RECURSIVE ESTIMATION OF DISCRETE TIME SYSTEM PARAMETERS AND TIME DELAY

By

SURESH K.S. RAO

Bachelor of Technology
Karnataka Regional Engineering College
Karnataka, India
1979

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1983

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1988

RECURSIVE ESTIMATION OF DISCRETE TIME
SYSTEM PARAMETERS AND TIME DELAY

Thesis Approved:

Gary E. Young
Thesis Adviser

Karl N. Reid

Lawrence L Hoberock

Martin T Hagan

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to express my gratitude to my adviser, Dr. G. E. Young, for his constant guidance and suggestions and I am indebted to him for his inspiration and encouragement. I am also indebted to my committee members, Dr. M. Hagan, Dr. L. L. Hoberock and Dr. K. N. Reid. Dr. Hagan and Dr. Reid's review and criticism, and Dr. Hoberock's support and funding efforts have all been extremely beneficial.

In addition, I would like to express my gratitude to my friends Mr. Sunkaran Mohan and Dr. Yousef Zurigat for their encouragement and helpful suggestions. I also wish to thank all the faculty and staff of the Mechanical and Aerospace Engineering Department at Oklahoma State University. Thanks go to Marsha Speer for doing an excellent job of typing the manuscript.

I would like to express my deepest gratitude to my parents, Dr. and Mrs. K. M. Subraya, my sister, Mrs. Nayanatara and my brother, Mr. Prakash Rao, for their support and encouragement throughout my graduate study at Oklahoma State University.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure		Page

Figure | Page

NOMENCLATURE

$A(q^{-1})$	polynomial matrix in q^{-1}
A^T	transpose of matrix A
$B(q^{-1})$	polynomial matrix in q^{-1}
C(j)	cosine $\frac{2\pi jk}{N}$
d	time delay
{e(t)}	white noise sequence
$\underline{E}$	Fourier Spectrum of residue
E^*	complex conjugate of E
$E[\cdot]$	expectation operator
{ f }	data sequence
{ F }	Fourier spectrum of f
i	$\sqrt{-1}$
J	performance measure
k	frequency argument
L	block size of data points
I	Identity matrix
n,t	time variable
N	number of DFT points
P(k)	covariance matrix
q^{-1}	backward shift operator

$S(j)$	sine $\frac{2\pi jk}{N}$
T	shifting transformation for basis vector
$u(t)$	input variable
$\underline{U}$	Fourier Spectrum of {u}
U_I	imaginary part of U
U_R	real part of U
V	objective function
V'	$\frac{dV}{dt}$
$\underline{W}_N^{kn}$	Discrete Fourier Polynomial (basis function)
$\underline{W}_N^{kn}$	Basis vector
$y(t)$	output variable
$\underline{Y}$	Fourier Spectrum of {y}
Y_I	imaginary part of Y
Y_R	real part of Y
Z	vector of components of Y
$\varepsilon(n)$	residue at time n
θ	parameter vector
θ^o	true value of θ
θ^*	least squares estimate of θ
$\hat{\theta}$	estimate of θ
Φ_N	information matrix

$\Phi_N^{\#}$	pseudo-inverse of Φ_N
ϕ	element of Φ_N
δ	$\frac{d}{dt}$ differential operator
$\lVert \cdot \rVert_2$	2 norm of a vector
$<.,.>$	inner product of vectors
ω	Frequency in radians/unit time
λ	Step size parameter

CHAPTER I

INTRODUCTION

A large number of systems can be described reasonably well by discrete-time linear models with time delays. However, the elusive nature of the time delays affect the usefulness of these models for prediction and controller design. Furthermore, if the system parameters are unknown, the presence of the unknown delay poses not only a problem for parameter estimation but also for the estimation of the time delay itself.

If the structure of a model describing the system and the time delay are known, a number of alternatives exist for parameter estimation (Goodwin and Sin, 1984; Ljung, 1987; Mendel, 1985; Astrom and Wittenmark, 1984). All the parameter estimation schemes are based on the assumption that the model order and the time delays are known. Among these variables, the model is most sensitive to the time delay.

The main purpose of this research is to develop a new and systematic procedure for the recursive estimation of constant or slowly varying system parameters and time delay. The usefulness of this algorithm lies in the ability to identify time delays without increasing the computational effort, unlike some of the other existing methods. The new technique provides consistent and unbiased estimates of the parameters and the time delay.

The proposed estimation scheme is based on the estimation of parameters and time delay in the frequency domain. The transformation to the frequency domain provides two advantages. First, it provides a good understanding of the system since many analyses, e.g., identifiability and persistency of excitation, are carried out in the frequency domain. Second, the transformation to the frequency domain parameterizes the delay term. For systems with no delay or known delay, the frequency domain model remains linear in the parameters if the time domain model is linear in the parameters. However, if the time delay is unknown, the transformed model is no longer linear in the time delay parameter. The Fourier Transform will be adopted to transform the time domain model into the frequency domain. The Fourier transform being complex, renders the transformed model complex. The complex equation describing the frequency domain model can be described by two real equations. The problem then is to develop a multi-input multi-output nonlinear recursive estimator in the frequency domain. A finite length of data is transformed into the frequency domain using the Fast Fourier Transform (FFT). The parameters and time delay are then recursively estimated at each frequency. Since most of the real signals are made up of a limited number of frequencies, their spectrum is finite and the Fourier Transform has a data compression effect.

The Frequency Domain estimator is equally applicable to systems with no delays, known delays and unknown delays. Typical discrete time systems described by linear difference equations (ARX models) with:

1. no delays and unknown parameters and
2. unknown delays and unknown parameters will be considered.

The examples presented in this thesis are:

1. frequency domain estimation of systems with no delays and unknown parameters.

2. frequency domain estimation of systems with unknown delays and unknown parameters.

Chapter II of this thesis is a summary of the literature review.

CHAPTER II

LITERATURE REVIEW

A considerable amount of information on various parameter estimation techniques is available in the literature (e.g., Ljung, 1987; Ljung and Soderstrom, 1985; Astrom and Wittenmark, 1984; Goodwin and Sin, 1984; Mendel, 1986). The major assumptions in all these schemes are:

1. The model structure is known (generally linear form).
2. The order of the model is known i.e., the structural indices are known.
3. The pure delay is known.

Given a model of the form

$$A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + e(t) \tag{2.1}$$

where

q^{-1} is the backward shift operator defined by

$q^{-1}y(t) = y(t-1)$

then d is the pure delay and is assumed to be known. The only unknowns in the model are the coefficients a_i and b_i where

$$A(q^{-1}) = 1 + a_1q^{-1} + \ldots + a_nq^{-n} \tag{2.2}$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \ldots + b_mq^{-m} \tag{2.3}$$

m, n are known.

Such a model can describe many processes reasonably well. But the usefulness of these models for parameter estimation is seriously hindered by the elusive nature of the time delays.

If the delay is unknown, two alternatives exist. First, a model of a certain order is chosen and fitted to the data by estimating the parameters. The next step is of model validation where diagnostic checks are applied to identify a lack of proper fit. This could be based on minimization of certain "goodness of fit" measures e.g., some function of error between the model output and that of the actual system. If the model is found inadequate, the model order is increased and the above procedure is repeated. The procedure is iterative and could be quite time consuming.

Alternately, time domain and frequency domain methods exist for estimating the plant order and transfer functions respectively. The time domain techniques for identifying the order of an Autoregressive Moving Average (ARMA) process is based on correlation analysis and computing the generalized partial autocorrelation (GPAC) arrays (Woodward and Gray, 1979; Bednar and Coberly, 1976). Once the model order has been established, the next step is to estimate the parameters. This again is a two step process and involves considerable effort.

The frequency domain and spectral methods, e.g., power spectrum analysis (Gabr, 1987; Gabr and Subbarao, 1984), essentially provide magnitude and phase information of the autoregressive moving average type of systems (ARMA). Models of different orders are then fitted iteratively until the errors are minimized. Both these approaches can be categorized as preliminary system identification procedures. Ljung (1987) presents a method for nonparameteric identification of a system. Called the Emperical Transfer Function Estimation (ETFE), it essentially computes the system transfer function at various frequencies.

If the model structure and the structural indices are assumed known, a number of alternatives exist. One approach is to use an overparameterized model (Kurz and Goedecke, 1981). For example, if the true system is

$$y(t) + a_1y(t-1) + a_2y(t-2) = b_3u(t-3) + b_4u(t-4) + e(t) \tag{2.4}$$

where n = 2, m = 1, and d = 3, choose the overparameterized model as

$$y(t) + a_1y(t-1) + a_2y(t-2) = b_0u(t) + b_1u(t-1) + b_2u(t-2) + b_3u(t-3) + b_4u(t-4) + b_5u(t-5) + e(t) \tag{2.5}$$

where the parameters b_0, b_1, b_2 and b_5 act as indicator functions. The delays are estimated closest to an integer by utilizing computationally elaborate means of rejecting extraneous estimated parameters. The major drawback of this approach is expanding the $B(q^{-1})$ polynomial to incorporate dummy parameters contributes significantly to the computation and cost overhead by increasing the dimension of the model. Furthermore, the upper bound on the delay is assumed to be known.

As an example, consider the third order system described by equation (2.4) where $a_1 = 0.5$, $a_2 = 0.25$, $b_3= 0.125$ and $b_4 = 1.0$. To use the overparameterized model for estimation of the unknown parameters and time delay, the maximal order of the delay has to be known. In this case it is d = 3. If the maximum value of the delay is incorrectly assumed, the parameters will be incorrectly estimated since a different model is being fitted to the data. A judicious choice of model for estimating the parameters of equation (2.4) is to use equation (2.5). However, this increases the computation burden considerably, since, in this case five parameters must be estimated instead of two. The resulting covariance matrix is also 5 by 5 instead of 2 by 2 thereby adding to the high cost in computation and time. Figure 1 shows the estimated parameters of the $B(q^{-1})$ polynomial when the noise to signal ratio is 10 percent. Parameters b_0, b_1, b_2, b_3 and b_5 are close to zero. The pure delay of the system cannot be inferred with confidence from the figure, since the true system could have had the leading coefficients of the $B(q^{-1})$ polynomial close to zero.

Ljung (1984) has shown that if there is a mismatch between the model order and the true system, the parameter estimates are biased. Lee and Hang (1984) have shown by simulation that the performance of the method of overparameterization used in the estimation of unknown or time varying delays is extremely sensitive to the presence of noise and step disturbances. Another major drawback of the estimation schemes based on overparameterization is that the upper bound on the delay has to be known in advance. If the delay drifts or falls outside this interval, this approach breaks down. Further, expanding the model to incorporate dummy parameters contributes significantly to the computational expense by increasing the dimension of the model. In the case of overparameterization, one has to pay the price of increasing computation as time delay increases.

In the case of continuous time systems (Agarwal and Canudas, 1986) the time delay and continuous time parameters are estimated by approximating the time delay in the frequency domain by a rational transfer function. Derivation of the process inputs and outputs are constructed using multiple filters. The system is discretized and the parameters estimated. The model becomes nonlinear in the desired parameters and hence a nonlinear estimation scheme has to be used. This approach is quite tedious and cannot be applied to systems that are discrete to begin with. Other attempts include assuming a constant delay and estimating the parameters in the time domain by minimizing certain cost functions. These get complicated because of multiple minima of the cost function that is minimized (Pupeikis, 1985).

Methods which circumvent this problem resort to such techniques as identification of several different model structures from the available data followed by the selection of one. These methods involve expensive computation and restrict themselves to off-line procedures (Rao and Sivakumar, 1976)

More recently, Juricic (1987) took the same approach with a stochastic setting. The basic idea is estimating different models that belong to a model set by observing their outputs simultaneously. The parameters and the time delay are deduced from the model that fits the process output in the best way. Again, the major assumption is that the unknown delay belongs to a finite set $D = (D_1, D_2 ..., D_n)$. The need to observe the outputs of all the models may be extremely costly since n estimators have to run in parallel.

In the proposed study, the unknown parameters and time delay will be estimated in the frequency domain. The main idea behind transforming the system to the frequency domain is to parameterize the delay term. In fact, a number of papers on the use of orthogonal polynomials for parameter estimation exist in the literature. King and Paraskevopoulos (1974), have used Laguerre polynomials for parameter identification; Chebyshev series and Legendre series have been utilized by Paraskevopoulos (1983, 1985); Jacobi series has been used by Liu and

Shin (1985); and the Walsh series has been used by Palanisamy and Battacharya (1981).

Chou and Horng (1986) have used the shifted Chebyshev series for identification and analysis of time-varying systems. Shih and Kung (1986) have extended the use of shifted Chebyshev series to analysis and parameter estimation of non-linear systems. Later, Horng and Ho (1987) developed a new and more convenient set of discrete orthogonal polynomials called discrete pulse orthogonal functions (DPOF's) and have used them in the analysis, parameter estimation and optimal control of linear time-varying digital systems.

LaMaire et al., (1987) have developed a robust frequency domain estimator for use in adaptive control systems which can identify both a nominal model of a plant as well as the frequency domain bounding function on the error associated with the model. Again their work is limited to transfer function type of models without time delay.

Among the estimation techniques based on the transform approach available in the literature, the real transforms $R^n \rightarrow R^n$ work well for systems with known or no time delays. However, for systems with time delays present it can be shown that the delay cannot be uniquely identified under such transformations. This is because the delay term becomes a coefficient of all the parameters of the input polynomial upon transformation. The time delay and the system parameters therefore cannot be uniquely identified. This problem can be overcome with complex transformations such as the Fourier transforms. Since, the transformed equations are complex, they provide two equations which uniquely identify the time delay and the system parameters. The Discrete Fourier

Transform which is a transformation to the complex domain is the basis of the estimation technique to be presented.

CHAPTER III

STATEMENT OF THE PROBLEM

A method for recursive estimation of the system parameters and time-delay based on nonlinear parameter estimation techniques is developed. The frequency domain estimation is based on transforming the original data sequence into the discrete frequency domain. The parameters and the time delay are recursively estimated using a nonlinear estimator which is to be developed for the frequency domain. Performance measures or "goodness of fit" will be developed upon which the estimation will be based. The advantage of using Fourier polynomials (Transforms) over the other orthogonal polynomials is that there are efficient and fast means of computing the coefficients using Fast Fourier Transforms (FFT) as compared to the other orthogonal polynomials. Furthermore, for systems to be identified, depending on the type of application, the Discrete Fourier Transform (DFT) may already be available as the output of some other data/signal processing. Coupled with the fact that the DFT can be computed efficiently and that most signals are band-limited and can be adequately described by a few frequencies, the present method results in an effective and practical parameter and delay estimation procedure. DFT is a linear transformation which parameterizes the time delay. However with unknown delays present, the equations are no longer linear in the transformed parameters. A nonlinear parameter estimation technique (multi-input multi-output) is developed in the frequency domain analogous to the recursive nonlinear parameter estimator for single-input single-output systems in the time domain (Goodwin and Sin, 1984). An attractive feature of the proposed methodology is that the size of the time-delay does not increase the computation as is the case with the other methods in the reviewed literature.

Finally, an on-line version based on infrequent parameter and delay updates in order to obtain a nominal set of parameters can be developed. On-line estimation and update at every time step can be accomplished but will involve considerable computation but is not important since most studies (Shimkin and Feuer, 1988) show that controllers can be guaranteed to be stable only if the controller parameters are batch adjusted.

The method can be applied to the preliminary identification (off-line) or the online estimation of slowly varying parameters on an infrequent basis.

In the next section of the thesis, the proposed research study is identified.

Problem Statement

The following is the statement of the proposed research study. Given a linear discrete-time system with an unknown parameters and time delay

> "Develop a recursive method to estimate discrete-time system parameters and delay in the frequency domain."

The degree of the input polynomial $B(q^{-1})$, and the output polynomial $A(q^{-1})$ are assumed known. The presence of the unknown delay make the equations nonlinear in the parameters and hence the estimators developed will be nonlinear. The background theory for the estimation procedure is developed. A performance measure for estimating the parameters is formulated which will enable the estimation problem involving complex variables to be reduced to a regular multi-input multi-output estimation problem with real variables. The next step is to establish some of the properties of the estimator. The parameters and time delay are also shown to be identifiable under transformation to the frequency domain. Finally, a number of simulation studies are conducted to substantiate the developed method. Simulation studies will also be used to show that the frequency domain estimator provides unbiased estimates in the presence of noise.

The next chapter will present the basic theory for the frequency domain estimation of parameters and delay.

CHAPTER IV

BASIC THEORY AND ALGORITHM DEVELOPMENT

Discrete Fourier Polynomials

The main idea of using a transformation is to simplify the original problem by parameterizing the delay term. The problem is then solved in the transformed domain, and if necessary inverted back to the original domain. Orthogonal transformations are especially useful and a number of papers on linear transforms can be found in literature. In this development the Discrete Fourier Polynomials will be used.

The Discrete Fourier Polynomials (DFP) are given by:

$$W_N^{kn} = e^{\frac{-i2\pi kn}{N}} \qquad n = 0, 1, \ldots, N-1 \qquad (4.1)$$

$$i = \sqrt{-1}$$

where N is the number of points or degree of the function, and k is the argument representing the various frequencies. The discrete polynomials W_N^{kn} constitute a complete set of orthogonal basis functions.

Any element f(n) of the data sequence { f } can be expressed in terms of the discrete Fourier polynomial as a weighted sum:

$$f(n) = \frac{1}{N} \sum_{k=0}^{N-1} F(k)\, W_N^{kn} \qquad (4.2)$$

The set { F } constitutes the 'weighting pattern' or the Fourier spectrum of the sequence { f }. The elements of { F } are given by (Oppenheim and Schafer,1975)

$$F(k) = \sum_{n=0}^{N-1} f(n)\, W_N^{-kn} \tag{4.3}$$

Equations (4.2) and (4.3) constitute a transform pair. The number of data points, N, does not add to the computational complexity, the choice of N is based on the following assumptions:

1. The observed data is processed in contiguous blocks each containing N data points.
2. Within a particular block, L, the parameters and time-delay are assumed constant.
3. The parameters and time-delay are permitted to change over blocks.
4. The parameters and time-delay are estimated recursively within a data block.

With these assumptions, the procedure can be used on-line, for updating the parameters on an infrequent basis.

Estimation Of Parameters

Two cases will be considered

1. Systems without delays and
2. Systems with delays

Consider the linear discrete time system described by the difference equation

$$A(q^{-1})y(n) = B(q\text{-}1)u(n) + \varepsilon(n) \tag{4.4}$$

where

$$A(q^{-1}) = 1 + a_1q^{-1} + \ldots + a_nq^{-n}$$
$$B(q^{-1}) = b_0+b_1q^{-1} + \ldots + b_mq^{-m} \qquad (4.5)$$

q^{-1} is the backward shift operator and $\varepsilon(n)$ is the residue at instant n and assumed to be a sequence of white noise independent of the input.

The input and output sequences can then be represented by the discrete fourier polynomials (the factor $\frac{1}{N}$ is neglected since it does not affect the computation).

$$u(n) = \sum_{k=0}^{N-1} U(k)\, W_N^{kn} = \langle \underline{U}, \underline{W}_N(n) \rangle \qquad (4.6)$$

where <.,.> denotes the inner product and

$$\underline{U} = [U(0)\ \ U(1) \ldots U(N-1)]^T$$
$$\underline{W}_N(n) = [W_N^{0n},\ W_N^{1n} \ldots W_N^{(N-1)n}]^T \qquad (4.8)$$

$\underline{U}$ is an N dimensional vector of the Discrete Fourier Spectrum of the input and $\underline{W}_N(n)$ is a vector of the discrete fourier basis functions. Similarly one can represent

$$y(n) = \sum_{k=0}^{N-1} Y(k)\ W_N^{kn} = \langle \underline{Y}, \underline{W}_N(n) \rangle \qquad (4.9)$$

From the definition (4.1) the following useful properties can be found.

If $\underline{W}_N(n)$ forms the basis vector for a function indexed by n, then the basis vector for the function advanced by j is given by $\underline{W}_N(n+j)$. $\underline{W}_N(n+j)$ is related to $\underline{W}_N(n)$ by the transformation

$$\underline{W}_N(n+j) = T^j\, \underline{W}_N(n) \qquad (4.10)$$

where the shifting transformation T has a diagonal form and is given by

$$T^j = \begin{bmatrix} W_N^{-oj} & 0 & 0 & . & 0 \\ 0 & W_N^{-1j} & 0 & . & 0 \\ 0 & 0 & W_N^{-2j} & . & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & . & W_N^{-(N-1)j} \end{bmatrix} \quad (4.11)$$

Then

$$y(n+j) = <\underline{Y}, \underline{W}_N(n+j)> = <\underline{Y}, T^j \underline{W}_N(n)> \quad (4.12)$$

$$u(n+j) = <\underline{U}, \underline{W}_N(n+j)> = <\underline{U}, T^j \underline{W}_N(n)> \quad (4.13)$$

Similarly, the basis vector for the function retarded by j is given by $\underline{W}_N(n-j)$. $\underline{W}_N(n-j)$ is related to $\underline{W}_N(n)$ by the transformation

$$\underline{W}_N(n-j) = T^{-j} \underline{W}_N(n) \quad (4.14)$$

where

$$T^{-j} = \begin{bmatrix} W_N^{oj} & 0 & 0 & . & 0 \\ 0 & W_N^{1j} & 0 & . & 0 \\ 0 & 0 & W_N^{2j} & . & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & . & W_N^{(N-1)j} \end{bmatrix} \quad (4.15)$$

Rewriting equation (4.4) with index t,

$$\begin{aligned} y(t) = & -a_1 y(t-1) \ldots -a_n y(t-n) + b_0 u(t) + b_1 u(t-1) \ldots \\ & + b_m u(t-m) + \varepsilon(t) \end{aligned} \quad (4.16)$$

Using (4.5) and (4.9) and the shifting property (4.14) gives,

$$< \underline{Y}, \underline{W}_N > \; = \; -[<\underline{Y}, T^{-1}\underline{W}_n(n)> \ldots <\underline{Y}, T^{-n}\underline{W}_N(n)>]\underline{a}$$
$$+ [<\underline{U}, \underline{W}_N(n)>\ldots <\underline{U}, T^{-m}\underline{W}_N(n)>]\underline{b}$$
$$+ <\underline{E}, \underline{W}_N(n)> \qquad (4.17)$$

at every instant t where

$$\underline{a} = [a_1 a_2 \ldots a_n]^T; \quad \underline{b} = [b_0 b_1 \ldots b_m]^T$$

and

$$\underline{E} = [E(0)\ E(1) \ldots E(N\text{-}1)]^T$$

is the transformed vector of the residue. In (4.17), if the coefficients of the same basis function $\underline{W}_N^k(n)$, k = 0, . . . , N-1 are equated, the following over-determined system of equations are obtained

$$\underline{y}_N = \underline{H}_N \underline{\theta} + \underline{V}_N \qquad (4.18)$$

where

$$\underline{y}_N = \underline{Y}$$
$$\underline{H}_N = [-<T^{-1},\underline{Y}>, \; \ldots \; <T^{-n},\underline{Y}>,\underline{U}, \; \ldots, <T^{-m},\underline{U}>]$$
$$\theta = [\underline{a}\ \underline{b}]^T$$
$$\underline{v}_N = \underline{E}$$

$\underline{y}_N$ can now be treated as the observation vector, $\underline{H}_N$ as the observation matrix and θ as the parameter vector. Since DFT is a complex transformation, $\underline{y}_N$, $\underline{H}_N$ and $\underline{E}_N$ are all complex. The problem then becomes an estimation problem in the complex domain.

The (n+m+1) parameters denoted as the (n+m+1) dimensional vector $\underline{\theta}$ have to be estimated from the N complex equations (or 2N real equations) given by (4.18). The transformed measurements become complex and is denoted by the N dimensional vector $\underline{Y}_N$. The (N X n+m+1) matrix $\underline{H}_N$ which forms the observation matrix is also complex. Its elements are complex but are assumed to be known. The number of spectral lines or the elements $\underline{Y}_N$ must be as large as twice the number of parameters (i.e., $\frac{N}{2}$ > n+m+1). In other words, $\underline{H}_N$ has minimal rank (n+m+1). The vector $\underline{V}_N$ represents the unknown transformed errors.

It is not possible to determine $\underline{\theta}$ uniquely from (4.18) because of the errors $\underline{V}_N$. However, if there are more frequencies (or measurements) than the unknown parameters, an attempt can be made to choose an estimator of $\underline{\theta}$ that minimizes in some arbitrarily chosen way the effect of the errors. For least-squares estimation, the estimator is chosen to minimize the sum of the square of the error. More precisely, $\underline{\hat{\theta}}_{LS}$ is defined as the least-square estimator of $\underline{\theta}$ given the data $\underline{Y}_N$ if it minimizes

$$J_{LS} = \sum_{k=0}^{N-1} v_k^2 = \underline{v}_N{}^T v_N{}^* \qquad (4.19)$$

where

$\underline{V}_N^*$ is the complex conjugate of $\underline{V}_N$.

The least-squares estimator $\underline{\hat{\theta}}_{LS}$ can be determined by locating the stationary point of (4.19). To do this, the first-order partial derivatives $\frac{\partial J_{LS}}{\partial \underline{\theta}}$ (i.e., the gradient) is formed and equated to zero. The resulting system of linear equations are solved for $\underline{\hat{\theta}}_{LS}$. From (4.19) it follows that:

$$J_{LS} = (\underline{Y}_N - \underline{H}_N\underline{\theta})^T (\underline{Y}_N^*\theta) \qquad (4.20)$$

$$\therefore \quad \frac{\partial J_{LS}}{\partial \underline{\theta}} = -(\underline{Y}_N - \underline{H}_N\theta)^T \underline{H}_N^* = 0 \qquad (4.21)$$

and the least-squares estimator satisfies the equation:

$$-(\underline{Y}_N - \underline{H}_N\hat{\underline{\theta}}_{LS})^T\underline{H}_N^* = 0 \tag{4.22}$$

This can be written as:

$$\underline{H}_N^{*T}\cdot(\underline{Y}_N - \underline{H}_N\hat{\underline{\theta}}_{LS}) = 0 \tag{4.23}$$

Note that the second partial derivative is:

$$\frac{\partial^2 J_{LS}}{\partial\underline{\theta}^2} = \underline{H}_N{}^T\underline{H}_N^* \tag{4.24}$$

This matrix is positive-definite as long as $\underline{H}_N$ has minimal rank. The solution of (4.21) is unique and minimizes J_{LS}. Note that equation (4.21) indicates that $\hat{\underline{\theta}}_{LS}$ must be chosen such that the residue $\underline{r}$:

$$\underline{r}_N = \underline{Y}_N - \underline{H}_N\hat{\underline{\theta}}_{LS} \tag{4.25}$$

is orthogonal to the columns of the observation matrix $\underline{H}_N^*$. Equation (4.22) can be written as:

$$[\underline{H}_N^{*T}\underline{H}_N]^{-1}\hat{\underline{\theta}}_{LS} = \underline{H}_N^{*T}\underline{Y}_N \tag{4.26}$$

This system is referred to as the normal equation. Since $\underline{H}_N$ has been assumed to have minimal rank, the inverse of $[\underline{H}_N^{*T}\underline{H}_N]$ exists and the least-squares estimator is found to be:

$$\hat{\underline{\theta}}_{LS} = [\underline{H}^*\underline{H}^T]^{-1}\underline{H}^{*T}\underline{Y}_N \tag{4.27}$$

<u>Properties of the Least-Squares Estimator</u>

1. The error in the estimator $\hat{\underline{\theta}}_{LS}$ is a linear function of the measurement errors $\underline{V}_N$. This follows since:

$$\underline{\tilde{\theta}}_{LS} = \underline{\theta} - \underline{\hat{\theta}}_{LS} = \underline{\theta} - [\underline{H}_N^{*T}\underline{H}_N]^{-1}\underline{H}_N^{*T}[\underline{H}_N\underline{\theta} + \underline{V}_N]$$

$$= -(H_N^{*T}H_N)^{-1}\underline{H}_N^{*T}\underline{V}_N \qquad (4.28)$$

2. Using (4.28) it follows immediately that the residue $\underline{r}_N$ can be written as:

$$\underline{r}_N = \underline{H}_N\underline{\tilde{\theta}}_{LS} + \underline{r}_N$$

$$= [I - \underline{H}_N(\underline{H}_N^{*T}\underline{H}_N)^{-1}\underline{H}_N^{*T}]\underline{V}_N \qquad (4.29)$$

But the matrix

$$[I - \underline{H}_N(\underline{H}_N^{*T}\underline{H}_N)^{-1}\underline{H}_N^{*T}]$$

is symmetric and idempotent. Thus it is an orthogonal projection matrix.

3. Since the left hand side of equation (4.37) is a real vector, it suggests that:

$$[H^{*T}H]^{-1}H^{*T}\underline{Y}_N$$

is a real matrix.

The information matrix $\underline{H}_N$, and the observation vector, $\underline{Y}_N$ being all complex introduces numerical and computational difficulties. For example, in equation (4.27) $\underline{H}_N^{*T}$ and $\underline{H}_N^{*}$ are complex. A complex matrix multiplication must be performed to compute the product of these two matrices. Next, the inverse of the product matrix has to be determined to obtain the pseudo-inverse of the information matrix. The pseudo-inverse matrix is then post multiplied by $\underline{H}_N^{*}$. Finally, the result obtained is post-multiplied by the observation vector to obtain the parameter estimates $\underline{\hat{\theta}}_{LS}$. Even though $\underline{H}_N$,

$\underline{H}_N^*$, and $\underline{Y}_N$ are complex, the right hand side of the equation (4.27), has to be a real quantity since the parameter vector $\underline{\theta}$ is real.

The difficulty of dealing with complex numbers can be avoided if each of the N complex equations is treated as two real equations. The following performance measure is proposed for estimating the parameters.

Performance Measure

The best estimates of θ can be obtained by minimizing the penalty function

$$J = \underline{v}_N{}^T \underline{v}_N$$

$$= \sum_{k=0}^{N-1} (E_R{}^2(k) + E_I{}^2(k)) \qquad (4.30)$$

where $\underline{v}_N{}^*$ is the complex conjugate of $\underline{v}_N$ and

$$E(k) = E_R(k) + i\, E_I(k)$$

By defining

$$D(k) = [E_R(k)\;\; E_I(k)]^T$$

then (4.19) can be written as

$$J = \sum_{k=0}^{N-1} D(k)^T D(k) \qquad (4.31)$$

It will now be shown how $E_R(k)$ and $E_I(k)$ can be determined. Considering the k^{th} equation of (4.18)

$$Y(k) = -a_1 W_N^k Y(k) - a_2 W_N^{2k} Y(k) \ldots - a_n W_N^{nk} Y(k) +$$

$$b_0 U(k) + b_1 W_N^k U(k) \ldots + b_m W_N^{mk} U(k) + E(k) \qquad (4.32)$$

Equating the real and imaginary parts of (4.32) gives:

$$E_R(k) = \sum_{j=0}^{n} [Y_R(k)C(j) + Y_I(k)S(j)]a_j$$

$$- \sum_{j=0}^{m} [U_R(k)C(j) + U_I(k)S(j)]b_j \qquad (4.33)$$

$$E_I(k) = \sum_{j=0}^{n} [Y_I(k)C(j) + Y_R(k)S(j)]a_j$$

$$- \sum_{j=0}^{n} [U_I(k)C(j) + U_R(k)S(j)]b_j \qquad (4.34)$$

with the following definitions

$$Y(k) = Y_R(k) + i\, Y_I(k)$$

$$U(k) = U_R(k) + i\, U_I(k)$$

$$W_N^{jk} = e^{-i2\pi jk/N} = C(j) + i\, S(j)$$

Putting (4.33) and (4.34) in the form

$$\underline{Z}(k) = \phi^T(k)\theta + \underline{D}(k)$$

where

$$\underline{Z}(k) = [Y_R(k) \;\; Y_I(k)]^T$$

we obtain

$$\phi = \begin{bmatrix} -(Y_R(k)\,C(1) + Y_I(k)\,S(1)) \ldots -(Y_R(k)\,C(n) + Y_I(k)\,S(n)) & U_R(k) \;\; (U_R(k)\,C(1) + U_I(k)\,S(1)) \ldots (U_R(k)\,C(m) + U_I(k)\,S(m)) \\ -(Y_I(k)\,C(1) - Y_R(k)\,S(1)) \ldots -(Y_I(k)\,C(n) - Y_R(k)\,S(n)) & U_I(k) \;\; (U_I(k)\,C(1) - U_R(k)\,S(1)) \ldots (U_I(k)\,C(m) - U_R(k)\,S(m)) \end{bmatrix}^T \quad (4.35)$$

and $\underline{D}(k)$ as previously defined. Clearly, we require (m+n+1)/2 distinct equations (or frequencies) to solve for the unknown parameters. This means that:

$$\text{rank}\, \underline{\phi}_N = \text{rank}\, [\phi(1) \ldots \phi(N)]^T = m+n+1$$

If N > (m+n+1)/2, then we have an overdetermined system of equations. It is clear that ϕ_N is of rank m+n+1 if we have (m+n+1)/2 distinct spectral lines for the input data sequence.

The best estimate of ϕ can be obtained by minimizing the cost function (4.30). This is accomplished by setting:

$$\delta J/\delta\theta = 0$$

whereupon the Least Squares Estimate of the parameter gives

$$\underline{\theta}^* = [\underline{\phi}_N^T \underline{\phi}_N]^{-1} \underline{\phi}^T Z_N = \underline{\phi}_N^{\#} Z_N \quad (4.36)$$

where $\underline{\phi}_N^{\#}$ is the pseudo-inverse of $\underline{\phi}_N$.

It is not difficult to see that a considerable data compression can be achieved by using the Fourier transform. Since any real signal consists of a finite number of frequency components, it can be accurately represented by the few spectral lines that describe the input and output signals. Hence, even if one had a large number of data points in the time domain, depending on the frequency content of the signal, the number of data points in the frequency domain will generally be fewer and equal to the number of spectral lines of the input sequence.

A great amount of effort is required in computing the pseudo-inverse for large spans of N. However, the recursive form of the least-square can be used. The recursive form of the least-squares estimates require the same number of computations as the batch form for the same number of data. The batch form however requires a large N by N matrix to be inverted, while the recursive form only requires a 2 by 2 matrix inversion. It is important to realize the the recursion now is not in time t, but in frequency k. The major advantage of using the recursive form is in estimating the system parameters and time delay. In this case, the system equations in the frequency domain are nonlinear and the nonlinear sequential least squares which is to be developed in the next chapter can be easily employed to estimate the parameters and time dealy.

Recursive Least-Squares Estimation

There is a technique for calculating the parameter estimates when one more frequency term is added to the frequency spectrum. Then (4.36) becomes:

$$\left[\frac{\underline{\Phi}_N}{\phi_{N+1}^T}\right] \theta_{N+1} = \left[\frac{\underline{Y}_N}{Y_{N+1}}\right] \tag{4.37}$$

Now

$$\left[\frac{\underline{\Phi}_N}{\phi_{N+1}^T}\right]^T \left[\frac{\underline{\Phi}_N}{\phi_{N+1}^T}\right] = [\underline{\Phi}_N^T \, \phi_{N+1}^T] \left[\frac{\underline{\Phi}_N}{\phi_{N+1}^T}\right]$$

$$= \underline{\Phi}_N^T \underline{\Phi}_N + \phi_{N+1}^T \, \phi_{N+1}^T \tag{4.38}$$

Defining

$$P^{-1}_{(N)} = \underline{\Phi}^T_N \underline{\Phi} \tag{4.39}$$

It follows that

$$P^{-1}_{(N+1)} = \underline{\Phi}^T_{N+1} \underline{\Phi}_{N+1}$$

$$= P^{-1}_{(N)} + \phi_{N+1} \phi^T_{N+1} \tag{4.40}$$

Using the matrix inversion Lemma (Sage, 1968), (4.40) can be written as:

$$P(N+1) = P(N) - P(N)\phi_{N+1} [I \, \phi^T_{N+1} P(N) \, \phi_{N+1}]^{-1} \phi^T_{N+1} P(N) \tag{4.41}$$

The new parameter estimate using (4.37) is

$$\underline{\hat{\theta}}_{(N+1)} = \underline{\Phi}^{\#}_{N+1} \underline{Y}_{N+1} = \left[\frac{\underline{\Phi}^T_{N+1}}{\underline{\Phi}_{N+1}}\right]^{-1} \underline{\Phi}^T_{N+1} \underline{Y}_{N+1} \tag{4.42}$$

$$= P(N+1) \, \underline{\Phi}^T_{N+1} \underline{Y}_{N+1} \tag{4.43}$$

$$= P(N+1) \, [\underline{\Phi}^T_N \, \phi^T_{N+1}] \left[\frac{\underline{Y}_N}{Y_{N+1}}\right] \tag{4.44}$$

$$= P(N+1) \, [\underline{\Phi}^T_N \underline{Y}_N + \phi_{N+1} Y_{N+1}] \tag{4.45}$$

Since

$$\underline{\Phi}^T_N \underline{Y}_N = P(N)^{-1} \underline{\hat{\theta}}(N)$$

$$\underline{\hat{\theta}}(N+1) = P(N+1) \, [P(N)^{-1} \underline{\hat{\theta}}(N) + \phi_{N+1} Y_{N+1}] \tag{4.46}$$

using (4.40), (4.46) becomes

$$\underline{\hat{\theta}}(N+1) = P(N+1) \, [\{P(N+1)^{-1} - \phi_{N+1} \phi^T_{N+1}\} \, \underline{\hat{\theta}}(N) + \phi_{N+1} Y_{N+1}] \tag{4.47}$$

$$= [I - P(N+1) \, \phi_{N+1} \phi^T_{N+1}] \, \underline{\hat{\theta}}(N) + P(N+1) \, \phi_{N+1} Y_{N+1} \tag{4.48}$$

$$= \hat{\underline{\theta}}(N) + P(N+1)\,\phi_{N+1}\,[Y_{N+1} - \phi_{N+1}^T\,\hat{\underline{\theta}}(N)] \tag{4.49}$$

The required recursion. In general for any frequency k, the equation can be written as

$$\underline{\theta}(k) = \underline{\theta}(k-1) - P(k)\phi(k)[Z(k) - \phi^T(k)\underline{\theta}(k-1)] \tag{4.50}$$

with the covariance P(k-1) given by

$$P(k) = P(k-1)-P(k-1)\phi(k)[\,\phi^T(k)P(k-1)\phi(k)+I]^{-1}\phi(k)P(k-1) \tag{4.51}$$

$P(-1) = P_0 = \alpha I$; α is any large positive number.

The above development can be applied with equal ease to parameter estimation problems with unknown parameters but known delays.

Generalized Least-Squares Problem

The least-squares cost function can be generalized by introducing a symmetric, positive-definite weighting matrix $\underline{W}$

$$J = \underline{V}_N^T\,\underline{W}\,\underline{V}_N^* \tag{4.52}$$

$$= \sum_{k=0}^{N-1} W(k)\,(E_R^{\,2}(k) + E_I^{\,2}(k)) \tag{4.53}$$

and $E_R(k)$ and $E_I(k)$ as previously defined.

W(k), the elements of $\underline{W}$ can be chosen to emphasize (or de-emphasize) the influence of the various frequency content of the data points. The introduction of the weighting matrix $\underline{W}$ makes no substantial differences in the development.

Equation (4.42) can be rewritten as:

$$J = \sum_{k=0}^{N-1} D^T(k)\,\underline{R}^{-1}\,D(k) \quad (4.54)$$

where

$$D(k) = [E_R(k)\;\; E_I(k)]^T$$

$$\underline{R} = \begin{bmatrix} W(k) & 0 \\ 0 & W(k) \end{bmatrix} \qquad (4.55)$$

Following a development similar to the least-squares problem it can be shown that the recursive equations for parameter and covariance update for:

$$\theta(k) = \theta(k\text{-}1) - P(k)\,\phi(k)\,[Z(k) - \phi^T(k)\,\theta(k\text{-}1)] \qquad (4.56)$$

$$P(k) = P(k\text{-}1) - P(k\text{-}1)\,\phi(k)\,[\phi^T(k)\,P(k\text{-}1)\,\phi(k) + R]^{-1}\phi(k)\,P(k\text{-}1) \qquad (4.57)$$

In the next chapter, the theory developed for the frequency domain estimation of parameters is extended to estimate unknown parameters and also the time delay. This however leads to a nonlinear estimation problem in the frequency domain.

CHAPTER V

IDENTIFICATION OF PARAMETERS AND DELAY

When the system parameters and time delay are unknown, the frequency domain estimator can be employed to estimate them. However, when the system time delay is unknown, the transformed frequency domain equations describing the system are non-linear in the time-delay parameters. A nonlinear parameter estimator in the frequency domain has to be developed. A brief review of the theory of nonlinear least-square theory is presented.

Nonlinear Least-Squares

A linear model is often too restrictive and it is therefore useful to examine a nonlinear model. Suppose the n parameters $\underline{\theta}$ are related to the measurement data $\underline{y}$ according the the equation

$$\underline{y} = \underline{h}(\underline{\theta}) + \underline{v} \tag{5.1}$$

where $\underline{h}(\underline{\theta})$ represents some nonlinear continuously differential function of $\underline{\theta}$. The least-squares estimator of $\underline{\theta}$ given $\underline{y}$ is defined as the value of $\hat{\theta}$ that is minimized

$$L_{LS} = [\underline{y} - \underline{h}(\hat{\theta})]^T[\underline{y} - \underline{h}(\hat{\theta})] \tag{5.2}$$

where, for convenience, the weighting matrix has been chosen as an identity matrix.

Assuming some reference value, $\underline{\theta}^*$, the objective function I_{LS} is expanded in a Taylor series

$$I_{LS}(\underline{\theta}) = I_{LS}(\underline{\theta}^*) + \left[\frac{\partial I_{LS}}{\partial \underline{\theta}}(\underline{\theta}^*)\right] (\underline{\theta} - \underline{\theta}^*)$$

$$+ \frac{1}{2}(\underline{\theta} - \underline{\theta}^*)^T \left[\frac{\partial^2 I_{LS}}{\partial \underline{\theta}^2}(\underline{\theta}^*)\right] (\underline{\theta} - \underline{\theta}^*)$$

$$+ o(||\underline{\theta} - \underline{\theta}^*||^2) \tag{5.3}$$

A necessary condition for $\underline{\theta}^*$ to minimize I_{LS} is

$$\frac{\partial I_{LS}}{\partial \underline{\theta}}(\underline{\theta}^*) = 0 \tag{5.4}$$

An additional necessary condition for minima is that the matrix $\left(\frac{\partial^2 I_{LS}}{\partial x^2}\right)$ evaluated at $\underline{\theta}^*$ must be positive-semidefinite. The sufficient conditions for $\underline{\theta}^*$ to minimize I_{LS} are $\frac{\partial^2 I_{LS}}{\partial x^2} = 0$ and $\frac{\partial^2 I_{LS}}{\partial \underline{\theta}^2}$ is positive-definite. To determine the least-squares estimate $\hat{\theta}$, set

$$\frac{\partial I_{LS}}{\partial \underline{\theta}} = -2[\underline{y} - \underline{h(\theta)}]^T \frac{\partial \underline{h}}{\partial \underline{\theta}} \hat{\theta} = 0 \tag{5.5}$$

which can be rearranged as

$$\left[\frac{\partial \underline{h}}{\partial \underline{\theta}} \quad \hat{\theta}\right]^T [\underline{y} - \underline{h(\theta)}] = 0 \tag{5.6}$$

The residual $[\underline{y} - \underline{h(\theta)}]$ must be orthogonal to the columns of $\frac{\partial \underline{h}}{\partial \underline{\theta}}$. In a linear problem, $\underline{h(\theta)} = \underline{H(\theta)}$ and $\frac{\partial}{\partial \underline{\theta}}(\underline{h(\theta)}) = \underline{H}$ and equation (5.2) becomes

$$\underline{H}^T[\underline{y} - \underline{H}\hat{\theta}] = 0 \tag{5.7}$$

Equation (5.7) represents a linear equation in $\underline{\theta}$ and has a solution

$$\hat{\theta}_{LS} = [\underline{H}^T\underline{H}]^{-1}\underline{H}^T\underline{y} \tag{5.8}$$

However, when $\underline{h}$ is nonlinear, (5.6) represents a system of nonlinear algebraic equations for which, in general, there is no closed form solution. One is, consequently, compelled to solve (5.6) numerically to find least-squares estimates when nonlinear systems are involved.

To obtain the estimates, suppose that the initial estimate $\underline{\theta}^*$ of $\underline{\theta}$ is available. For example, the expected value $E[\underline{\theta}]$ would provide such an estimate. Using $\underline{\theta}^*$, the system equations (5.1) is expanded out in a Taylor series

$$\underline{y} = \underline{h}(\underline{\theta}^*) + \left[\frac{\partial \underline{h}}{\partial \underline{\theta}}(\underline{\theta}^*)\right](\underline{\theta} - \underline{\theta}^*) + o\,(||\underline{\theta} - \underline{\theta}^*||) + \underline{v} \tag{5.9}$$

For the norm $||\underline{\theta} - \underline{\theta}^*||$ sufficiently small, the higher order terms can be neglected and (5.9) can be written as

$$\underline{y} = \underline{h}(\underline{\theta}^*) + \left[\frac{\partial \underline{h}}{\partial \underline{\theta}}(\underline{\theta}^*)\right](\underline{\theta} - \underline{\theta}^*) + \underline{v} \tag{5.10}$$

By defining

$$\underline{H} \equiv \frac{\partial \underline{h}}{\partial \underline{\theta}}(\underline{\theta}^*) \tag{5.11}$$

Since (5.10) represents linear system, the least-squares estimate of $\underline{\theta}$ is given by

$$\underline{\hat{\theta}} = \underline{\theta}^* + [\underline{H}^T\underline{H}]^{-1}\underline{H}^T[\underline{y} - \underline{h}(\underline{\theta}^*)] \tag{5.12}$$

Thus, an approximation of the least-squares estimate is provided by (5.12). In addition, equation (5.12) suggests an interative procedure for refining the estimate. Suppose that the linearization (5.9) is performed relative to $\underline{\theta}^*$ and the new estimate $\underline{\hat{\theta}}$ is computed using (5.12). A necessary condition for $\underline{\hat{\theta}}$ to be the least-squares estimate is that the orthogonality condition (5.6) must be satisfied. In general, $\underline{\hat{\theta}}$ may not satisfy (5.6) but, hopefully, $\underline{\hat{\theta}}$ will provide an improvement over $\underline{\theta}^*$. If it does provide an improvement, the $\underline{\hat{\theta}}$ can be used to replace $\underline{\theta}^*$ and the linearization of (5.9) repeated. Then, a new estimate $\underline{\hat{\theta}}$ is computed, the orthogonality condition tested, and if necessary, the procedure is repeated. Since equation (5.12) is derived from (5.6), the orthogonality condition is also contained in equation (5.12). The iterations are terminated when $\underline{H}^T[\underline{y} - \underline{h}(\underline{q}^*)]$ vanishes (i.e., the gradient vanishes).

The iteration procedure can be modified by introducing a step size parameter λ so that

$$\underline{\hat{\theta}} = \underline{\theta}^* + \lambda[\underline{H}^T\underline{H}]^{-1}\underline{H}^T[\underline{y} - \underline{h}(\underline{\theta}^*)] \tag{5.13}$$

The above procedure is known as Gauss' method and the step size parameter λ serves as a control on the correction made to $\underline{\theta}^*$. It can greatly influence the convergence or divergence of the iterations.

Instead of linearizing the measurement equation (5.1), the objective function I_{LS} can be approximated as a quadratic cost function and then minimized. Ignoring the higher order terms than the second in (5.3), $\underline{\hat{\theta}}$ that minimizes the resulting quadratic cost function

$$l_{LS}(\underline{\theta}) = l_{LS}(\underline{\theta}^*) + \left[\frac{\partial l_{LS}}{\partial \underline{\theta}}(\underline{\theta}^*)\right](\underline{\theta} - \underline{\theta}^*)$$

$$+ \frac{1}{2}(\underline{\theta} - \underline{\theta}^*)^T \left[\frac{\partial l^2_{LS}}{\partial \underline{\theta}^2}(\underline{\theta}^*)\right](\underline{\theta} - \underline{\theta}^*) \tag{5.14}$$

satisfies

$$\frac{\partial l_{LS}}{\partial \underline{\theta}}(\underline{\theta}^*) + (\hat{\underline{\theta}} - \underline{\theta}^*)^T \left[\frac{\partial l^2_{LS}}{\partial \underline{\theta}^2}(\underline{\theta}^*)\right] = 0 \tag{5.15}$$

when solved for $\hat{\underline{\theta}}$ gives

$$\hat{\underline{\theta}} = \underline{\theta}^* - \left[\frac{\partial l^2_{LS}}{\partial \underline{\theta}^2}(\underline{\theta}^*)\right]^{-1} \left[\frac{\partial l_{LS}}{\partial \underline{\theta}}(\underline{\theta}^*)\right]^T \tag{5.16}$$

Since

$$\frac{\partial l_{LS}}{\partial \underline{\theta}}(\underline{\theta}^*) = -2[\underline{y} - \underline{h}(\underline{\theta}^*)]^T \frac{\partial h}{\partial \underline{\theta}}(\underline{\theta}^*) \tag{5.17}$$

the second derivation of the objective function is

$$\frac{\partial l^2_{LS}}{\partial \underline{\theta}^2}(\underline{\theta}^*) = -2\left\{\sum_{i=1}^{n} [y_i - h_i(\theta^*)] \frac{\partial^2 h_i}{\partial \underline{\theta}^2}(\underline{\theta}^*)\right\}$$

$$+ 2\left[\frac{\partial h}{\partial \underline{\theta}}(\underline{\theta}^*)\right]^T \left[\frac{\partial h}{\partial \underline{\theta}}(\underline{\theta}^*)\right] \tag{5.18}$$

Since (5.16) is only an approximation, it is reasonable to introduce a step size parameter to insure that the estimate $\hat{\underline{\theta}}$ actually reduces the value of the cost function relative to $\underline{\theta}^*$. Then letting $\underline{H} = \frac{\partial \underline{h}}{\partial \underline{\theta}}$ evaluated at $\underline{\theta}^*$, equation (5.16) becomes

$$\hat{\underline{\theta}} = \underline{\theta}^* + \lambda\,[\underline{H}^T\underline{H} - \sum_{i=1}^{n} [y_i - h_i(\underline{\theta}^*)] \frac{\partial^2 h_i}{\partial \underline{\theta}^2}(\underline{\theta}^*)]^{-1} H^T [\underline{y} - \underline{h}(\underline{\theta}^*)] \quad (5.19)$$

The above least-squares estimate provides the basis for an iterative search procedure. It is called the Newton or Newton-Raphson procedure (Sorenson, 1980; Luenberger, 1973). If the terms $[y_i - h_i(\underline{\theta}^*)] \frac{\partial^2 h_i}{\partial \underline{\theta}^2}(\underline{\theta}^*)$ are neglected, the two methods are identical. If the reference value $\underline{\theta}^*$ is a "good" approximation of $\hat{\underline{\theta}}$, the residual should be small so that neglecting this term may be justified. Further, to ensure that the sufficient condition for minimization of $I_{LS,}$ $\frac{\partial^2 I_{LS}}{\partial \underline{\theta}^2}$ has to be positive definite. Since equation (5.18) may not always be positive-definite (Speedy, Brown and Goodwin, 1970). Neglecting $[y_i - h_i(\underline{\theta}^*)]^T \frac{\partial^2 h_i}{\partial \underline{\theta}^2}(\underline{\theta}^*)$ ensures that $\frac{\partial^2 I_{LS}}{\partial \underline{\theta}^2}$ is positive definite.

Another well-known procedure is obtained when $\frac{\partial^2 I_{LS}}{\partial \underline{\theta}^2}$ is replaced by the identity matrix in (5.16). In this case, the estimate $\hat{\underline{\theta}}$ is obtained from the reference value $\underline{\theta}^*$ by a step taken in the negative gradient direction. Since the gradient defines the direction in which the cost increases most rapidly in the neighborhood of $\underline{\theta}^*$, the search direction defined by

$$\underline{\theta} = \underline{\theta}^* + \lambda \left[\frac{\partial I_{LS}}{\partial \underline{\theta}}\right]$$

$$\lambda < 0 \quad (5.20)$$

is called the steepest descent method. The step size is chosen to ensure that the cost is reduced.

A hierarchy of search procedures have been defined in which the steepest descent or gradient search method is the simplest and the Newton method the most complicated. In fact, in the Gauss and Newton's methods, the gradient vector is automatically scaled by $\left[\frac{\partial^2 I_{LS}}{\partial \underline{\theta}^2}(\theta^*)\right]^{-1}$ so that the search for the next estimate moves in the optimal direction. The Newton method is quite complicated since it requires the computation of $\frac{\partial^2 \underline{h}}{\partial \underline{\theta}^2}$. When the initial estimate $\underline{\theta}^*$ is poor, the steepest descent method often exhibits superior performance in the sense that the estimates approach the true value rapidly. However, the gradient method is known to converge very slowly in the vicinity of the minima. Whereas the Newton method converges very rapidly near the minimum.

In the next section, the methodology for sequentially implementing the nonlinear least-squares in the frequency domain is presented.

Estimation Of System Parameters And Delay

The frequency domain parameter estimation technique developed in the previous chapter will be extended in this section for estimating the unknown parameters and time delay of a discrete time system. Consider the class of linear discrete time systems described by

$$A(q^{-1})\, y(t) \;=\; q^{-d} B'(q^{-1})\, u(t) \;+\; \varepsilon(t) \tag{5.22}$$

where d is the pure unknown time delay, and

$$A(q^{-1}) = 1 + a_1q^{-1} + \ldots + a_nq^{-n}$$

$$B'(q^{-1}) = b_0 + b_1q^{-1} + \ldots + b_mq^{-m}$$

which makes the number of unknown "parameters" m+n+2. Following a development similar to (4.17) of the last chapter the following equation is obtained

$$\begin{aligned} <\underline{Y}, \underline{W}_N(n)> = & -[<\underline{Y}, T^{-1}\underline{W}_N(n)> \ldots <\underline{Y}, T^{-n}\underline{W}_N(n)>]\,\underline{a} \\ & [<\underline{U}, T^{-d}\underline{W}_N(n)> \ldots <\underline{U}, T^{-(m+d)}\underline{W}_N(n)>]\,\underline{b} \\ & + <\underline{E}, \underline{W}_N(n)> \end{aligned} \tag{5.23}$$

or

$$\begin{aligned} <\underline{Y}, \underline{W}_N(n)> \quad & -[<\underline{Y}, T^{-1}\underline{W}_N(n)> \ldots <\underline{Y}, T^{-n}\underline{W}_N(n)>]\,\underline{a} \\ & + T^{-d}[<\underline{U}, \underline{W}_N(n)> \ldots <\underline{U}, T^{-m}\underline{W}_N(n)>]\,\underline{b} \\ & + <\underline{E}, \underline{W}_N(n)> \end{aligned} \tag{5.24}$$

However the shifting transformation matrix $T^{-(d+j)}$, (j = 0,m) contains the unknown delay parameter d, which makes the equation (5.23) and hence the problem nonlinear. If, instead of the DFT, any other real orthogonal transformation is used then, Y, T^{-j}, U, and E are all real. Since T^{-d} is a product matrix of the $\underline{b}$ vector, unique identification of d and $\underline{b}$ is not possible due to loss of identifiability. Only estimates of the product of d and b_j = (0,. . m) can be obtained. With DFT this problem does not arise. Since, for each frequency k, a complex equation is obtained, a unique solution for d and $\underline{b}$

d and $\underline{b}$ exists and the estimation of d and $\underline{b}$ is possible as shown below. The k^{th} equation of (5.22) can be written as

$$
\begin{aligned}
Y(k) = & - a_1 W_N^{k} Y(k) - a_2 W_N^{2k} Y(k) \ldots - a_n W_N^{nk} Y(k) \\
& + b_0 W_N^{dk} U(k) + b_1 W_N^{(d+1)k} U(k) \ldots + b_m W_N^{(d+m)k} U(k) \\
& + E(k) \qquad (5.25)
\end{aligned}
$$

Equating the real and imaginary parts of (5.25) gives

$$
\begin{aligned}
E_R(k) = & \sum_{j=0}^{n} [Y_R(k)C(j) + Y_I(k)S(j)]a_j \\
& - \sum_{j=0}^{m} [U_R(k)C(d+j) + U_I(k)S(d+j)]b_j \qquad (5.26)
\end{aligned}
$$

$$
\begin{aligned}
E_I(k) = & \sum_{j=0}^{n} [Y_I(k)C(j) - Y_R(k)S(j)]a_j \\
& - \sum_{j=0}^{m} [U_I(k)C(d+j) - U_R(k)S(d+j)]b_j \qquad (5.27)
\end{aligned}
$$

with the equations (5.26) and (5.27) now being nonlinear in the parameters. Again defining

$$D(k) = [E_R(k) \;\; E_I(k)]^T$$

$$\underline{Z}(k) = [Y_R(k) \;\; Y_I(k)]^T$$

and

$$\underline{Z}(k,\theta) = \begin{bmatrix} -\sum_{j=0}^{n} [Y_R(k)C(j) + Y_I(k)S(j)]a_j \\ -\sum_{j=0}^{n} [Y_I(k)C(j) - Y_R(k)S(j)]a_j \\ -\sum_{j=0}^{m} [U_R(k)C(d+j) + U_I(k)S(d+j)]b_j \\ -\sum_{j=0}^{m} [U_I(k)C(d+j) + U_R(k)S(d+j)]b_j \end{bmatrix} \tag{5.28}$$

Performance Measure

To obtain an estimate of the parameters, the following performance measure is proposed. The best estimate of $\theta = (\underline{a}, \underline{b}, d)$ is obtained by minimizing the the penalty function

$$\begin{aligned} J &= \underline{E}^T \underline{E}^* \\ &= \sum_{k=0}^{N-1} D(k)^T D(k) \\ &= \sum_{k=0}^{N-1} [Z(k) - Z(k,\theta)]^T [Z(k) - Z(k,\theta)] \end{aligned} \tag{5.29}$$

Any nonlinear minimization technique can be adopted to minimize the performance measure J. However for computational ease over large spans of frequencies, a procedure similar to the Sequential Prediction Error method (Goodwin and Sin, 1984) is developed. The important difference being that the data is being processed sequentially at various frequencies k (instead of time t). The error D(k) being a vector, the algorithm has to be developed for a vector case.

Nonlinear Frequency Domain Parameter Estimation Algorithm

In the development of the nonlinear algorithm both the system equations and the Performance measure are linearized using the Taylor's series. The nonlinear frequency domain estimation scheme is developed here. Following a development similar to the Nonlinear Sequential Prediction Error method in the time domain the Nonlinear Frequency Domain Estimator (NFDE) will be developed for a multi-input multi-output case. Consider the scalar criterion of goodness of fit

$$V_N(\theta) = 1/N \sum_{k=0}^{N} l[Z(k,\theta),Z(k)] \qquad (5.30)$$

with

$$l[Z(k,\theta),Z(k)] = \frac{1}{2}[Z(k,\theta) - Z(k)]^T [Z(k,\theta) - Z(k)]$$

$$= \frac{1}{2} \| Z(k,\theta) - Z(k) \|_2^2 \qquad (5.31)$$

where

$$Z(k) \in R_{2x1}$$

$$Z(k,\theta) \in R_{2x1}$$

with

$$Z(k) = [Y_R(k) \;\; Y_I(k)]^T$$

and

$$Z(k,\theta) = \begin{bmatrix} -\sum_{j=0}^{n} [Y_R(k)\, C(j) + Y_I(k) S(j)] a_j \\ -\sum_{j=0}^{n} [Y_I(k)\, C(j) - Y_R(k) S(j)] a_j \\ + \sum_{j=0}^{m} [U_R(k) C(d+j) + U_I(k) S(d+j)] b_j \\ + \sum_{j=0}^{m} [U_I(k) C(d+j) - U_R(k) S(d+j)] b_j \end{bmatrix}$$

Then (5.30) can be expressed as

$$(N+1) V_{N+1}(\theta) = N V_N(\theta) + \frac{1}{2} [Z(N+1,\theta) - Z(N+1)]^T [Z(N+1,\theta) - Z(N+1)] \quad (5.32)$$

Approximating $Z(N+1,\theta)$ by using a first order Taylor's series about $\theta(N)$ gives

$$Z(N+1,\theta) = Z(N+1,\theta(N)) + Z'(N+1)[\theta - \theta(N)] \quad (5.33)$$

where

$$Z'(N+1) = \begin{bmatrix} \delta Z_R/\delta\theta_1 & \delta Z_R/\delta\theta_2 \\ \delta Z_R/\delta\theta_1 & \delta Z_R/\delta\theta_2 \end{bmatrix} \quad (5.34)$$

at $\theta = \theta(N)$

Substituting (5.33) into (5.32) gives

$$(N+1)V_N(\theta) = NV_N(\theta) + \frac{1}{2}[Z(N+1,\theta(N)) + Z'(N+1)$$

$$[\theta - \theta(N)] - Z(N+1)]^{T*}$$

$$Z(N+1,\theta(N)) + Z'(N+1)[\theta - \theta(N)] - Z(N+1)] \qquad (5.35)$$

$$+ NV_N(\theta) + \frac{1}{2}[\psi(N)^T\theta - Z(N+1)]^T[\psi^T(N)\theta - Z(N+1)]$$

where

$$\psi(N) = Z'(N+1)$$

$$= \begin{vmatrix} \delta Z_{R(N+1)}/\delta\theta_1 & \ldots & \delta Z_{R(N+1)}/\delta\theta_P \\ \delta Z_{R(N+1)}/\delta\theta_1 & \ldots & \delta Z_{R(N+1)}/\delta\theta_P \end{vmatrix}$$

$$\text{at} \quad \theta = \theta(N) \qquad (5.36)$$

with $p = m+n+2$ and

$$X(N+1) = Z(N+1) - Z(N+1, \theta(N)) + \psi(N)^T\theta(N) \qquad (5.37)$$

Differentiating (4.35) with respect to θ gives

$$(N+1)V_N'(\theta) = NV_N'(\theta) + \psi(N)[\psi(N)^T\theta - X(N+1)] \qquad (5.38)$$

where

$$V_N'(\theta) = d/d\theta \; V_N(\theta) \qquad (5.39)$$

Expanding $V_N'(\theta)$ in a first order Taylor's series about $\theta(N)$ gives

$$V_N'(\theta) = V_N'(\theta(N)) + V_N''[\theta - \theta(N)] \qquad (5.40)$$

with the Hessian matrix

$$VN'' = \begin{bmatrix} \delta^2 V_N/\delta\theta_1{}^2 & \cdots & \delta^2 V_N/\delta\theta_1\delta\theta_P \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \delta^2 V_N/\delta\theta^P\delta\theta_1 & \cdots & \delta^2 V_N/\delta\theta_P{}^2 \end{bmatrix}$$

Substituting (5.40) into (5.38) and noting that $V_N'(\theta)$ is zero in view of optimality of θ(N) gives

$$(N+1)V_N' = NV_N'' [\theta - \theta (N)] - \psi(N) [X(N+1) - \psi(N)^T\theta] \qquad (5.41)$$

Now the value, θ (N+1) of θ optimizing $V_{N+1}(\theta)$ gives $V'_{N+1}(\theta) = 0$ and hence from (5.41) satisfies

$$NV_N'' [\theta (N+1) - \theta (N)] - \psi(N) [Z(N+1) - \psi(N)^T\theta (N+1)] = 0 \qquad (5.42)$$

that is

$$[NV_N'' + \psi(N)\, \psi(N)^T]\, \theta(N+1) = NV_N''\, \theta(N) + \psi(N)\;\; X(N+1) \qquad (5.43)$$

$$= [NV_N'' + \psi(N)\, \psi(N)^T]\, \theta(N) + \psi(N)\, [X(N+1) - \psi(N)^T\theta(N)]$$

or

$$\theta(N+1) =$$
$$\theta(N) + [NV_N'' + \psi(N)\, \psi(N)^T]^{-1}\, \psi(N)\, [X(N+1) - \psi(N)^T\theta(N)] \qquad (5.44)$$

This is the basic form of the algorithm. For computational reasons, it can be shown that V_N an be computed iteratively. Differentiating (5.38) w.r.t. θ gives

$$(N+1)V_{N+1}'' = NV_N'' + \psi(N)\, \psi(N)^T \qquad (5.45)$$

defining

$$P^{-1}(N) = (N+1)V_{N+1}$$

and using the Matrix Inversion Lemma

$$[A + BC]^{-1} = A^{-1} - AB^{-1}[I + CA^{-1}B]^{-1}CA^{-1}$$

gives

$$\begin{aligned} P(N) &= [P^{-1}(N-1) + \psi(N)\, \psi(N)^T]^{-1} \\ &= P(N-1) - P(N-1)\, \psi(N) \\ &* [I + \psi(N)^T P(N-1)\, \psi(N)]^{-1}\, \psi(N)^T\, P(N-1) \end{aligned} \tag{5.46}$$

Finally the algorithm can be summarized as

$$\theta(k) = \theta(k-1) - P(k)\, \psi(k)\, [Z(k) - \hat{Z}(k,\theta)\, \theta(k-1)] \tag{5.47}$$

$$P(k) = P(k-1) - P(k-1)\, \psi(k)\, [\psi^T(k)\, P(k-1)\, \psi(k) + I]^{-1}\, \psi(k)\, P(k-1) \tag{5.48}$$

Equations (5.47) and (5.48) provide recursive estimates of the parameter estimates and the covariance of these estimates at the various frequencies. The parameter and covariance update equations for the nonlinear case are structurally similar to the equation (4.50) and (4.51) developed in the previous chapter. The only difference being that the information vector ϕ has now been replaced by ψ the derivative of $Z(k,\theta)$ w.r.t θ in the nonlinear case.

In some cases, it is not feasible to utilize the least-squares based algorithms due to the computations involved in updating and storing the covariance matrix. This is especially so when the number of parameters is large. In such cases, it is possible to replace the covariance matrix by an identity matrix to obtain a variation of gradient algorithm called the Normalized Least-Mean-Square Algorithm.

$$\hat{\theta}(k) = \hat{\theta}\,(K\text{-}1) + \frac{\mu(k)\ \phi(k)}{c+\phi^T(k)\phi(k)}[y(k+1) - \phi^T(k)\ \hat{\theta}\,(K\text{-}1)] \tag{5.49}$$

where $\mu(k)$ is a scaler gain, usually taken to be a small positive constant.

This algorithm is suggested by Albert and Gardiner (1967), Nagumo and Noda (1967) and others. It is also called as the projection algorithm (Goodwin and Sin, 1984). The Normalized Least-Mean-Square (NLMS), has been developed for a system linear in the parameters. The variant obtained in this thesis is a nonlinear version of the algorithm

$$\hat{\theta}(k) = \hat{\theta}\,(K\text{-}1) + \frac{\psi(k)}{l+\psi^T(k)\psi(k)}[Z(k) - \hat{Z}(k, \hat{\theta}\,(K\text{-}1))] \tag{5.50}$$

where

$$\psi(k) = \frac{d\hat{Z}(k,\theta)}{d\theta}\Big|\ \theta = \hat{\theta}\,(K\text{-}1) \tag{5.51}$$

It is possible to simplify the algorithm even further by removing the normalization all together. This leads to a variation of the Least-Mean-Square Algorithm (LMS) (Widrow and Hoft, 1960)

$$\hat{\theta}(k) = \hat{\theta}\,(K\text{-}1) + \mu(k)\ \phi(k)\ \ [y(k) - \phi^T(k)\ \hat{\theta}\,(K\text{-}1)] \tag{5.52}$$

The nonlinear variant is obtained by replacing $\phi(k)$ by the gradient $\psi(k)$

$$\hat{\theta}(k) = \hat{\theta}\,(K\text{-}1) + \mu(k)\ \psi(k)\ \ [Z(k) - \hat{Z}(k, \theta\,(K\text{-}1)] \tag{5.53}$$

This is similar to the stochastic approximation scheme proposed by Robbins and Tomro (1951).

Getting back to the problem on hand, the transformed equation for the discrete time system with unknown parameters and time delay (5.28) is in the same form as the equation used in the development of the nonlinear frequency domain estimator. The parameter and delay estimates can therefore be recursively computed by solving (5.47) and (5.48) sequentially. Even though the original system with unknown parameters and time delay is described by a linear difference equation in the time domain, the transformed equation becomes nonlinear in the time delay parameter. The estimation procedure being nonlinear, the convergence of the parameter estimates depends on the how close the initial estimates are to the actual values. On startup, it might therefore be necessary to iterate a few times until good estimates are obtained.

In the next chapter some important properties of the estimator will be derived and presented.

CHAPTER VI

PROPERTIES OF THE FREQUENCY DOMAIN ESTIMATOR

An important question that arises is how do we know whether or not the results obtained from the an estimator are good? A good parameter estimator should possess certain properties. The properties of the estimator must therefore be studied. An important property desired in any estimator is unbiasedness. The estimates provided by a unbiased estimator will converge to their true values on the average.

Further, the identifiability conditions of the system must be preserved under any transformation, otherwise the true parameters of the system can never be estimated. Since the transformation is to the frequency domain, the identifiability conditions can be readily established. It will be shown that the system that is identifiable does not suffer loss of identifiability of parameter and delay when transformed to the frequency domain. It should be noted that other real transformations like Laguerre polynomials cannot be used to explicitly identify the parameters and delay.

The following definitions and theorems are presented.

Theorem I

Given that $E[x(n)] = \theta$ where E is the expectation operator, the expected value of the Fourier transform of the signal i.e. ,

$E\,[X(e^{i\omega})] = 0$

Proof

By definition

$$X(e^{i\omega}) = \sum_{n=-\infty}^{\infty} x(n)\, e^{i\omega n} \tag{6.1}$$

$$E[X(e^{i\omega})] = E\left[\sum_{n=-\infty}^{\infty} x(n)\, e^{i\omega n}\right] \tag{6.2}$$

The order of expectation and summation may be interchanged since they are linear operators and it is assumed that the series converges uniformly. Interchanging the order of expectation and summation,

$$E\,[X(e^{i\omega})] = \sum_{n=-\infty}^{\infty} E\,[x(n)]\, e^{i\omega n}] \tag{6.3}$$

$$= 0$$

Definition I: (Mendel,1984)

Estimator $\hat{\theta}(n)$ is an unbiased estimator of deterministic θ if

$$E[\hat{\theta}(n)] = \theta \quad \text{for all } n$$

or of random θ if

$$E[\hat{\theta}(n)] = E[\,\theta] \quad \text{for all } n$$

In terms of the estimation error, $\theta(n)$, unbiasedness means, that

$$E[\tilde{\theta}(n)] = \theta \quad \text{for all } n$$

Theorem II

The Least-Squares Frequency Domain Estimator

$$\hat{\underline{\theta}}_{LS} = [\underline{\Phi}_N^T \underline{\Phi}_N]^{-1} \underline{\Phi}_N^T \underline{y}_N \tag{6.4}$$

is unbiased if $\underline{v}_N$ is zero mean and if $\underline{v}_N$ and $\underline{H}_N$ are statistically independent.

Proof

Since $\underline{Y}_N = \underline{\Phi}_N \underline{\theta} + v_N$ and the least-squares estimator in the frequency domain is given by (6.4), $\hat{\underline{\theta}}_{LS}$ can be written as

$$\hat{\underline{\theta}}_{LS} = [\underline{\Phi}_N^T \underline{\Phi}_N]^{-1} \underline{\Phi}_N^T [\underline{\Phi}_N \underline{\theta} + v_N] \tag{6.5}$$

$$= \underline{\theta} + [\underline{\Phi}_N^T \underline{\Phi}_N]^{-1} \underline{\Phi}^T \underline{v}_N \tag{6.6}$$

Taking the expectation on both sides of (6.2), it follows that

$$E[\hat{\underline{\theta}}_{LS}] = \underline{\theta} + E[[\underline{\Phi}_N^T \underline{\Phi}_N]^{-1} \underline{\Phi}_N^T \underline{V}_N] \tag{6.7}$$

for $\underline{\Phi}_N$ and $\underline{V}_N$ are statistically independent,

$$E[\hat{\underline{\theta}}_{LS}] = \underline{\theta} + E[[\underline{\Phi}_N^T \underline{\Phi}_N]^{-1} \underline{\Phi}_N] \, E[\underline{v}_N] \tag{6.8}$$

In deriving (6.8) the fact that $\underline{\Phi}_N$ and $\underline{v}_N$ are statistically independent has been used. Recall that if two random variables, a and b, are statistically independent, the probability density function $p(a,b) = p(a)p(b)$; thus the expected value $E[a,b] = E[a]\,E[b]$. The second term in (6.8) is zero since $E[\underline{V}_N] = 0$ and therfore

$$E[\hat{\underline{\theta}}_{LS}] = \underline{\theta} \quad \text{for all } N \tag{6.9}$$

In the case of estimation of parameters depending on the interrelationships between the parameter vector $\underline{\theta}$, the observation matrix $\underline{H}$ and the measurement noise vector $\underline{V}$, a number of situations can occur as shown in Table I. Given the system

$$A(q^{-1})y(n) = B(q^{-1})u(n) + \varepsilon(n) \tag{6.10}$$

with E [ε(n)] = 0, the frequency domain model is given by equation

$$\underline{Y}_N = \underline{\Phi}_N \underline{\theta} + \underline{D}_N$$

where $E[\underline{D}(k)] = 0$ by Theorem I.

The observation matrix $\underline{\Phi}_N$ is random and $\underline{\theta}$ is deterministic. This is a case of A.2.b of Table I. The least-squares estimate in the frequency domain is given by

$$\hat{\underline{\theta}}(k) = [\underline{\Phi}_K^T \underline{\Phi}_K]^{-1} \underline{\Phi}_K^T \underline{Y}_K$$

Now

$$E[\hat{\underline{\theta}}(k)] = E[\underline{\Phi}_K^T \underline{\Phi}_K]^{-1} \underline{\Phi}_K^T [\underline{\Phi}_K \underline{\theta} + V_{\underline{k}}]$$

$$= \underline{\theta} + E[\underline{\Phi}_K^T \underline{\Phi}_K]^{1} [\underline{\Phi}_K^T \underline{V}_K]$$

However, since $\underline{\Phi}_K$ and $\underline{V}_K$ are statistically dependent and are related by a complicated relationship, the estimator will be biased.

The next definition and theorem establish that the unknown parameters and time delay of the frequency domain model are identifiable.

The next definition and theorem establish the identifiability properties in the frequency domain. Identifiability is a concept that is central to identification problems. Loosely speaking, the problem is whether the identification procedure will yield a unique value of the parameter $\underline{\theta}$ and/or whether the resulting model is equal to the true system.

Definition II: (Ljung, 1987)

The system

$$y(z) = G(z,\theta)\, U(z) + H(z,\theta)\, V(z)$$

is globally identifiable at θ^* if and only if

$$G(z,\theta) \equiv G(z,\theta^*)$$

and

$$H(z,\theta) \equiv H(z,\theta^*) \tag{6.9}$$

for almost all z. For local identifiability, only θ confined to a sufficiently small neighborhood of θ^* is considered.

Theorem III

Given the Equation Error Model structure

$$A(q^{-1})\, y(t) = B(q^{-1})\, u(t) + \varepsilon(t)$$

the model structure is strictly globally identifiable. The above equation can be written as

$$y(t) = \sum_{n=1}^{\infty} g(n)\,[q^{-n}\, u(t)] + \sum_{n=0}^{\infty} b(n)\,[q^{-n}\, \varepsilon(t)] \tag{6.10}$$

$$= G(q)\, u(t) + H(q)\, \varepsilon(t) \tag{6.11}$$

with

$$G(z,\theta) = \frac{B(z)}{A(z)}$$

$$H(z,\theta) = \frac{1}{A(z)}$$

$$\underline{\theta} = [a_1, \ldots a_n, b_0 \ldots b_m]^T \tag{6.12}$$

Equality of H in (6.9) implies that the A-polynomials must coincide, which in turn implies that the B-polynomials must coincide for G to be equal. It is thus immediate to verify that equation (6.9) holds for all θ^* in the model structure (6.12). Consequently, the structure (6.12) is strictly globally identifiable.

In case of system with unknown delay, the model structure is

$$A(q^{-1})\, y(t) = q^{-d} B(q^{-1})\, u(t) + \varepsilon(t) \tag{6.13}$$

with

$$G(z,\theta) = \frac{q^{-d} B(z)}{A(z)}$$

$$H(z,\theta) = \frac{1}{A(z)}$$

and

$$\underline{\theta} = [a_1, \ldots a_n, b_o, \ldots b_m, d] \tag{6.14}$$

Once again, equality of H in (6.12) implies that the A-polynomials must coincide. For equality of the G transfer function, the B-polynomials have to coincide and

$$z^{-d} = z^{-d^*} \tag{6.15}$$

i.e.,

$$e^{-i\omega d} = e^{-i\omega d^*} \tag{6.16}$$

Equation (6.13) and (6.14) can be satisfied only if d = d*. Hence equation (6.9) holds good for all θ* in the model structure and the Equation Error Model structure with unknown time delay is globally and locally identifiable.

CHAPTER VII

EXAMPLE PROBLEMS

The frequency domain estimation technique will be applied to estimate two classes of problems. One will be to estimate unknown parameters. The other will be to estimate the unknown parameters and also the time delay. In simulating both cases the effect of noise on the estimates is investigated. The length of the sequences used in the transforming the data to the frequency domain is also considered. Examples showing the effect of various input sequences on the estimation of parameters and the time delay will be presented.

Estimation of Parameters

In the first set of six examples, the unknown parameters of a second order system are estimated. The effects of the various input sequences and measurement noise on the parameter estimates are investigated. A summary of these example problems is presented in Table II.

Consider the system described by the difference equation:

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_0 u(t) + \varepsilon(t) \tag{7.1}$$

where $a_1 = 0.125$, $a_2 = 0.5$ and $b_0 = 1.0$ and are assumed constant. For efficient computation using the Fast Fourier Transform (FFT), the number of DFT points $N = 2^r$, where r is any positive integer. Using equation (4.1) and (4.2), the system (7.1) can be written as

$$\begin{bmatrix} Y_R(k) \\ Y_I(k) \end{bmatrix} = \begin{bmatrix} -[Y_R(k)\,C1) + Y_I(k)\,S(1)] - [Y_R(k)\,C(2) + Y_I(k)\,S(2)] - U_R(k) \\ -[Y_I(k)\,C(1) - Y_R(k)\,S(1)] - [Y_I(k)\,C(2) - Y_R(k)\,S(2)] - U_I(k) \end{bmatrix}$$

$$\begin{pmatrix} \theta 1 \\ \theta 2 \\ \theta 3 \end{pmatrix} + \begin{bmatrix} E_R(k) \\ \varepsilon_I(k) \end{bmatrix} \qquad (7.2)$$

where θ_1, θ_2, and θ_3 are the parameters to be estimated. The initial estimates of the parameters are $\hat{\underline{\theta}}$ = [0.10, 0.40, 0.75] and the covariance matrix P_o is initialized to 10^3*I.

In Example 1, the input sequence u(t) = t exp(-t/5) is used for identification purposes. The measurement noise ε(t) is assumed to be zero. Figure 3 shows the estimate of the parameters obtained for (7.2) as the transformed data at the various frequencies is processed sequentially when the block size or the number of points for the FFT used is N = 64. After the initial transients, the parameters quickly converge to their true values.

Example 2 illustrates the effect of noise on the parameter estimates. In this case, the same problem of Example 1 is considered. However, the measurement noise is assumed to be present. To illustrate the effect of noise, the system is corrupted with noise ε(t) where ε(t) is a uniformly distributed random sequence whose strength is 10 percent of the maximum value of the output y(t) (i.e., $\varepsilon(t) \in ymax * [-0.1, 0.1]$). Figure 4 shows the estimate of the parameters as the data at the various frequencies is processed recursively. In this case, the estimates converge to their true values in a mean sense.

In the next two examples, a square-wave input is used for identification. The square wave input has unit amplitude and a period of 32 sampling intervals. For Example 3, the measurement noise is assumed absent. Figure 5 shows the estimated parameters as the data is processed sequentially. As in Example 1, except for the initial transients, the parameters converge very quickly to their true values and remain constant.

With Example 4, measurement noise is assumed present. As with Example 2, the strength of the noise is assumed to be 10 percent of the signal. With the same square-wave input sequence, Figure 6 shows the estimate of the parameters. Notice that after the initial transients, the parameters converge to their true values in a mean sense. The estimates obtained in this case are superior to those obtained using an exponential input. This is due to the fact that the square wave has more frequency content than the exponential input.

Example 5 investigates the estimation of parameters of the second order system with a sinusoidal input. The input sequence used for identification is a sum of sinusoids

$$u(t) = \sin\frac{2\pi}{N}t + \sin\frac{4\pi}{N}t + \sin\frac{8\pi}{N}t + \sin\frac{3\pi}{N}t + \sin\frac{5\pi}{N}t + \sin\frac{7\pi}{N}t \qquad (7.3)$$

Since the number of unknown parameters is three, to ensure that they can be estimated, the input sequence is chosen as a sum of three sinusoid sequences. This prevents the iobservation matrix from being singular. Figure 7 shows the estimate of the system parameters at the various frequencies when measurement noise is absent. Figure 8 shows the estimate of the same parameters with sinusoidal input with measurement noise, which has been defined previously.

Estimation of Parameters and Delay (First Order)

In this set of six examples, the estimation of both the system parameters and time delay of a first order system is considered. The effects of the various input sequences and measurement noise on the estimates are investigated. The number of DFT points used for transforming the data (block size) in N = 64. Table III presents a summary of the various examples.

Consider the system described by the difference equation:

$$y(t) + a_1 y(t-1) = b_0 u(t-d) \tag{7.4}$$

where $a_1 = 0.5$, $b_0 = 1.0$ and $d = 3$ and are assumed constant. Using equation (5.6), the system (7.4) can be transformed to the frequency domain and written as:

$$Z(k) = \begin{bmatrix} Y_R(k) \\ Y_I(k) \end{bmatrix}$$

and

$$\hat{Z}(k,\theta) = \begin{bmatrix} -[Y_R(k)C(1) + Y_I(k)S(1)]\,\theta_1 + [U_R(k)C(\theta_3) + U_I(k)S(\theta_3)] \\ -[Y_I(k)C(1) - Y_R(k)S(1)]\,\theta_1 + [U_I(k)C(\theta_3) - U_I(k)S(\theta_3)] \end{bmatrix} \tag{7.5}$$

with θ_1, θ_2 and θ_3 are the parameters to be estimated. The parameters and the time delay are estimated using the Nonlinear Least-Square Frequency Domain Estimator given by equations (5.14), (5.25) and (5.26). The initial estimates of the parameters are $\theta = (0.4, 0.75, 2.0)$ and the covariance matrix was initialized to $P_0 = \text{diag}\,(10,10,10^3)$. When measurement noise is assumed present, the system output is corrupted with noise $\varepsilon(t)$ where $\varepsilon(t)$ is a uniformly distributed random sequence whose strength is 10 percent of the maximum value of the output y(t) (i.e., $\varepsilon(t) \in ymax * [-0.1, 0.1]$).

Example 7 uses the exponential input u(t) = t exp(-t/5) and has no measurement noise. Figure 9 shows the estimate of the parameters obtained for (7.5) as the data at the various frequencies is processed sequentially. After the initial transients, the parameters converge to their true values. An interesting observation is that the parameters converge only after the time delay has been correctly identified.

Figure 10 shows the estimate of the parameters and time delay for example 8 which uses the exponential input and has measurement noise, as the data is processed sequentially. With noise present, the parameters converge to their true values in a mean sense. Since a block of data of length N is being processed, the estimates d(k) of the time delay within a block need not be an integer. However, the final estimate for a block of data, d(N-1) must be an integer.

In the next two examples, a square-wave input is used for identification of the system parameters and time delay. The square wave has unit amplitude and a period of 32 sampling intervals. Figure 11 shows the estimate of the parameters and time delay with no measurement noise. A comparison of Figure 11 and Figure 9 shows that the estimates and time delay obtained using the square-wave input converge faster than those obtained using an exponential input. This can be attributed to the richer frequency content of the square wave. The effect of measurement noises are included in example 10. Figure 12 shows the estimate of the parameters and time delay in a noisy environment using the square wave. Again the estimates now converge in a mean sense. The estimates obtained with noise for square wave converge faster as compared to a decaying exponential input sequence.

Example 11 investigates the estimation of parameters and time delay of the first order system using sinusoidal inputs. The input sequence used for identification is a sum of sine terms given by (7.3). The input with three frquencies is chosen to ensure that the three parameters a_1, b_0 and d can be estimated. Figure 13 shows the estimate of the system parameters and time delay. When the parameters and time delay are estimated with sinusoidal input in the presence of noise, the estimate obtained are noisy and converge to their true values in a mean sense.

The next six examples investigate the effect of increasing the number of DFT points. The same first-order system without time delay (7.2) and with time-delay described by (7.4) is considered. The number of DFT points or the block size used is N = 128. Table IV presents a summary of the various examples using different inputs for the system with no delay and Table V presents the summary for system with unknown delay.

Figure 14 shows the estimates of the parameters with an exponential input. The estimates converge in about 32 frequency steps. Figure 15 presents the estimated parameters in the presence of noise.

With a square wave input of period of 32 sampling intervals, estimate of parameters obtained are as shown in Figure 16. Figure 17 shows the estimated parameters and time delay in a noisy environment.

When the sinusoidal input (7.3) is used the parameter estimates obtained are as in Figure 18 and the estimated parameters and time delay are shown in Figure 19.

In the final six examples of this chapter, the effect of increasing the number of DFT points on the parameter and delay estimates for the first order system with time delay. For an exponential input, Figure 19 presents the estimates for a noiseless case and Figure 20 provides the estimates in the presence of noise. Figures 21 and 22 provide the corresponding estimates for a square wave input. Finally, Figures 23 and 24 present the estimates when a sinusoidal input is used for estimation.

From the above examples the following conclusions can be drawn. The quality of the estimates depends on the strength of signal to measurement noise ratio. In the presence of measurement noise, the estimates deteriorate. The estimates obtained are nevertheless unbiased and value of the estimates wanders about the true value. Also, the mean value of the estimates converge to the true values.

In the case of estimation of parameters and time delay, the parameter estimates begin to converge to their true values after the delay has been identified. This is quite reasonable to expect since a wrong delay value implies a wrong structural indices for the model and hence a different set of parameters are fitted to minimize the error between the true and incorrectly estimated model.

The increase in the number of DFT points does not make a significant improvement in the quality of estimates since it merely increases the frequency resolution of the signal spectrum. A slight improvement in the estimates may be expected since finite signals are better represented in the frequency domain with more DFT points.

One of the problems with nonlinear minimization problems is that of parameter convergence. The convergence of the estimates depend on the starting points or initial guesses for the estimates. Since the estimation of parameters and time delay is inherently nonlinear, the convergence of the estimates depend on the initial guesses for these parameters. Initial values should be close to the true parameter otherwise the estimates may never converge at all.

Further, in the nonlinear estimation problem the choice of the initial covariance matrix plays an important role in the quality of the estimates. Simulation studies show that setting the initial covariance to an arbitrarily large value may lead to biased estimates. Careful consideration has to be given to the gradients of the Performance Measure hypersurface in adjusting the initial values of the covariance matrix P_0. Examining equation (7.5) shows the Performance Hypersurface has the smallest gradient with respect to the delay parameter. hence to accelerate convergence of the delay estimates over the parameter estimates, the element of the initial covariance matrix corresponding to the delay term must be chosen much larger than corresponding parameter covariances. Further justification to the above mentioned choice of covariances can be argued based on the magnitude of the parameters and delay since delay magnitudes are generally larger than parameter magnitudes for stable systems.

The quality of the estimated parameters depend on the type of input signal. An input signal that is rich and contains a large number of frequencies will provide better estimates. The good input signal accelerates parameter convergence.

Finally, the number of data points also has an effect on the parameter estimates. As the number of data points increase better estimates are obtained if the unknown parameters are constant. However, if the unknown parameters are slowly varying, a proper choice of the datapoints has be made so as to satisfy the assumptions made in developing the procedure.

CHAPTER VIII

IMPROVING PARAMETER CONVERGENCE

In estimating the parameters and time delay, the unknown parameters converge to their true values only after the delay has converged to its true value. If the correct value of the delay is not estimated, the parameters estimates will never converge. This is quite reasonable to expect because without the correct time delay the order of the system identified is different from the true system which leads to a different set of parameters being estimated.

When the time delays are identified due to presence of disturbance, measurement noises or even round off errors, generally, the time delays estimated are not exact but approach the true value. Since the time delay is the most critical parameter, small errors can cause the parameter estimates to deviate from their true values significantly.

To improve the parameter estimates the following procedure is recommended. When the variance of the delay estimates (which can be obtained from the covariance matrix P(k)) is small and less than a predefined bound δ a small positive number, the delay estimate d(k) is rounded off to the nearest integer. The elements in the covariance matrix corresponding to the delay term are then set to zero. This reduces the rank of the covariance matrix and prevents the delay term from being updated. The covariance matrix is then reset and the estimation continued. This keeps the delay bound to an integer. The resetting of the covariance matrix revitalizes the estimation algorithm. Once the delay is fixed the estimation now reduces to the case of a known delay. A further restriction is imposed that for causal systems, the time delay $d(k) \geq 0$.

The time delay and the parameters example problem 7 of the last chapter are estimated with these modifications. Figure 25 shows the estimates of the time delay and the parameters at the various frequencies as they are estimated recursively. Referring to Figure 25, at about $k = 40$, the delay has been identified and rounded off to 3, the nearest integer. The elements in the covariance matrix corresponding to the time delay are set to zero and the covariance matrix reset to rejuvenate the estimation process. Notice that once the delay has been fixed, and the covariance matrix reset, the system parameters "jump" to their true values.

To investigate the effect of noise, the system is Example problem 6 which is a noise corrupted version of Example 7 is considered. The noise level is the same as in the previous examples. Figure 26 shows the estimated time delay and parameters. The parameter estimates are deteriorated as compared to the noise free case. However, the estimates converge in a mean sense.

Figure 27 shows the improved estimates of parameters and time delay of the example problem 9 where the input is a square wave and the measurement noises are assumed absent. In this case the time delay and the parameters converge much faster than the previous case (Figure 26) with the exponential input. This is due to the fact that a square wave input is more suitable for identification purposes than an exponential signal.

The above problem,is corrupted with measurement noise to obtain Example 10. With the improved estimation scheme, the estimate of the parameters are as shown in Figure 28. Again as in the noiseless case, a square wave input provides better estimates with faster convergence as compared to an exponential input signal. With a sinusoidal input and when the improved algorithm is used, the parameter estimates when no measurement noise is present are shown in Figure 29. Figure 30 shows the corresponding parameter estimates in the presence of measurement noise.

Finally, the improved technique is applied to the estimation of parameters and delay of the second order system. The second order system considered is described by the difference equation

$$y(t) + a_1\, y(t-1) + a_2\, y(t-2) \;=\; b_0\, u(t-d) + b_1\, u(t-d-1) \qquad (7.6)$$

where $a_1 = 0.25$; $a_2 = 0.5$; $b_0 = 1.0$; and $b_1 = 0.75$. and the unknown time delay, d=3, are all assumed constant. Table VII present a summary of the example problems studied for this case.

Figure 31 shows the improved estimates in an ideal case whereas Figure 32 provides the estimates in a noisy environment for an exponential input. With a square wave input, the noise free estimates are provided by Figure 33. Finally, with noise present, the improved estimates are as shown in Figure 34.

A number of techniques have been proposed in this chapter for improving the parameter estimates. On comparing the results of the the various simulation studies, the modified estimation techniques definitively provide better estimates. Further, the figures show that the estimates converge more quickly to their true values with the improved estimation technique.

CHAPTER IX

SUMMARY AND CONCLUSIONS

A method for recursive identification of the discrete time system parameters and time delay has been presented. The method is based on transforming the original time domain model and the data into the frequency domain. Transformation into the frequency domain results in the parameterization of the delay term. However, the transformed equations are nonlinear in the delay term. An efficient method for recursively estimating the time delay and the parameters from these nonlinear equations is presented. This led to the development of the sequential linear and nonlinear frequency domain estimator. In addition parameter convergence was improved and also accelerated by incorporating enhancements in to the estimation technique.

Special cases of the nonlinear estimator were developed which lead to variants of the Normalized Least Mean Square (LMS) in the frequency domain. These algorithms are required greatly reduced computational effort.

Some important properties of the estimator were also established. It was shown by simulation studies that, as in the time domain, the frequency domain estimator provides unbiased estimates in the presence of white noise. In the presence of measurement noise, the parameters converge to their true values in a mean sense. It is shown that the identifiability of the parameters and time-delay is preserved when the model is transformed to the frequency domain. The time delay and the parameters were shown to be identifiable.

The simulation studies showed that the parameter estimates depend upon the type of input signals. A signal which is rich with a number of frequency components is shown to provide better estimates than an input with few frequency components. In the ideal case with no noise, the estimates quickly converged to their true values. The estimates deteriorated in the presence of noise. However they converged to their true values in the mean sense. In the case of estimation of time delays and parameters, the estimation of the correct time delay was critical to obtain good parameter estimates. If the time delay was not identified, the true parameters were never identified. In estimation, the time delay converges to its true value before the parameter estimates. Also, if more data points were used in computing the DFT, slightly better estimates were obtained.

Finally, the present method provides advantages in terms of computation as compared to some of the other methods. Overparameterization or multiple models in a class require the covariance matrix has to be expanded or several estimation algorithms run simultaneously and additional computations done to extract the time delay. Moreover the computational expense in these methods is proportional to the size of the delay. With the proposed method, the computational effort remains the same, independent of the size of the delay. Transformation to the frequency domain has the effect of data compression. This requires greatly reduced computation since data at only a finite number of frequencies (which make up the input and output signals) have to be processed to estimate the parameters. In recursively estimating the parameters, the transformed data does not have to be processed sequentially at the various frequencies. Instead, it can be processed in any convenient order.

Recommendations For Future Work

In the development of the theory for the frequency domain estimator, the assumption made was that the parameters and time delay were slowly varying and that for a given block of data they remained constant. If however, the parameters were changing, the parameter sequence can be simply treated as another time sequence multiplying the data sequence. Hence the parameters can no longer can be factored out while computing the Fourier transforms. Instead, in taking the Fourier transform a convolution of the parameter sequence and the data sequence is obtained. For such cases similar parameter estimation techniques need to be developed.

One of the problems of nonlinear minimization is that of parameter convergence. The convergence of the estimates depend on the inital guesses and also the choice of the initial covariance matrix. The initial covariance matrix plays an important role in the quality of the estimates. Careful consideration has to be given to the gradients of the Performance Hypersurface in adjusting the initial values of the covariance matrix for nonlinear estimation problem requires further study.

REFERENCES

Agarwal, M. and Canudas, C. 1986. American Control Conference, Minneapolis, Minnesota, Vol.1, pp. 728-733

Albert, A. E. and Gardner, L. A. 1967. Stochastic Approximation and Nonlinear Regression, Cambridge, Massachusetts: The MIT Press.

Astrom, K. J. and Wittenmark, B. 1984. Computer Controlled Systems: Theory and Design, Prentice-Hall.

Bee Bednar, J. and Coberly, W.A. 1976. Applied Time Series Analysis II, University Of Tulsa, Tulsa, Oklahoma.

Chou, J. H. and Horng, J. R. 1986. "Shifted-Chebyshev-Series analysis and identification of time-varying bilinear systems," International Journal of Control, Vol. 44, No. 3, pp. 847-866

Gabr, M. M. 1987. "Estimation of time delay and lag relationship of quadratic systems," Int. J. Control, Vol. 45, No.5, pp. 1549-1557.

Gabr, M. M. and Subba Rao, T. 1984. International Journal of Control, Vol. 40, No. 121.

Goodwin, G. C. and Sin, K. S. 1984. Adaptive Filtering Prediction and Control, Prentice-Hall.

Horng, J. R. and Ho, S. J. 1987. "Discrete pulse orthogonal functions for the analysis, parameter estimation and optimal control of linear time-varying digital systems," International Journal of Control, Vol. 45, No. 6, pp. 1975-1984.

Juricic, Dj. 1987. "Recursive Estimation Of Systems With Time-Varying Parameters and Delays", IFAC 10 World Congress on Automatic Control, Vol. 10, Munich, pp. 265-270.

King, R. E. and Paraskevopoulos, P. N. 1979. "Parameter Identification of Discrete-Time SISO Systems," Int. J. Control, Vol. 30, pp. 1023-1029.

Kurz, H. and Goedecke, W. 1981. "Digital Parameter-Adaptive Control of Processes with Unknown Dead Time," Automatica, Vol.17, No.1, pp. 245-252.

LaMarie, R. O., Valvani, L., Athans, M., and Stein, G. 1987. A Frequency-Domain Estimator for use in Adaptive Control Systems," ACC June 10-12, Minneapolis, Minnesota, pp. 238-244.

Ljung, L. 1987. System Identification Theory for the User, Prentice-Hall Inc.

Ljung, L., and Soderstrom, T. 1983. Theory and Practice of Recursive Identification, Cambridge, Massachusetts: The MIT Press.

Luenberger, D. G. 1973. Introduction to Linear and Nonlinear Programming, Reading, Massachusetts: Addison-Wesley.

Lui, C. and Shih, Y. P. 1985. "System analysis, parameter estimation and optimal regulator design of linear systems via Jacobi series," Int. J. Control, Vol. 42, pp. 211-224.

Mendel, J. M. 1986. Lessons in Digital Estimation Theory, Prentice-Hall Inc.

Nagumo, J. I. and Noda, A. 1967. "A learning method for system identification," IEEE Trans. on Automatic Control, Vol. AC-12, No. 3, pp. 282-287.

Oppenheim, A. V. and Schafer R. W. 1975. Digital Signal Processing, Prentice-Hall, Inc.

Palanisamy, K. R. and Bhattacharya, D. K. 1981. "System identification via block-pulse functions," Int. J. System Science, Vol. 12, pp. 643-647.

Paraskevopoulos, P. N. 1983. "Chebychev series approach to system identification, analysis and optimal control," J. Franklin Inst., Vol. 316, pp. 135-157.

Paraskevopoulos, P. N. 1985. "Legendre series approach to identification and analysis of linear systems," IEE Trans. Automatic Control, Vol. AC-30, pp. 585-589.

Pupeikis, R. 1985. "Recursive estimation of the parameters of linear systems with time delay," Proceedings of 7th IFAC Symposium on Identification and System Parameter Estimation, York, UK.

Rao, G. P. and Sivakumar, L. 1976. "Identification of deterministic time-lag systems," IEE Trans. on Automatic Control, August, pp. 527-529.

Robbins, H. and Monro, S. 1951. "A stochastic approximation method," Ann. Math. Stat., Vol. 22, pp. 400-407.

Sage, A. P. 1968. Optimum Systems Control, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Shih, D. H. and Kung, F. C. 1986. "Analysis and parameter estimation of non-linear systems via shifted Chebyshev expansions," International Journal of System Science, Vol. 17, No. 2, pp. 231-240.

Shimkin, N. and Feuer, A. 1988. "On the necessity of "block invariance" for the convergence of adaptive pole-placement algorithm with persistently exciting input," IEEE Trans. Automatic Control, Vol. 33, No. 8, pp. 775-780.

Sorenson, H. W. 1980. Parameter Estimation Principles and Problems, Control and Systems Theory, Vol. 9, Marcel Dekker, Inc.

Speedy, C. B., Brown, R. E. and Goodwin, G. C. 1970. Control Theory, Oliver S. Boyd, Edinburg.

Widrow, B. and Hoff, M. E. 1960. "Adaptive switching circuits," IRE Western Conv. Rec., Part 4, pp. 96-104.

Woodward, W. A. and Gray, H. L. 1979. "On relationship between the R and S arrays and the Box-Jenkins method of ARMA identification," ONR Tech. Report No.134, SMU Dept. of Statistics, July.

APPENDIXES

APPENDIX A

TABLES

TABLE I

INTERRELATIONSHIPS BETWEEN θ, $\underline{\Phi}(k)$ AND $\underline{V}(K)$

A. θ is deterministic.

1. $\underline{\Phi}(k)$ is deterministic.
2. $\underline{\Phi}(k)$ is random.
   a. $\underline{\Phi}(k)$ and $\underline{V}(k)$ are statistically independent.
   b. $\underline{\Phi}(k)$ and $\underline{V}(k)$ are statistically dependent.

B. θ is random.

1. $\underline{\Phi}(k)$ is deterministic.
2. $\underline{\Phi}(k)$ is random.
   a. $\underline{\Phi}(k)$ and $\underline{V}(k)$ are statistically dependent.
   b. $\underline{\Phi}(k)$ and $\underline{V}(k)$ are statistically dependent.

TABLE II

ESTIMATION OF PARAMETERS OF SYSTEM WITH NO DELAY,
NUMBER OF DFT POINTS, N = 64

Example	Order	Noise	Input
1	2	No	Exponential
2	2	Yes	Exponential
3	2	No	Square
4	2	Yes	Square
5	2	No	Sinusoids
6	2	Yes	Sinusoids

TABLE III

ESTIMATION OF SYSTEM PARAMETERS AND TIME DELAY, NUMBER OF DFT POINTS, N = 64

Example	Order	Noise	Input
7	1	No	Exponential
8	1	Yes	Exponential
9	1	No	Square
10	1	Yes	Square
11	1	No	Sinusoids
12	1	Yes	Sinusoids

TABLE IV

ESTIMATION OF PARAMETERS OF SYSTEM WITH NO DELAY, NUMBER OF DFT POINTS, N = 128

Example	Order	Noise	Input
13	2	No	Exponential
14	2	Yes	Exponential
15	2	No	Square
16	2	Yes	Square
17	2	No	Sinusoids
18	2	Yes	Sinusoids

TABLE V

ESTIMATION OF SYSTEM PARAMETERS AND TIME DELAY,
NUMBER OF DFT POINTS, N = 128

Example	Order	Unknown Delay	Noise	Input
19	1	No	Yes	Exponential
20	1	No	Yes	Exponential
21	1	No	Yes	Square
22	1	Yes	Yes	Square
23	1	Yes	Yes	Sinusoids
24	1	Yes	Yes	Sinusoids

TABLE VI

ESTIMATION OF SYSTEM PARAMETERS AND TIME DELAY, IMPROVED ALGORITHM, NUMBER OF DFT POINTS, N = 64

Example	Order	Noise	Input
25	1	No	Exponential
26	1	Yes	Exponential
27	1	No	Square
28	1	Yes	Square
29	1	No	Sinusoid
30	1	Yes	Sinusoid

TABLE VII

ESTIMATION OF SYSTEM PARAMETERS AND TIME DELAY, IMPROVED ALGORITHM, NUMBER OF DFT POINTS, N = 64

Example	Order	Noise	Input
31	2	No	Exponential
32	2	Yes	Exponential
33	2	No	Square
34	2	Yes	Square

APPENDIX B

FIGURES

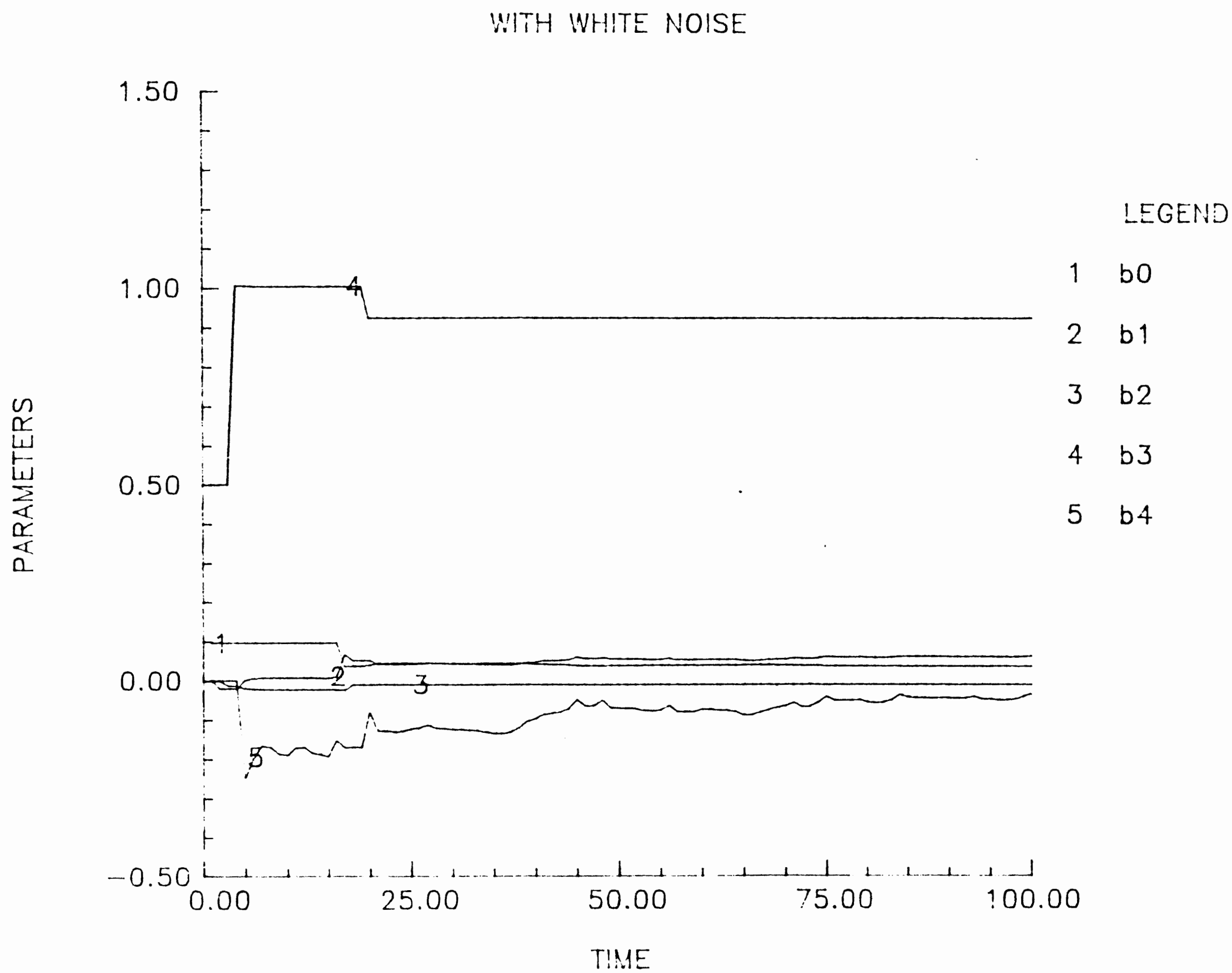

Figure 1. Parameter Estimates for System with Unknown Delay in the Time Domain Using an Overparameterized Model, Noise Present

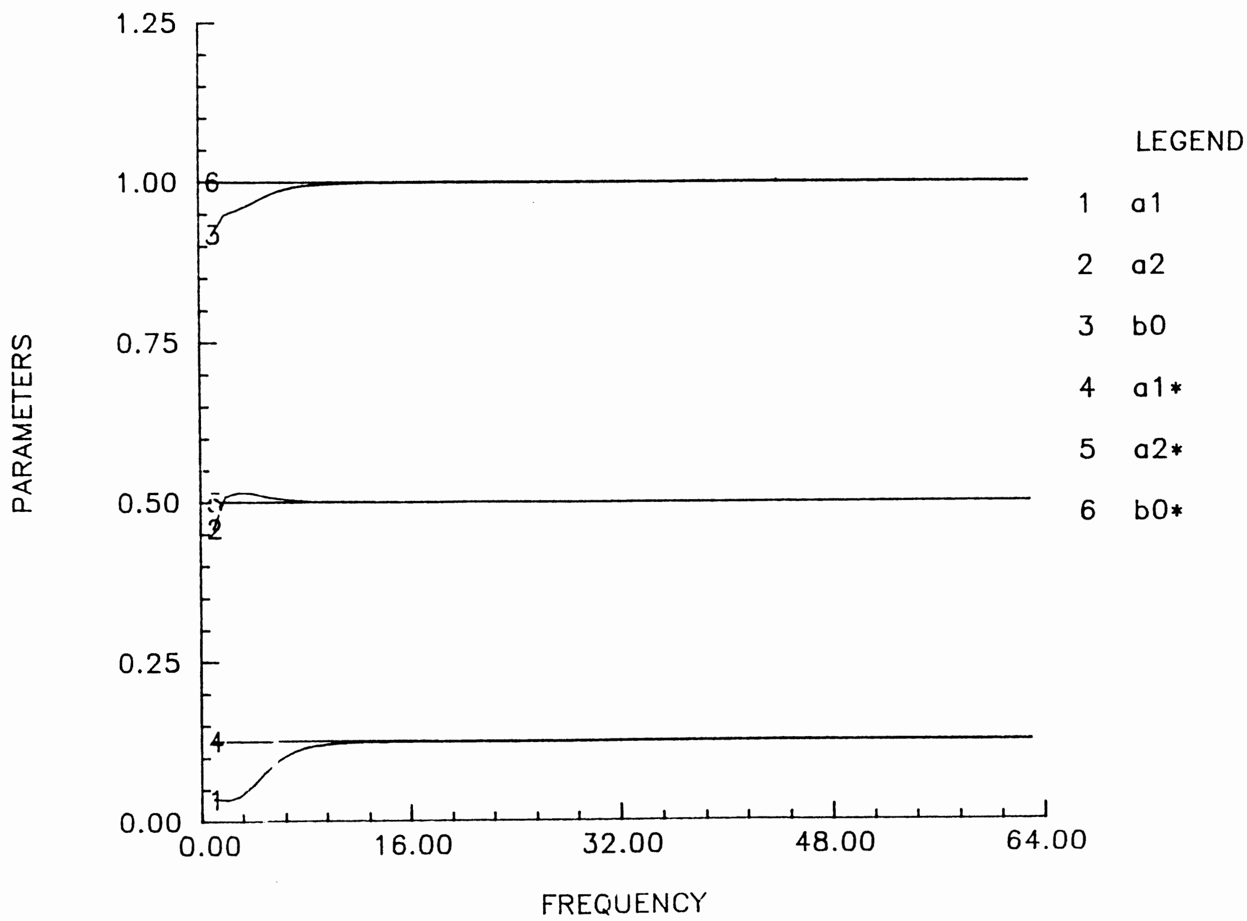

Figure 2. Parameter Estimates for System with no Delay, Exponential Input, N=64

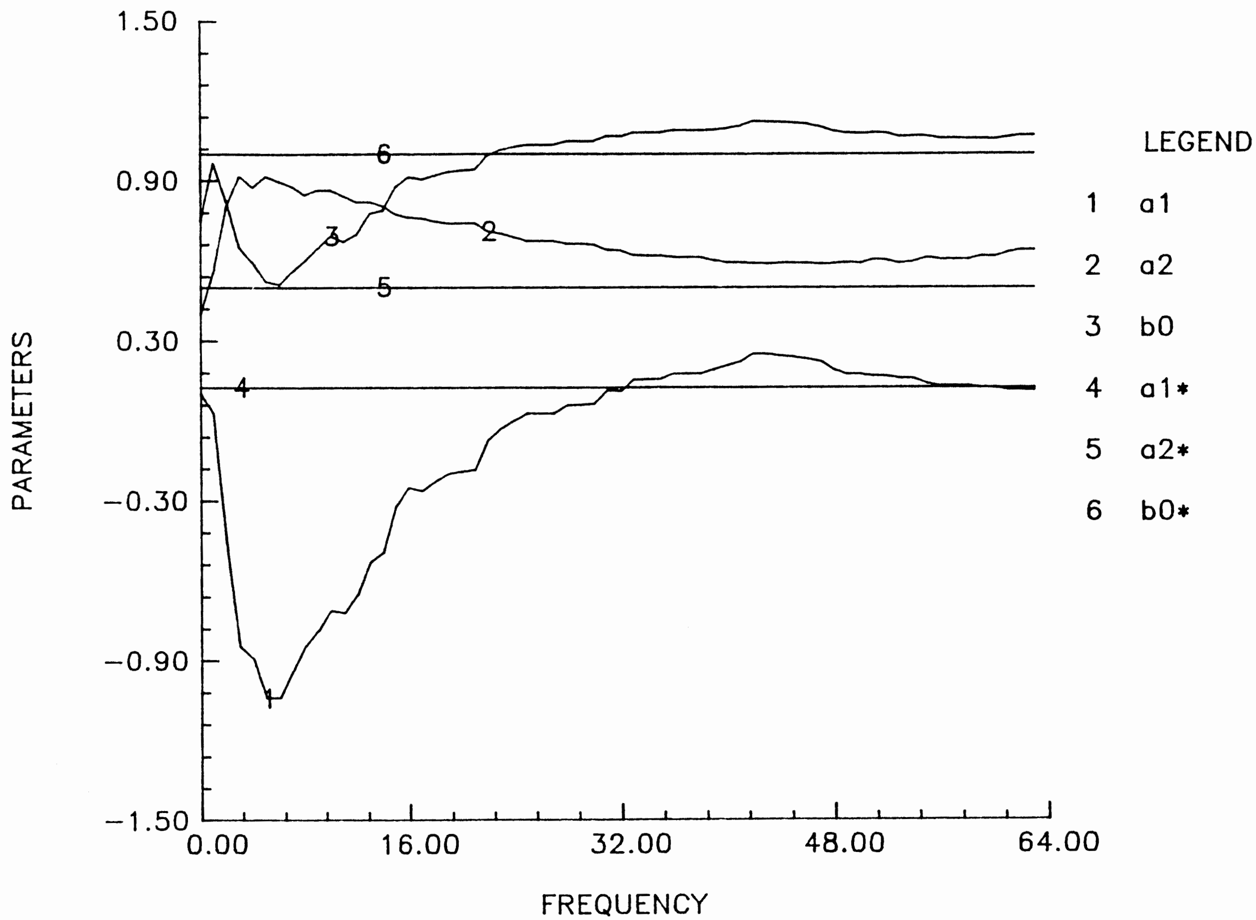

Figure 3. Parameter Estimates for System with no Delay, Exponential Input and Noise, N=64

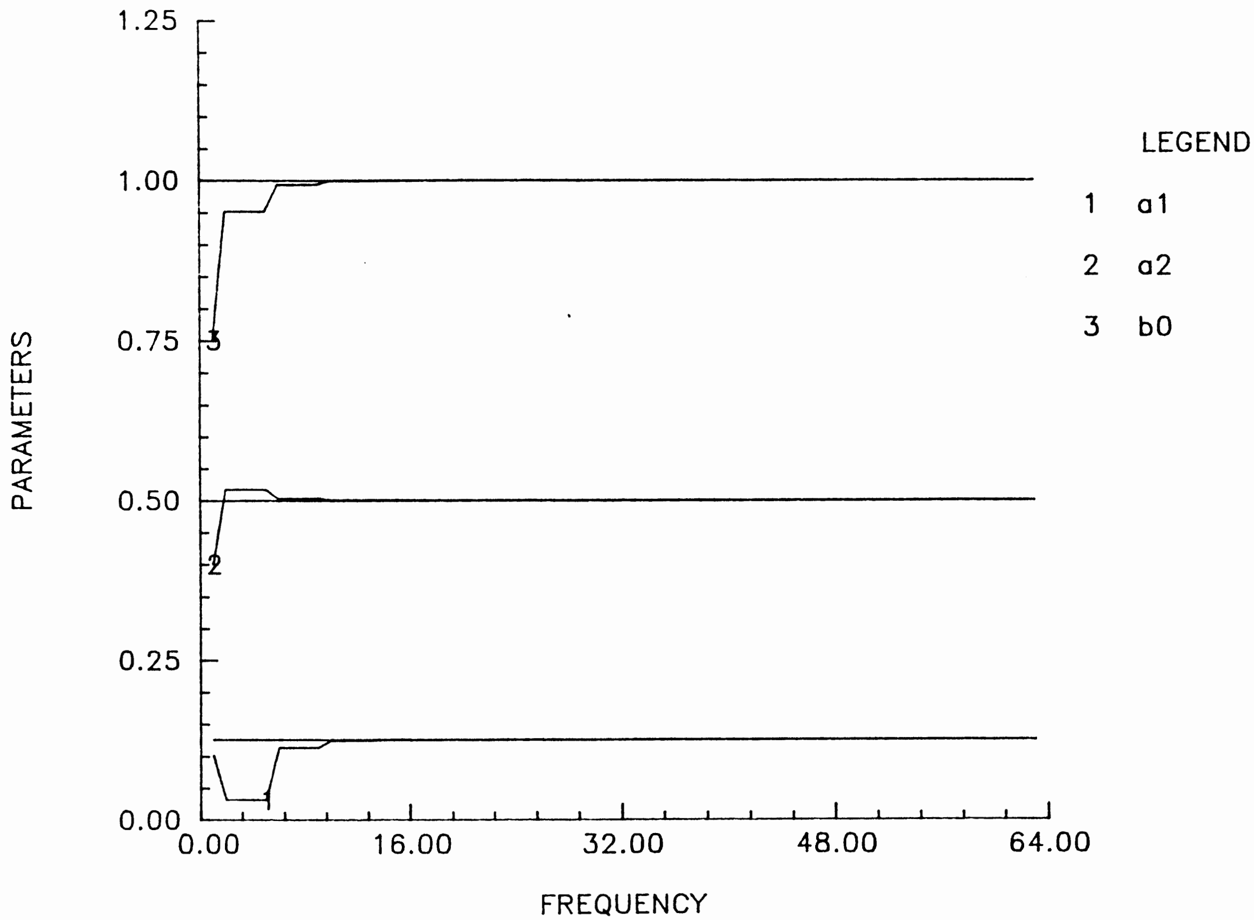

Figure 4. Parameter Estimates for System with no Delay, Square Wave Input, N=64

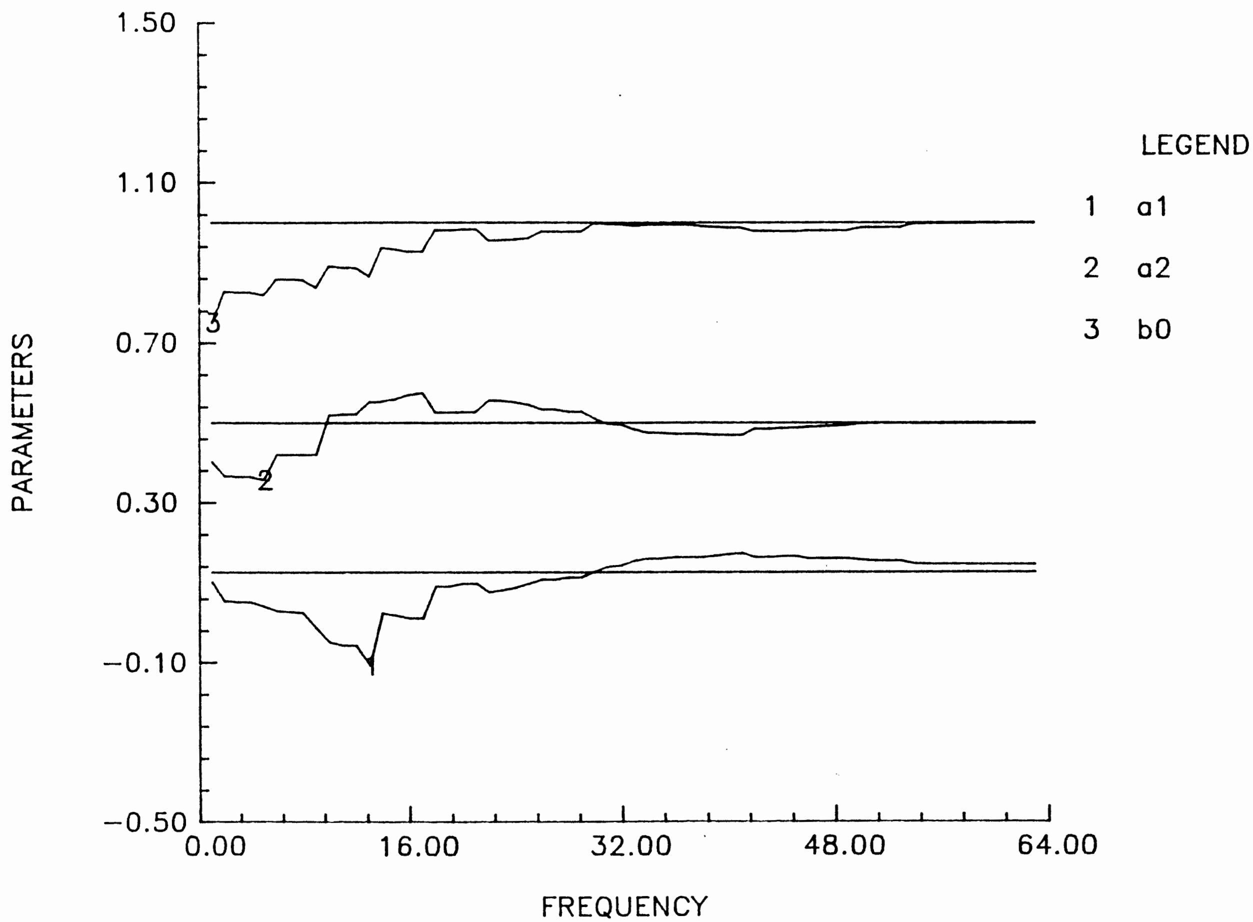

Figure 5. Parameter Estimates for System with no Delay, Square Wave Input and Noise, N=64

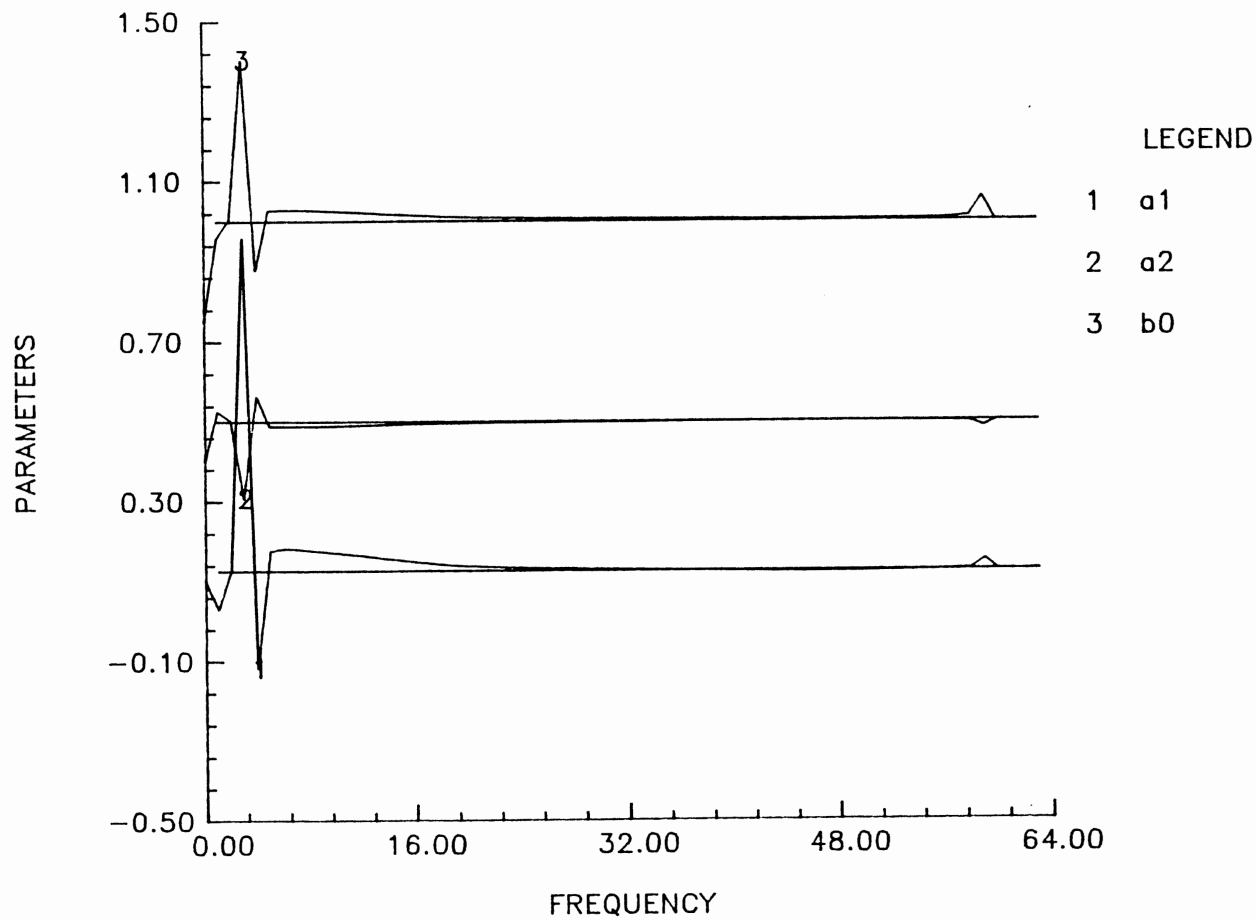

Figure 6. Parameter Estimates for System with no Delay, Sinusoidal Input, N=64

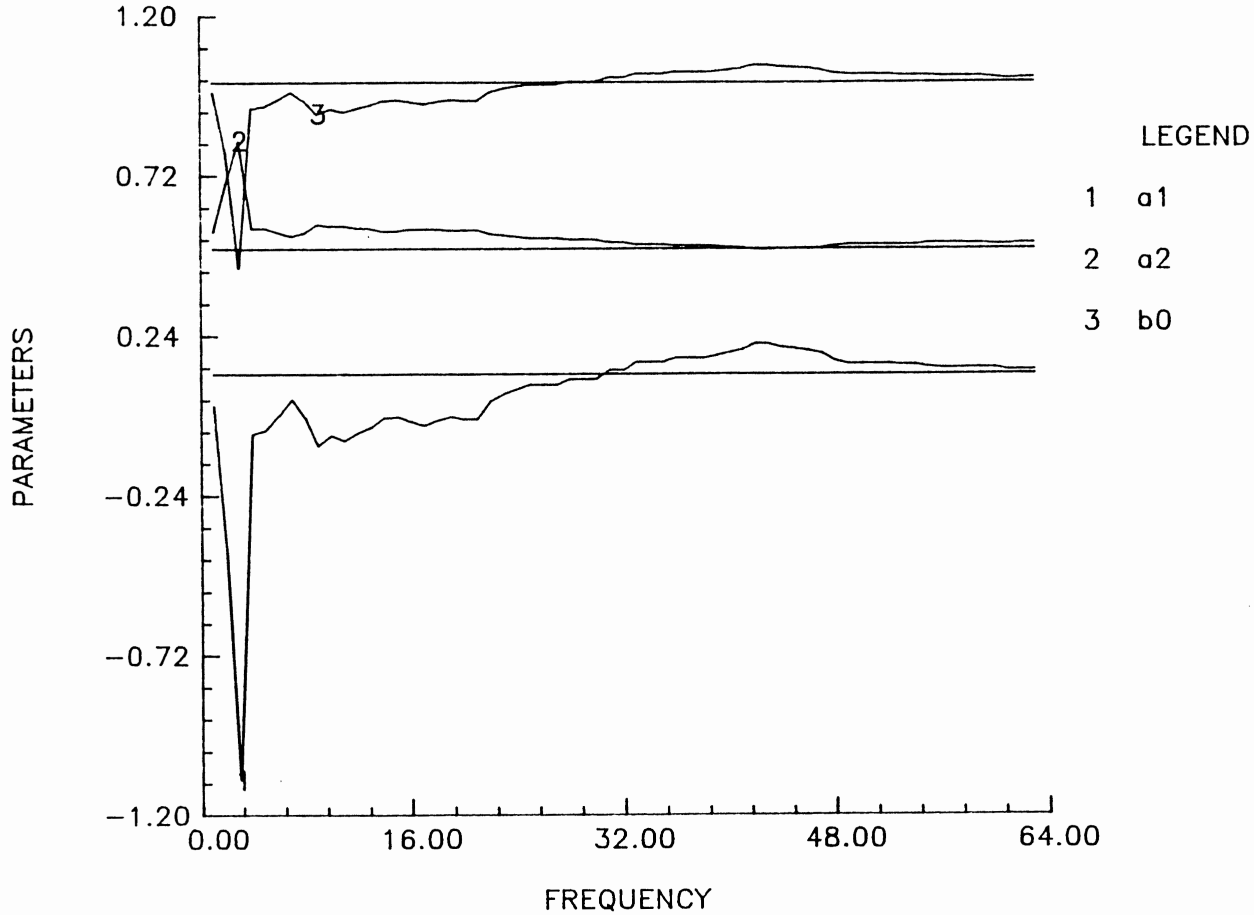

Figure 7. Parameter Estimates for System with no Delay, Sinusoidal Input and Noise , N =64

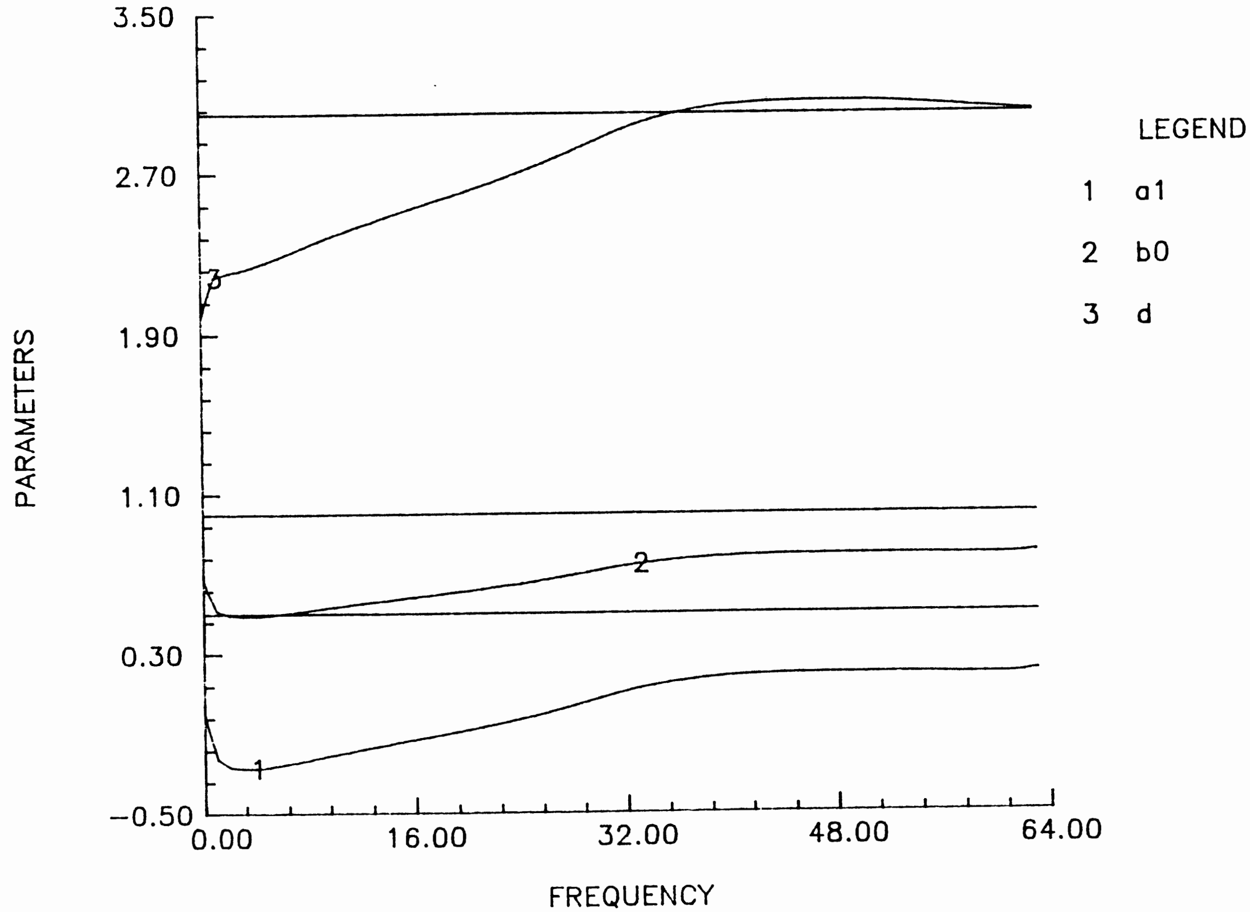

Figure 8. Parameter Estimates for System with Unknown Delay, Exponential Input, N=64

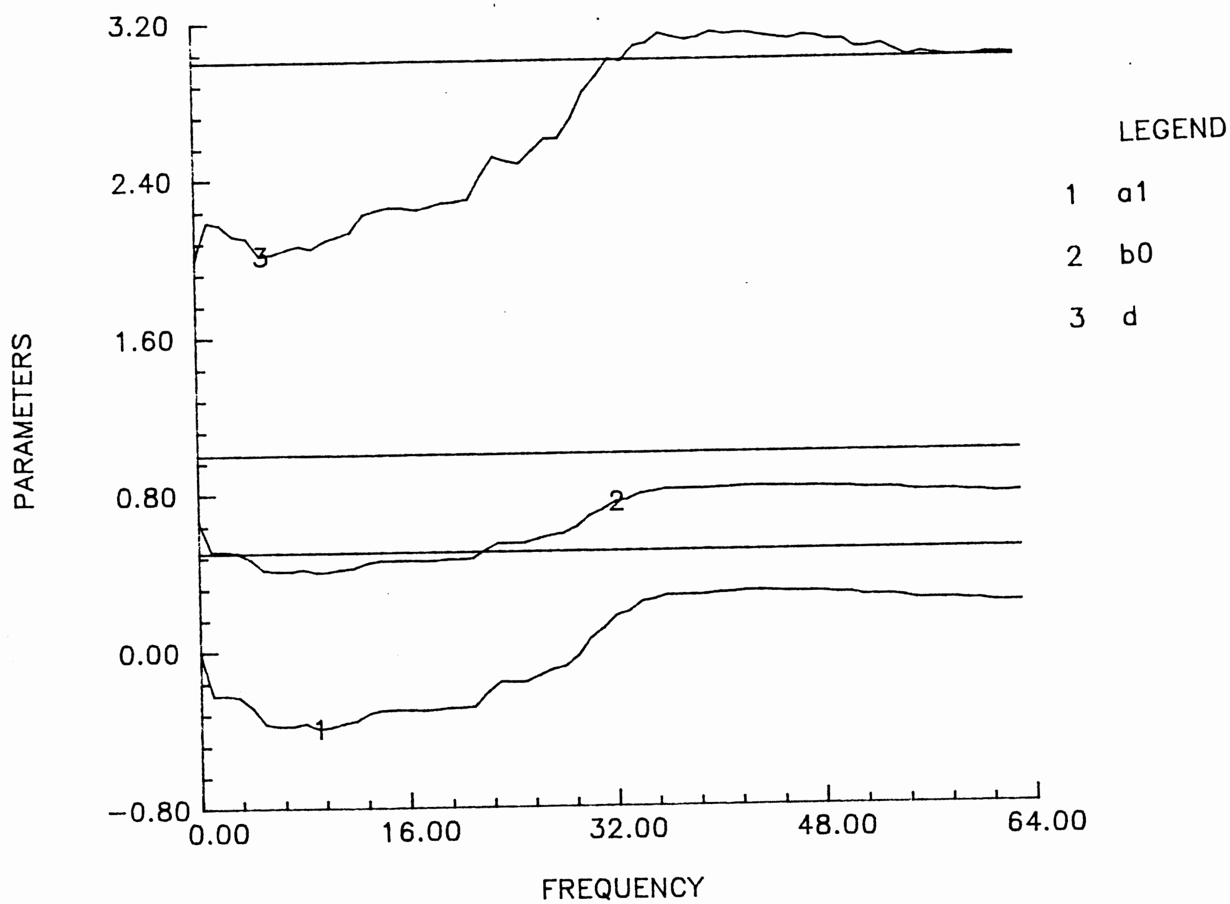

Figure 9. Parameter Estimates for System with Unknown Delay, Exponential Input and Noise, N=64

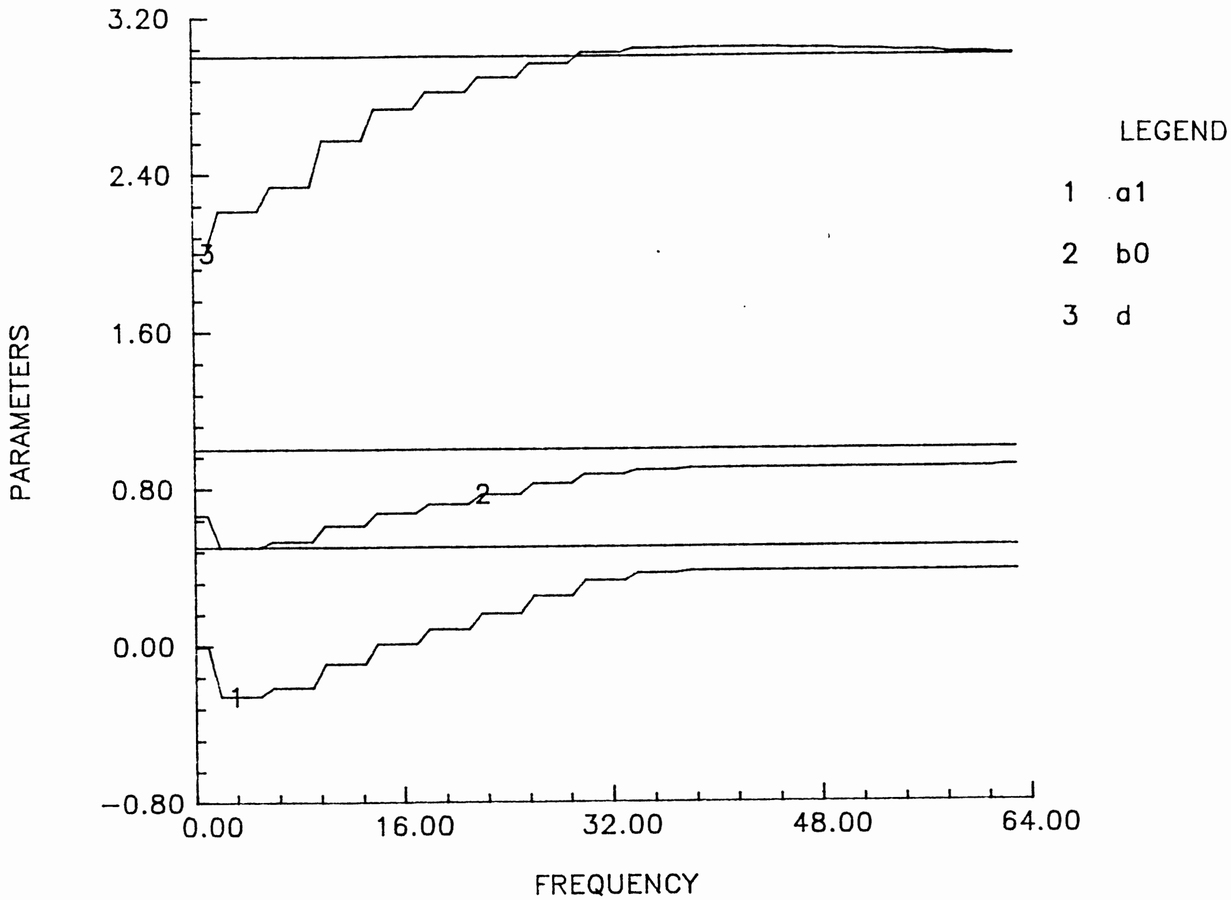

Figure 10. Parameter Estimates for System with Unknown Delay, Square Wave Input, N=128

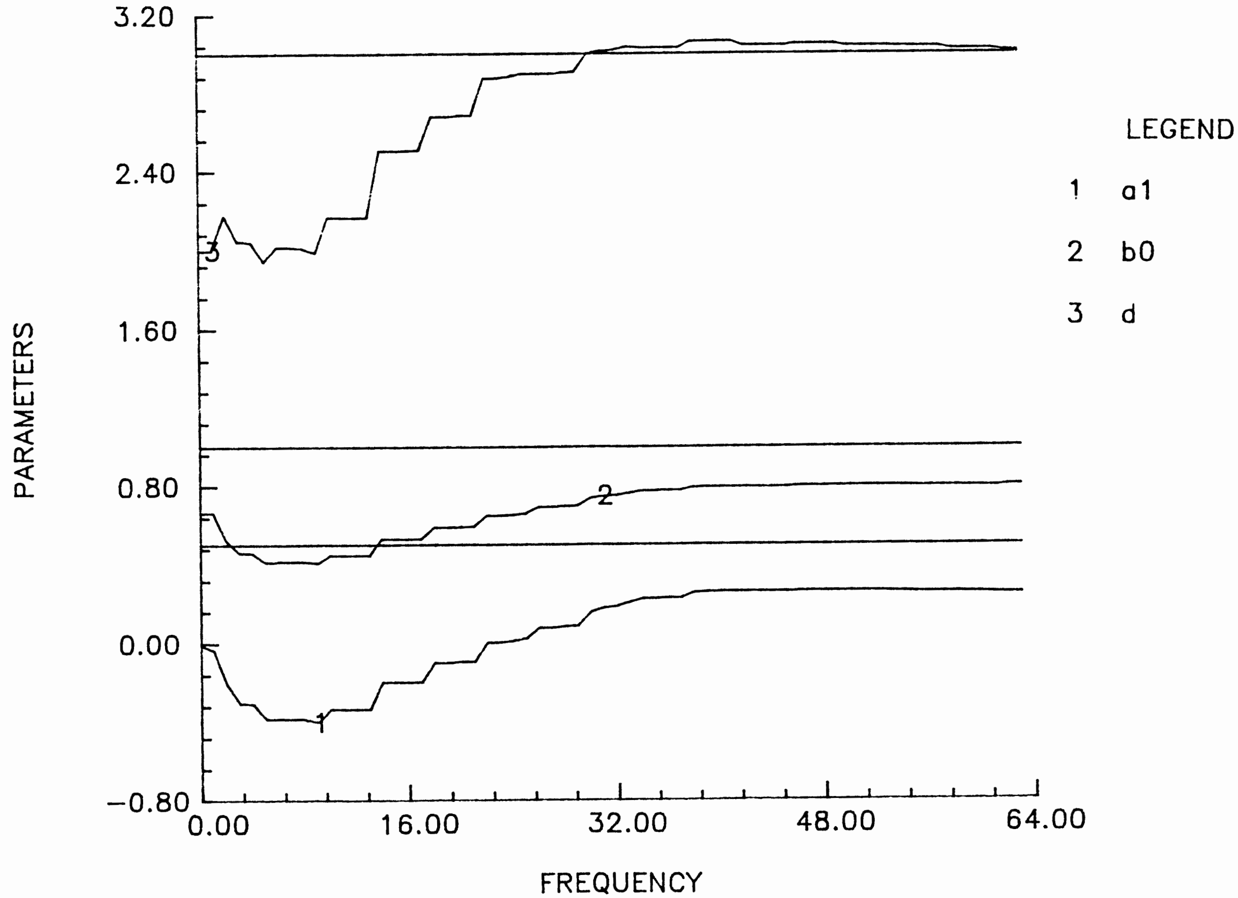

Figure 11. Parameter Estimates for System with Unknown Delay,
Square Wave Input and Noise, N=128

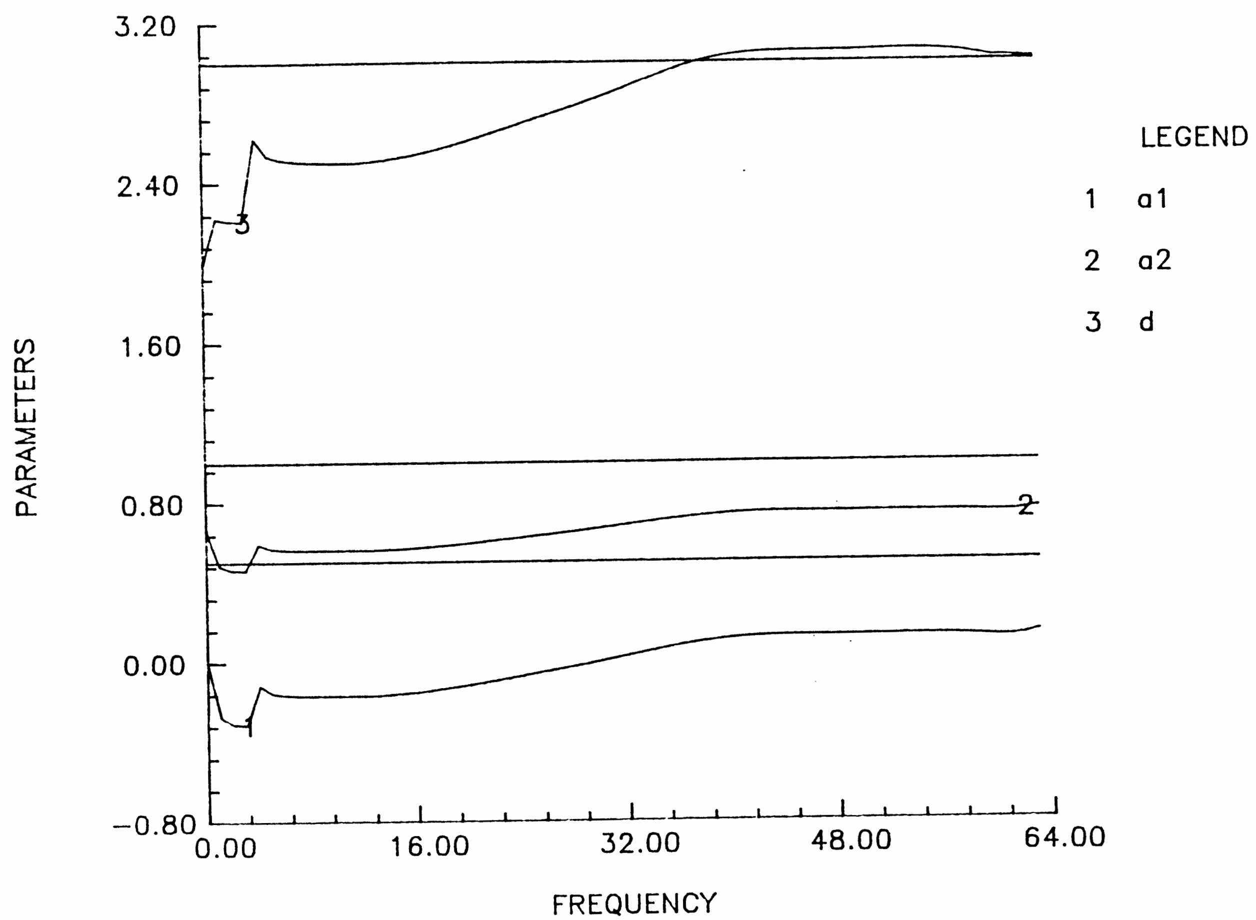

Figure 12. Parameter Estimates for System with Unknown Delay, Sinusoidal Input, N=64

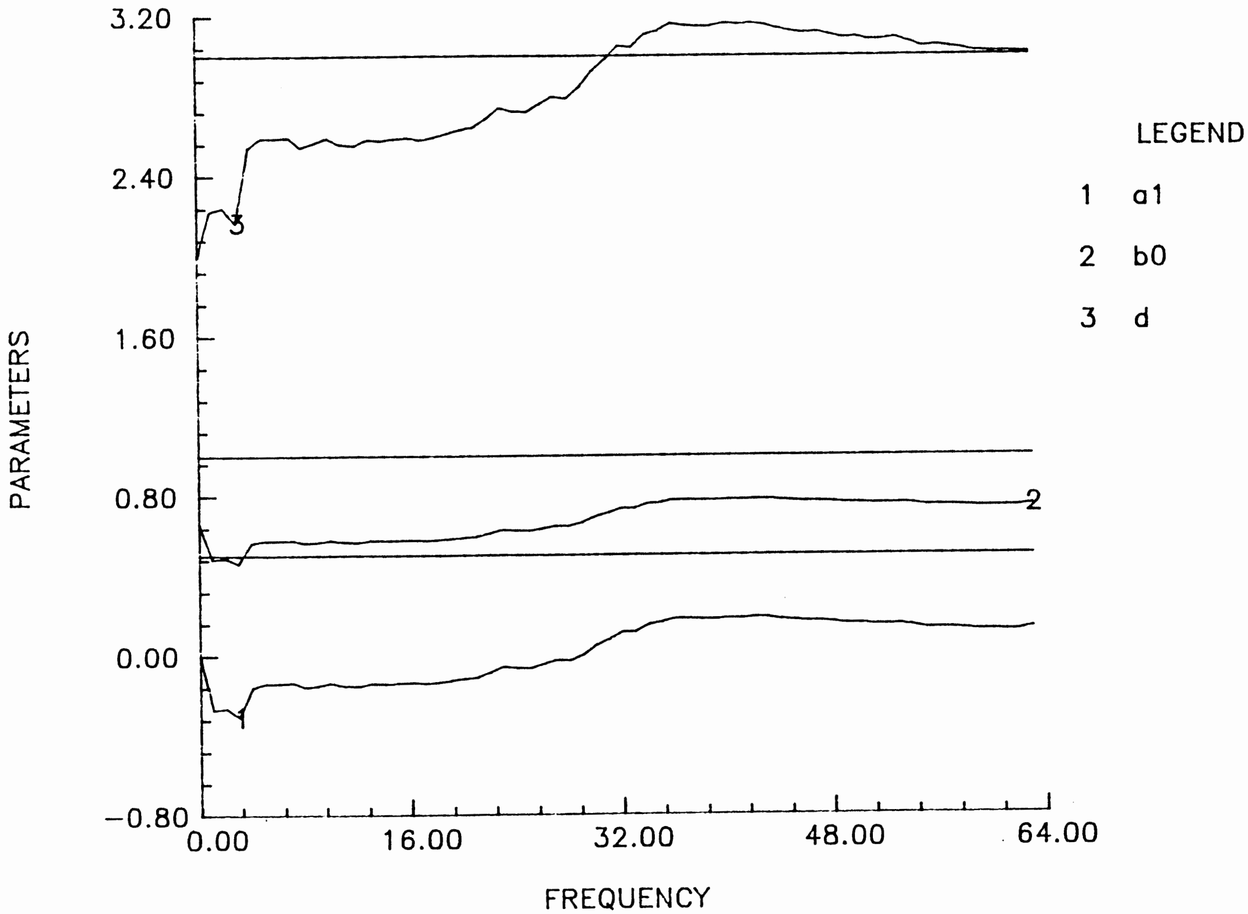

Figure 13. Parameter Estimates for System with Unknown Delay, Sinusoidal Input and Noise, N=64

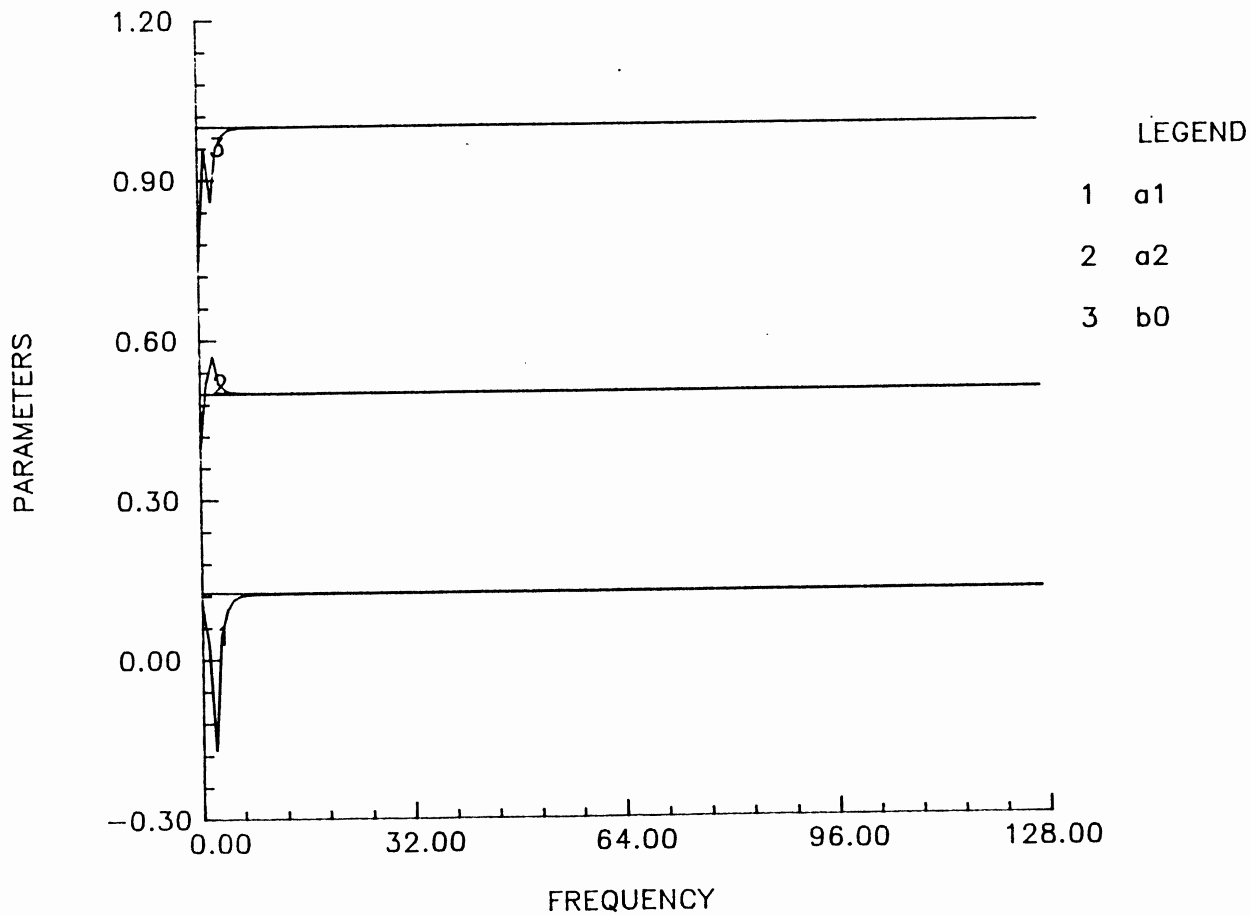

Figure 14. Parameter Estimates for System with no Delay, Exponential Input, N=128

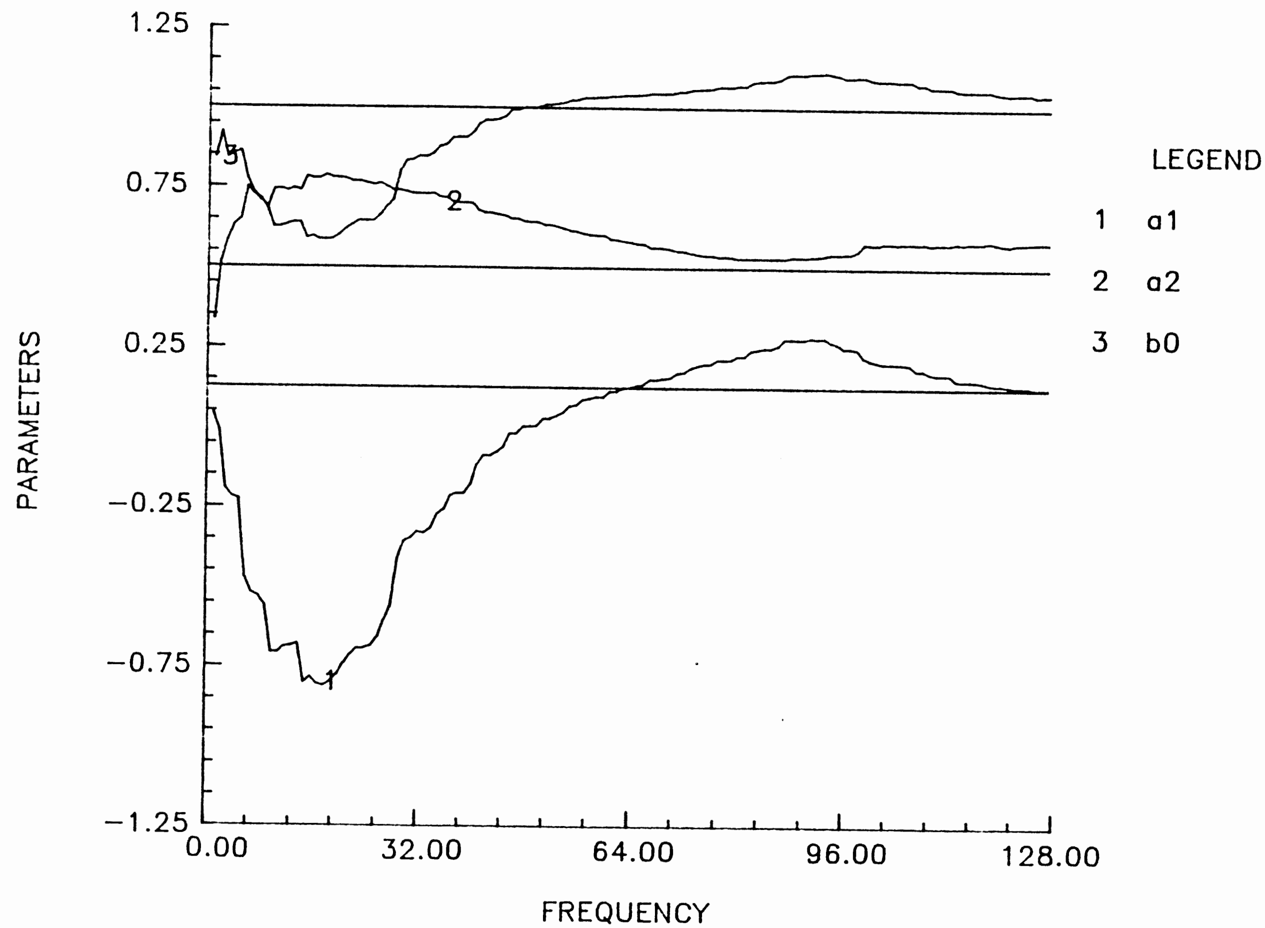

Figure 15. Parameter Estimates for System with no Delay, Exponential Input and Noise, N=128

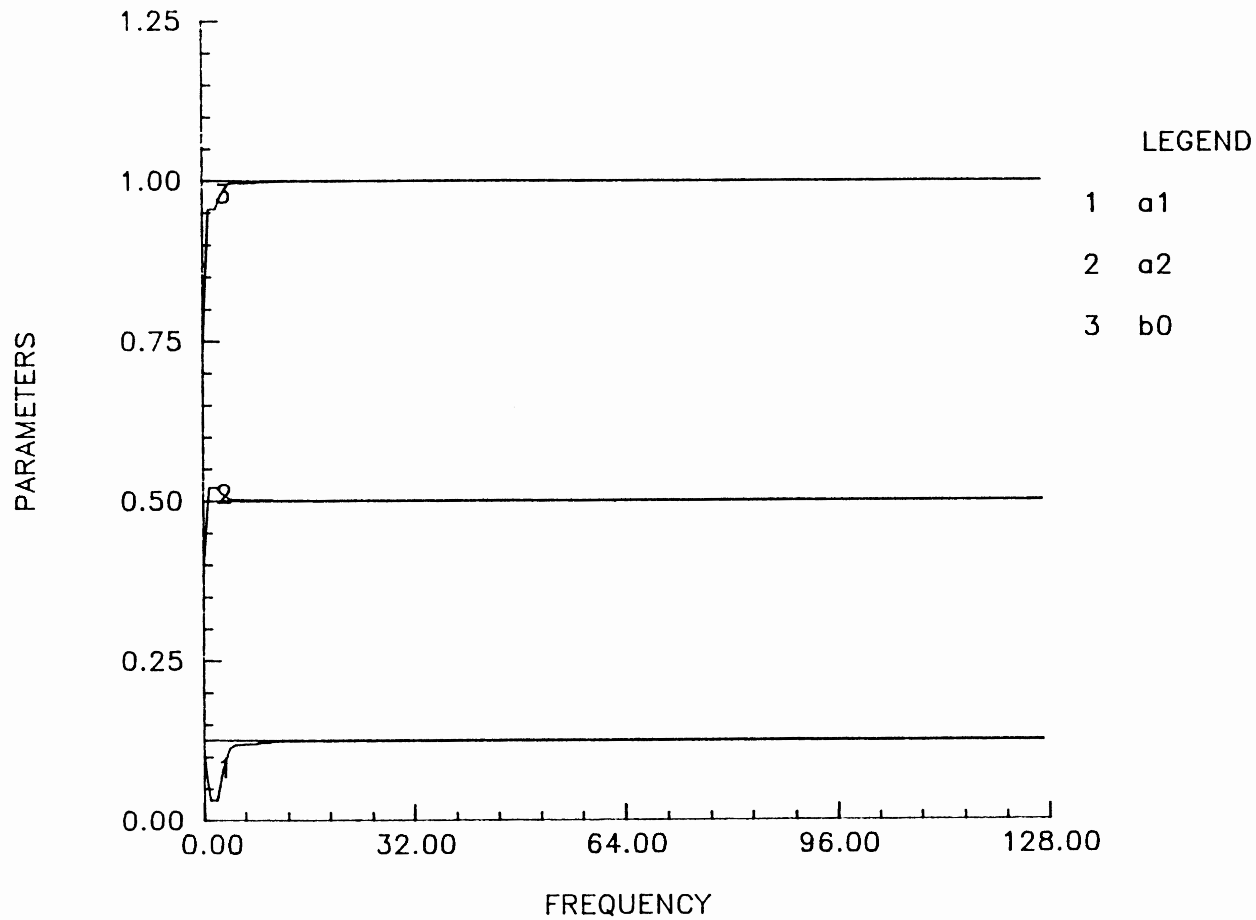

Figure 16. Parameter Estimates for System with no Delay, Square Wave Input , N=128

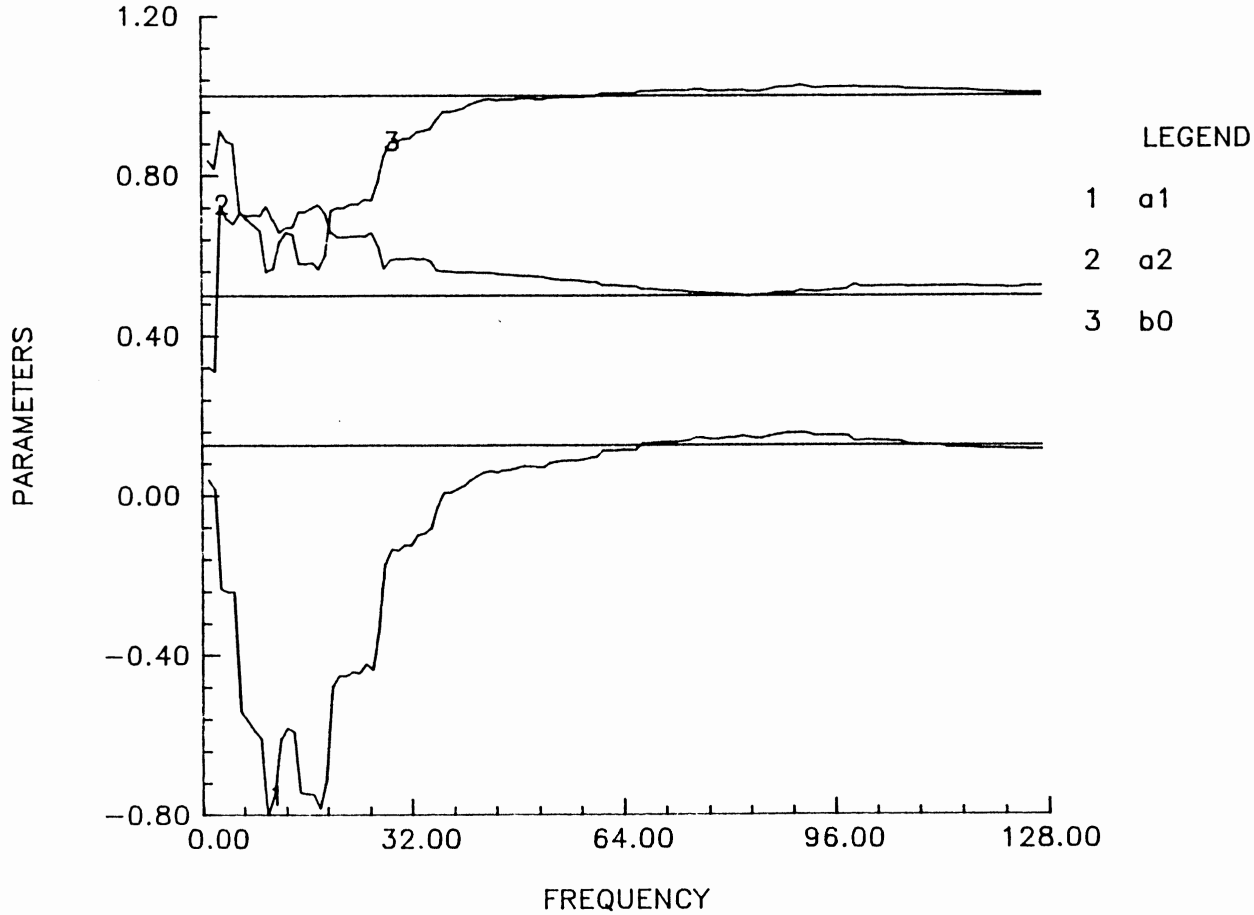

Figure 17. Parameter Estimates for System with no Delay, Square Wave Input and Noise, N=128

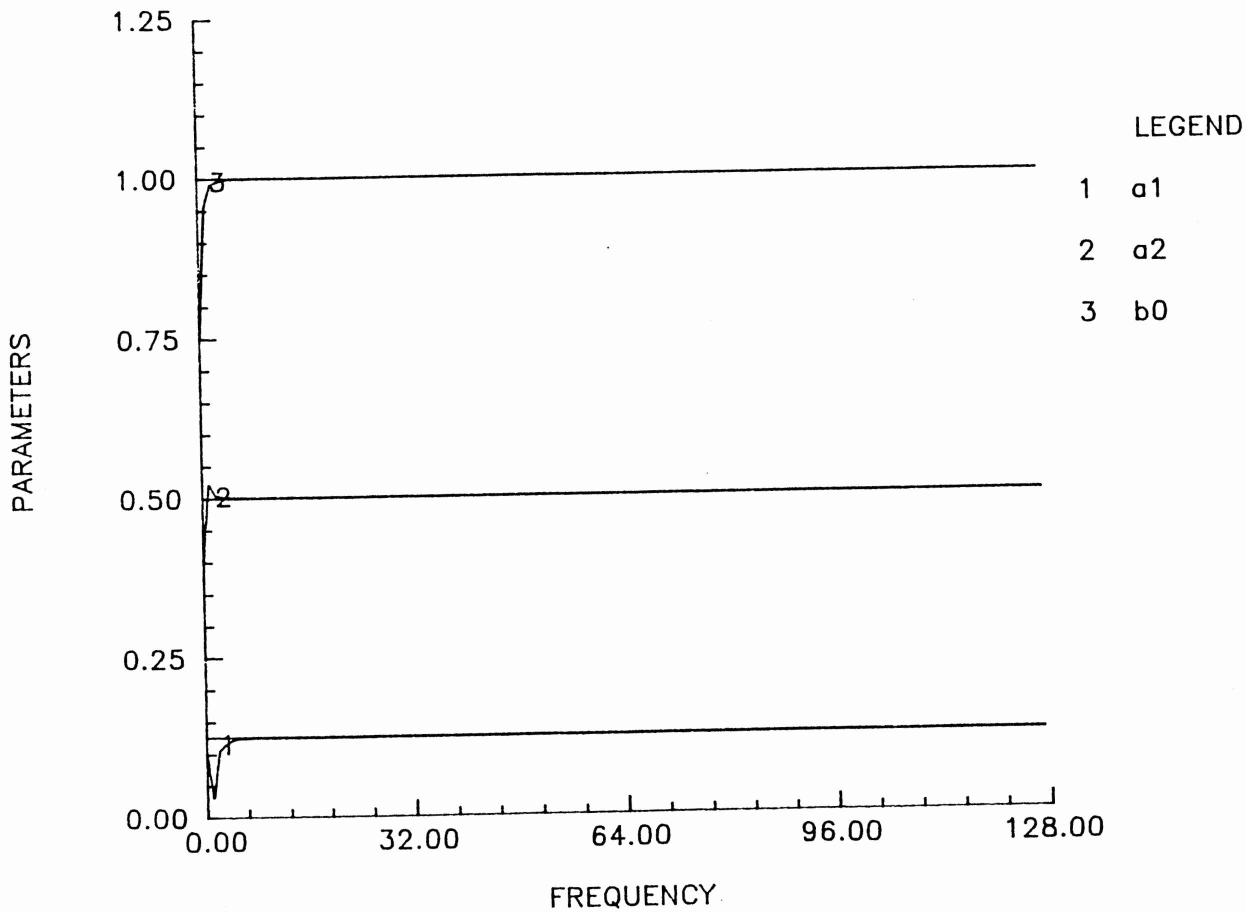

Figure 18. Parameter Estimates for System with no Delay, Sinusoidal Input, N=128

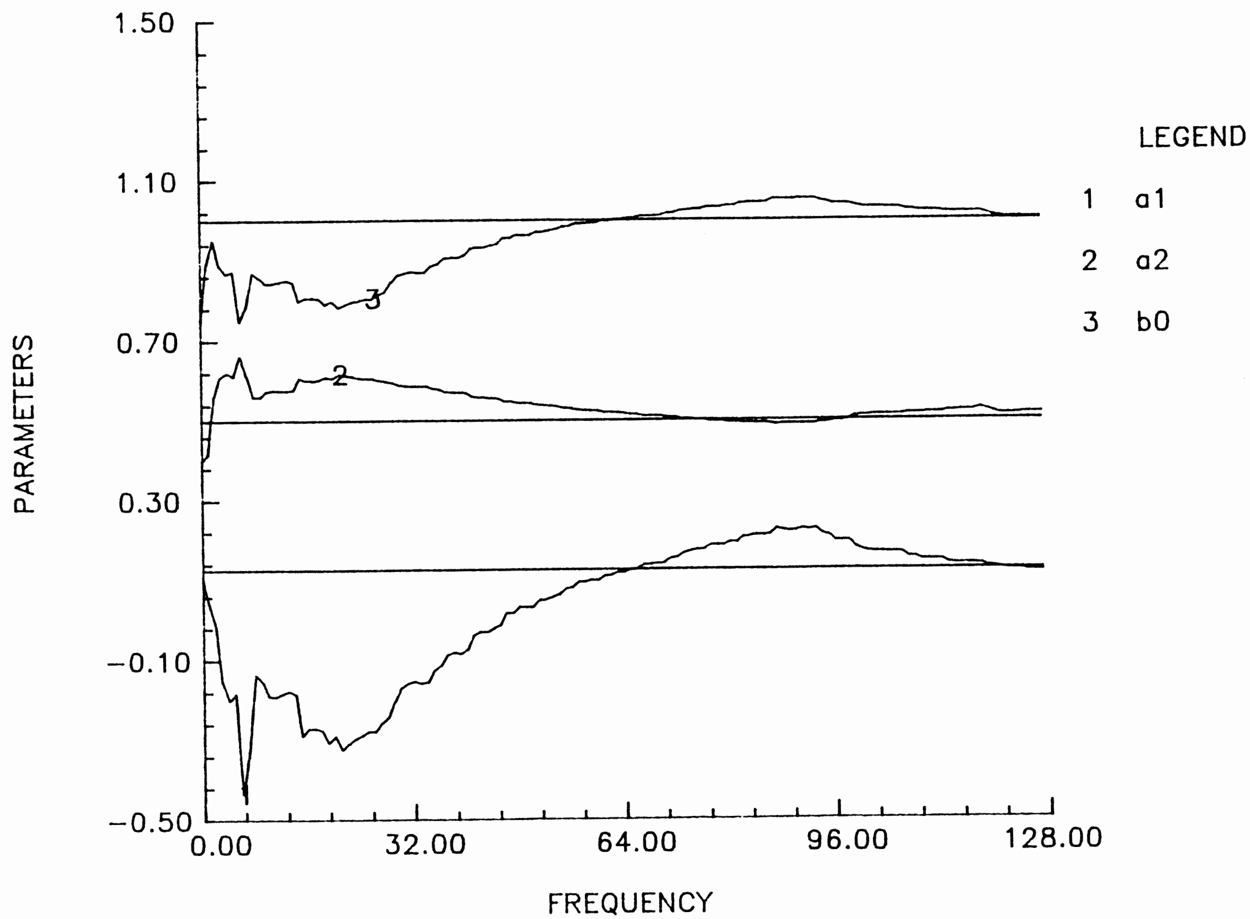

Figure 19. Parameter Estimates for System with no Delay, Sinusoidal Input and Noise, N=128

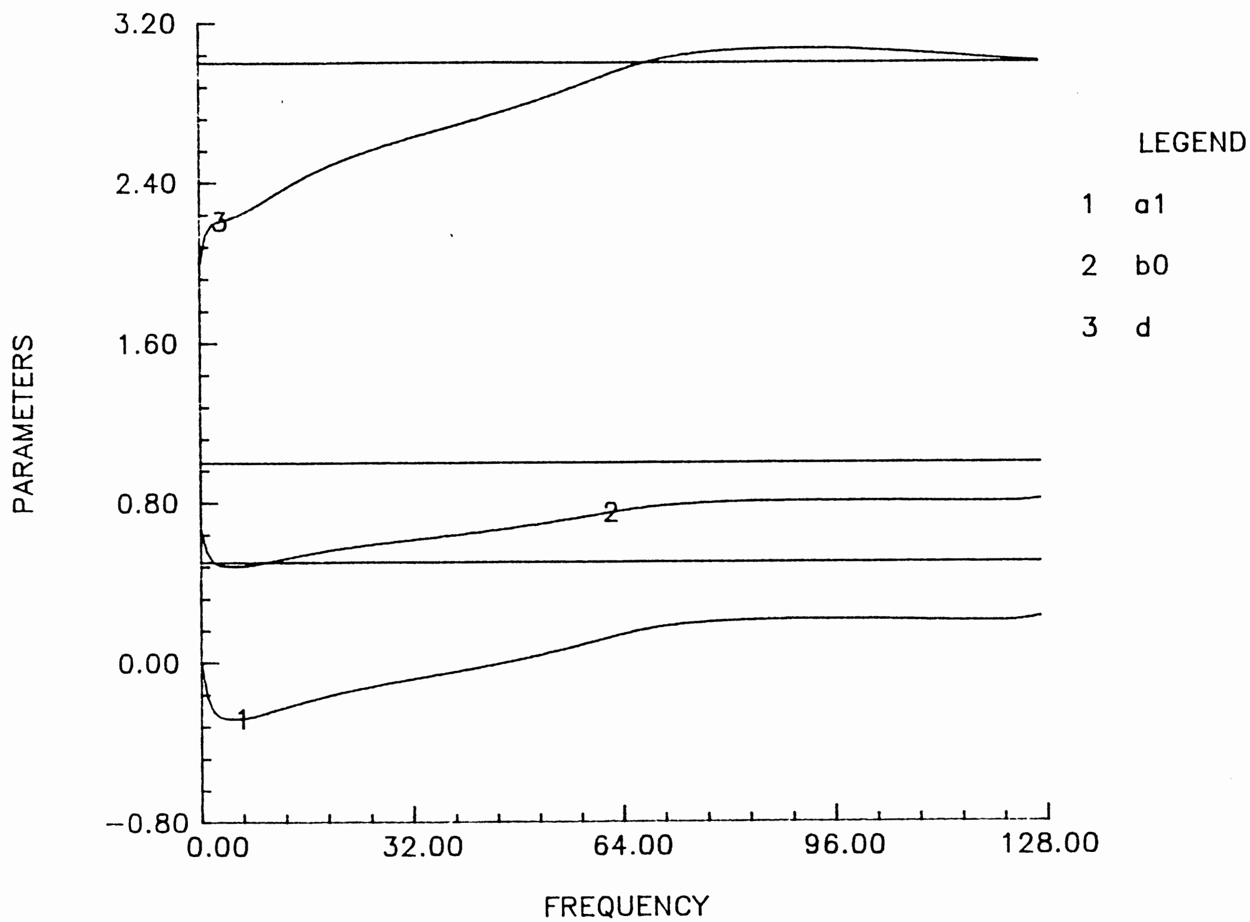

Figure 20. Parameter Estimates for System with Unknown Delay, Exponential Input, N=128

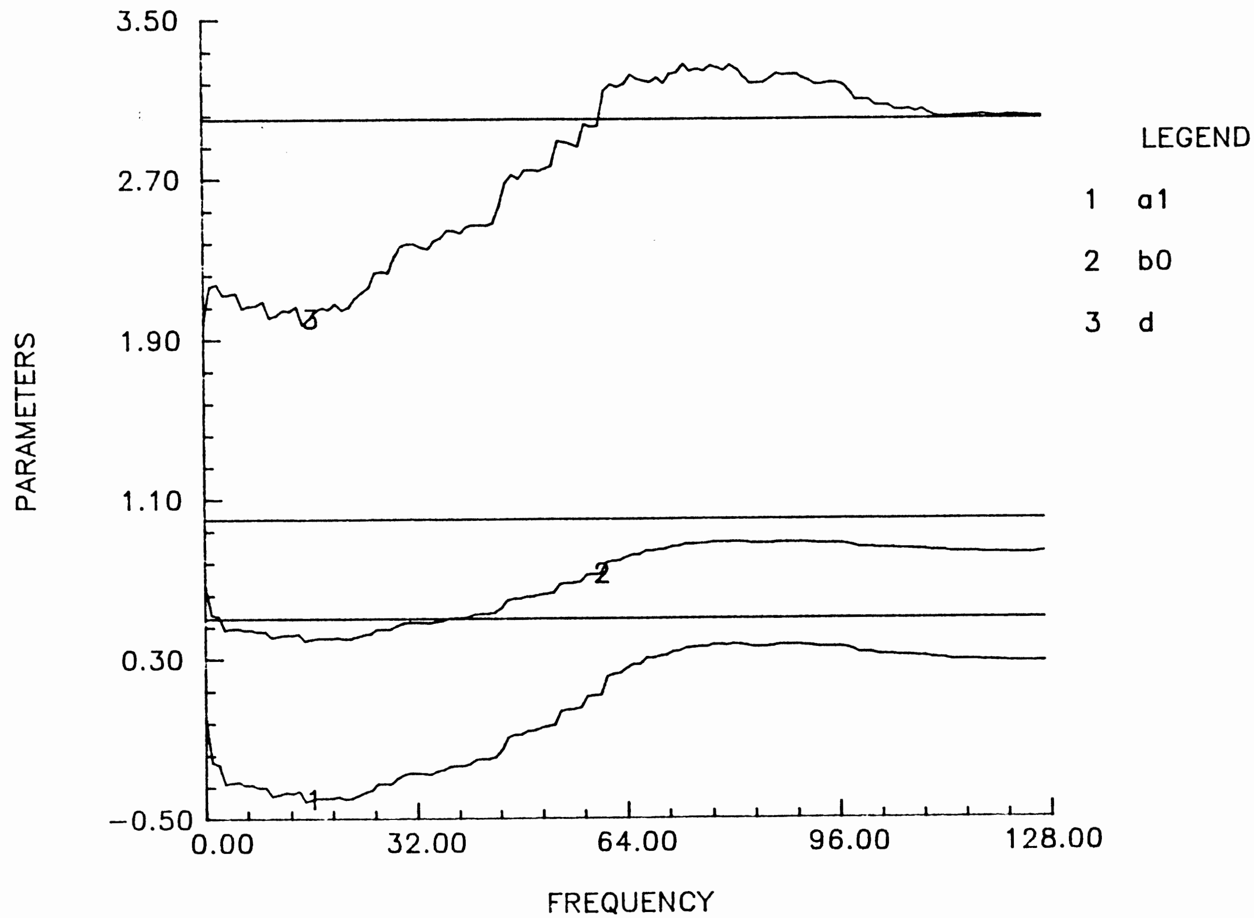

Figure 21. Parameter Estimates for System with Unknown Delay, Exponential Input and Noise, N=128

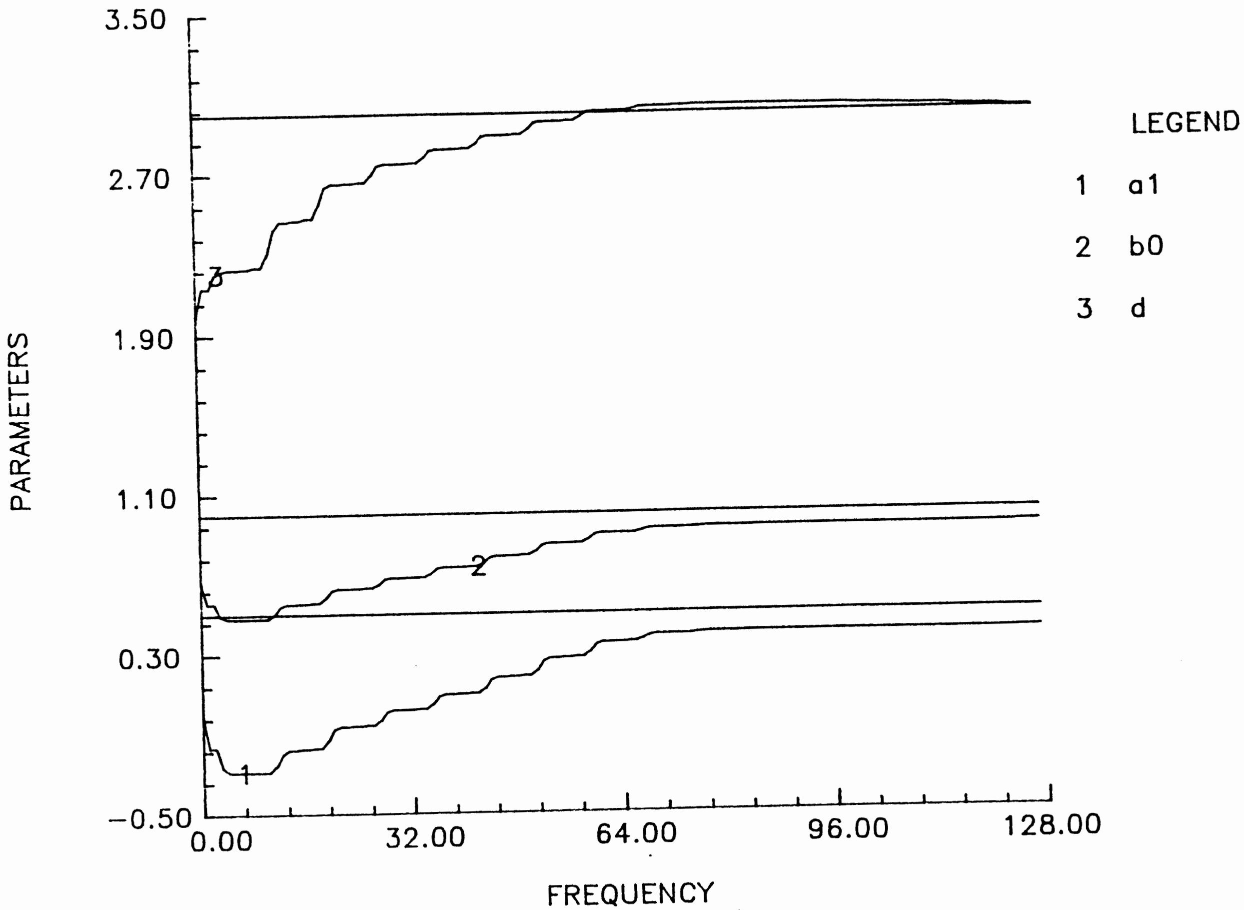

Figure 22. Parameter Estimates for System with Unknown Delay, Square Wave Input, N=128

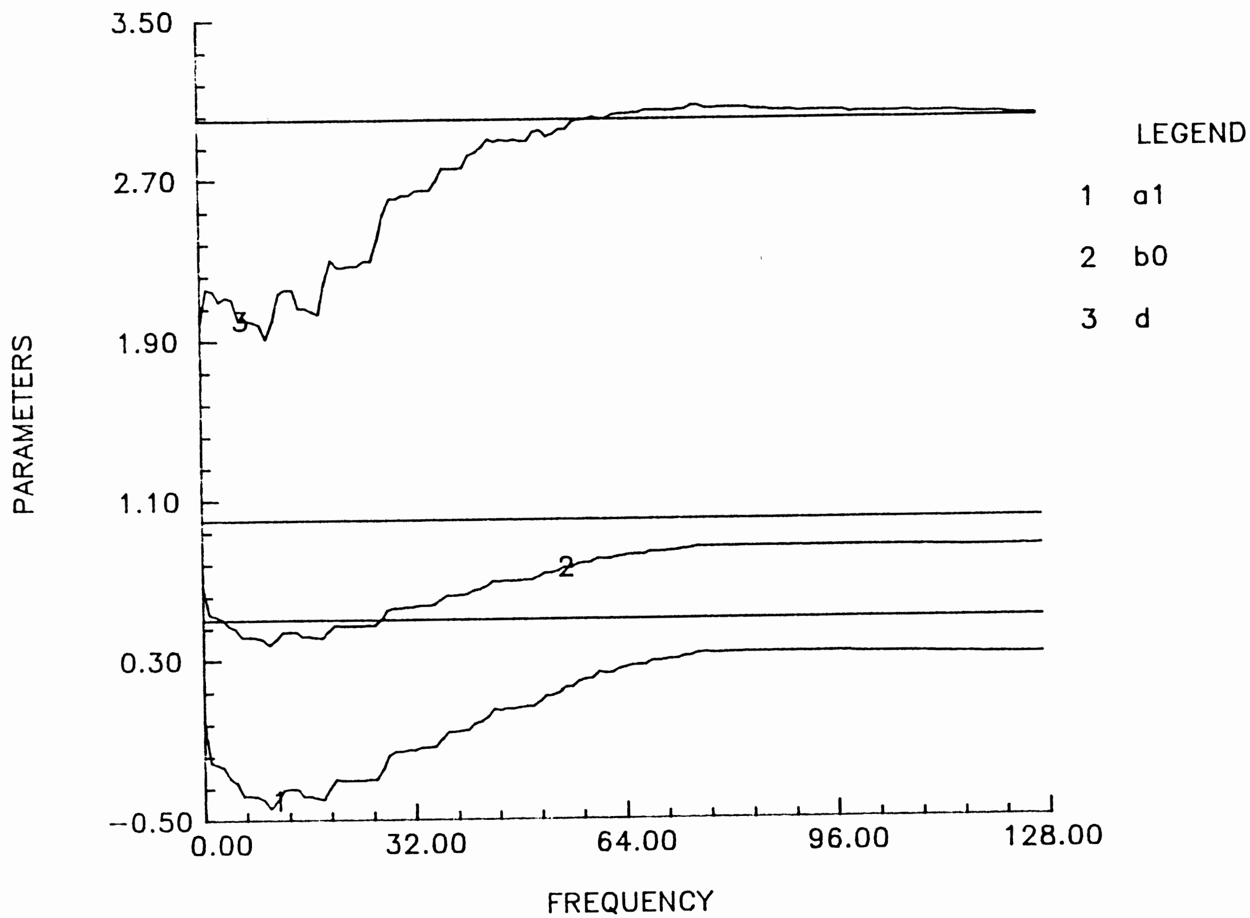

Figure 23. Parameter Estimates for System with Unknown Delay,
Square Wave Input, N=128

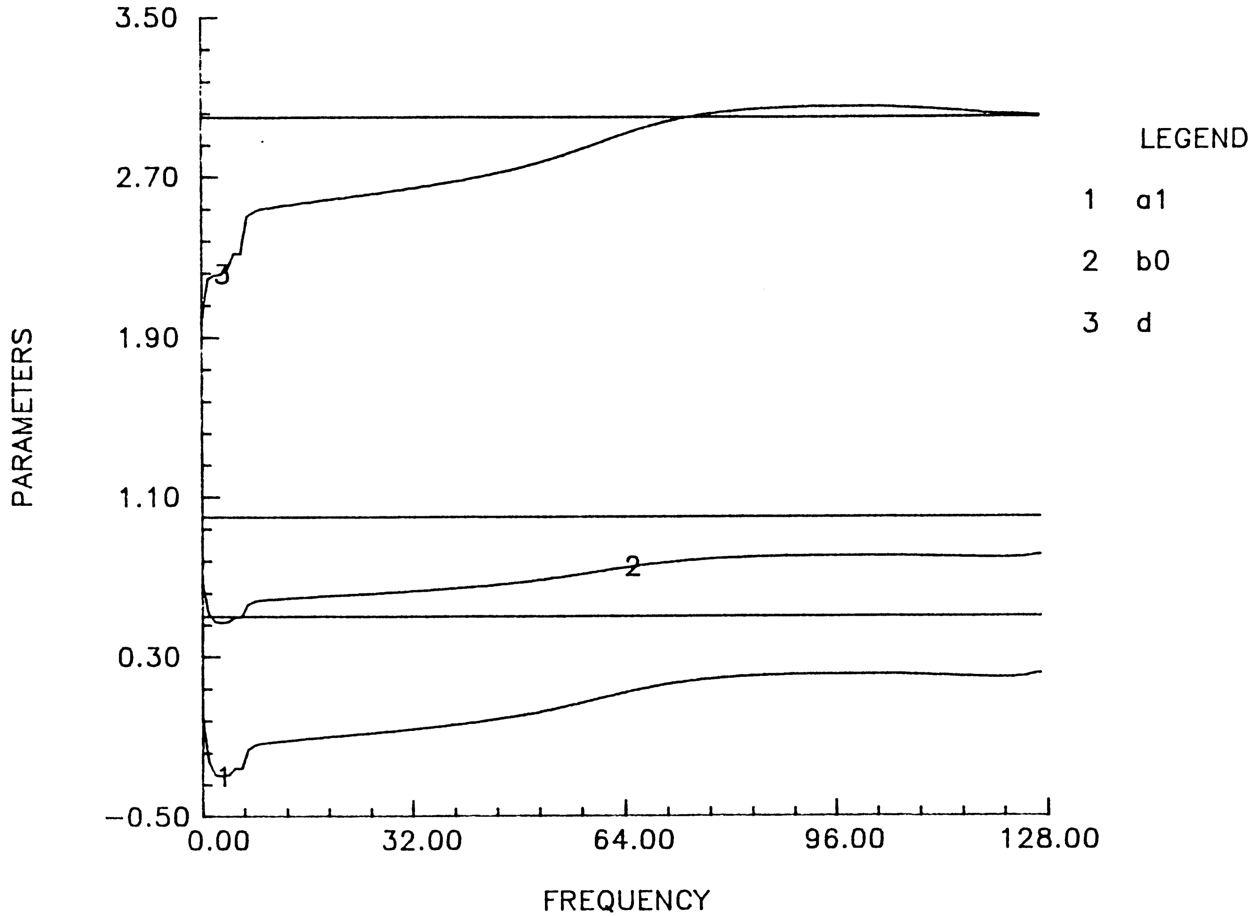

Figure 24. Parameter Estimates for System with Unknown Delay,
Sinusoidal Input, N=128

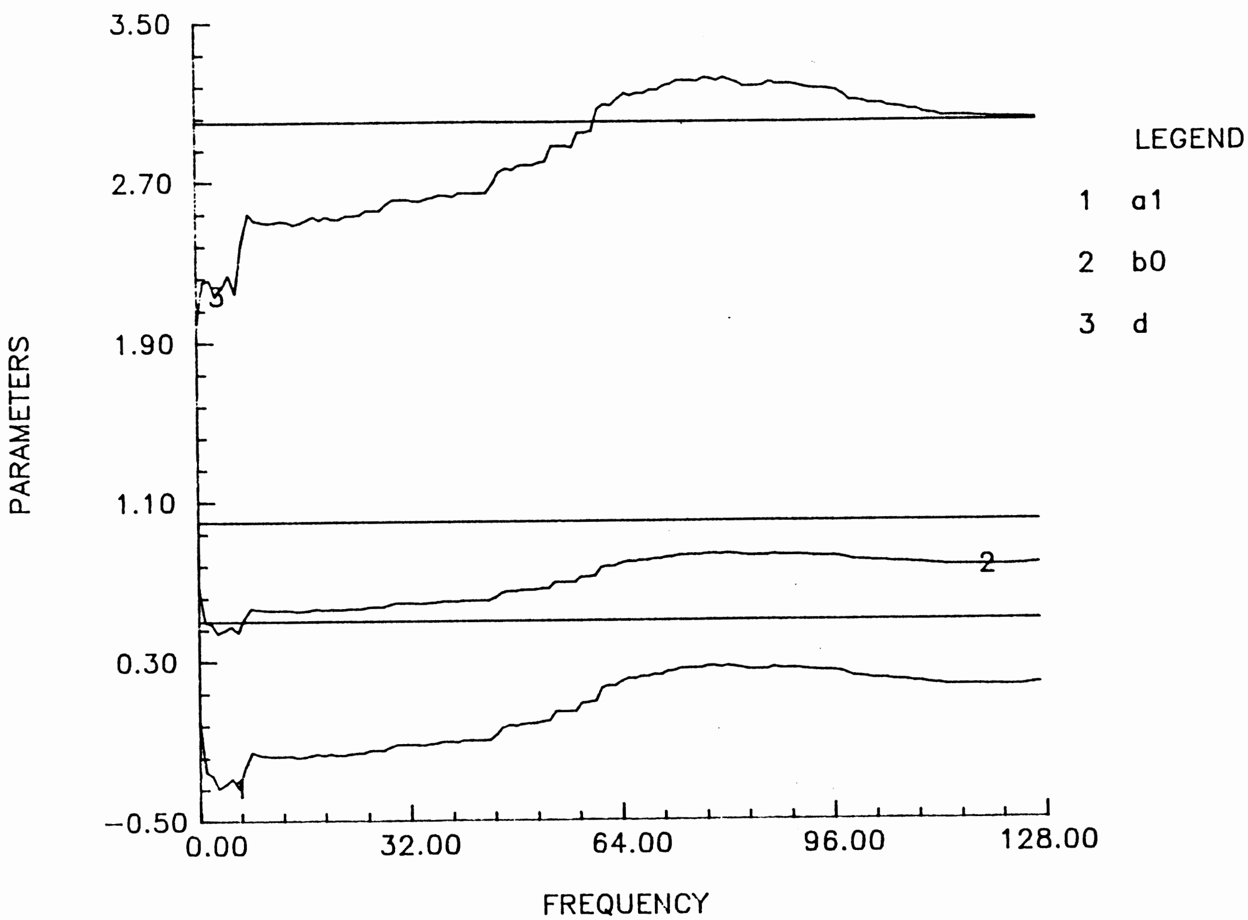

Figure 25. Parameter Estimates for System with Unknown Delay, Sinusoidal Input and Noise, N=128

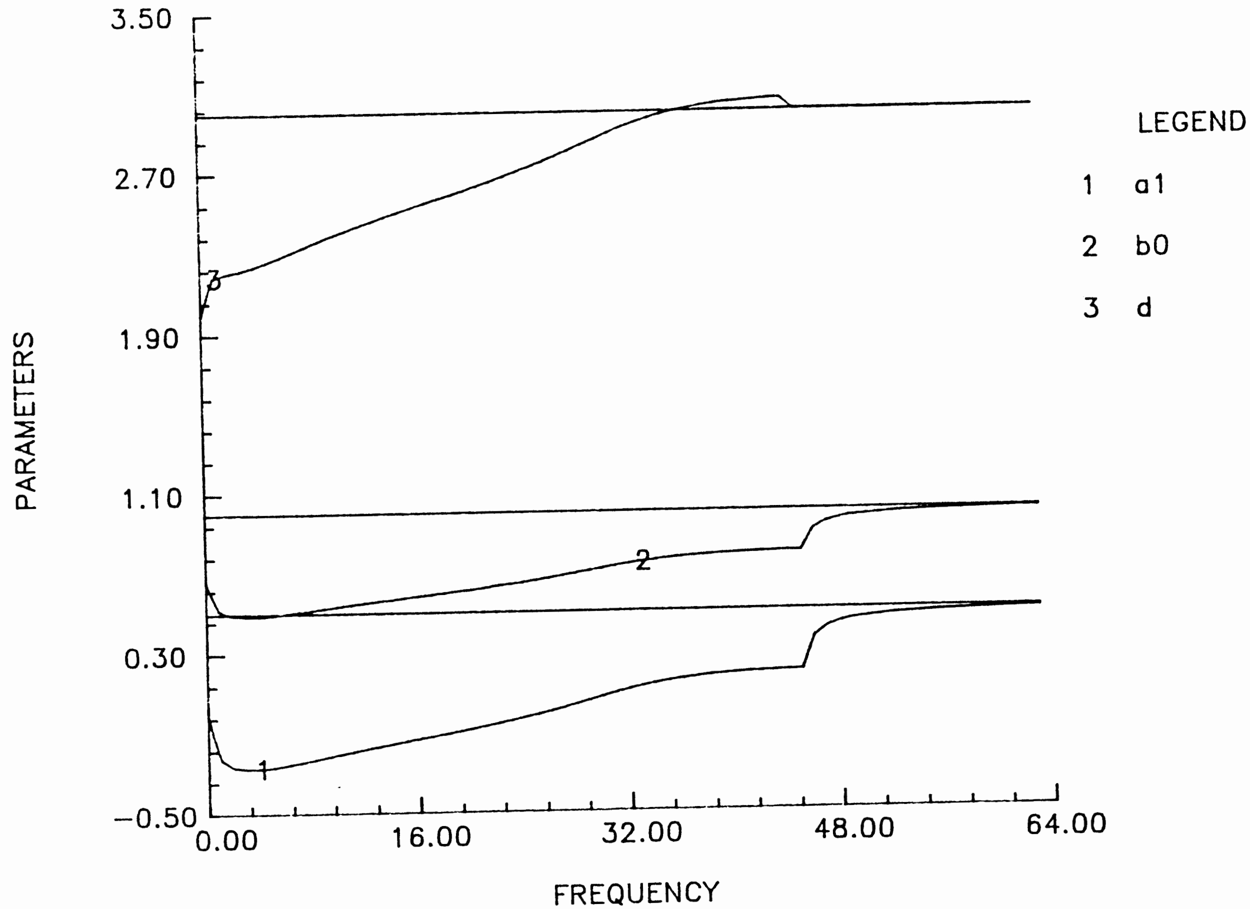

Figure 26. Parameter Estimates for System with Unknown Delay, Exponential Input, N=64 (Improved Algorithm)

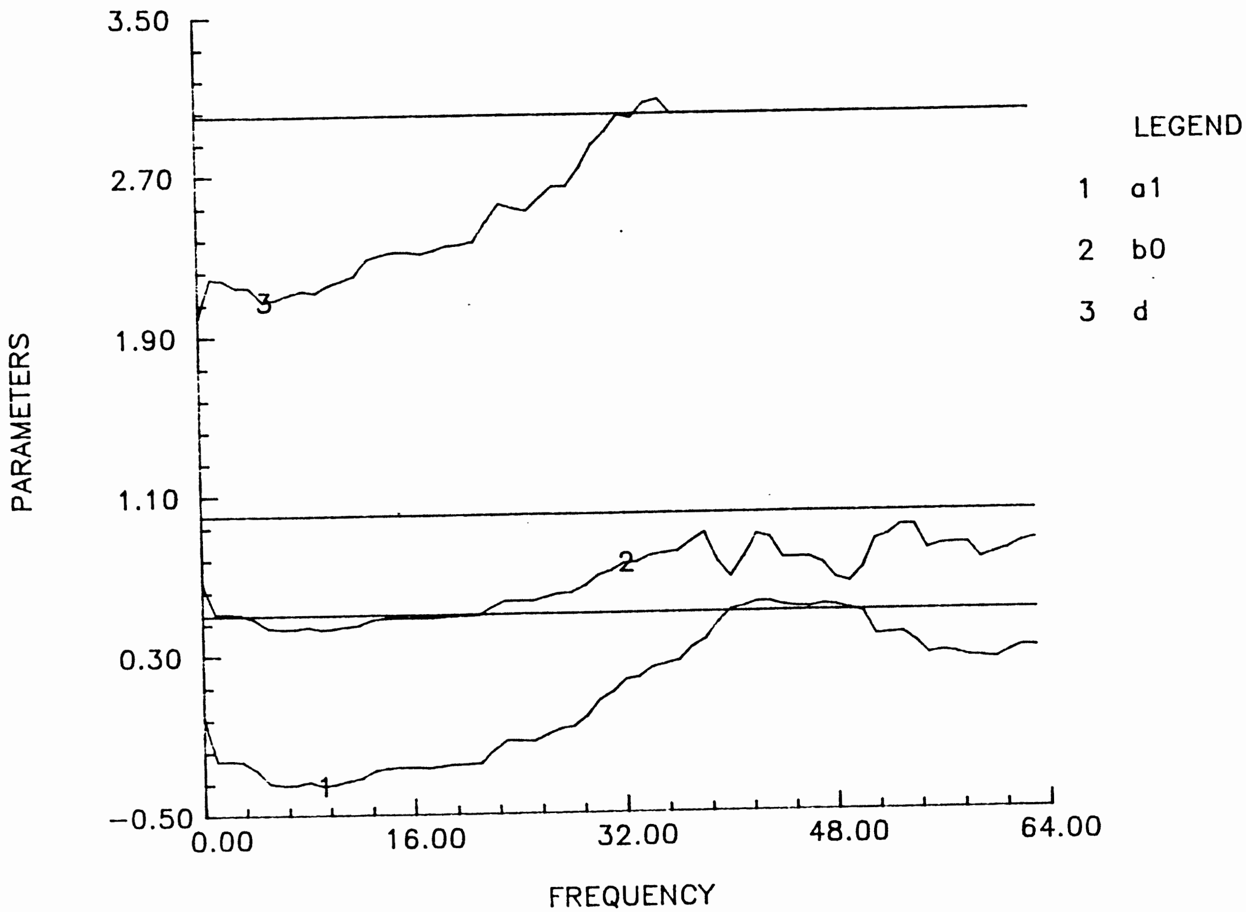

Figure 27. Parameter Estimates for System with Unknown Delay, Exponential Input and Noise, N=64 (Improved Algorithm)

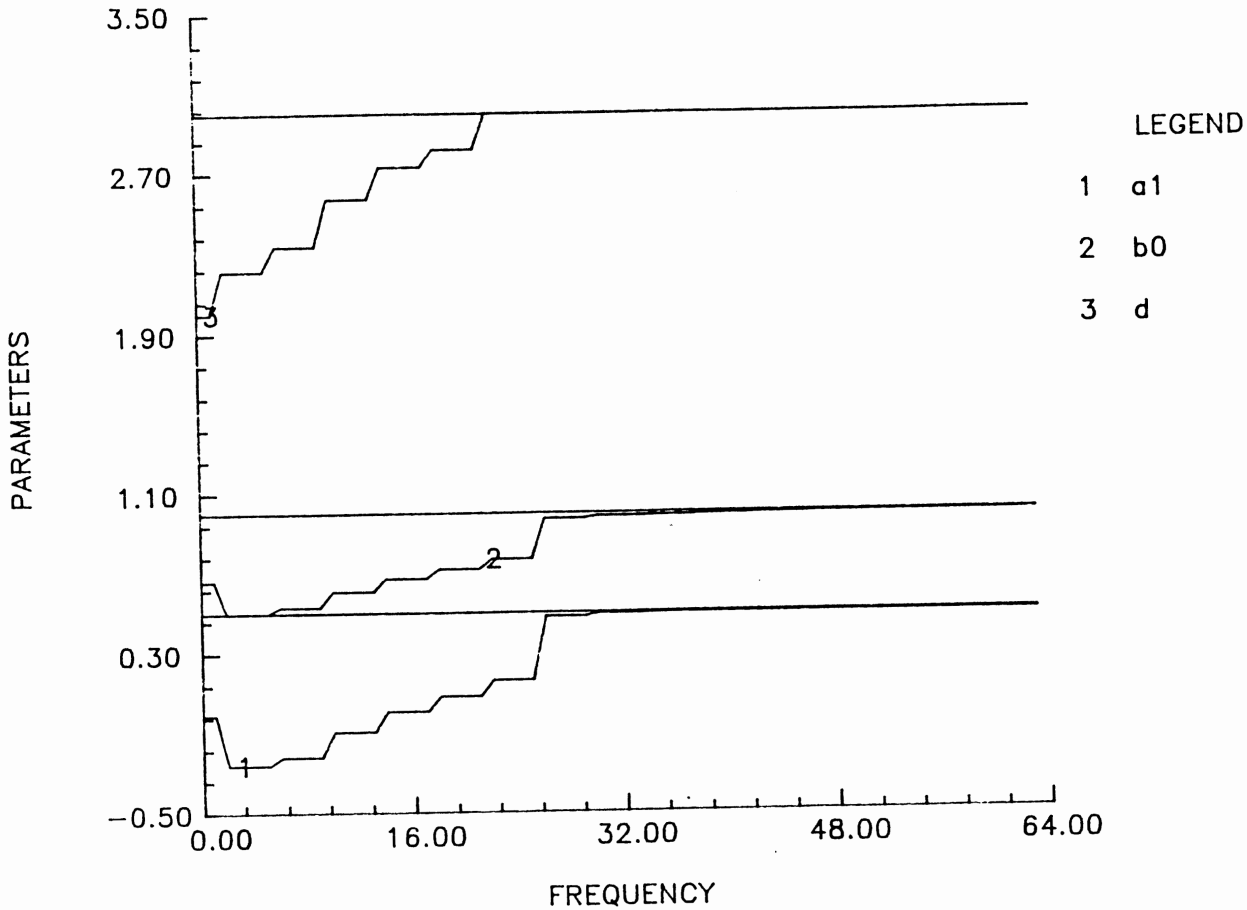

Figure 28. Parameter Estimates for System with Unknown Delay, Square Wave Input, N=64 (Improved Algorithm)

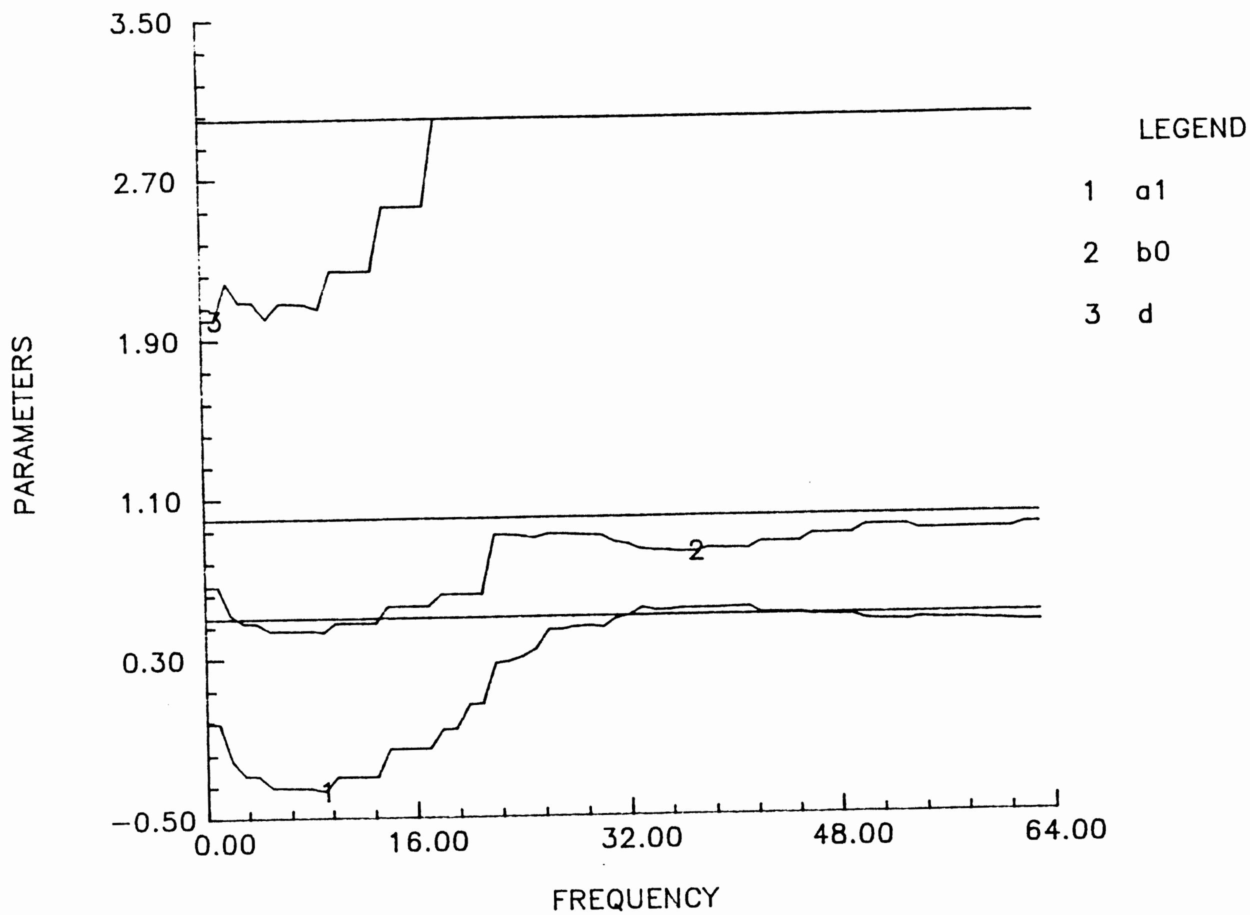

Figure 29. Parameter Estimates for System with Unknown Delay, Square Wave Input, N=64 (Improved Algorithm)

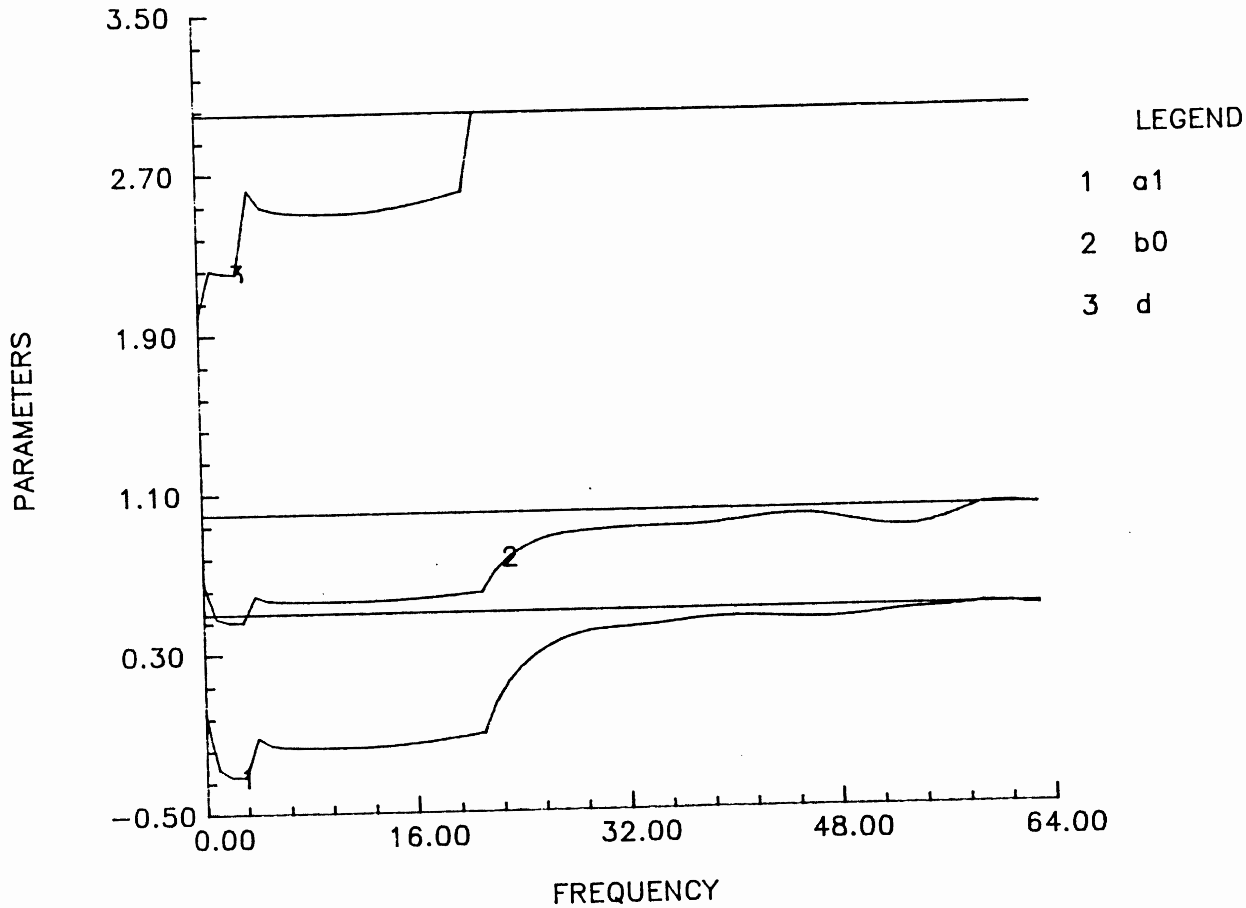

Figure 30. Parameter Estimates for System with Unknown Delay, Sinusoidal Input, N=64 (Improved Algorithm)

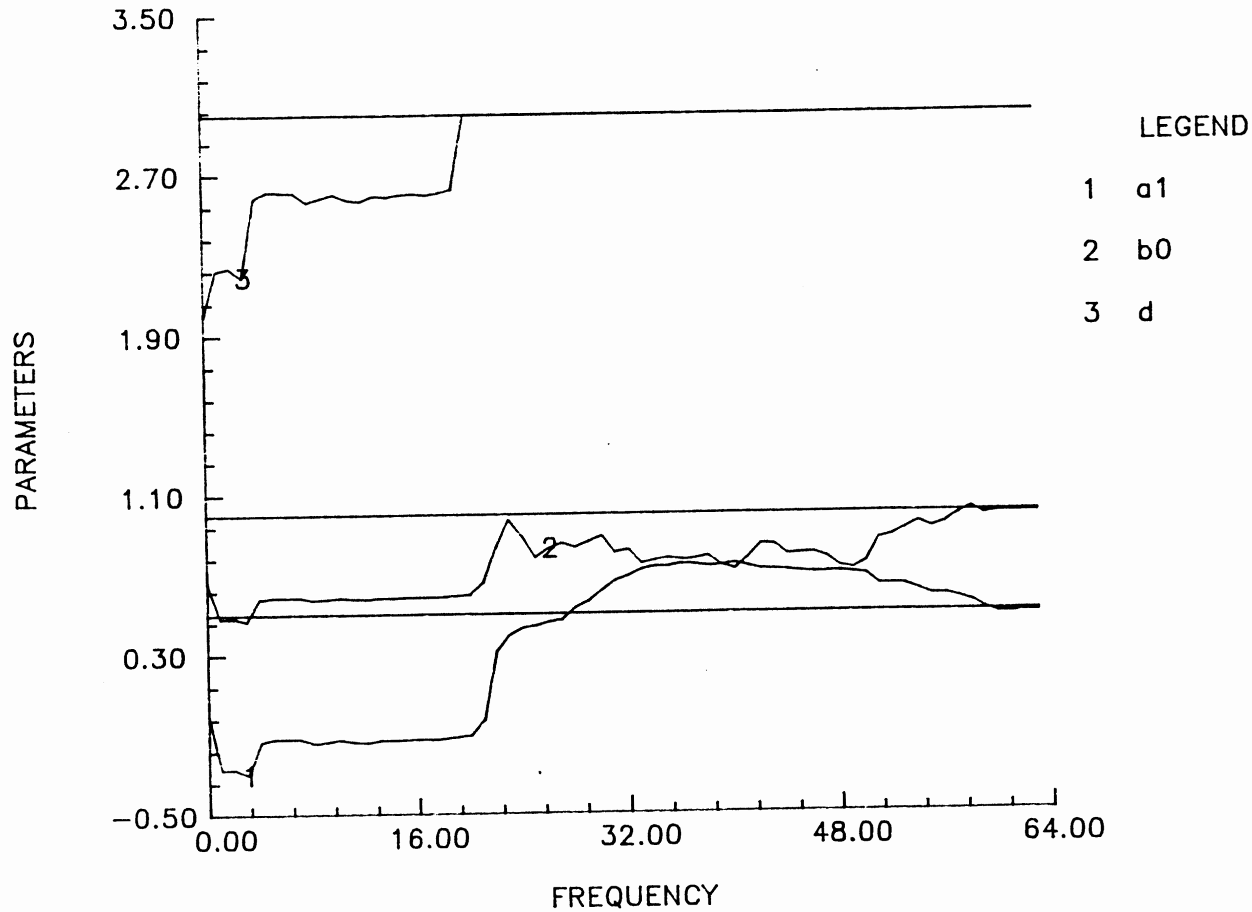

Figure 31. Parameter Estimates for System with Unknown Delay, Sinusoidal Input and Noise, N=64 (Improved Algorithm)

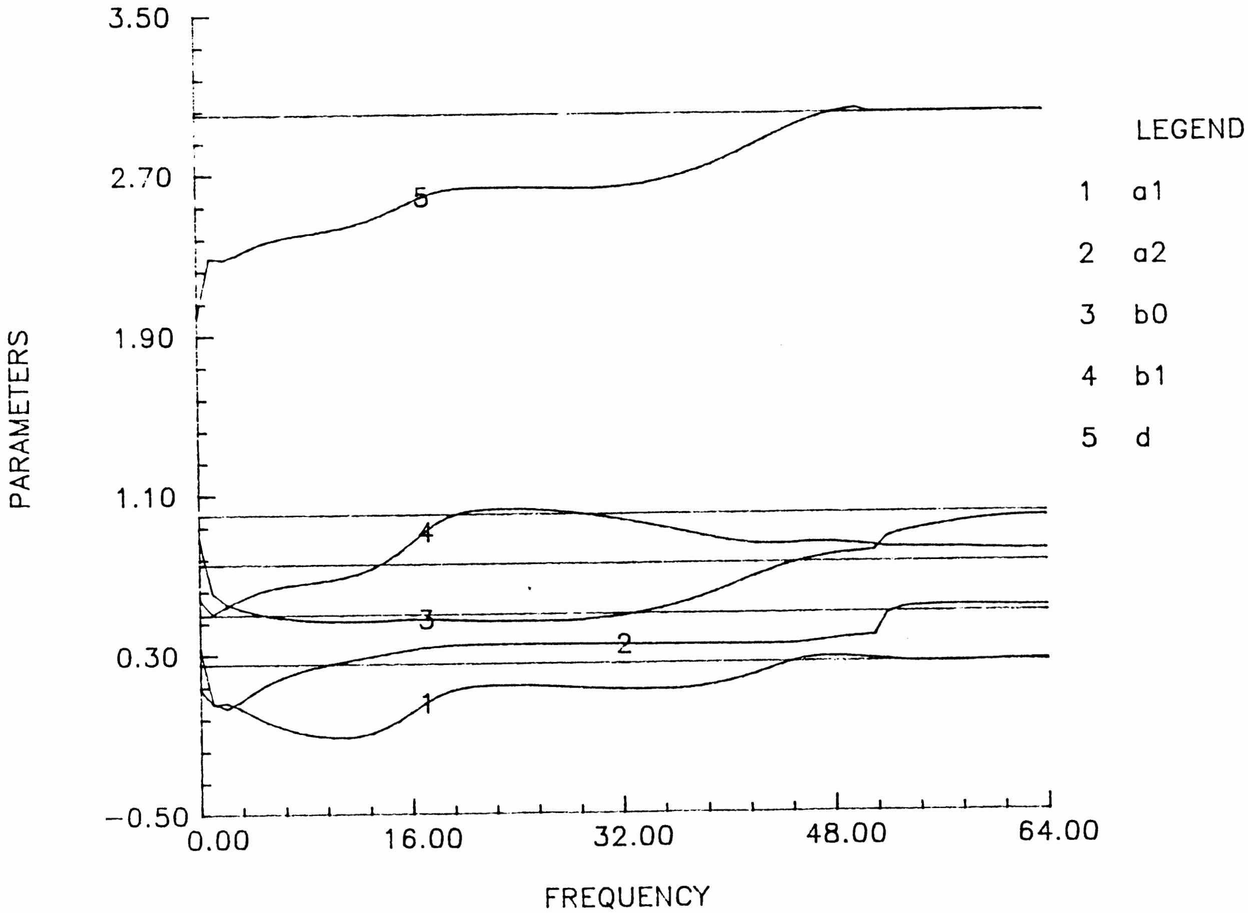

Figure 32. Parameter Estimates for Second Order System with Unknown Delay, Exponential Input, N=64 (Improved Algorithm)

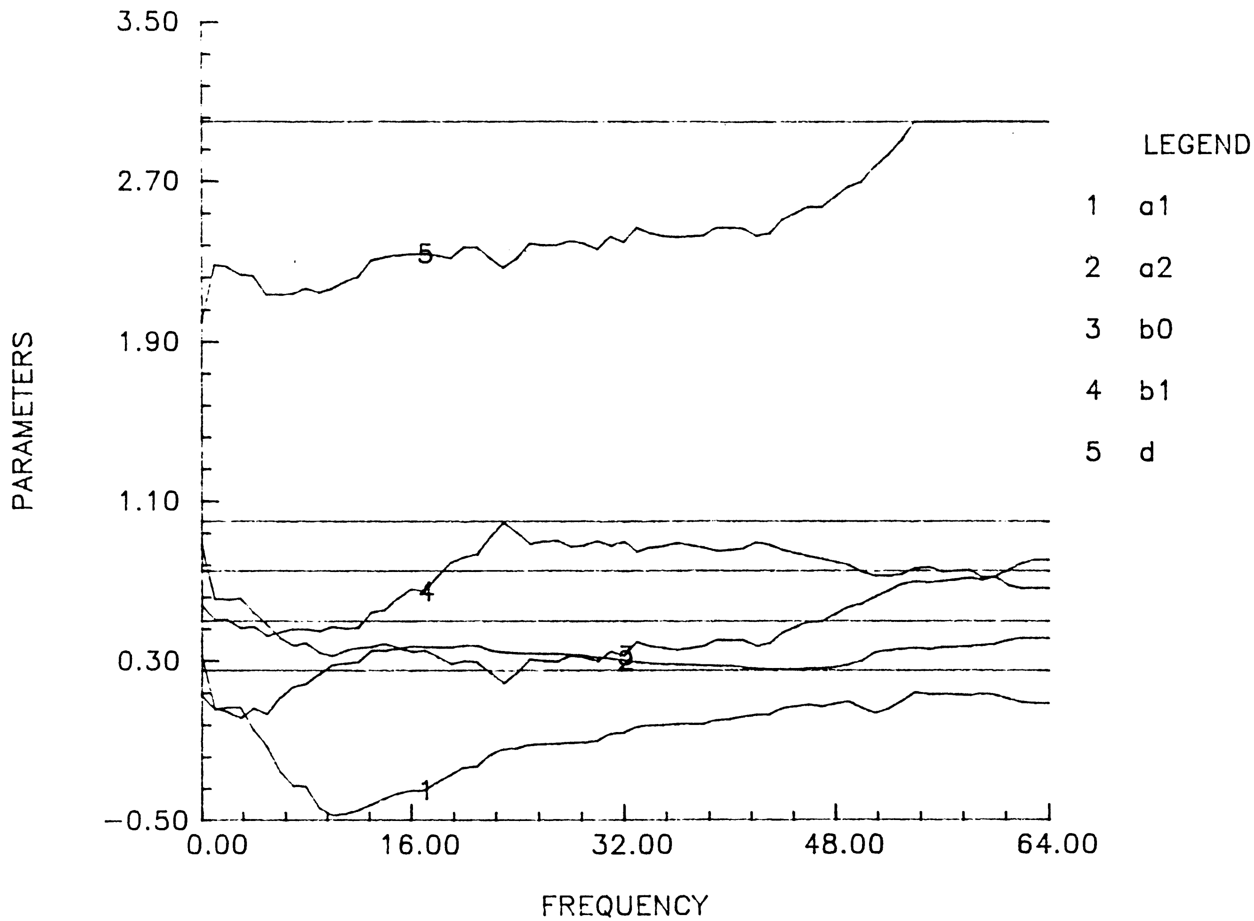

Figure 33. Parameter Estimates for Second Order System with Unknown Delay, Exponential Input and Noise, N=64 (Improved Algorithm)

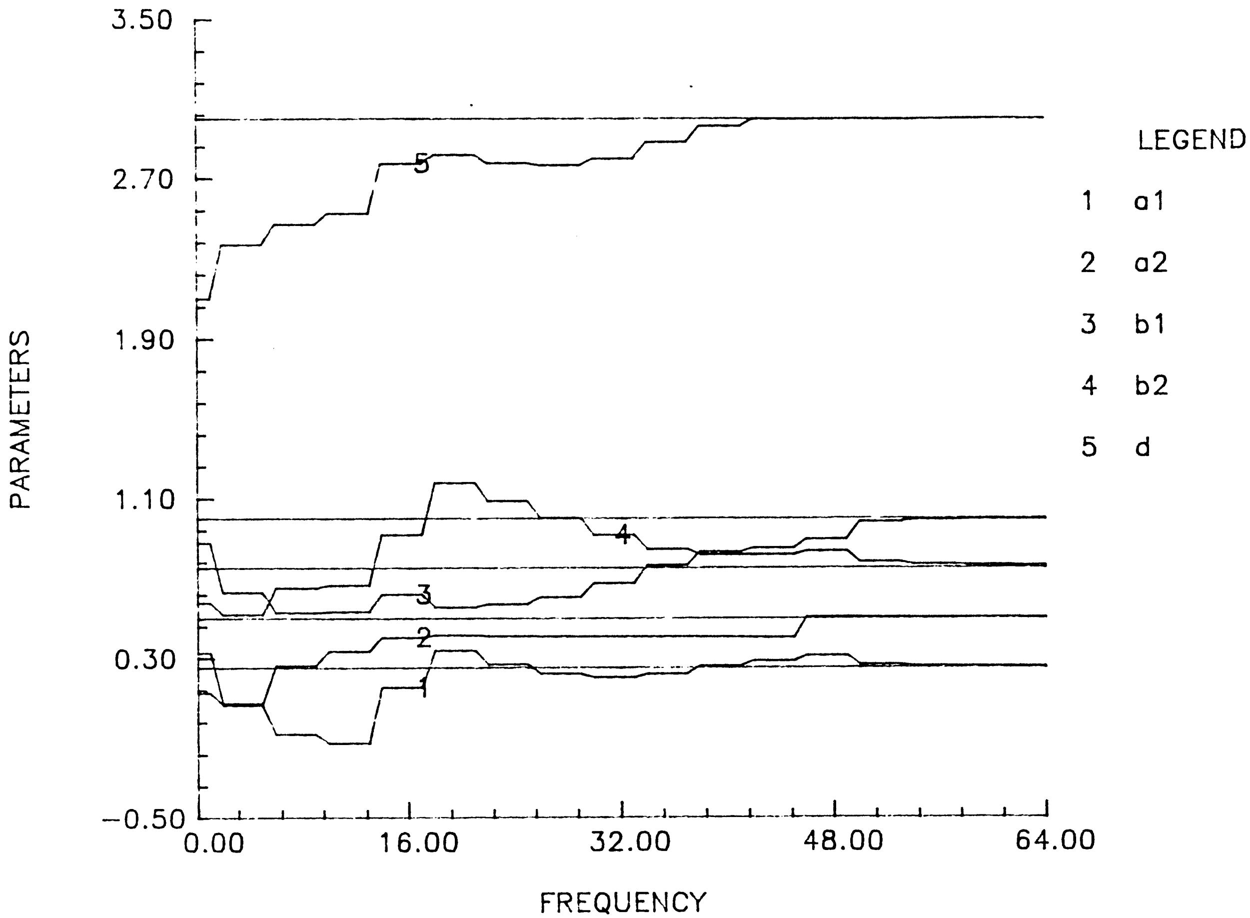

Figure 34. Parameter Estimates for Second Order System with Unknown Delay, Square Wave Input, N=64 (Improved Algorithm)

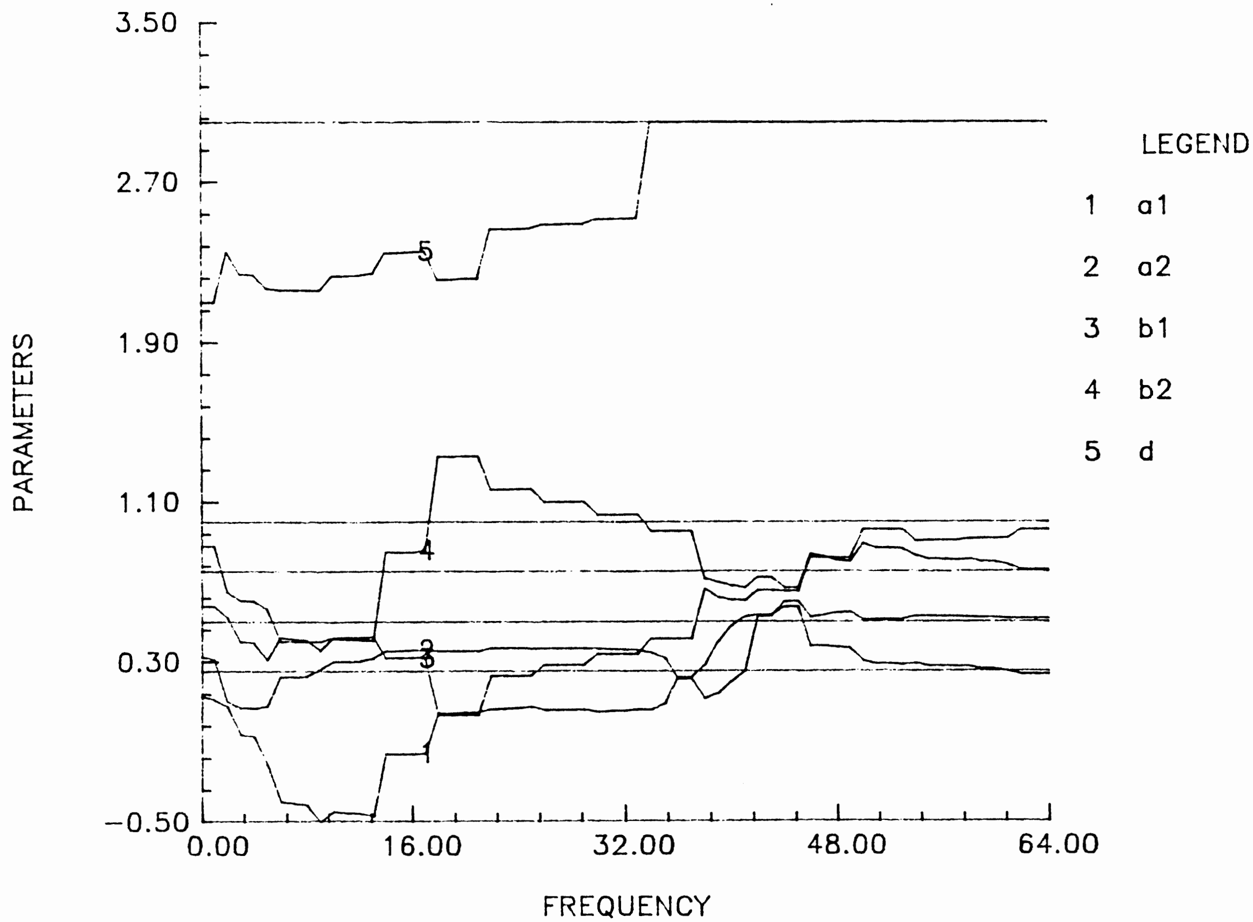

Figure 35. Parameter Estimates for Second Order System with Unknown Delay, Square Wave Input and Noise, N=64 (Improved Algorithm)

VITA

Suresh K. S. Rao

Candidate for the Degree of

Doctor of Philosophy

Thesis: RECURSIVE ESTIMATION OF DISCRETE TIME SYSTEM PARAMETERS AND TIME DELAY

Major Field: Mechanical Engineering

Biographical:

Personal Data: Born in Puttur, Karnataka, India, August 1, 1956, the son of Dr. and Mrs. K. M. Subraya.

Education: Graduated from St. Aloysius College, Mangalore, India, in July, 1972; received the Bachelor of Technology Degree in Mechanical Engineering from Karnataka Regional Engineering College in July, 1979; received the Master of Science Degree in Mechanical Engineering from Oklahoma State University in December, 1983; completed the requirements for the Doctor of Philosophy Degree at Oklahoma State University in December, 1988.

Professional Experience: Sales and Service Engineer, Godrej and Boyce Manufacturing Co., India, 1979-1980. Teaching and Research Associate, School of Mechanical and Aerospace Engineering, Oklahoma State University, 1981-1988; Instructor, School of Mechanical and Aerospace Engineering, Oklahoma State University, 1987.

Professional Societies: Tau Beta Pi (Engineering Honor Society); National Society of Professional Engineers.