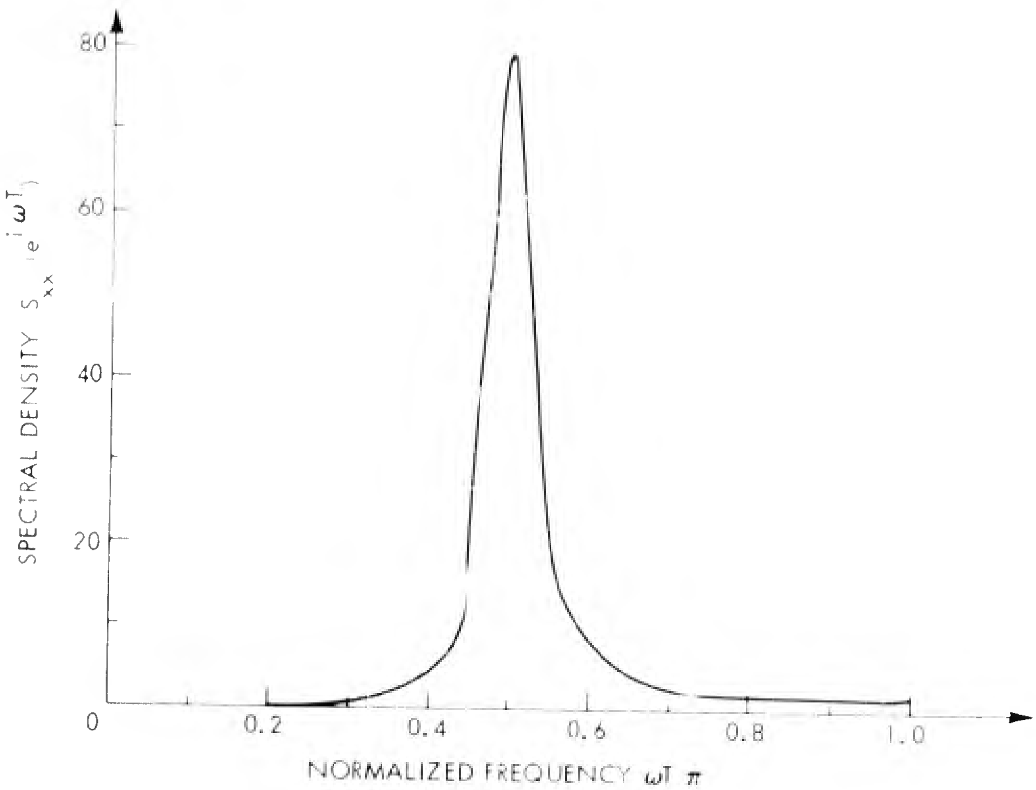


FIG. A-7 NOMINAL SAMPLED POWER SPECTRAL DENSITY FOR STANDARD EXAMPLE #3

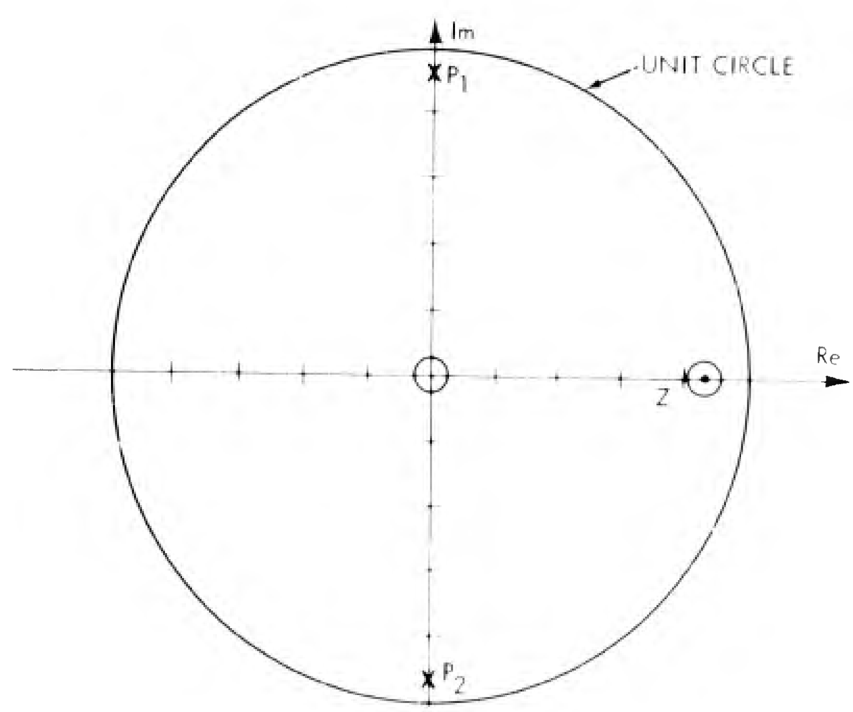


FIG. A-8 Z-PLANE POLE-ZERO PATTERN FOR THE SYSTEM FUNCTION OF EXAMPLE #3

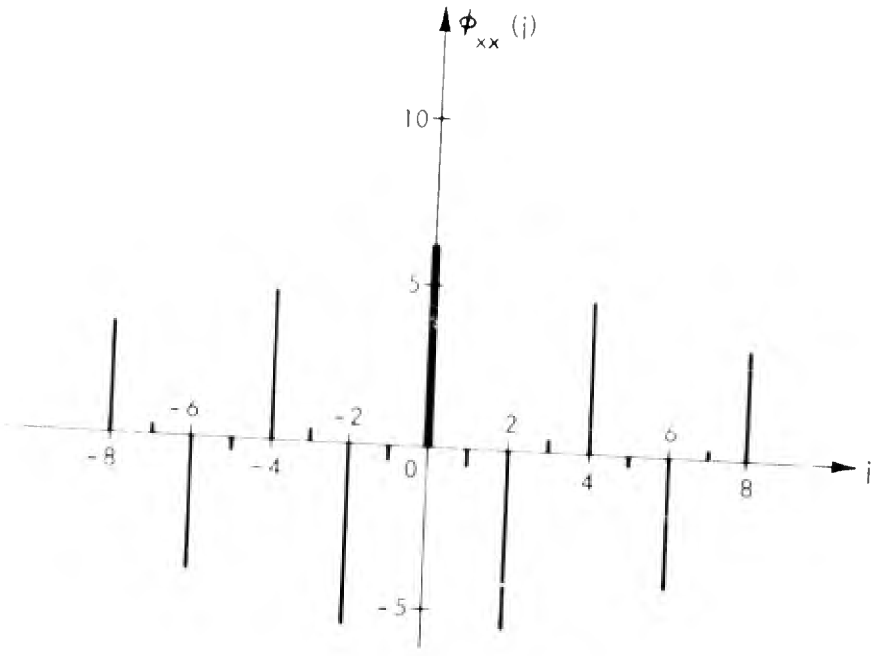


FIG. A-9A THEORETICAL AUTOCORRELATION SEQUENCE FOR EXAMPLE #3

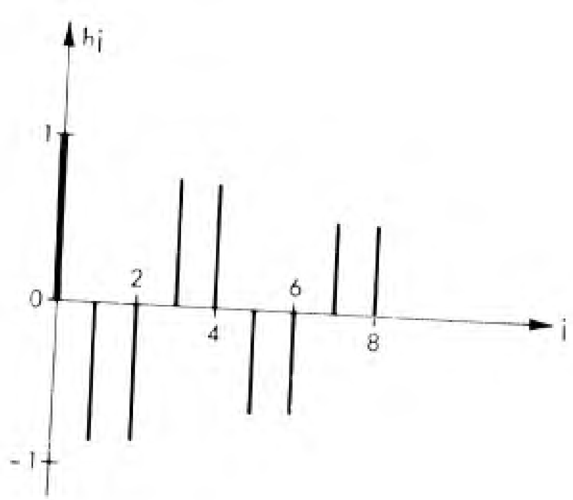


FIG. A-9B UNIT SAMPLE RESPONSE FOR THE SYSTEM ASSOCIATED WITH EXAMPLE #3

$$\{\phi_{xx}(j)\}_{j=0}^{\infty} = 6.38011, -0.46357, -5.44220, 0.39543, 4.64222, -0.33730, -3.95981, 0.28772, 3.37772, \ldots\ldots\ldots$$

$$\{h_j\}_{j=0}^{\infty} = 1.0, -0.859, -0.853, 0.73272, 0.72761, -0.62502, -0.62065, 0.53314, 0.52941, \ldots\ldots\ldots$$

and these are plotted in Figure A-9.

4. Standard Example # 4

The fourth example is constructed as the tandem connection of two digital systems of the type of the third example to yield a spectrum with two resonant peaks. This is shown in Figure A-10. The system and recursion equations are

$$H(z) = \frac{A(z)}{B(z)} = \frac{1 - 1.7835z^{-1} + 0.793z^{-2}}{1 - 1.337z^{-1} + 1.632z^{-2} - 0.987z^{-3} + 0.660z^{-4}}$$

and

$$x_t = n_t - 1.7835n_{t-1} + 0.793n_{t-2} + 1.337x_{t-1} - 1.632x_{t-2} + 0.987x_{t-3} - 0.660x_{t-4}$$

and the important poles and zeros are given by

$$z_{Z_1} = 0.8590$$

$$z_{Z_2} = 0.9245$$

$$z_{P_1}, z_{P_2} = \pm i0.8590$$

$$z_{P_3}, z_{P_4} = 0.667 \pm i0.667$$

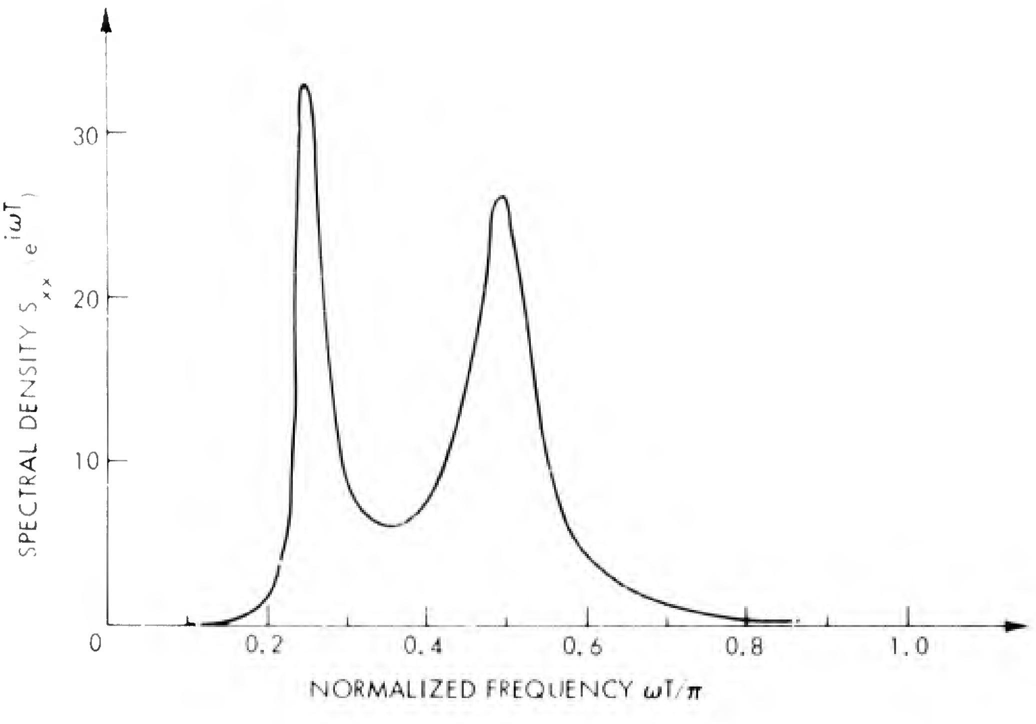


FIG. A-10 NOMINAL SAMPLED POWER SPECTRAL DENSITY FOR STANDARD EXAMPLE #4

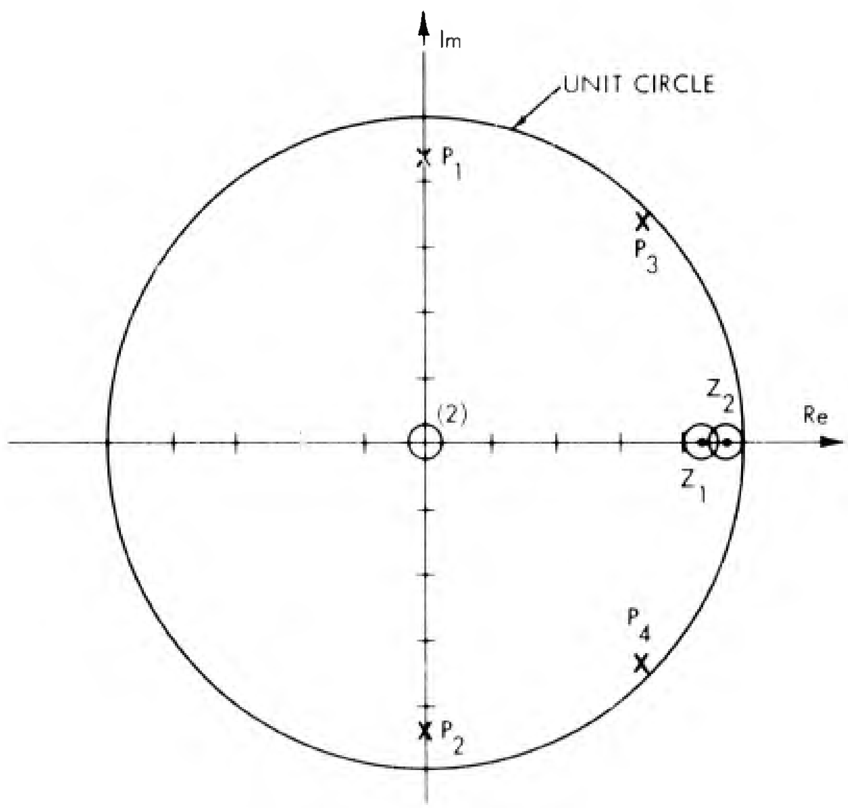


FIG. A-11 Z-PLANE POLE-ZERO PATTERN FOR THE SYSTEM FUNCTION OF EXAMPLE #4

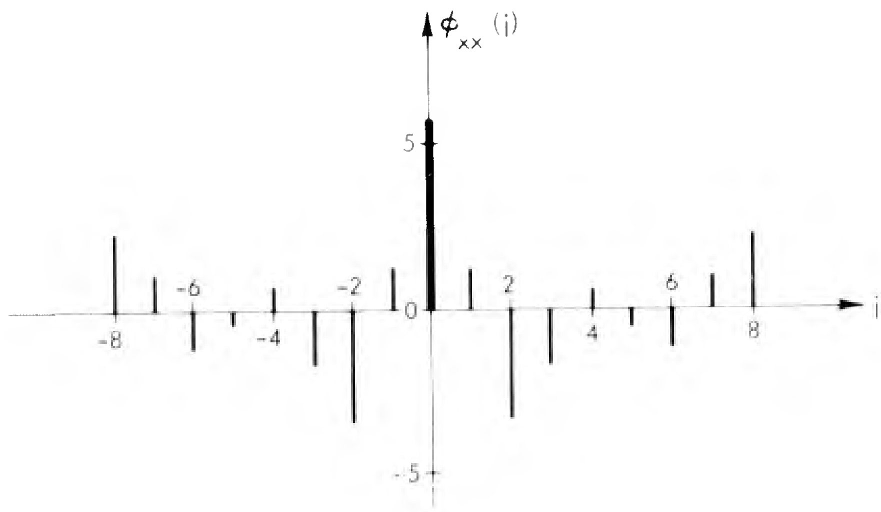


FIG. A-12A THEORETICAL AUTOCORRELATION SEQUENCE FOR EXAMPLE #4

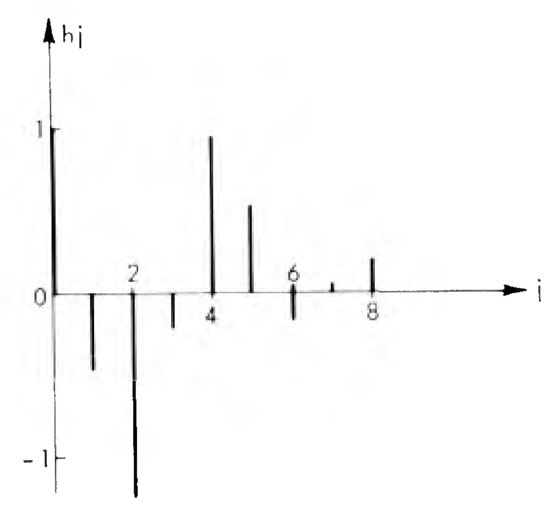


FIG. A-12B UNIT SAMPLE RESPONSE FOR THE SYSTEM ASSOCIATED WITH EXAMPLE #4

These are shown in Figure A-11. Note that there is also a double order zero at the origin. The present example displays resonant peaks at $v = 0.25$ and $v = 0.5$, with respective Q's of 7 and 5. As shown in Figure A-12, the autocorrelation and unit sample response sequences are given by

$$\{\phi_{xx}(j)\}_{j=0}^{\infty} = 5.58831,\ 1.20165,\ -3.33402,\ -1.69611,\ 0.67117,\ -0.41836,\ -1.12832,\ 0.95610,\ 2.26384,\ \ldots\ldots$$

$$\{h_j\}_{j=0}^{\infty} = 1.0,\ -0.44650,\ -1.43597,\ -0.20421,\ 0.96979,\ 0.50725,\ -0.15830,\ 0.05247,\ 0.18910,\ldots\ldots$$

5. Standard Example # 5

The last example is similar to the preceding one in that it has the same number of poles and zeros and since these are arrayed in the z-plane in roughly the same way. The root positions have been juggled, however, so that the two spectral peaks of the 4^{th} example have coalesced into a single relatively broad peak, giving the spectrum of Figure A-13. This can be considered to be a band-pass spectrum centered at $v = 0.4$. The system and recursion equations are

$$H(z) = \frac{A(z)}{B(z)} = \frac{1 - 1.74z^{-1} + 0.81z^{-2}}{1 - 1.352z^{-1} + 1.338z^{-2} - 0.662z^{-3} + 0.240z^{-4}}$$

$$x_t = n_t - 1.74n_{t-1} + 0.81n_{t-2} + 1.352x_{t-1} - 1.338x_{t-2} + 0.662x_{t-3} - 0.240x_{t-4}$$

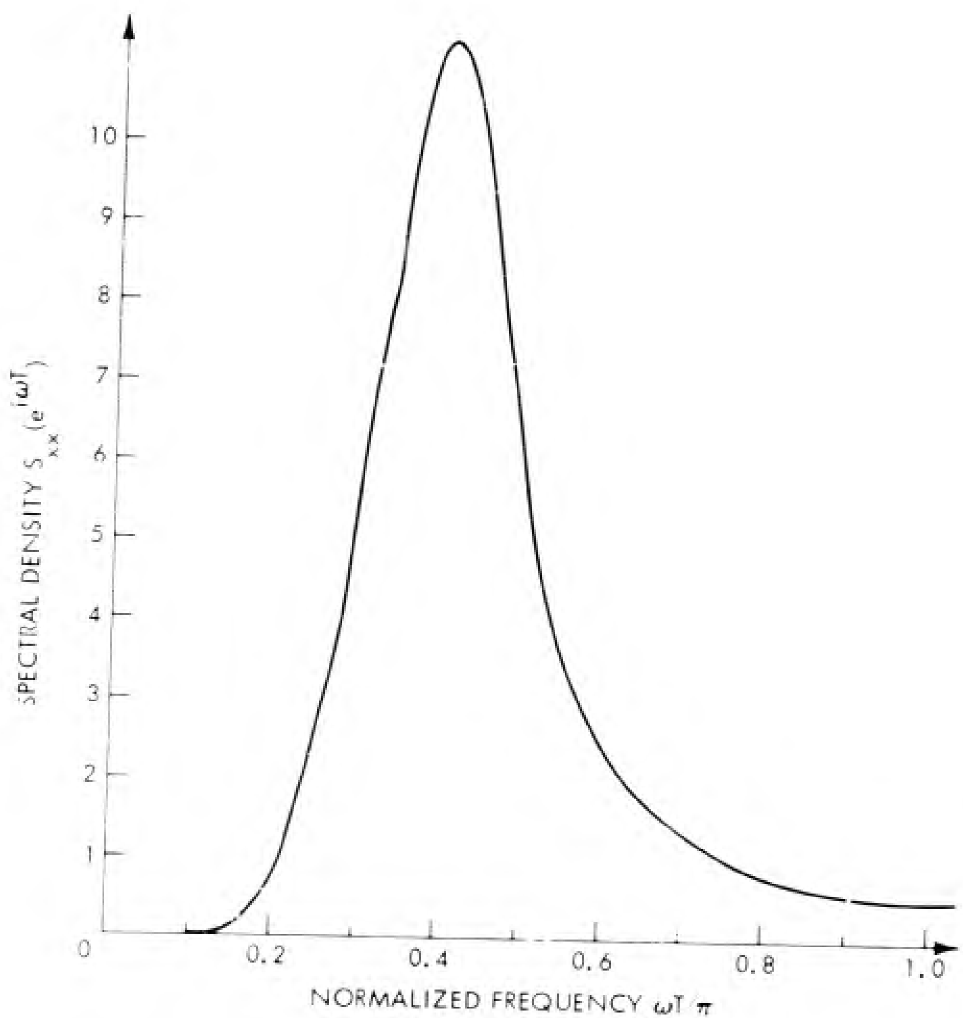


FIG. A-13 NOMINAL SAMPLED POWER SPECTRAL DENSITY FOR STANDARD EXAMPLE #5

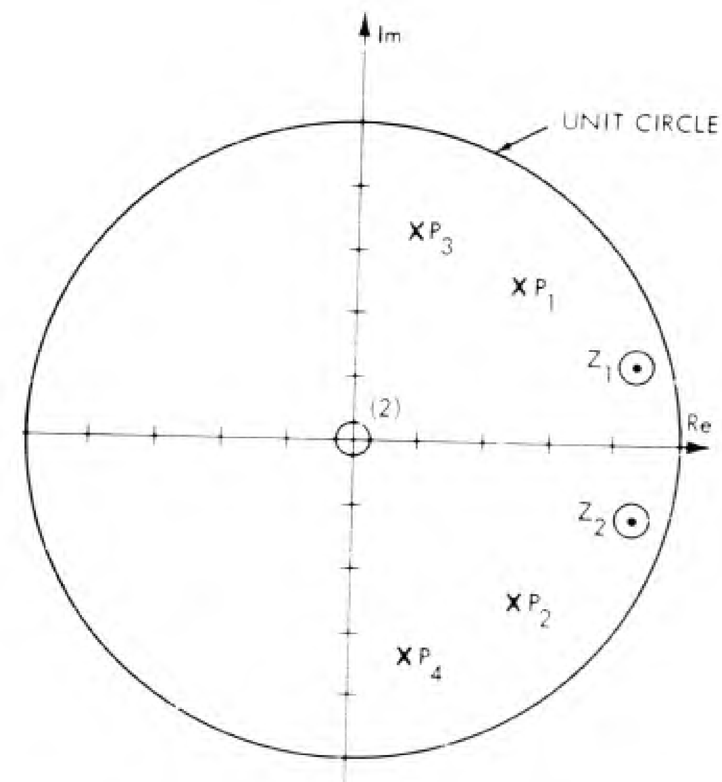


FIG. A-14 Z-PLANE POLE-ZERO PATTERN FOR THE SYSTEM FUNCTION OF EXAMPLE #5

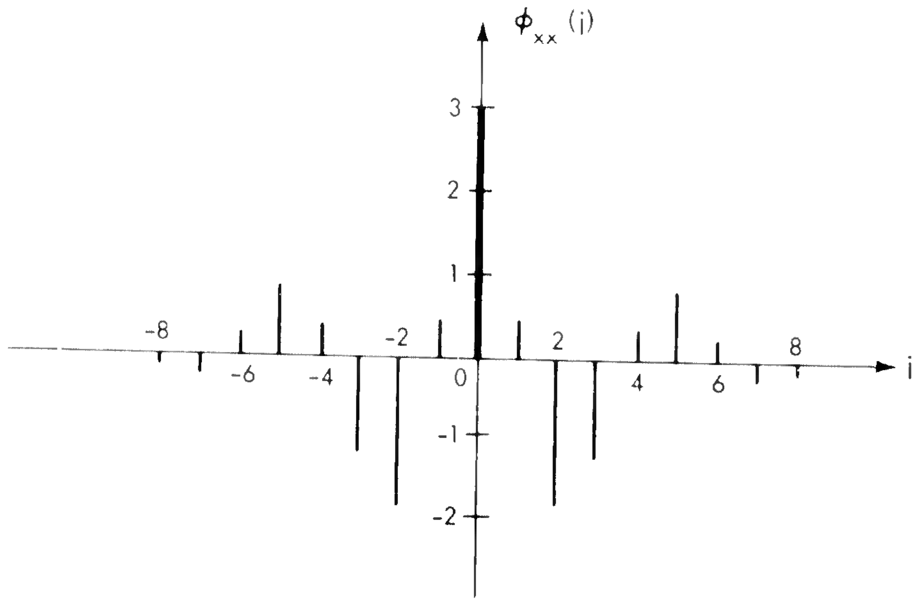


FIG. A-15A THEORETICAL AUTOCORRELATION SEQUENCE FOR EXAMPLE #5

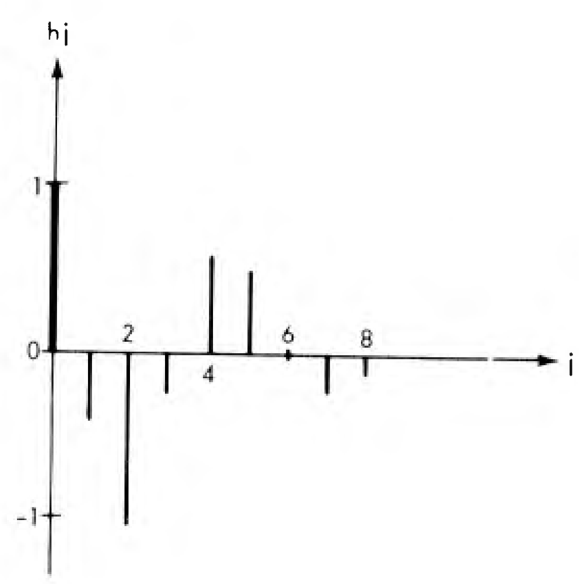


FIG. A-15B UNIT SAMPLE RESPONSE FOR THE SYSTEM ASSOCIATED WITH EXAMPLE #5

The system poles and zeros are given by

$$z_{Z_1}, z_{Z_2} = 0.869 \pm i0.233$$

$$z_{P_1}, z_{P_2} = 0.495 \pm i0.495$$

$$z_{P_3}, z_{P_4} = 0.181 \pm i0.676$$

with an additional double order zero at the origin (See Figure A-14.). Again using the methods of Chapter II, the following autocorrelation and unit sample response sequences are found:

$$\{\phi_{xx}(j)\}_{j=0}^{\infty} = 2.99383,\ 0.45912,\ -1.83151,\ -1.21878,\ 0.38819,\ 0.83291,\ 0.23943,\ -0.24124,\ -0.18829,\ \ldots$$

$$\{h_j\}_{j=0}^{\infty} = 1.0,\ -0.38800,\ -1.05258,\ -0.24194,\ 0.58439,\ 0.51023,\ 0.00023,\ -0.23731,\ -0.12369,\ \ldots$$

These are plotted in Figure A-15.

6. Generation of Sample Sequences and Mean Lagged Products

The sample sequences from which the mean lagged products were computed were generated on a General Electric System 265 digital computer by means of a remote time sharing station from which programs in the BASIC language[a] could be executed. One of the built-in functions of this language is a random number generator which on call will yield a random number uniformly distributed between 0.0 and 1.0. These were converted to approximately Gaussian random numbers of zero mean and unit variance using the method of Juncosa[b]. This uncorrelated Gaussian sequence was used as the variable n_t to "drive" the recursive equations which generated the process sample sequences. The first thousand values of x_t produced after the onset of computation were thrown away to eliminate any "starting transients" that might have been present. Then, mean lagged products for up to 8 lags were computed using formula (2-70) for 100, 200, 500, 1,000, 2,000, and 5,000 samples, and these were tabulated as raw material for the spectral analyses. To check the quality of this data, the theoretical standard deviations for the mean lagged products were calculated using the equations of Section F of Chapter II and used to see if the measured values lay within one standard deviation of the mean of the sampling distribution. Except for a few out-liers, whose occasional appearance is to be expected, this was found to be true.

Tables A-1 through A-5 present the computed mean lagged products for the five standard examples, along with the theoretical auto- and cross-correlation sequences out to 8 lags. Also indicated in the tables are the standard deviations for $\hat{\phi}_{xx}(0)$ for the several sample sizes, and these can be used as a quick means of evaluating the statistical reliability of the data.

A smaller set of mean lagged products for the same process system functions was generated using the uniformly distributed number sequence (with normalization to unit variance)

Table A-1

Computed Mean Lagged Products for

Standard Example # 1

$$H(z) = \frac{1}{1 - 2.036z^{-1} + 1.812z^{-2} - 0.770z^{-3} + 0.129z^{-4}}$$

Input sequence: Gaussian, zero-mean, $\sigma^2 = 1.0$

j	$\phi_{nx}(j)$ Theo.	$\phi_{xx}(j)$ Theo.		Mean Lagged Products $\hat{\phi}_{xx}(j)$ N=100	N=200	N=500	N=1,000	N=2,000	N=5,000
0	1.0	15.24370		14.78741	13.45371	16.19344	15.32550	14.94273	15.16452
1	2.036	13.15450		12.10923	11.22343	13.90091	13.18281	12.87591	13.16343
2	2.33330	8.22859		5.96995	6.10584	8.47392	8.13402	8.02005	8.45853
3	1.83136	2.95809		-0.37074	0.97149	2.64332	2.73565	2.82404	3.42416
4	0.93943	-0.72498		-4.05518	-1.98769	-1.40683	-0.99685	-0.82833	-0.11669
5	0.12826	-2.19703		-4.11032	-2.10885	-2.89257	-2.36560	-2.35030	-1.58000
6	-0.33197	-1.94326		-1.44492	-0.20396	-2.30851	-1.82850	-2.22507	-1.41604
7	-0.42117	-0.91528		1.89754	2.11454	-0.74323	-0.42321	-1.33611	-0.51073
8	-0.27841	0.05949		3.92719	3.39261	0.83666	0.90150	-0.39607	0.36006
Standard Deviation of $\hat{\phi}_{xx}(0)$:				3.88	2.74	1.74	1.23	0.87	0.55

Table A-2

Computed Mean Lagged Products for

Standard Example # 2

$$H(z) = 1 + 2.036z^{-1} + 1.812z^{-2} + 0.770z^{-3} + 0.129z^{-4}$$

Input sequence: Gaussian, zero-mean, $\sigma^2 = 1.0$

j	$\phi_{nx}(j)$ Theo.	$\phi_{xx}(j)$ Theo.		Mean Lagged Products $\hat{\phi}_{xx}(j)$ N=100	N=200	N=500	N=1,000	N=2,000	N=5,000
0	1.0	9.03818		8.71429	8.18513	9.50483	9.03687	8.88721	9.01019
1	2.036	7.21980		6.61596	6.25417	7.57782	7.20885	7.06717	7.25115
2	1.812	3.61347		2.43863	2.55259	3.71219	3.58305	3.46744	3.78580
3	0.770	1.03264		-0.65570	0.05869	0.85751	0.96480	0.90688	1.33456
4	0.129	0.12900		-1.62021	-0.56999	-0.19295	0.01785	0.03715	0.47674
5	0.0	0.0		-1.13021	-0.14361	-0.23974	-0.07123	-0.06663	0.31441
6	0.0	0.0		0.07309	0.72159	0.01396	[illegible]	-0.09510	0.25614
7	0.0	0.0		1.27620	1.52510	0.25614	0.27813	-0.19247	0.20619
8	0.0	0.0		1.86544	1.75538	0.43543	0.40113	-0.27603	0.16571
Standard Deivation of $\phi_{xx}(0)$:				2.067	1.468	0.930	0.657	0.464	0.293

Table A-3

Computed Mean Lagged Products for

Standard Example # 3

$$H(z) = \frac{1 - 0.859z^{-1}}{1 + 0.853z^{-2}}$$

Input sequence: Gaussian, zero-mean, $\sigma^2 = 1.0$

	$\phi_{nx}(j)$	$\phi_{xx}(j)$		Mean Lagged Products $\hat{\phi}_{xx}(j)$					
j	Theo.	Theo.		N=100	N=200	N=500	N=1,000	N=2,000	N=5,000
0	1.0	6.38010		6.06867	6.79562	6.06990	6.56377	6.52532	6.56915
1	-0.859	-0.46357		-0.55375	-0.60126	-0.46536	-0.52455	-0.47059	-0.49042
2	-0.853	-5.44222		-5.12886	-5.75304	-5.16660	-5.60531	-5.54682	-5.60657
3	0.73273	0.39543		0.90604	0.65261	0.36367	0.48098	0.33114	0.41410
4	0.72761	4.64222		4.34338	5.05528	4.55324	4.91535	4.78229	4.81313
5	-0.62502	-0.33730		-1.52285	-1.02829	-0.43304	-0.56023	-0.25745	-0.34577
6	-0.62065	-3.95981		-3.16[illegible]2	-4.13565	-3.92259	-4.26414	-4.14338	-4.13151
7	0.53314	0.28772		1.782[illegible]0	1.24826	0.44581	0.64483	0.23614	0.27742
8	0.52942	3.37772		2.22215	3.16928	3.36351	3.65927	3.59548	3.56833
Standard Deviation of $\hat{\phi}_{xx}(0)$:				2.26	1.60	1.02	0.72	0.51	0.32

Table A-4

Computed Mean Lagged Products for
Standard Example # 4

$$H(z) = \frac{1 - 1.7835z^{-1} + 0.793z^{-2}}{1 - 1.337z^{-1} + 1.632z^{-2} - 0.987z^{-3} + 0.660z^{-4}}$$

Input sequence: Gaussian, zero-mean, $\sigma^2 = 1.0$

j	$\phi_{nx}(j)$ Theo.	$\phi_{xx}(j)$ Theo.		Mean Lagged Products $\hat{\phi}_{xx}(j)$ N=100	N=200	N=500	N=1,000	N=2,000	N=5,000
0	1.0	5.58830		7.19060	6.53263	5.44004	5.51625	5.33161	5.39452
1	-0.44650	1.20165		2.17533	1.67891	1.16801	1.17999	1.0745	1.07801
2	-1.43597	-3.33402		-3.11472	-3.36426	-3.11366	-3.17705	-3.19314	-3.28089
3	-0.20421	-1.69611		-2.44584	-1.85488	-1.57103	-1.57189	-1.57575	-1.55615
4	0.96979	0.67117		-1.03167	-0.15392	0.38125	0.43323	0.64016	0.78354
5	0.50725	-0.41836		-2.13622	-1.71304	-0.61942	-0.71985	-0.33677	-0.34037
6	-0.15830	-1.12830		-1.13034	-1.17128	-0.80567	-0.88032	-0.88864	-1.11214
7	0.05247	0.95610		2.44809	2.18631	1.15258	1.27706	0.92555	0.82528
8	0.18910	2.26380		3.27991	2.92175	1.81757	1.99834	1.84783	2.07245
Standard Deviation of $\hat{\phi}_{xx}(0)$:				1.35	0.96	0.60	0.43	0.30	0.19

Table A-5

Computed Mean Lagged Products for

Standard Example # 5

$$H(z) = \frac{1 - 1.740z^{-1} + 0.810z^{-2}}{1 - 1.352z^{-1} + 1.338z^{-2} - 0.662z^{-3} + 0.240z^{-4}}$$

Input sequence: Gaussian, zero-mean, $\sigma^2 = 1.0$

j	$\phi_{nx}(j)$ Theo.	$\phi_{xx}(j)$ Theo.		Mean Lagged Products $\hat{\phi}_{xx}(j)$ N=100	N=200	N=500	N=1,000	N=2,000	N=5,000
0	1.0	2.99383		3.31871	3.08134	3.07722	2.99020	3.03731	2.95026
1	-0.38800	0.45912		0.50500	0.45109	0.47408	0.43139	0.44762	0.41820
2	-1.05258	-1.83151		-1.83220	-1.83848	-1.87840	-1.80222	-1.83243	-1.80584
3	-0.24194	-1.21878		-1.34794	-1.08189	-1.23581	-1.16127	-1.24523	-1.15962
4	0.58439	0.38819		0.23536	0.36839	0.27634	0.36396	0.35582	0.41318
5	0.51012	0.83291		0.68754	0.51879	0.78330	0.72742	0.86431	0.80605
6	0.00023	0.23943		0.29792	0.14287	0.37005	0.30068	0.31672	0.20484
7	-0.23731	-0.24124		-0.10121	-0.04916	-0.14513	-0.16862	-0.23016	-0.24824
8	-0.12369	-0.18829		-0.17103	-0.01214	-0.29456	-0.29686	-0.25596	-0.16272
Standard Deviation of $\hat{\phi}_{xx}(0)$:				0.65	0.46	0.29	0.20	0.15	0.09

directly as input to the recursive equations. This data was used to provide examples of the spectral analysis of non-Gaussian processes.

References for Appendix A:

a.) Kemeny, J.G. and Kurtz, T.E., A Manual for BASIC, Dartmouth College, Hanover, New Hampshire, 1965.

b.) Juncosa, M.L., "Random Number Generation of the BRL High-Speed Computing Machines", U.S. Army, Aberdeen Proving Grounds, Aberdeen, Maryland, Ballistics Research Laboratory, Report # 855, May 1953.

APPENDIX B

A Linear Least Squares Approach to All-Pole Approximation

In this derivation of the least squares all-pole approximation, it is assumed initially that the general process x_t can be written in the form

$$x_t = \varepsilon_t - \sum_{j=1}^{M} d_j x_{t-j} \qquad \text{(B-1)}$$

where ε_t is a random variable of zero mean and unknown variance. This is to say that x_t can be expressed as a linear combination of its own past values and a random perturbation that can be interpreted as a prediction error. In the least squares approach, the coefficients $\{d_j\}$ are chosen such that the mean square value of the random component of x_t is minimized. Once this is done, x_t can be predicted on the basis of its M most recent values with the smallest possible mean square error.

If, from equation (B-1)

$$\varepsilon_t = \sum_{j=0}^{M} d_j x_{t-j} \; , \; d_0 = 1 \qquad \text{(B-2)}$$

then

$$\varepsilon_t^2 = \sum_{j=0}^{M} \sum_{k=0}^{M} d_j d_k x_{t-j} x_{t-k} \qquad \text{(B-3)}$$

and taking the expected value yields

$$\overline{\varepsilon_t^{\,2}} = \sum_{j=0}^{M} \sum_{k=0}^{M} d_j d_k \phi_{xx}(j-k) \qquad \text{(B-4)}$$

The set $\{d_j\}$ which minimizes $\overline{\varepsilon_t^{\,2}}$ must satisfy the system

$$\frac{\partial \overline{\varepsilon_t^{\,2}}}{\partial d_j} = 0, \; j = 1,2,\ldots,M \qquad \text{(B-5)}$$

and using equation (B-4),

$$\sum_{k=0}^{M} d_k \phi_{xx}(k-j) = 0, \; j = 1,2,\ldots,M \qquad \text{(B-6)}$$

This is the same set of equations derived in the main text.

To find the minimum mean square value of the random component when equation (B-6) is satisfied, equation (B-4) is written as

$$\overline{\varepsilon_t^{\,2}} = \sum_{j=0}^{M} d_j \sum_{k=0}^{M} d_k \phi_{xx}(j-k) \qquad \text{(B-7)}$$

The inner summation vanishes except when $j = 0$, and hence

$$\overline{\varepsilon_t^{\,2}}_{min} = \sum_{k=0}^{M} d_k \phi_{xx}(k) \qquad \text{(B-8)}$$

It is instructive to investigate the correlation properties of the random component ε_t. $\phi_{x\varepsilon}(k)$ is found by multiplying equation (B-2) by x_{t-k} and taking the expected value. The result is

$$\phi_{x\varepsilon}(k) = \sum_{j=0}^{M} d_j \phi_{xx}(j-k) \qquad \text{(B-9)}$$

By inspection,

$$\phi_{x\varepsilon}(0) = \sum_{j=0}^{M} d_j \phi_{xx}(j) = \overline{\varepsilon_{t_{min}}^{2}} \qquad \text{(B-10)}$$

and by equation (B-6),

$$\phi_{x\varepsilon}(k) = 0, \ k = 1,2,\ldots,M \qquad \text{(B-11)}$$

Thus, the random component of x_t is uncorrelated with the last M values of the process. $\phi_{\varepsilon\varepsilon}(k)$ is found by multiplying equation (B-2) by ε_{t+k} and taking the expected value:

$$\phi_{\varepsilon\varepsilon}(k) = \sum_{j=0}^{M} d_j \phi_{x\varepsilon}(k+j) \qquad \text{(B-12)}$$

$\phi_{\varepsilon\varepsilon}(0)$ is, of course, equal to $\overline{\varepsilon_{t_{min}}^{2}}$ and thus

$$\phi_{\varepsilon\varepsilon}(0) = \sum_{j=0}^{M} d_j \phi_{xx}(j) \qquad \text{(B-13)}$$

For a general process, equation (B-12) indicates that ε_t is not an uncorrelated sequence; $\phi_{\varepsilon\varepsilon}(k)$ is non-zero when k is non-zero. Only if x_t is indeed an M^{th} order autoregression can ε_t be shown to be uncorrelated. In the general case, however, it can be shown that ε_t becomes uncorrelated in the limit of large M when the magnitude of $\phi_{xx}(j)$ decreases rapidly enough with its argument. This requirement is essentially that of stability for the process.

If a mixed type process is represented as a finite order autoregression with coefficients $\{d_j\}$ and input noise variance equal to $\overline{\epsilon_t^2}_{min}$, the approximation error of the model can be interpreted as arising from the assumption (inherent in the autoregressive representation) that the random component is an uncorrelated sequence. As we have seen, this is not true except where x_t is an M^{th} order autoregression to begin with.

Equations (B-6) and (B-8) are identical with those derived for the all-pole approximation in the main text and also with those found for Gaussian processes by maximum likelihood arguments.

APPENDIX C

RECURSIVE CALCULATION OF AUTOREGRESSIVE MODELS

The results to be developed here are presented in both Bartlett[2] and Whittle[20], where reference is made to sources that seem generally unavailable. This Appendix presents a derivation of the equations describing the relationship between the coefficients of the $(M + 1)^{th}$ and M^{th} autoregressive models for a process and thus shows how a recursive estimation scheme can be carried out. The result is then extended to indicate the relationship between successive values of the input variance estimate, and it is demonstrated that these must be monotonically decreasing.

If the theoretical autocorrelation sequence for a discrete time series is denoted $\phi(j)$, and the coefficients of the M^{th} order autoregressive model are denoted $\left\{d_k^{(M)}\right\}_{k=0}^{M}$ (where $d_0^{(M)} = 1$), then

$$\sum_{k=0}^{M} d_k^{(M)} \phi(j-k) = 0, \quad j = 1,2,\ldots,M \tag{C-1}$$

For the $(M + 1)^{th}$ order model,

$$\sum_{k=0}^{M+1} d_k^{(M+1)} \phi(j-k) = 0, \quad j = 1,2,\ldots,M+1 \tag{C-2}$$

The first M equations of the latter can be written as

$$\sum_{k=1}^{M} d_k^{(M+1)} \phi(j-k) + d_{M+1}^{(M+1)} \phi(j-M-1) + \phi(j) = 0, \; j = 1,2,\ldots,M \tag{C-3}$$

Similarly, the equations of (C-1) can be written

$$\sum_{k=1}^{M} d_k^{(M)} \phi(j-k) + \phi(j) = 0 \;, \quad j = 1,2,\ldots,M \tag{C-4}$$

and when they are subtracted from the equations of (C-3), the result is

$$\sum_{k=1}^{M} \left[d_k^{(M+1)} - d_k^{(M)} \right] \phi(j-k) + d_{M+1}^{(M+1)} \phi(j-M-1) = 0 \qquad j = 1,2,\ldots,M \tag{C-5}$$

Defining

$$f_k \equiv d_k^{(M+1)} - d_k^{(M)} \tag{C-6}$$

equation (C-5) can be expressed as

$$\left[\quad \Phi \quad \right] \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_M \end{bmatrix} = -d_{M+1}^{(M+1)} \begin{bmatrix} \phi(M) \\ \phi(M-1) \\ \cdot \\ \cdot \\ \cdot \\ \phi(1) \end{bmatrix} \tag{C-7}$$

where Φ is the theoretical process covariance matrix defined in the main body. The solution to (C-7) is written directly as

$$\begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_M \end{bmatrix} = -d_{M+1}^{(M+1)} \begin{bmatrix} \Phi^{-1} \end{bmatrix} \begin{bmatrix} \phi(M) \\ \phi(M-1) \\ \cdot \\ \cdot \\ \cdot \\ \phi(1) \end{bmatrix} \tag{C-8}$$

and this should be compared with the solution of the system of equation (C-3):

$$\begin{bmatrix} d_1^{(M)} \\ d_2^{(M)} \\ \cdot \\ \cdot \\ \cdot \\ d_M^{(M)} \end{bmatrix} = - \begin{bmatrix} \Phi^{-1} \end{bmatrix} \begin{bmatrix} \phi(1) \\ \phi(2) \\ \cdot \\ \cdot \\ \cdot \\ \phi(M) \end{bmatrix} \tag{C-9}$$

Because of the symmetry of Φ^{-1}, equation (C-9) can be written

$$\begin{bmatrix} d_M^{(M)} \\ d_{M-1}^{(M)} \\ \cdot \\ \cdot \\ \cdot \\ d_1^{(M)} \end{bmatrix} = - \begin{bmatrix} \Phi^{-1} \end{bmatrix} \begin{bmatrix} \phi(M) \\ \phi(M-1) \\ \cdot \\ \cdot \\ \cdot \\ \phi(1) \end{bmatrix} \tag{C-10}$$

and combining (C-8) and (C-10),

$$f_k = d_{M+1}^{(M+1)} \; d_{M+1-k}^{(M)} \qquad \text{(C-11)}$$

or

$$d_k^{(M+1)} = d_k^{(M)} + d_{M+1}^{(M+1)} d_{M+1-k}^{(M)} \quad , \; k = 1,2,\ldots,M \qquad \text{(C-12)}$$

There remains the problem of finding $d_{M+1}^{(M+1)}$. The last equation of the system (C-2) is first written

$$\sum_{k=0}^{M} d_k^{(M+1)} \phi(M+1-k) + d_{M+1}^{(M+1)} \phi(0) = 0 \qquad \text{(C-13)}$$

The partial result of equation (C-12) is now substitued into equation (C-13) to yield

$$\sum_{k=0}^{M} d_k^{(M)} \phi(M+1-k) + d_{M+1}^{(M+1)} \sum_{k=1}^{M+1} d_{M+1-k}^{(M)} \phi(M+1-k) = 0 \qquad \text{(C-14)}$$

The second summation is identical with $\sum_{k=0}^{M} d_k^{(M)} \phi(k)$, and hence

$$d_{M+1}^{(M+1)} = \frac{-\sum_{k=0}^{M} d_k^{(M)} \phi(M+1-k)}{\sum_{k=0}^{M} d_k^{(M)} \phi(k)} \qquad \text{(C-15)}$$

Together, equations (C-12) and (C-15) provide the means for computing the $d_k^{(M+1)}$ from the $d_k^{(M)}$.

The variance estimates for the M^{th} and $(M+1)^{th}$ order models have the form respectively

$$v_M = \sum_{k=0}^{M} d_k^{(M)} \phi(k) \qquad \text{(C-16a)}$$

$$v_{M+1} = \sum_{k=0}^{M+1} d_k^{(M+1)} \phi(k) \qquad \text{(C-16b)}$$

Substituting equation (C-12) into equation (C-16b) and using (C-16a) gives

$$v_{M+1} = v_M + d_{M+1}^{(M+1)} \sum_{k=1}^{M+1} d_{M+1-k}^{(M)} \phi(k) \qquad \text{(C-17)}$$

and after manipulating the indices of summation,

$$v_{M+1} = v_M + d_{M+1}^{(M+1)} \sum_{k=0}^{M} d_k^{(M)} \phi(M+1-k) \qquad \text{(C-18)}$$

Now it is noted that equation (C-15) can be written

$$d_{M+1}^{(M+1)} = \frac{-\Delta_M}{v_M} \qquad \text{(C-19)}$$

where we have defined

$$\Delta_M \equiv \sum_{k=0}^{M} d_k^{(M)} \phi(M+1-k) \qquad \text{(C-20)}$$

Now equation (C-18) becomes

$$v_{M+1} = v_M - \frac{\Delta_M^2}{v_M} \qquad \text{(C-21)}$$

and if Δ_M^2 is non-zero,

$$v_{M+1} < v_M \qquad \text{(C-22)}$$

Furthermore, by considering the form of the system of equation (C-1), it becomes apparent that Δ_M can not be zero if a solution for the $d_j^{(M)}$ exists. Thus, v_M must be monotonically decreasing for increasing M. The only exception to this condition is when the process of interest is truly an M^{th} order autoregression, in which case $d_{M+1}^{(M+1)}$ will vanish, and by equation (C-18),

$$v_{M+1} = v_M \tag{C-23}$$

LITERATURE CITED:

[1] Hannan, E.J., Time Series Analysis, London, Methuen, 1960.

[2] Bartlett, M.S., An Introduction to Stochastic Processes, Cambridge, Cambridge University Press, 1955.

[3] Grenander, U. and Rosenblatt, M., Statistical Analysis of Stationary Time Series, New York, John Wiley & Sons, 1957.

[4] Blackman, R.B. and Tukey, J.W., The Measurement of Power Spectra, New York, Dover, 1959.

[5] Bendat, J.S. and Piersol, A.G., Measurement and Analysis of Random Data, New York, John Wiley & Sons, 1966.

[6] Cooley, J.W. and Tukey, J.W., "An Algorithm for the Machine Calculation of Complex Fourier Series", Math. Comp., Vol. 19, pp. 297-301, April, 1965.

[7] Cochran, W.T. et al (G-AE Subcommittee on Measurement Concepts), "What is the Fast Fourier Transform?", IEEE Trans. on Audio and Electroacoustics, Vol. AU-15, No. 2, pp. 45-55, June 1967.

[8] Welch, P.D., "The Use of the Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Trans. on Audio and Electroacoustics, Vol. AU-15, No. 2, pp. 70-73, June 1967.

[9] Cooley, J.W., Lewis, P.A.W., and Welch, P.D., "Application of the Fast Fourier Transform to Computation of Fourier Integrals, Fourier Series, and Convolution Integrals", IEEE Trans. on Audio and Electroacoustics, Vol. AU-15, No. 2, pp. 79-84, June 1967.

[10] Rosenberg, A.E., and Shen, D.W.C., "Regression Analysis and its Application to the System Identification Problem", Proc. 4th JACC, pp. 446-461, June 1963.

[11] Steiglitz, K. and McBride, L.E., "A Technique for the Identification of Linear Systems", IEEE Trans. on Automatic Control, Vol. AC-10, No. 4, pp. 461-464, Oct. 1965.

[12] Åström, K.J., and Bohlin, T., "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", Proc. 2nd IFAC Sym. on Th. of Self-Adaptive Control Sys., (Teddington, England), pp. 96-111, Sept. 1965.

[13] Rogers, A.E., and Steiglitz, K., "Maximum Likelihood Estimation of Rational Transfer Function Parameters", IEEE Trans. on Automatic Control, Vol. AC-12, No. 5, pp. 594-597, Oct. 1967.

[14] Sakrison, D.J., "The Use of the Stochastic Approximation to Solve the System Identification Problem", IEEE Trans. on Automatic Control, Vol. AC-12, No. 5, pp. 563-567, Oct. 1967.

[15] Joseph, P., Lewis, J., Tou, J., "Plant Identification in the Presence of Disturbances and Application to Digital Adaptive Systems", Trans. AIEE (Applications and Industry) Vol. 80, pp. 18-24, March 1961.

[16] Levin, M.J., "Estimation of a System Pulse Transfer Function in the Presence of Noise", IEEE Trans. on Automatic Control, Vol. AC-9, No. 3, pp. 229-235, July 1964.

[17] Mann, H.B. and Wald, A., "On the Statistical Treatment of Linear Stochastic Difference Equations", Econometrica, Vol. 11, pp. 173-220, July 1943.

[18] Steiglitz, K., "Power Spectrum Identification for Adaptive Systems", IEEE Trans. on Application and Industry, Vol. 83, pp. 195-197, May 1964.

[19] Wold, H., A Study in the Analysis of Stationary Time Series, Stockholm, Almquist & Wiksell, 1938.

[20] Whittle, P., Prediction and Regulation by Linear Least Square Methods, Princeton, Van Nostrand & Co., 1963.

[21] Tretter, S.A., and Steiglitz, K., "Power Spectrum Identification in Terms of Rational Models", IEEE Trans. on Automatic Control. Vol. AC-12, No. 2, pp. 185-188, April 1967.

[22] Whittle, P., "Estimation and information in stationary time series", Ark. Mat. Astr. Fys., Vol. 2, pp. 423-434, 1953.

[23] Whittle, P., Appendix 2 in Wold (reference 19).

[24] Walker, A.M., "Large-sample Estimation of Parameters for Moving Average Models", Biometrika, Vol. 48, pp. 343-357, 1961.

[25] Zetterberg, L.H., "Estimation of the Parameters of a Linear Difference Equation with Application to EEG Analysis", Royal Institute of Technology, Stockholm, Tech. Rpt. No. 18, April 1968.

[26] Walker, A.M., "Large-sample Estimation of Parameters for Autoregressive Processes with Moving-Average Residuals", Biometrika, Vol. 49, pp. 117-131, 1962.

[27] Tou, J.T., Digital and Sampled-data Control Systems, New York, John Wiley & Sons, 1959.

[28] Ragazzini, J.R., and Franklin, G.F., Sampled-data Control Systems, New York, McGraw-Hill, 1958.

[29] Lee, Y.W., Statistical Theory of Communication, New York, John Wiley & Sons, 1960.

[30] Parzen, E., Stochastic Processes, San Francisco, Holden-Day, 1962.

[31] Walker, A.M., "The Asymptotic Distribution of Serial Correlation Coefficients for Autoregressive Processes with Dependent Residuals", Proc. Camb. Phil. Soc., Vol. 50, pp. 60-64, 1954.

[32] Anderson, R.L., "Distribution of the Serial Correlation Coefficient", Ann. Math. Statist., Vol. 13, pp. 1-13, 1942.

[33] Tretter, S.A., "On the Relationship of Maximum Likelihood Sampled-Data Power Spectrum Identification and Optimum Prediction Filters", IEEE Trans. on Automatic Control, Vol. AC-13, No. 3, pp. 303-304, June 1968.

[34] Cramer, H., Mathematical Methods of Statistics, Princeton Univ. Press, Princeton, 1946.

[35] Hsia, T.C. and Landgrebe, D.A., "On a Method for Estimating Power Spectra", IEEE Trans. on Instrumentation and Measurement, Vol. IM-16, No. 3, pp. 255-257, September 1967.

[36] Durbin, J., "Efficient Estimation of Parameters in Moving-Average Models", Biometrika, Vol. 46, pp. 306-316, 1959.

[37] Faddeeva, V.N., Computational Methods of Linear Algebra, Dover Publications, New York, 1959.

[38] Nash, R.A., Jr., and Tuteur, F.B., "The Effect of Uncertainties in the Noise Covariance Matrices on the Maximum Likelihood Estimate of a Vector", IEEE Trans. on Automatic Control, Vol. AC-13, No. 1, pp. 86-88, February 1968.

[39] Lee, R.C.K., Optimal Estimation, Identification, and Control, M.I.T. Press, Cambridge, Mass., 1964.

[40] Durbin, J., "Estimation of Parameters in Time Series Regression Models", J. Roy. Statist. Soc., B.22, pp. 139-153, 1960.

[41] Tretter, S.A. and Steiglitz, K., "Adaptive Signal Reconstruction", Proc. NEC, Vol. 21, pp. 487-492, October 1965.

[42] Steiglitz, K. and Thomas, J.B., "A Class of Adaptive Matched Digital Filters", Proc. NEC, Vol. 20, pp. 641-646, October 1964.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) U. S. Naval Ordnance Laboratory White Oak, Silver Spring, Maryland 20910	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE
THE SPECTRAL ANALYSIS OF DISCRETE TIME SERIES IN TERMS OF LINEAR REGRESSIVE MODELS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
NOLTR

5. AUTHOR(S) (First name, middle initial, last name)
Edward C. Whitman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
23 June 1970	258	42

8a. CONTRACT OR GRANT NO. WEPTASK MAT 03L-000/ZF17-312-001 b. PROJECT NO. Problem 009 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) NOLTR 70-109 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale, its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Material Washington, D.C.

13. ABSTRACT

This paper considers several methods for the spectral analysis of discrete time series modeled as linear regressive processes. The most general of these models is the so-called "mixed-type" process in which the present value of the series is given as a weighted sum of both its own past values and those of an uncorrelated random sequence. Special cases of this class are the autoregression and the moving average. In the present approach, the determination of the weighting coefficients of the model is equivalent to specifying the sampled power spectrum of the process, and the central problem treated here is that of estimating such a set of parameters on the basis of a sample sequence of the series.

After a development of the necessary mathematical background, the estimation problem is formulated and solved from several alternative points of view, with particular attention to statistical stability, sample size requirements, and possible approximation errors. For each method, the results are demonstrated by a computational analysis of the spectra of a set of computer generated examples. A general purpose spectral analysis algorithm for digital computation is proposed and discussed.

Methods for extending an existing spectral estimate to take into account newly available data are briefly explained, and suggestions for further research are presented.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED
Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Spectral analysis						
Discrete time series						
Linear regressive models						
Correlation						
Adaptive processing						
Stochastic processes						
Statistical estimation						
Sampled power spectra						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

UNCLASSIFIED
Security Classification