

BIAS REMOVAL APPROACH IN SYSTEM IDENTIFICATION  
AND ARMA SPECTRAL ESTIMATION

By

MOHAMMAD VERNOSFADERANI MALAKOOTI

Bachelor of Science  
University of Tulsa  
Tulsa, Oklahoma  
1978

Master of Science  
Virginia Polytechnic Institute and State University  
Blacksburg, Virginia  
1980

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, 1987

Thesis  
1987D  
M236b  
cop. 2

BIAS REMOVAL APPROACH IN SYSTEM IDENTIFICATION  
AND ARMA SPECTRAL ESTIMATION

Thesis Approved:

*Keith A. Jensen*  
\_\_\_\_\_  
Thesis Adviser

*C.M. Bacon*  
\_\_\_\_\_

*Naw Yuhyaedden*  
\_\_\_\_\_

*Martin J. Kuen*  
\_\_\_\_\_

*Norman N. Deukham*  
\_\_\_\_\_  
Dean of the Graduate College

COPYRIGHT

by

Mohammad Vernosfaderani Malakooti

December, 1987

DEDICATION

This disseration is dedicated to  
the memory of my mother  
HAJIEH-GOHAR VERNOSFADERANI KABIRI  
who passed away and joined the Lord  
on April 1, 1987

## ACKNOWLEDGMENTS

In the name of Allah, the most beneficent, the most merciful. I thank the Lord whose worship makes me strong in faith, and whose obedience makes me closer to Him. It would have been impossible to complete this dissertation without the support of Allah, especially in the stressful situation of my mother's sudden death.

I would like to thank my major adviser, Dr. R. K. Yarlagadda, for accepting the responsibility of being my dissertation chairman. His suggestions and constructive criticism made my work satisfying and enjoyable. Thanks and appreciation is also given to the other members of my doctoral committee, Dr. C. M. Bacon, Dr. M. S. Keener, and Dr. K. A. Teague for their assistance, time, and effort spent throughout my doctoral program. I am especially appreciative to my thesis adviser, Dr. K. A. Teague, who aided me in conducting this research. I wish to extend my appreciation to him for the excellent comments, suggestions, close supervision, and support in the form of a graduate research assistantship. I cannot express the depth of gratitude felt for him.

I wish to express my gratitude to Dr. J. E. Baker, head of Oklahoma State University School of Electrical and Computer Engineering, for his encouragement and support in the form of a graduate teaching assistantship. I am indeed grateful and owe a word of appreciation to my inspired teacher and former adviser, Dr. J. A. Cadzow, who taught me various methods of signal processing while I was his graduate research assistant at Arizona State University, Department of Electrical

Engineering. I wish to thank my colleague, Mr. D. M. Smith for his useful and stimulating discussion. I am grateful to Ms. Barbara Caldwell for her tireless assistance in the preparation of this manuscript.

I would like to express my appreciation to the ministry of Science and Education of the Islamic Republic of Iran who facilitated the transfer of funds provided by my father. I am grateful to my family for their love, confidence, support, and continual encouragement which was provided as I worked to finish this dissertation. Special thanks is given to my father for his patience, sacrifice, and financial support provided during the entire course of my education. I wish him a long life, and ask the Lord to give him patience following my mother's death.

Finally, I seek direction from the Lord of mankind, the king of mankind, and the God of mankind for me to be constructive rather than destructive in the comfort of mankind. May Allah reward me and benefit my mother. May her spirit be happy.

## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION . . . . .	1
Motivation . . . . .	1
Overview . . . . .	6
II. CLASSICAL METHODS OF SYSTEM IDENTIFICATION. . . . .	9
Background . . . . .	9
Mathematical Models. . . . .	14
Least Squares Identification . . . . .	19
Generalized Least Squares Identification . . . . .	24
Whitening Filter and GLS Algorithm . . . . .	26
Instrumental Variable Method of Identification . . . . .	30
III. ARMA SYSTEM IDENTIFICATION . . . . .	33
Introduction . . . . .	33
Modified Equation Error Identification . . . . .	34
Singular Value Decomposition . . . . .	43
Stability of Cross Product of $A^*A$ . . . . .	46
SVD Identities . . . . .	48
Data Matrix Based Identification . . . . .	51
Minimum Error Solution: Data Matrix Approach . . . . .	56
Correlation Matrix Based Identification. . . . .	61
Minimum Norm Solution: Correlation Matrix Approach . . . . .	65
Bias Removal in Combined Correlation Method. . . . .	68
Optimum Dimension of the Correlation Matrix. . . . .	73
Determination of the AR Model Order. . . . .	75
Numerical Example. . . . .	77
IV. CARMA MODEL METHOD OF TWO-DIMENSIONAL SHAPE CLASSIFICATION: AN EIGENSYSTEM APPROACH VS THE $L_p$ NORM . . . . .	81
Motivation . . . . .	81
Introduction . . . . .	82
Mathematical Formulation . . . . .	83
Results and Conclusions. . . . .	88



Chapter	Page
V. ESTIMATION OF THE VOCAL TRACT PARAMETERS FROM ARMA MODEL: AN EIGENSYSTEM APPROACH VS LPC . . . . .	89
Motivation . . . . .	89
Introduction . . . . .	90
Pre-processing of the Speech Samples . . . . .	92
Mathematical Formulation . . . . .	93
Forward-Backward Prediction of Speech. . . . .	95
Results and Conclusions. . . . .	98
VI. SPEAKER-INDEPENDENT WORD RECOGNITION: AN LPC FEATURE EXTRACTION APPROACH VS NON-LINEAR SPECTRAL MATCHING. . . .	100
Motivation . . . . .	100
Introduction . . . . .	101
Pre-processing of the Isolated Word. . . . .	103
Endpoint Detection. . . . .	104
Feature Extraction Using LPC . . . . .	105
Distance Measure Calculation . . . . .	106
Dynamic Time Warping . . . . .	107
Nonlinear Spectral Matching . . . . .	108
Results and Conclusions. . . . .	110
VII. ESTIMATING THE FREQUENCIES OF MULTIPLE SINUSOIDS IN WHITE NOISE: A CARMA MODEL APPROACH VS LINEAR PREDICTION . . . . .	115
Motivation . . . . .	115
Introduction . . . . .	116
Mathematical Model . . . . .	118
Results and Conclusions. . . . .	121
VIII. SUMMARY AND CONCLUSIONS. . . . .	123
Suggestions for Future Research. . . . .	126
REFERENCES. . . . .	132
APPENDIXES. . . . .	139
APPENDIX A. System Transfer Function. . . . .	140
APPENDIX B. M-Transform Algorithm . . . . .	146

## LIST OF TABLES

Table	Page
I. Comparison of LS and GLS for ARMA(2,2) . . . . .	30
II. Comparison of EEI and Modified EEI for a MA process . . . . .	42
III. Comparison of Scaling and Noise Removal Technique, SNR Output = 3 dB SNR Input = 6 dB . . . . .	72
IV. Comparison of the Resolution of Correlation Matrices and Combined Correlation Matrix. . . . .	75
V. The Length of 5 Different Words Spoken by Speaker No. 1 (Male) . . . . .	111
VI. The Length of 5 Different Words Spoken by Speaker No. 2 (Male) . . . . .	111
VII. The Length of 5 Different Words Spoken by Speaker No. 3 (Male) . . . . .	112
VIII. The Length of 5 Different Words Spoken by Speaker No. 4 (Female) . . . . .	112
IX. The Length of 5 Different Words Spoken by Speaker No. 5 (Female) . . . . .	113
X. Speaker No. 2 (Male): Word Recognition Failure of 1. LPC Using Euclidean Norm 2. ARMA Using New Distance Measure Criterion. . . . .	113
XI. Speaker No. 4 (Female): Word Recognition Failure of LPC Using Euclidean Norm . . . . .	114
XII. Speaker No. 5 (Female): Word Recognition Failure of LPC Using Euclidean Norm . . . . .	114

## LIST OF FIGURES

Figure	Page
1. Block Diagram Representation of Equation Error Identification . . . . .	3
2. Block Diagram Representation of Generalized Least Squares. . . . .	25
3. Block Diagram Representation of Whitening Filter . . . . .	26
4. Comparison of Equation Error and Modified Equation Error Identification With True System; p=q=2, SNR Input = SNR Output = 6 dB. . . . .	161
5. Magnitude of 10 Transfer Functions of the Least Squares Identification; p=q=5, SNR Input=SNR Output = 6 dB . . . . .	162
6. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Least Squares Identification; p=q=5, SNR Input = SNR Output = 6 dB . . . . .	163
7. Magnitude of 10 Transfer Functions of the Instrumental Variable Identification; p=q=5, SNR Input = SNR Output = 6 dB . . . . .	164
8. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Instrumental Variable Identification; p=q=5, SNR Input = SNR Output = 6 dB . . . . .	165
9. Magnitude of 10 Transfer Functions of the Generalized Least Squares Identification; p=q=5, SNR Input = SNR Output = 6 dB . . . . .	166
10. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Generalized Least Squares Identification; p=q=5, SNR Input = SNR Output = 6 dB . . . . .	167
11. Magnitude of 10 Transfer Functions of the Correlation Based Identification With Scaling and No Noise Subtraction; p=q=5, $T_1 = T_2 = 29$ SNR Input = SNR Output = 6 dB . . .	168

Figure	Page
12. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Correlation Based Identification With Scaling and No Noise Subtraction; p=q=5, $T_1 = T_2 = 29$ , SNR Input = SNR Output = 6 dB . . .	169
13. Comparison of the LS, GLS, IV, and EIGSP Solutions With the True System; SNR Input = SNR Output = 6 dB. . . . .	170
14. Correlation Based Identification Via SVD With Scaling and Noise Subtraction; p=q=5, $T_1 = T_2 = 29$ , SNR Input = SNR Output = 6 dB . . .	171
15. Correlation Based Identification Via SVD With Scaling and No Noise Subtraction; p=q=5, $T_1 = T_2 = 29$ , SNR Input = SNR Output = 6 dB . . .	172
16. Correlation Based Identification Via SVD With Scaling and Noise Subtraction; p=q=5, $T_1 = 8 T_2 = 56$ , SNR Input = SNR Output = 6 dB . .	173
17. Correlation Based Identification Via SVD With Scaling and No Noise Subtraction; p=q=5, $T_1 = 8, T_2 = 56$ , SNR Input = SNR Output = 6 dB. .	174
18. Correlation Based Identification Via SVD With Scaling and Noise Subtraction; p=q=5, $T_1 = 6, T_2 = 28$ , SNR Input = SNR Output = 6 dB. .	175
19. Correlation Based Identification Via SVD With Scaling and No Noise Subtraction; p=q=5, $T_1 = 6, T_2 = 28$ , SNR Input = SNR Output = 6 dB. .	176
20. Power Spectrum of AR Vs CAR for a Two Dimensional Shape; Top(AR), Bottom(CAR), p=12, q=0, $\bar{p}=4, \bar{q}=0, N=128, L=64$ . . . . .	177
21. Power Spectrum of ARMA Vs CARMA for a Two Dimensional Shape; Top(ARMA), Bottom(CARMA), p=q=12, $\bar{p}=\bar{q}=4, N=128, L=64$ . . . . .	178
22. Power Spectrum of Scaled CAR by 1.0, 0.5, and 0.25 for a Two-Dimensional Shape; p=12, q=0, $\bar{p}=4, \bar{q}=0, N=128, L=64$ . . . . .	179
23. Power Spectrum of Scaled CARMA by 1.0, 0.5, and 0.25 for a Two-Dimensional Shape; p=q=12, $\bar{p}=\bar{q}=4, N=124, L=64$ . . . . .	180

Figure	Page
24. Power Spectrum of Rotated CAR by 0.0, $\pi/4$ , and $\pi/2$ for a Two-Dimensional Shape; p=12, q=0, $\bar{p}=4$ , $\bar{q}=0$ , N=128, L=64. . . . .	181
25. Power Spectrum of Rotated CARMA by 0.0, $\pi/4$ , and $\pi/2$ for a Two-Dimensional Shape; p=q=12, $\bar{p}=\bar{q}=4$ , N=128, L=64 . . . . .	182
26. Power Spectrum of Scaled $L_1$ by 1.0, 0.5, and 0.25 for a Two-Dimensional Shape; p=12, q=0, $\bar{p}=4$ , $\bar{q}=0$ , N=128, L=64 . . . . .	183
27. Power Spectrum of Rotated $L_1$ by 0.0, $\pi/4$ , and $\pi/2$ for a Two-Dimensional Shape; p=12, q=0, $\bar{p}=4$ , $\bar{q}=0$ , N=128, L=64. . . . .	184
28. Power Spectrum of Vowel Portion /a/ of "Cat" by a Male Speaker Using EIGSP Solution; p=q=12, $\bar{p}=\bar{q}=8$ , N=128, L=64, 25% Overlap. . . . .	185
29. Power Spectrum of Vowel Portion /a/ of "Cat" by a Male Speaker Using EIGSP Solution; p=12, $\bar{p}=8$ , q=q=0, N=128, L=64, 25% Overlap . . . . .	186
30. Power Spectrum of Vowel Portion /a/ of "Cat" by a Male Speaker Using LPC; p=12, N=128, L=64, 25% Overlap . . . . .	187
31. Power Spectrum of Vowel Portion /a/ of "Cat" by a Female Speaker Using EIGSP Solution; p=q=12, $\bar{p}=\bar{q}=8$ , N=128, L=64, 25% Overlap. . . . .	188
32. Power Spectrum of Vowel Portion /a/ of "Cat" by a Female Speaker Using EIGSP Solution; p=12, $\bar{p}=8$ , q=q=0, N=128, L=64, 25% Overlap . . . . .	189
33. Power Spectrum of Vowel Portion /a/ of "Cat" by a Female Speaker Using LPC; p=12, N=128, L=64, 25% Overlap . . . . .	190
34. Comparison of the Power Spectrum of AR and Long AR for a Male Speaker; p=12, $\bar{p}=8$ , q=q=0, N=128, L=64, 25% Overlap . . . . .	191
35. Comparison of the Power Spectrum of AR and Long AR for a Female Speaker; p=12, $\bar{p}=8$ , q=q=0, N=128, L=64, 25% Overlap . . . . .	192
36. Block Diagram Representation of Word Recognition System . . . . .	102

Figure	Page
37. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using LP Solution; $f_1 = 0.15, f_2 = 0.185, \text{SNR} = 30 \text{ dB}, N=8, p=4.$ . . . . .	193
38. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using LP Solution; $f_1 = 0.15, f_2 = 0.185, \text{SNR} = 30 \text{ dB}, N=8, p=6$ . . . . .	194
39. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using Tuft-Kumaresan Method; $f_1 = 0.15, f_2 = 0.185, \text{SNR} = 30 \text{ dB}, N=8, L=6$ . . . . .	195
40. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution; $f_1 = 0.15, f_2 = 0.185, \text{SNR} = 30 \text{ dB}, p=6, \bar{p}=4, q=\bar{q}=0, N=8, L=7$ . . . . .	196
41. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution; $f_1 = 0.15, f_2 = 0.185, \text{SNR} = 30 \text{ dB}, p=6, \bar{p}=4, q=\bar{q}=1, N=8, L=7.$ . . . . .	197
42. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of CFBLP; $f_1 = 0.15, f_2 = 0.185, \text{SNR} = 30 \text{ dB}, p=5, \bar{p}=4, q=\bar{q}=0, N=8$ . . . . .	198
43. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of CFBLP; $f_1 = 0.15, f_2 = 0.185, \text{SNR} = 30 \text{ dB}, p=q=5, \bar{p}=\bar{q}=4, N=8$ . . . . .	199
44. Spectral Estimates of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of Correlation Matrix; $f_1 = 0.20, f_2 = 0.21, \text{SNR} = 0 \text{ dB}, p=q=12, \bar{p}=\bar{q}=4, N=128, L=64.$ . . . . .	200
45. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of Circular Correlation Matrix; $f_1 = 0.20, f_2 = 0.21, \text{SNR} = 0 \text{ dB}, p=q=12, \bar{p}=\bar{q}=4, N=128, L=64.$ . . . . .	201
46. Spectral Estimate for Two Sinewaves in White Gaussian Noise Using EIGSP Solution of CFBLP Data Matrix; $f_1 = 0.20, f_2 = 0.21, \text{SNR} = 0 \text{ dB}, p=q=12, \bar{p}=\bar{q}=4, N=128$ . . . . .	202
47. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of FBLP Data Matrix; $f_1 = 0.20, f_2 = 0.21, \text{SNR} = 0 \text{ dB}, p=q=12, \bar{p}=\bar{q}=4, N=128$ . . . . .	203

Figure	Page
48. Exact Autocorrelation of Sum of Two Cosinewaves; $f_1=0.2, f_2=0.21, N=100$ . . . . .	204
49. Regular Autocorrelation of Sum of Two Cosinewaves; $f_1=0.2, f_2=0.21, N=100$ . . . . .	205
50. Circular Autocorrelation of Sum of Two Cosinewaves; $f_1=0.2, f_2=0.21, N=100$ . . . . .	206
51. Exact Autocorrelation of Sum of Two Cosinewaves; $f_1=0.2, f_2=0.25, N=100$ . . . . .	207
52. Regular Autocorrelation of Sum of Two Cosinewaves; $f_1=0.2, f_2=0.25, N=100$ . . . . .	208
53. Circular Autocorrelation of Sum of Two Cosinewaves; $f_1=0.2, f_2=0.25, N=100$ . . . . .	209

## CHAPTER I

### INTRODUCTION

#### Motivation

The fields of system identification and spectral estimation have grown rapidly since the 1960's. Advancing technology and the introduction of more powerful and reliable digital computers have influenced growth in the fields of system identification and estimation theory. Scientists and engineers have developed many sophisticated algorithms and made other significant contributions to facilitate the rapid growth of system identification technology.

System identification is the fundamental problem of system engineers and system analysts. It is generally referred to as the determination of a mathematical model for a plant (process) from observation of its input-output measurements. The application of the system identification technology goes beyond the fields of engineering and physical sciences. A variety of identification methods are used to achieve a mathematical model for the systems arising in diverse fields, such as chemical processes, biomedical systems, biological sciences, medicine, and economics.

System identification is used to determine the essential characteristics of the system from a set of input-output measurements. It is understood that for some applications the set of input signals are not known with a high degree of certainty. In addition, they may be



unobservable. In these cases, the system characteristic is only obtained from a set of output measurements which is called the time series. The model is referred to as a stochastic model, due to the inherent uncertainty in the unobservable inputs.

The problem of system identification is best illustrated by Figure 1. A set of known inputs (excitations),  $u(k)$ , are applied to the system, and the outputs (responses),  $y(k)$ , are measured. These input-output measurements are used to identify the model for the underlying system. The observed input and output are usually corrupted with measurement noise  $w(k)$  and  $v(k)$ , respectively.

It is desired to obtain a linear model that relates the noisy input-output measurements and minimizes the residual,  $e(k)$ , as shown in Figure 1. Several restrictions are imposed on the model due to the assumed digital computer application for identification and control. The process is assumed to be a single input/single output linear shift-invariant discrete-time system, and an autoregressive moving average (ARMA) model of order  $(p,q)$  is used to describe the actual system. Furthermore, the set of input-output measurements may not be exactly related to an ARMA model due to model inaccuracies, i.e., input-output are not linearly related. A residual term,  $e(k)$ , is introduced so as to compensate for model inaccuracies.

A classical method to obtain the estimated parameters of an ARMA model is based on minimization of the mean square error. This method is known as the least squares (LS) method [4], [16], [18], [22], [49], [68]. The LS algorithm can be used to obtain the unbiased parameter estimates of an ARMA model when the residuals are uncorrelated or signal-to-noise ratio (SNR) is very high [45].

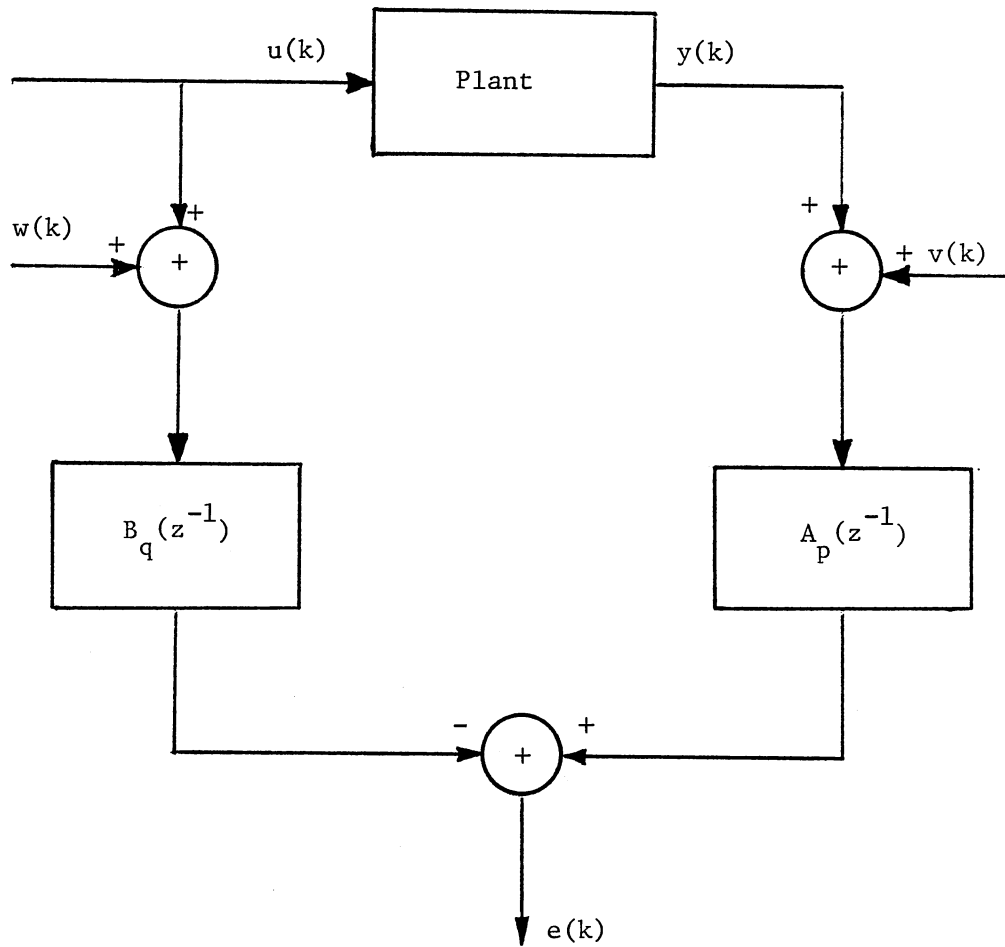


Figure 1. Block Diagram Representation of Equation Error Identification.

It has been shown (see Equation (2.22)) that the unbiased parameter estimates of a model can be obtained if the input-output are noise free or the residuals are uncorrelated (white noise). However, in practice the data are noise-contaminated, signal-to-noise-ratio (SNR) is low, or the residuals are correlated (colored noise). In this case the estimated parameters obtained from the LS method are biased. To overcome the difficulty caused by correlated residuals, Soderstrom [69] has proposed several instrumental variables (IV) methods that generate a new data vector, input-output measurements, which is uncorrelated with the residuals and gives rise to better parameter estimates. The generalized least square (GLS) method [10], and maximum likelihood (ML) method [3] are other alternative methods. A good survey of least squares related problems can be found in [9], [11], [21], [28], [35], [66], [69], [74], [80].

A brief discussion of the LS, GLS, and IV methods are covered in Chapter II. The ML method is based on maximizing the probability density of the measurements, often assumed Gaussian. In this research the probability density of the measurements is assumed to be unknown a priori. Thus, the ML method is not used.

It has been shown [45] that the LS, GLS, and IV methods cannot be used to obtain the unbiased parameter estimates when the additive noise is strongly correlated and SNR is low. A new identification procedure is proposed to remove the bias effect from the input-output measurements. As a result, the noise reduced measurements, at least in theory, give an unbiased parameter estimate and produce better model performance.

This method consists of recursive and non-recursive ARMA system identification which removes the effect of artifacts due to modeling error and input-output measurement noise. The recursive bias removal algorithm is based on the modified equation error identification (MEEI). The equation error identification (EEI) is modified to remove the noise effect and adapt the model parameters in order to account for any variation in plant parameters. This algorithm is capable of accurately estimating the model parameters if characteristics of the additive noise are known a priori. Otherwise, the variance of the additive noise must be estimated. The non-recursive bias removal algorithm, which is based on the eigenspace solution of the combined correlation matrix, can be used to reduce the noise effect and obtain unbiased parameters with or without a priori knowledge of the variance of the additive noise. Of course, knowledge of the noise variance will result in a better ARMA model.

Should the additive noise be of zero mean and white, its effects only appear at the first  $p+1$  diagonal elements of the correlation matrix. If the variance of the additive noise is known, it can be subtracted from the first  $p+1$  diagonal elements of the correlation matrix. Otherwise, the first  $p+1$  rows of the correlation matrix can be deleted so as to remove the noise effect. One always should be aware of the trade-off between the  $p+1$  row elimination, to compensate the noise effect, and the bias introduced by higher lag indices, especially for the fast decaying autocorrelation processes. The simulated results in Chapter III show the superiority of the proposed methods over the classical methods of identification.

## Overview

Chapter II describes the classical methods of system identification and explains three quantities which characterize an identification problem. A discrete-time model is used rather than a continuous time model to represent the underlying system due to the application of digital computers for identification and control. Limitations and problems associated with the classical methods, i.e., LS, GLS, and IV, are also discussed along with some procedures performed to improve the quality of the LS estimates.

In Chapter III the recursive and nonrecursive based identification are introduced and the necessary theory is given. In the first section, the recursive identification technique is developed based on the modification of EEI. A second order moving average (MA) process is simulated and the significant improvement of MEEI over EEI is shown. The majority of Chapter III is devoted to the eigenspace solution of the combined correlation matrix. First, the correlation matrices of the input-output measurements are formed. Then, the singular value decomposition (SVD) is used to obtain the minimum norm solution based on the linear combination of those right singular vectors that span a basis for the solution space. The dimension of the subspace is obtained from a new nullity algorithm based on the singular values of the correlation matrix. A second order ARMA process is simulated and improvement over the classical methods of identification is shown. A new algorithm is developed for determining the order of the model from the singular values.

In Chapter IV-VII several applications of stochastic modeling are covered and the superiority of the proposed method over linear

prediction (LP) is shown. Chapter IV describes the eigenspace approach to the texture boundary detection problem. A new circular ARMA (CARMA) model is proposed to represent the time series obtained from the shape classification. This model is compared with a regular ARMA model and its high resolution and accuracy are tested for several two dimensional objects.

In Chapter V, a pole-zero model is proposed to obtain the vocal tract parameters from sampled speech. This model satisfies the requirement of the acoustic theory for nasal and fricative sounds. The speech samples are pre-processed and the long correlation method is used to calculate the correlation lag indices.

In Chapter VI, two methods of isolated word recognition are presented. The first method is based on the feature vectors obtained from the linear prediction coefficients along with zero crossing rate (ZC), energy (ENG), normalized residual error (ERRN), and the normalized correlation coefficients. The second method is based on the eigenspace (EIGSP) solution of the correlation matrix via SVD. A new distance measure is proposed to match the input words with the reference words stored in the dictionary. The improvement of the proposed distance measure over the Euclidean distance measure is shown.

In Chapter VII, the estimation of frequencies of multiple sinusoids corrupted with white noise is discussed. A CARMA model, based on the modified forward-backward linear prediction, is proposed to improve the spectral resolution of the estimated frequencies. The estimated frequencies obtained from a short data record are compared with a method based on minimization of the sum of absolute value of the error,  $L_1$ .

The empirical studies and simulation results obtained from the eigenspace solution of the input-output correlation matrices, in Chapter III clearly indicate that the proposed method of system identification is superior over the classical method. Furthermore, the spectral estimation techniques developed in Chapter IV-VII show that minimum norm solution derived from the nullspace of the correlation matrix can be used as a powerful method to obtain the high resolution spectra.

In Chapter VIII, a summary of the results of this research along with several recommendations for future research are presented. A new orthogonal transform, M-transform, is also suggested for future research.

There are two appendices. Appendix A describes the numerical calculation of the variance of the output using Parseval's theorem. Some useful identities are developed to calculate the variance of the additive noise based on the required SNR. In Appendix B, the M-transform algorithm is described and its eigenvalues are obtained from a proposed recursive algorithm. The possible applications of M-transform are also discussed.

## CHAPTER II

### CLASSICAL METHODS OF SYSTEM IDENTIFICATION

#### Background

The problem of system identification and modeling has attracted considerable attention since the 1960's. In the 1960's, there was a large number of man-made control systems in diverse fields, such as biological science, physical science, chemical process, medicine, and economics. The design and implementation of both simple controllers and complicated multiloop systems required the identification of the underlying system. Moreover, the theory of automatic control was unable to model complicated practical problems such as satellite trajectories which required updating the position and velocity from large amounts of data sequentially accumulated with each pass of the satellite over a tracking station. Classical control theory was soon augmented by the state space representation following the development of Kalman filtering and Bellman's dynamic programming. The realization of the system function from the state space representation was an alternative problem in system identification. In addition, the idea of replacing a conventional analog controller with a digital controller was first considered in the late 1960's when computer technology began to grow rapidly in capability and availability. In each case, a mathematical model of the plant to be controlled was required for understanding and predicting the behavior of the system and designing a proper



controller. The mathematical model helps the system analyst find an elegant way to solve complicated optimization problems from the observed static and dynamic properties of the system to be controlled.

The problem of system modeling falls into two closely related categories: system identification and stochastic modeling. In the first type one can associate a number of measurable causes (inputs) and a number of measurable effects (outputs) to each physical phenomenon. The inputs and outputs are related through a set of mathematical equations, usually nonlinear partial differential equations. For all practical purposes, we assume the underlying system is approximately related to a linear system. The determination of a mathematical model which relates the inputs and outputs in a linear fashion is called system identification.

The second type of system modeling, in which the measurable effects or outputs can be identified while the causes are not well defined, is called stochastic modeling. Some typical examples of stochastic modeling are the hourly measure of the heart beat of a patient in abnormal condition, the radar received signal of a moving target at the tracking station, the annual population of China, and the average flow of blood in a vessel. In all these cases, the outputs, which are referred to as the time series, are available but the inputs are unknown in addition to often being unobservable. Of course, one can make some assumptions about the statistics of the input in order to make the model feasible. The name, stochastic modeling, is derived from the inherent uncertainty in the unobservable inputs.

In both cases we must satisfy some criteria in order to obtain the optimum solution. The criterion often refers to minimization of a

scalar cost function. The cost function is often chosen based on some assumption, i.e., sum of the squared error, when the problem is formulated as an identification problem. More often the criterion is expressed as the mean square error where the error is deviation of the model output from the process output and is referred to as output error. The optimum solution of the parametric identification problem can be obtained from minimizing the cost function. The minimization can be done in many different ways via the gradient method [73], steepest descent method [73], Newton's method [73], conjugate gradient method [42], or stochastic approximation method [78]. A good survey of optimization techniques can be found in [78], [73], [42], [62].

Often the main purpose of identification is to design a digital controller for a particular system. However, there are also some situations where the primary goal of identification is to analyze the properties of a system. In such cases the determination of parameters of the underlying system will be the final goal of identification. If the purpose of identification is to design a digital controller for a specific system, an accurate or a crude model of the dynamic system is required depending upon the nature of the control problem. In most practical problems there is not enough a priori information about the system and its environment. It is necessary to conduct some experiments and observe the outputs of the system while it is perturbed by the input signals. This set of input/output measurements can be processed to obtain a model which can be used for an optimum closed loop design.

In most applications it is necessary to identify the system in a short time and update the controller parameters so as to account for variations in the model parameters. This type of identification is

called "on-line" identification. A recursive algorithm is used for adjusting the estimates of the parameters for each time interval/sampling instant. An identification method is said to be "off-line" if a large amount of data corresponding to inputs and outputs of a system is stored on a disk or magnetic tape and non-recursive identification (batch solution) is used to obtain the best estimate of the model parameters so as to minimize a prescribed cost function. With off-line identification there is no real restriction on the computing time so a variety of methods can be used to obtain an accurate estimate of the system parameters. But in some applications one cannot afford to wait the required time to collect enough data necessary for determination of an optimum solution. This situation arises in cases such as controlling the blood pressure of a patient in a critical condition, tracking a moving target, etc. Thus, the on-line method must be used in spite of the fact that it may not likely lead to parameter estimates as accurate as those produced by off-line methods which use large amounts of data.

System identification can often be considered as finding the extrema of a functional. The form of the functional is given by the mathematical model of the system and the criterion to be minimized. To achieve the desired extrema, different methods of identification can be applied. These methods can be divided into two distinct classes of identification: direct methods, and model adjusting methods.

In direct methods, considerable amounts of memory are required, and the solution can be obtained explicitly from the mathematical model without using the physical realization of the system. The representations such as correlation functions, transfer functions,

impulse responses, and spectral densities are good examples of the direct or open loop methods. The model-adjusting methods, often called iterative methods, require less memory. They use the physical realization of the system, and the solution is obtained by a self-correcting procedure. The model-adjusting or closed loop methods can be characterized by the state space representation [16-17], model reference adaptive techniques [35-36], or recursive parametric models.

It has been shown [45] that the nonparametric models, i.e. correlation function methods, do not impair the estimated parameters when the extended model ordering is used, and the order of the process need not be specified explicitly. It is known that parametric models lead to biased estimates if the order of the model does not agree with the order of the process.

The input signals are not known for all phenomena and some assumption was made about their statistics to make stochastic modeling feasible. It has been shown [45] that significant improvement in modeling can be achieved if the statistics of the input signals are known a priori. The typical input signals used in the system identification and stochastic modeling are impulse functions, step functions, uncorrelated noise (white noise), correlated noise (colored noise), pseudo-random binary noise, and sinusoidal signals. In system identification it is highly desirable to use techniques that are independent of the input signals. However, often in stochastic modeling the input signals are unknown and some restrictions must be imposed on their statistics. According to Zadeh [83] an identification problem can be characterized by three quantities: a class of models, a class of input signals, and a criterion.

### Mathematical Models

The computer revolution in the last two decades has greatly influenced system identification and stochastic modeling. The advanced technology and inexpensive, powerful, and reliable microcomputers have brought a major revolution in design and development of digital controllers.

In spite of the fact that most systems are of the continuous type, the mathematical models of a dynamic system should be defined by discrete-time models due to application of digital computers for identification and control. The most important mathematical models of the discrete shift-invariant single-input/single-output linear dynamic system are the state space representation and the autoregressive moving average (ARMA) model. The state space representation is formulated as follows,

$$\begin{aligned}x(k+1) &= Ax(k) + Bu(k) + w(k) \\y(k) &= Cx(k) + Du(k) + v(k),\end{aligned}\tag{2.1}$$

where  $x(k)$  is called the state variable, and  $u(k)$  and  $y(k)$  are the input and output, respectively. The input noise,  $w(k)$ , and output noise,  $v(k)$ , in Equation (2.1) are assumed to be white Gaussian noise.  $A$ ,  $B$ ,  $C$ , and  $D$  are matrices of constant parameters.

The ARMA model is formulated as

$$y(k) = - \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j u(k-j) + v(k) \quad 1 \leq k \leq N,\tag{2.2}$$

where the  $a_i$  and  $b_j$  parameters are constants and the residual,  $v(k)$ , is assumed to be white Gaussian. This model is often used to obtain the estimated parameters of a model-adjusting or closed loop method. The

integers  $p$  and  $q$  are the orders of the ARMA model. The focus of this thesis is the ARMA model representation for system identification, rather than the state space representation. The ARMA model will only represent the essential properties of the dynamic system in a suitable form. This means that an exact mathematical representation of the underlying physical system is not expected and that the model will be derived to be optimum in some, usually limited, sense.

The mathematical techniques used in identification problems are either of the deterministic or stochastic type. In the deterministic method the noise is assumed to be negligible and an error cost function is to be minimized. In this method the number of equations set up for the identification problem is equal to the model parameters. Therefore, it is sufficient to set the gradient with respect to the unknown parameters equal to zero. The resulting equations are used to obtain the unknown parameters.

The stochastic method of identification is more important than the deterministic method and has special features of its own. In this method, a large number of data corresponding to the input-output measurements of a time varying system need to be processed so that potential application of a digital computer is almost mandatory. The number of input-output measurements must always exceed the number of estimated parameters in order to make the stochastic method of identification possible. This leads to an overdetermined system of linear equations.

It is also assumed that input-output measurements are corrupted with measurement noise. To be more precise, it is assumed that the additive noise is zero mean white Gaussian with unknown variance. The

exact parameters of the model can never be obtained from these noisy input-output measurements. Only estimates can be determined. The quality of these estimates can be evaluated from some statistical procedures, including consistency, sufficiency, or unbiasedness of the estimated parameters. By increasing the number of input-output measurements, the quality of these estimates should successively increase so that an unbiased estimate will be obtained for an infinite number of sampled data. However, in practice the number of sampled data are finite and never can have infinite precision. Thus the true values of the parameters can never be found in general.

Since the input-output measurements are noise contaminated and only a finite number of measurements are available, the exact solution can never be obtained. The best procedure is to minimize a cost function subject to some desired constraint. The method of Lagrange multipliers can be used to obtain the optimum solution for the above constrained optimization problem.

Stochastic methods of identification are categorized according to the particular cost function which is used to evaluate the quality of the estimation. Some traditional methods of system identification based on minimization of a least square error criterion are [72]

1. Ordinary least squares [2], [4], [13], [16], [22], [40]
2. Weighted least squares [13], [22]
3. Stochastic approximation [7], [15], [34], [61], [65]
4. Markov estimate [13]
5. Kalman-Bucy filtering [29], [30]
6. Instrumental variable method [33], [69], [82]
7. Generalized least squares [10], [17], [20], [22]

8. Maximum likelihood estimate [2], [5], [13], [60], [75]

9. Bayes' estimation [56], [57].

The identification problem defined in Equation (2.2) can be expressed as

$$\underline{y}(k) = \underline{X}^T(k)\underline{\theta} + v(k) \quad 1 \leq k \leq N, \quad (2.3)$$

where

$$\underline{\theta}^T = [ a_1, a_2, \dots, a_p, b_0, b_1, \dots, b_q ] \quad (2.4)$$

$$\underline{X}^T(k) = [ -y(k-1), \dots, -y(k-p), u(k), \dots, u(k-q) ] \cdot \quad (2.5)$$

Using matrix format Equation (2.3) can be expressed as

$$\begin{bmatrix} y(1) \\ y(2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y(N) \end{bmatrix} = \begin{bmatrix} -y(0) & \dots & -y(1-p) & u(1) & \dots & u(1-q) \\ -y(1) & \dots & -y(2-p) & u(2) & \dots & u(2-q) \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ -y(N-1) & \dots & -y(N-p) & u(N) & \dots & u(N-q) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_p \\ b_0 \\ b_1 \\ \cdot \\ \cdot \\ b_q \end{bmatrix} + \begin{bmatrix} v(1) \\ v(2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ v(N) \end{bmatrix} \quad (2.6)$$

where the input-output measurements outside the interval  $[1, N]$  are identically zero. Equation (2.6) can be written as

$$\underline{Y} = \begin{bmatrix} -Y_p & | & U_{q+1} \end{bmatrix} \underline{\theta} + \underline{V} \quad (2.7)$$



or more compactly as

$$\underline{Y} = H \underline{\theta} + \underline{V}, \quad (2.8)$$

where  $Y_p$  is the  $N \times p$  Toeplitz structured output matrix,  $U_{q+1}$  is the  $N \times (q+1)$  Toeplitz structured input matrix,  $\underline{\theta}$  is the  $(p+q+1) \times 1$  parameter vector,  $\underline{V}$  is the  $N \times 1$  residual vector, and  $\underline{Y}$  is the  $N \times 1$  output measurements.

The least squares method of identification is concerned with determining the best estimate,  $\hat{\underline{\theta}}$ , of  $\underline{\theta}$ , which minimizes the weighted sum of square of the residuals,  $\underline{e}^T W \underline{e}$ , where

$$\underline{e} = \underline{Y} - H \underline{\theta}. \quad (2.9)$$

Thus, the parameter vector can be obtained so that the cost function  $J$ ,

$$\begin{aligned} J &= [\underline{Y} - H \underline{\theta}]^T W [\underline{Y} - H \underline{\theta}] \\ &= \underline{e}^T W \underline{e}, \end{aligned} \quad (2.10)$$

is minimized. The elements of the weighting matrix  $W$  determine the degrees of freedom that can be placed on the individual measurements. The matrix  $W$  is an identity for the ordinary least squares problem, but it has some specific form for the weighted least squares and Markov estimate problems.

One can easily verify that least squares identification leads to an unbiased estimate if the residual vector is uncorrelated, i.e., white noise. Multiplying both sides of Equation (2.8) by  $H^T$  and taking the expected value leads to

$$\begin{aligned}
E [ H^T \underline{Y} ] &= E [ H^T (H\underline{\theta} + \underline{V}) ] \\
&= E [ H^T H\underline{\theta} ] + E [ H^T \underline{V} ] \\
&= E [ H^T H ] \underline{\theta} + E [ H^T ] E [ \underline{V} ] \\
&= E [ H^T H ] \underline{\theta} \quad \begin{array}{l} \nearrow \\ 0 \end{array}
\end{aligned} \tag{2.11}$$

where  $E[\cdot]$  is the expected value operator, i.e.,

$$E[X] = \sum_{k=-\infty}^{\infty} x_k p_k \tag{2.12}$$

and  $p_k$  is the discrete probability density of  $x_k$ .

Thus,

$$\hat{\underline{\theta}} = E [ H^T H ]^{-1} E [ H^T \underline{Y} ], \tag{2.13}$$

which is unbiased. However, in practice the residual vector  $\underline{V}$  is correlated, and least squares identification leads to biased estimates.

Several techniques have been suggested to take into account the error caused by the correlated residuals. Generalized least squares, instrumental variables and the maximum likelihood method are the most popular methods of system identification in the presence of correlated residuals. In the following sections the LS, GLS, and IV algorithms and the nature of bias and bias removal techniques are explained briefly. The ML method requires a priori knowledge of the probability density of the measurements and is not covered in this thesis.

#### Least Squares Identification

The LS error solution is the most favorable recursive technique among system analysts when the disturbance is a sequence of zero mean white noise. It leads to a mathematical model which can achieve the

best fit to input-output measurements in the sense of minimum-squared error.

Assume that noise contaminated input and output measurements are linearly related by an ARMA (p,q) process and that a residual,  $v(k)$ , compensates the inaccuracies of the model as shown in Equation (2.2). Furthermore, the noise free excitation - response are related by the following plant difference equation,

$$y(k) = - \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j u(k-j) \quad 1 \leq k \leq N . \quad (2.14)$$

The criterion is selected to be

$$V(\underline{\theta}) = 1/2 E[(y(k) - \underline{X}^T(k)\underline{\theta})^2], \quad (2.15)$$

where  $V(\underline{\theta})$  is a quadratic function of the  $a_i$  and  $b_j$  parameters, and  $\underline{\theta}$  and  $\underline{X}^T(k)$  are defined in Equations (2.17) and (2.18), respectively. It is easy to find its minimum analytically.

Upon taking the gradient of  $V(\underline{\theta})$  with respect to  $\underline{\theta}$  and using the linear property of the expected value, the following equations can be derived,

$$\begin{aligned} \frac{\partial V(\underline{\theta})}{\partial \theta_i} &= E [ y(k) - \underline{X}^T(k)\underline{\theta} ] E [ x_i(k) ] & 1 \leq k \leq N & \quad (2.16) \\ &= 0 & 1 \leq i \leq p + q + 1 & \end{aligned}$$

where

$$\underline{\theta}^T = [ \theta_1, \theta_2, \dots, \theta_{p+q+1} ] \quad (2.17)$$

and

$$\underline{X}^T(k) = [ x_1(k), x_2(k), \dots, x_{p+q+1}(k) ] . \quad (2.18)$$

Therefore, we conclude that

$$E [ y(k) ] = E [ \underline{X}^T(k) \underline{\theta} ] . \quad (2.19)$$

Multiplying both sides of Equation (2.14) by  $\underline{X}(k)$  and taking the expected value gives

$$\underline{\theta}_{\text{opt}} = E[\underline{X}(k) \underline{X}^T(k)]^{-1} E[\underline{X}(k) y(k)], \quad (2.20)$$

where  $\underline{\theta}_{\text{opt}}$  is the optimum value of the parameter vector  $\underline{\theta}$ . By substituting the stochastic approximation of  $E[\underline{X}(k)y(k)]$  and  $E[\underline{X}(k)\underline{X}^T(k)]$  into Equation (2.20), the minimum variance solution can be found. Most industrial processes are recursive in nature and require on-line identification rather than a batch solution. Therefore, only the recursive solution of least squares identification will be considered.

The recursive LS solution utilizes all incoming data and updates the parameters without incorporating all past data, but it is highly sensitive to noise. The form of solution is [22]

$$\hat{\underline{\theta}}(k+1) = \hat{\underline{\theta}}(k) + L(k) [y(k) - \underline{X}^T(k)\hat{\underline{\theta}}] \quad (2.21)$$

where  $L(k)$  is the correction gain. It can be shown that the recursive solution leads to an unbiased solution when the residuals in Equation (2.2) is zero mean white noise,

$$\begin{aligned} E[ \hat{\underline{\theta}}(k) ] &= E [ \underline{\theta} ] \\ &= E \{ [\underline{X}(k)\underline{X}^T(k)]^{-1} \underline{X}(k)y(k) \} \\ &= E \{ [\underline{X}(k)\underline{X}^T(k)]^{-1} \underline{X}(k)[\underline{X}^T(k)\underline{\theta} + v(k)] \} \\ &= E \{ \underline{\theta} + [\underline{X}(k)\underline{X}^T(k)]^{-1} \underline{X}(k)v(k) \} \\ &= E [ \underline{\theta} ] + [\underline{X}(k)\underline{X}^T(k)]^{-1} \underline{X}(k)E[v(k)] \\ &= \underline{\theta} . \end{aligned} \quad (2.22)$$

Therefore, one can conclude that the method of least squares is applicable and leads to unbiased estimates when the residuals (disturbances) are zero mean white noise, or signal-to-noise ratio (SNR) is high. If the residuals are not zero mean white noise, the non-zero term in Equation (2.22) will affect the parameters and will lead to biased estimation. Soderstrom [69] has proposed the IV technique to estimate parameters of an ARMA (p,q) process when the residuals are correlated with the data vector. Clarke [10] has suggested the GLS method when both the excitation and response are corrupted with correlated noise. It has been shown that the modified GLS estimate improved the parameter estimation when additive noise is correlated. The simulated results tabulated in Table I show this improvement.

To overcome the difficulty caused by the correlated residuals, two alternative solutions are proposed. The first method is based on the modification of the equation error identification (EEI) which is covered in Chapter III. The second method utilizes the singular value decomposition (SVD) to obtain the minimum norm solution of the parameter estimates based on the linear combination of right singular vectors of the correlation matrix which span an orthogonal basis for the solution space. Chapter III contains a review of some properties of SVD. The details of the problem will be discussed in Chapter III.

In addition to the bias effect due to correlated noise, one needs to consider the solution of ill-conditioned problems arising very often in least squares identification (see Equation (2.13)). It is easy to see that the smaller eigenvalues of  $H^T H$  make the least squares solution unstable. By using the SVD of  $H$ , unstable modes can be removed by setting to zero all smaller eigenvalues and approximating the

$N \times (p+q+1)$  rectangular matrix  $H$  by its best rank approximation using the Frobenius norm technique or nullity criterion covered in Chapter III. The SVD representation of  $H$  is

$$H = \sum_{k=1}^{p+q+1} \sigma_k \underline{u}_k \underline{v}_k^T, \quad (2.23)$$

where  $\sigma_k$ ,  $\underline{u}_k$ ,  $\underline{v}_k$  are the singular values, left singular vectors, and right singular vectors of  $H$ , respectively. Similarly,  $H^T$  can be written as

$$H^T = \sum_{k=1}^{p+q+1} \sigma_k \underline{v}_k \underline{u}_k^T. \quad (2.24)$$

The best rank approximation of  $H$  and  $H^T$  can be formed by setting the  $r$  smallest singular values to zero resulting in

$$H = \sum_{k=1}^{p+q+1-r} \sigma_k \underline{u}_k \underline{v}_k^T \quad (2.25)$$

$$H^T = \sum_{k=1}^{p+q+1-r} \sigma_k \underline{v}_k \underline{u}_k^T. \quad (2.26)$$

Thus,

$$H^T H = \sum_{k=1}^{p+q+1-r} \sum_{j=1}^{p+q+1-r} \sigma_j \sigma_k \underline{v}_k \underline{u}_k^T \underline{u}_j \underline{v}_j^T. \quad (2.27)$$

Using the orthogonal property of  $\underline{u}_k^T$ , i.e.,

$$\underline{u}_k^T \underline{u}_j = \begin{cases} 1 & j=k \\ 0 & j \neq k \end{cases}, \quad (2.28)$$

gives,

$$H^T H = \sum_{k=1}^{p+q+1-r} \sigma_k^2 \underline{v}_k \underline{v}_k^T. \quad (2.29)$$

The inverse of  $H^T H$ ,  $[H^T H]^{-1}$ , can be obtained from Equation (2.29)

$$[H^T H]^{-1} = \sum_{k=1}^{p+q+1-r} \frac{1}{\sigma_k^2} \underline{v}_k \underline{v}_k^T \quad (2.30)$$

Therefore, the stable least square solution can be written as

$$\begin{aligned} \underline{\theta} &= [H^T H]^{-1} H^T \underline{Y} \\ &= \sum_{k=1}^{p+q+1-r} \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^T \underline{Y} \\ &= H_r^+ \underline{Y} \end{aligned} \quad (2.31)$$

where  $H_r^+$  is called the pseudoinverse of the best rank approximation of  $H$ . It has been shown [45] that the parameter estimates obtained from Equation (2.31) are very close to the exact values even at low SNR. The source of improvement is the signal information contained in the first  $p+q+1-r$  large singular values and their corresponding right and left singular vectors while the noise information is reduced or removed.

#### Generalized Least Squares Identification

The generalized least squares algorithm is proposed by Clarke [10] to overcome the difficulty with correlated residuals. It can be used to estimate the parameters of an ARMA model more accurately than LS when the input-output are corrupted with colored noise or when the SNR is very low. It was shown that the LS algorithm leads to a biased parameter estimate when the additive noise is colored (i.e. correlated noise). The relaxation method of GLS with a whitening filter can improve the parameter estimation.

Assume that noisy input-output data are related by the following block diagram

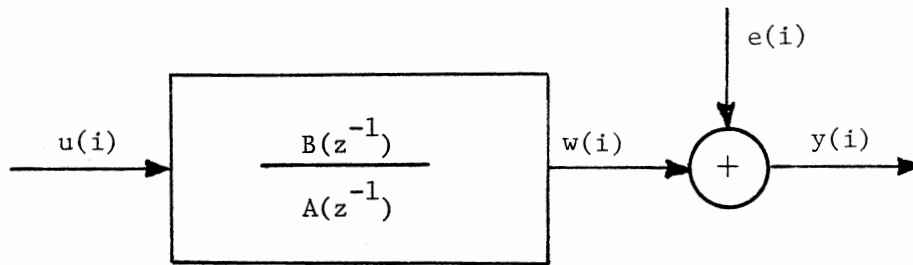


Figure 2. Block Diagram Representation of the GLS

where  $A(z^{-1})$  and  $B(z^{-1})$  are defined as

$$A(z^{-1}) = 1 + a_1 z^{-1} + \dots + a_p z^{-p} \quad (2.32)$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + \dots + b_q z^{-q} \quad (2.33)$$

and  $z^{-1}$  is the unit delay operator, i.e.,  $z^{-1}[y(k)] = y(k-1)$ .

Therefore, the GLS block diagram can be represented by the following equations

$$A(z^{-1})w(i) = B(z^{-1})u(i) \quad (2.34)$$

$$y(i) = w(i) + e(i). \quad (2.35)$$

Using (2.34) and (2.35), gives

$$A(z^{-1})y(i) = B(z^{-1})u(i) + A(z^{-1})e(i). \quad (2.36)$$

Assume that  $v(i)$  can be represented as

$$v(i) = A(z^{-1})e(i), \quad (2.37)$$



where  $\{e(i)\}$  is a sequence of uncorrelated random variables. If  $v(i)$  is substituted in Equation (2.36), the identification process can be governed by the following relationship

$$A(z^{-1})y(i) = B(z^{-1})u(i) + v(i). \quad (2.38)$$

Should  $v(i)$  be white, Equations (2.3) and (2.38) will be identical and the available LS algorithm can be used to obtain the unbiased parameter estimates of the given ARMA model. However, in more practical cases,  $v(i)$  is not a white process in spite of the fact that  $e(i)$  can be assumed to be white. Therefore, a whitening filter technique is suggested to convert the correlated residuals  $v(i)$  into a white residuals  $e(i)$ .

#### Whitening Filter and GLS Algorithm

As mentioned earlier, Equations (2.3) and (2.38) would be identical if the residuals,  $v(i)$ , were white. This suggests that the input-output measurements be processed by a whitening filter prior to the identification. The whitening filter is an  $l$ th-order moving average,  $MA(l)$ , process whose input is the correlated residuals,  $v(i)$ , and whose output is a sequence of white noise,  $e(i)$ ,

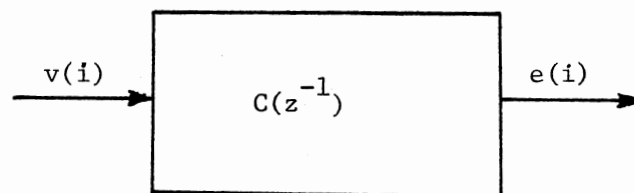


Figure 3. Block Diagram Representation of the Whitening Filter.

where

$$C(z^{-1}) = 1 + c_1 z^{-1} + \dots + c_\ell z^{-\ell} \quad (2.39)$$

is called the filter transfer function (see Figure 3).

Multiplying both sides of Equation (2.38) by  $C(z^{-1})$  gives

$$A(z^{-1})C(z^{-1})y(i) = B(z^{-1})C(z^{-1})u(i) + C(z^{-1})v(i). \quad (2.40)$$

Let  $e(i)$  be given by

$$e(i) = C(z^{-1})v(i).$$

Therefore,

$$A(z^{-1})C(z^{-1})y(i) = B(z^{-1})C(z^{-1})u(i) + e(i). \quad (2.41)$$

It is desired to obtain the parameter vectors  $\underline{a}$ ,  $\underline{b}$ , and  $\underline{c}$

$$\underline{a}^T = [ a_1, a_2, \dots, a_p ] \quad (2.42)$$

$$\underline{b}^T = [ b_0, b_1, \dots, b_q ] \quad (2.43)$$

$$\underline{c}^T = [ c_1, c_2, \dots, c_\ell ] \quad (2.44)$$

so as to minimize the sum of the squared error  $\sum_i e^2(i)$ , where

$$e(i) = A(z^{-1})C(z^{-1})y(i) - B(z^{-1})C(z^{-1})u(i). \quad (2.45)$$

Equation (2.45) is a nonlinear polynomial in  $A(z^{-1})$ ,  $B(z^{-1})$ , and  $C(z^{-1})$ . An analytical method cannot be used to estimate the  $\underline{a}$ ,  $\underline{b}$ , and  $\underline{c}$  parameter vectors. The following relaxation method can be used to determine the LS solution. The LS solution can then be used to obtain the first estimate of the whitening filter coefficients. The white-like residuals obtained from the MA whitening filter along with the new input-output data generated from the pre-whitening process can be used

to find a new estimate of the parameters. This procedure can be carried on until the change in parameters is less than some predetermined limit for successive iterations. The minimization in each step is done with the LS algorithm. The algorithm consists of the following four steps.

Step 1: Set

$$C(z^{-1}) = 1 \quad (2.46)$$

and solve the least square problem to estimate the a and b coefficients

$$\hat{\underline{\theta}}(N) = \left[ \sum_{k=\max(p,q)+1}^N \underline{X}(k)\underline{X}^T(k) \right]^{-1} \sum_{k=\max(p,q)+1}^N \underline{X}(k)y(k) . \quad (2.47)$$

Step 2: With  $\hat{A}(z^{-1})$  and  $\hat{B}(z^{-1})$  estimated, define the correlated residual  $v(k)$  by

$$v(k) = \hat{A}(z^{-1})y(k) - \hat{B}(z^{-1})u(k) \quad (2.48)$$

or

$$\begin{aligned} v(k) &= y(k) - \left[ - \sum_{i=1}^p \hat{a}_i y(k-i) + \sum_{j=0}^q \hat{b}_j u(k-j) \right] \\ &= y(k) - \underline{X}^T(k) \hat{\underline{\theta}}. \end{aligned} \quad (2.49)$$

The criterion to be minimized is the sum of the squared error  $\sum_i e^2(i)$ , where

$$e(k) = v(k) + \sum_{i=1}^l c_i v(k-i) . \quad (2.50)$$

To estimate the  $c_i$  parameters, such that  $e(k)$  approaches white noise, solve the standard least square problem.

Step 3: From the estimated filter coefficients,  $c_i$ , generate new excitation-response data according to

$$\tilde{y}(k) = C(z^{-1})y(k) \quad (2.51)$$

$$= \sum_{i=0}^{\ell} c_i y(k-i) \quad \ell+1 \leq k \leq N$$

$$\tilde{u}(k) = C(z^{-1})u(k) \quad (2.52)$$

$$= \sum_{i=0}^{\ell} c_i u(k-i) \quad \ell+1 \leq k \leq N .$$

Therefore,

$$A(z^{-1})\tilde{y}(k) = B(z^{-1})\tilde{u}(k) + \varepsilon(k) . \quad (2.53)$$

The objective is to minimize the sum of squared error  $\underline{\varepsilon}^T \underline{\varepsilon}$ . The criterion to be minimized is

$$V(\underline{\theta}) = \sum_{k=\ell+1}^N \varepsilon^2(k) . \quad (2.54)$$

The solution to (2.54) is

$$\hat{\underline{\theta}} = \left[ \sum_{k=\max(\bar{p}, \bar{q})+1+\ell+1}^N \bar{X}(k)\bar{X}^T(k) \right]^{-1} \sum_{k=\max(\bar{p}, \bar{q})+1+\ell+1}^N \bar{X}(k)\tilde{y}(k) , \quad (2.55)$$

and the estimated parameters have been improved.

Step 4: Test for convergence of parameters  $a_i$ ,  $b_j$ , and  $c_i$  and return to step 2 if the convergence has not been obtained.

A comparison of GLS with LS using a standard second order system shows the superiority of GLS. The simulated results in Table I show the parameter improvement of GLS over LS. The first set of parameters in Table I is the initial set of parameters of GLS estimated by LS and the rest are the parameters improved by the GLS algorithm.

Since the criterion which minimizes the error in the GLS algorithm is a nonlinear function of the parameters  $a_i$ ,  $b_i$ , and  $c_i$ , the estimated parameters may not be the global minimum if the initial condition

derived from LS is far beyond the actual value. In order to overcome this difficulty, the stable least squares solution developed in Equation (2.31) is suggested.

TABLE I  
COMPARISON OF LS AND GLS FOR ARMA(2,2)

iteration	Actual Parameters					
	$A_1=-1.5$ A(1)	$A_2=0.5$ A(2)	$B_0=1$ B(0)	$B_1=0.5$ B(1)	$B_2=0.0$ B(2)	
1	0.540392	-0.091164	1.041906	1.323327	0.935114	LS
6	-0.697277	0.058547	0.959168	0.985259	0.730484	GLS
11	-1.031692	0.338076	0.987557	0.720019	0.610862	
16	-1.402333	0.637571	1.000992	0.462327	0.248112	
21	-1.470351	0.682612	0.986839	0.479235	0.029028	
26	-1.474624	0.685013	0.985158	0.483477	0.010129	
31	-1.475156	0.685308	0.984964	0.484022	0.007730	
36	-1.475223	0.685345	0.984941	0.484085	0.007424	

#### Instrumental Variable Method of Identification

It has been shown [45] that LS identification leads to biased estimation when the residuals,  $v(k)$ , in Equation (2.3) are correlated.

Assume that the correlated residuals are given by  $e(k)$ . Therefore, the identification problem in Equation (2.3) can be written as

$$y(k) = \underline{X}^T(k)\underline{\theta} + e(k) . \quad (2.56)$$

Using matrix notation, Equation (2.53) can be written as

$$\underline{Y} = H\underline{\theta} + \underline{e} \quad (2.57)$$

where  $\underline{e}$  is the  $N \times 1$  correlated residual vector and  $\underline{Y}$ ,  $H$ , and  $\underline{\theta}$  are as defined in Section 2.2.

In the instrumental variable method, both sides of equation (2.57) are multiplied by a matrix  $W$  which is called the instrumental matrix which has the following properties

$$(a) \quad E [ WH ] \text{ is nonsingular} \quad (2.58)$$

$$(b) \quad E [ W \underline{e} ] = \underline{0} . \quad (2.59)$$

The instrumental variable equation then can be written as

$$\tilde{y}(k) = - \sum_{i=1}^p a_i \tilde{y}(k-i) + \sum_{j=0}^q b_j \tilde{u}(k-j) + v(k) \quad (2.60)$$

where elements of  $\tilde{u}(k)$  and  $\tilde{y}(k)$  are called the instrumental variables input-output measurements and elements of  $v(k)$  are the uncorrelated residual vector.

The ordinary LS algorithm can be used to obtain the unbiased parameter estimates from Equation (2.60). Although the IV method is very effective in removing the asymptotic bias from the parameter estimates, the derivation of an optimum IV matrix with properties given above is impractical, especially when the residuals are highly correlated. Moreover, the statistical efficiency of the solution is

dependent on the degree of correlation between  $\underline{\tilde{Y}} = W \underline{Y}$  and  $\underline{Y}$ . In particular, the most efficient, i.e. low variance, estimates can be obtained if the newly generated data vector  $\underline{\tilde{Y}}$  is equal to  $\underline{Y}$ , which is the noise-free case. Thus, the method of IV is not so practical when the input-output measurements are corrupted with strong additive noise.

In summary, the IV method of identification can be used to generate a new set of input-output data so that the residuals are uncorrelated. This method seems to be much simpler than the GLS method of identification which is time consuming, requiring about 30 or more iterations to converge. In addition, GLS requires the solution of two LS problems at each iteration. The only disadvantage of the IV algorithm is selection of the instrumental variables themselves with such specific statistical properties. A comparison of IV and GLS is shown in Figure 8 and Figure 10.

## CHAPTER III

### ARMA SYSTEM IDENTIFICATION

#### Introduction

Several classical methods of identification, i.e., LS, GLS, and IV, were briefly covered in Chapter II. These methods can be used to obtain the unbiased parameter estimates of an ARMA model if the model residuals are zero mean white noise or the SNR is sufficiently high. Should the residuals be correlated, the SNR be low, or strong noise be present on both input and output measurements, the estimated parameters will be biased. New techniques are proposed to overcome the difficulty caused by correlated residuals thus producing unbiased parameter estimates of an ARMA process using either recursive or non-recursive based identification. The recursive technique uses the modified equation error identification which is capable of estimating the parameters of a time varying system. Should the variance of the additive noise be known a priori, the unbiased parameter estimates would be obtained.

The non-recursive noise cancellation approach is based on the eigenanalysis of the combined input-output correlation matrix. The SVD is used to calculate the null space solution of the correlation matrix. The invariant right singular vectors of the correlation matrix are used as an orthogonal basis for the solution space. The dimension of the spanned space is calculated from a proposed nullity algorithm. The minimum norm, here the Euclidean norm of the parameter vector,



solution is obtained from the linear combination of this basis so as to satisfy the constraint that the first AR coefficient is equal to one. To illustrate the modeling improvement, a second-order system is simulated. The comparison of proposed methods with the performance of other recursive identification techniques is discussed. Simulations clearly indicate that the resolution capability of the new methods compare favorably with existing methods.

#### Modified Equation Error Identification

The equation error identification is a recursive method of parameter estimation. It can be used to obtain the parameter estimates of a process without using all past input-output observations at each step. It is also capable of identifying the parameters of a time-varying system (tracking problem). Many investigators have studied its asymptotic stability [25], [27], [36-38]. This method has a special configuration of a direct link between the parameter error and observed output and may be shown to lead to the LS solution. Landau [37] has proposed a recursive method that is based on the model reference adaptive techniques. This method gives good speed of convergence by using an adaptive algorithm with decreasing gain. A new method is proposed based on the EEI which removes the bias from the adaptation algorithm. This method is called the Modified Equation Error Identification.

Assume that the input and output measurements are related by an ARMA model of order (p,q). The process to be identified is described by

$$y(k) = - \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j u(k-j), \quad (3.1)$$

where  $u(k)$  and  $y(k)$  are the excitation and response at time instant  $k$ , respectively. Taking the Z-transform of both sides of (3.1) gives

$$Y(z) \left[ 1 + \sum_{i=1}^p a_i z^{-i} \right] = U(z) \sum_{j=0}^q b_j z^{-j}. \quad (3.2)$$

The associated transfer function

$$H_{p,q}(z) = \frac{\sum_{j=0}^q b_j z^{-j}}{1 + \sum_{i=1}^p a_i z^{-i}}, \quad (3.3)$$

is strictly positive real (SPR), i.e.,

$$\operatorname{Re} \{ H_{p,q}(z) \} > 0 \text{ for } |z| = 1. \quad (3.4)$$

Furthermore, assume that the plant transfer function has no common pole-zero cancellations, giving a reduced order system.

The equation error identification of the estimator can be described by

$$\hat{y}(k) = - \sum_{i=1}^p \hat{a}_i(k) y(k-i) + \sum_{j=0}^q \hat{b}_j(k) u(k-j), \quad (3.5)$$

or more compactly

$$\hat{y}(k) = \underline{X}^T(k) \hat{\underline{\theta}}(k), \quad (3.6)$$

where

$$\hat{\underline{\theta}}^T(k) = [\hat{a}_1(k), \dots, \hat{a}_p(k), \hat{b}_0(k), \dots, \hat{b}_q(k)] \quad (3.7)$$

$$\underline{X}^T(k) = [-y(k-1), \dots, -y(k-p), u(k), \dots, u(k-q)]. \quad (3.8)$$

Similar notations for the plant difference equation are given in Equations (2.2-2.4).

It has been shown [50] that the estimated parameter vector,  $\hat{\underline{\theta}}(k)$ , will approach the plant parameter vector,  $\underline{\theta}$ , for large  $k$  when both excitation and response are noiseless. However, in practice, both input and output are corrupted with additive noise and their actual values can't be observed. Assume that scalar output  $y(k)$  is corrupted with zero mean white noise  $v(k)$ ,

$$z(k) = y(k) + v(k), \quad (3.9)$$

and the  $(p+q+1) \times 1$  input-output measurements  $\underline{X}(k)$  are contaminated with a sequence of uncorrelated, zero mean white noise  $\underline{N}(k)$ ,

$$\underline{R}(k) = \underline{X}(k) + \underline{N}(k). \quad (3.10)$$

Thus, the output of the estimator can be written as

$$\hat{y}(k) = \underline{R}^T(k) \hat{\underline{\theta}}(k). \quad (3.11)$$

The estimated parameters are updated as follows,

$$\hat{\underline{\theta}}(k+1) = \hat{\underline{\theta}}(k) + \text{correction term}. \quad (3.12)$$

The error between  $z(k)$  and  $\hat{y}(k)$  is  $e(k)$ ,

$$e(k) = z(k) - \hat{y}(k). \quad (3.13)$$

A direct method to obtain  $\hat{\underline{\theta}}(k)$  is to minimize the Euclidean norm,  $\| \hat{\underline{\theta}}(k) - \underline{\theta} \|^2$ . Since the parameter vector is not known to the identifier, a less direct method is to minimize the square of the error via a gradient search algorithm. The criterion to be minimized is

$$J = 0.5 e^2(k). \quad (3.14)$$

The parameters will be updated in the negative direction of the gradient,

$$\hat{\underline{\theta}}(k+1) = \hat{\underline{\theta}}(k) - h(k) \frac{\partial J}{\partial \underline{\theta}(k)} \quad (3.15)$$

or

$$\hat{\underline{\theta}}(k+1) = \hat{\underline{\theta}}(k) + h(k) \underline{R}(k) [z(k) - \hat{y}(k)]. \quad (3.16)$$

Stability requires restriction on correction gain as

$$0 \leq h(k) \leq \frac{2}{\underline{R}^T(k) \underline{R}(k)}, \quad (3.17)$$

and it has been selected to be

$$h(k) = \frac{2}{1 + \underline{R}^T(k) \underline{R}(k)}, \quad (3.18)$$

so that abrupt changes in the gain are prevented when  $\underline{R}(k)$  approaches the zero vector. It is assumed that there is no correlation between  $\hat{\underline{\theta}}(k+l)$  and  $\hat{\underline{\theta}}(k)$  for any  $l \neq 0$ . Therefore, Equation (3.16) can be written as

$$\hat{\underline{\theta}}(k+l) = \hat{\underline{\theta}}(k) + h(k) \underline{R}(k) [z(k) - \hat{y}(k)]. \quad (3.19)$$

The following assumptions are also made

$$E [\underline{N}(k) \underline{X}^T(j)] = \underline{\mathbf{0}} \quad \text{for all } k, j \quad (3.20)$$

$$E [\underline{N}(k) \underline{N}^T(j)] = \underline{\mathbf{0}} \quad \text{for all } k \neq j \quad (3.21)$$

$$E [\underline{X}(k) \underline{v}(j)] = \underline{\mathbf{0}} \quad \text{for all } k, j \quad (3.22)$$

$$E [\underline{v}(k) \underline{v}(j)] = \underline{\mathbf{0}} \quad \text{for all } k \neq j \quad (3.23)$$

$$E [\underline{v}(k) \underline{N}(j)] = \underline{\mathbf{0}} \quad \text{for all } k \neq j \quad (3.24)$$

$$E [\underline{N}(k) \underline{N}^T(k) | \hat{\underline{\theta}}(k)] = C_{nn}. \quad (3.25)$$

Taking the expected value of both sides of Equation (3.19) and using the linear property of the expected value operator gives

$$\begin{aligned} E [\hat{\underline{\theta}}(k+1)] &= E \{ \hat{\underline{\theta}}(k) + h(k) \underline{R}(k) [z(k) - \hat{y}(k)] \} \\ &= E [\hat{\underline{\theta}}(k)] + E \{ h(k) \underline{R}(k) [z(k) - \hat{y}(k)] \}. \end{aligned} \quad (3.26)$$

From (2.3), (3.9), and (3.11), Equation (3.26) can be written as

$$\begin{aligned} E [\hat{\underline{\theta}}(k+1)] &= E [\hat{\underline{\theta}}(k)] + h(k) E [\underline{R}(k) \underline{X}^T(k) \underline{\theta} + \underline{R}(k) v(k) - \\ &\quad \underline{R}(k) \underline{R}^T(k) \hat{\underline{\theta}}(k)] \\ &= E [\hat{\underline{\theta}}(k)] + h(k) E [\underline{R}(k) \underline{X}^T(k) \underline{\theta}] + E [\underline{R}(k) v(k)] - \\ &\quad E [\underline{R}(k) \underline{R}^T(k) \hat{\underline{\theta}}(k)]. \end{aligned} \quad (3.27)$$

By definition of the conditional expectation,

$$E [\underline{R}(k) \underline{R}^T(k) \hat{\underline{\theta}}(k)] = E [E [\underline{R}(k) \underline{R}^T(k) \hat{\underline{\theta}}(k) | \hat{\underline{\theta}}(k)] \hat{\underline{\theta}}(k)]. \quad (3.28)$$

Substituting (3.10) into (3.28) gives

$$E [\underline{R}(k) \underline{R}^T(k) \hat{\underline{\theta}}(k)] = E [(E [\underline{X}(k) \underline{X}^T(k) | \hat{\underline{\theta}}(k)] + E [\underline{N}(k) \underline{N}^T(k) | \hat{\underline{\theta}}(k)]) \hat{\underline{\theta}}(k)]. \quad (3.29)$$

Since  $\hat{\underline{\theta}}(k)$  and  $\underline{X}(k)$  are independent

$$\begin{aligned} E [\underline{X}(k) \underline{X}^T(k) | \hat{\underline{\theta}}(k)] &= E [\underline{X}(k) \underline{X}^T(k)], \\ &= C_{xx}, \end{aligned} \quad (3.30)$$

where  $C_{xx}$  is the covariance of  $\underline{X}(k)$ .

Similarly,  $\hat{\underline{\theta}}(k)$  and  $\underline{N}(k)$  are independent giving

$$\begin{aligned} E [\underline{N}(k) \underline{N}^T(k) | \hat{\underline{\theta}}(k)] &= E [\underline{N}(k) \underline{N}^T(k)], \\ &= C_{nn}, \end{aligned} \quad (3.31)$$

where  $C_{nn}$  is the covariance of  $\underline{N}(k)$ .

From (3.30) and (3.31), (3.28) can be written as

$$E [\underline{R}(k)\underline{R}^T(k)\hat{\underline{\theta}}(k)] = (C_{xx} + C_{nn}) E [\hat{\underline{\theta}}(k)]. \quad (3.32)$$

Since the parameter vector  $\underline{\theta}$  is constant

$$\begin{aligned} E[\underline{R}(k)\underline{X}^T(k)\underline{\theta}] &= \underline{\theta} E[\underline{R}(k)\underline{X}^T(k)] \\ &= \underline{\theta} E[(\underline{X}(k) + \underline{N}(k))\underline{X}^T(k)] \\ &= \underline{\theta} E[\underline{X}(k)\underline{X}^T(k)] + \underline{\theta} E[\underline{N}(k)\underline{X}^T(k)] \\ &= \underline{\theta} C_{xx}. \end{aligned} \quad (3.33)$$

Similarly,

$$\begin{aligned} E [\underline{R}(k)\underline{v}(k)] &= E [(\underline{X}(k) + \underline{N}(k))\underline{v}(k)] \\ &= E [\underline{X}(k)\underline{v}(k)] + E [\underline{N}(k)\underline{v}(k)] \\ &= \underline{0}. \end{aligned} \quad (3.34)$$

Therefore, Equation (3.27) can be written as

$$E[\hat{\underline{\theta}}(k+l)] = E[\hat{\underline{\theta}}(k)] + h(k) \{ \underline{\theta} C_{xx} - (C_{xx} + C_{nn}) E[\hat{\underline{\theta}}(k)] \}. \quad (3.35)$$

It is assumed that

$$E [\hat{\underline{\theta}}(k+l)] = E [\hat{\underline{\theta}}(k)]. \quad (3.36)$$

as  $k \rightarrow \infty$

Define

$$\lim_{k \rightarrow \infty} \hat{\underline{\theta}}(k) = \lim_{k \rightarrow \infty} E [\hat{\underline{\theta}}(k)]. \quad (3.37)$$

as  $k \rightarrow \infty$

Therefore

$$\underline{\theta} C_{xx} - (C_{xx} + C_{nn}) \hat{\underline{\theta}}(\infty) = 0 \quad (3.38)$$

$$\hat{\underline{\theta}}(\infty) = [(C_{xx} + C_{nn})^{-1} C_{xx}] \underline{\theta} \quad (3.39)$$

or

$$\lim_{k \rightarrow \infty} E[\hat{\underline{\theta}}(k)] = [(C_{xx} + C_{nn})^{-1} C_{xx}] \underline{\theta} \quad (3.40)$$

as  $k \rightarrow \infty$

Thus, noise contaminated input-output leads to biased parameter estimates. Johnson, Hamm, and Triechler [26] have shown that Equation (3.26) is incomplete for use with AR processes. The effect of noise contribution can be removed to obtain the unbiased parameter estimates by modifying Equation (3.19) according to

$$\hat{\underline{\theta}}(k+1) = [1+h(k)\sigma_{nn}^2]I\hat{\underline{\theta}}(k) + h(k) \underline{R}(k)[z(k)-\hat{y}(k)], \quad (3.41)$$

where  $\sigma_{nn}^2$  is the variance of additive noise.

Should the system be an AR or MA process, the additive noise is only contributed to the output or input, respectively. But, in an ARMA process the noise is contributed on both input and output. Therefore, the modified EEI can be written as

$$z(k) = - \sum_{i=1}^p a_i z(k-i) + \sum_{j=0}^q b_j r(k-j) \quad (3.42)$$

and

$$\hat{z}(k) = - \sum_{i=1}^p \hat{a}_i(k) z(k-i) + \sum_{j=0}^q \hat{b}_j(k) r(k-j), \quad (3.43)$$

where  $z(k)$  and  $r(k)$  are the noise corrupted input and output, respectively. Then,

$$\hat{a}_i(k+1) = \hat{a}_i(k) + \mu_i z(k-i)h(k)[z(k)-\hat{z}(k)] + h(k)\sigma_{vv}^2 \hat{a}_i(k) \quad (3.44)$$

$$1 \leq i \leq p$$

$$\hat{b}_j(k+1) = \hat{b}_j(k) + \rho_j r(k-j)h(k)[z(k)-\hat{z}(k)] + h(k)\sigma_{ww}^2 \hat{b}_j(k) \quad (3.45)$$

$$1 \leq j \leq q$$

where  $\mu_i$  and  $\rho_j$  are positive constant gain and  $h(k)$  is

$$h(k) = \frac{1}{1 + \sum_{i=1}^p \mu_i z^{2(k-i)} + \sum_{j=0}^q \rho_j r^{2(k-j)}} . \quad (3.46)$$

As an example, the MEEI algorithm is used to estimate the parameters of a second order MA process corrupted with zero mean white noise and adjustable variance. The plant difference equation is

$$y(k) = 0.2 u(k-1) + 0.6 u(k-2) + n(k), \quad (3.47)$$

and the variance of input noise is selected to be 0.0, 0.5, and 1.0. The MA parameters,  $b_1$  and  $b_2$ , are estimated using EEI and MEEI. These estimates are tabulated in Table II for three different selections of noise variance. The first column of Table II is the variance of the additive noise. The second and third columns of Table II show the estimated parameters,  $b_1$  and  $b_2$ , using EEI while the fourth and fifth columns show these estimates using MEEI. These results clearly demonstrate the significant parameter improvement of modified EEI over EEI.

The theoretical results and numerous simulations have shown the advantage of modified EEI over the classical methods of identification. This method leads to unbiased parameter estimates when the additive noise is zero mean white and with known variance. Obviously, this assumption can't be made for every system and the presence of colored noise and unknown noise variance will result in biased parameter estimates. An  $\ell$ -th order MA whitening filter is suggested prior to identification to convert the correlated residuals



TABLE II  
COMPARISON OF EEI AND MODIFIED EEI FOR A MA PROCESS

$\sigma_{nn}^2$	EEI		Modified EEI	
	b1	b2	b1	b2
0.0	0.199967	0.599946	0.199967	0.599946
0.5	0.136384	0.401071	0.208057	0.593800
1.0	0.104664	0.308619	0.205714	0.5947701

into white residuals. Furthermore, one can estimate the variance of the additive noise by some prior knowledge of the true system.

Malakooti and Baltas [45] have shown that both noisy input and output can be scaled by constants to obtain equal variance, where the constant parameters are the standard deviations of the noisy output and input calculated from their noise contaminated measurements, respectively. The effect of scaling is to reduce the noise power significantly when bias removal techniques have not been used. Furthermore, even with prior knowledge of additive noise power and when applying the bias removal techniques, the scaling is suggested.

In this thesis, two different methods of scaling based on the input-output measurements and input measurement are suggested to improve the parameter estimates. Simulation results will be presented later indicating that the method of scaling provides a significant improvement in identification of the system parameters.

In the following section the SVD algorithm and its application are described briefly, and some useful identities based on SVD are derived to provide a mathematical basis for the non-recursive based identification covered in this chapter.

### Singular Value Decomposition

Singular value decomposition (SVD) is one of the most powerful and stable algorithms in both analytical and numerical analysis of linear algebra which provides the quantitative information about the structure of a system of linear equations. Application of SVD covers a variety of areas such as system identification, spectral estimation, deconvolution, and adaptive filtering due to the fact that the decomposition is accomplished by an efficient and stable algorithm. It can be used to improve the stability of ill-conditioned problems by calculating their best rank approximation.

The SVD representation of any arbitrary, complex  $M \times N$  matrix  $A$  can be given as

$$A = \sum_{k=1}^N \sigma_k \underline{u}_k \underline{v}_k^* , \quad (3.48)$$

where  $\sigma_k$ ,  $\underline{u}_k$ , and  $\underline{v}_k$  are called the singular values, the left singular vectors, and the right singular vectors of  $A$ , respectively. The asterisk symbol (\*) denotes the operation of complex conjugate transposition, and it has been implicitly assumed that  $M \geq N$ . Using the matrix form, Equation (3.48) can be written as

$$A = U \Sigma V^* , \quad (3.49)$$

where

$$U = [\underline{u}_1, \underline{u}_2, \dots, \underline{u}_M] \text{ and} \quad (3.50)$$

$$V = [\underline{v}_1, \underline{v}_2, \dots, \underline{v}_N] \quad (3.51)$$

are  $M \times M$  and  $N \times N$  unitary matrices, i.e.;

$$U^*U = UU^* = I \quad (3.52)$$

$$V^*V = VV^* = I \quad (3.53)$$

and

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \vdots & \vdots & \vdots & \sigma_N \\ \vdots & \vdots & \vdots & 0 \\ 0 & \dots & & 0 \end{bmatrix} \text{ is } M \times N. \quad (3.54)$$

The singular values are non-negative real valued quantities and are ordered in a monotonically nondecreasing fashion,

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0. \quad (3.55)$$

The columns of the  $M \times M$  unitary matrix  $U$ ,  $\underline{u}_k$ , are linearly independent with the orthogonal property,

$$\underline{u}_k^* \underline{u}_j = \begin{cases} 1 & k=j \\ 0 & k \neq j \end{cases}. \quad (3.56)$$

Similarly, the  $\underline{v}_k$  are linearly independent with the orthogonal property,

$$\underline{v}_k^* \underline{v}_j = \begin{cases} 1 & k=j \\ 0 & k \neq j \end{cases}. \quad (3.57)$$

From (3.48), (3.56), and (3.57) it follows that

$$\begin{aligned}
A^*A &= \left( \sum_{k=1}^N \sigma_k \underline{u}_k \underline{v}_k^* \right)^* \sum_{j=1}^N \sigma_j \underline{u}_j \underline{v}_j^* \\
&= \sum_{k=1}^N \sum_{j=1}^N \sigma_k \sigma_j \underline{v}_k \underline{u}_k^* \underline{u}_j \underline{v}_j^* \\
&= \sum_{k=1}^N \sigma_k^2 \underline{v}_k \underline{v}_k^* \\
&= V (\Sigma^* \Sigma) V^*
\end{aligned} \tag{3.58}$$

where  $\Sigma^* \Sigma$  is an  $N \times N$  diagonal matrix. It is easy to see that the diagonal elements of  $\Lambda = (\Sigma^* \Sigma)$  are the square of the singular values,

$$\Lambda = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \dots & & \dots & \vdots \\ \vdots & & & 0 \\ 0 & \dots & 0 & \sigma_N^2 \end{bmatrix} \tag{3.59}$$

and

$$A^*A = V \Lambda V^*. \tag{3.60}$$

This expression clearly indicates that the eigenvalues of  $A^*A$  and the associated eigenvectors are just the squares of the singular values and the right singular vectors of  $A$ , respectively. Similarly, it can be shown that

$$\begin{aligned}
AA^* &= U (\Sigma \Sigma^*) U^* \\
&= U \Lambda U^*,
\end{aligned} \tag{3.61}$$

where the eigenvalues of  $AA^*$  and the associated eigenvectors are just the square of the singular values and the left singular vectors of  $A$ , respectively. Although Equations (3.60) and (3.61) provide some useful

relationships between the singular values, left singular vectors, and right singular vectors of  $A$  with the eigenvalues and associated eigenvectors of  $A^*A$  and  $AA^*$ , they are not numerically accurate and lead to an unstable solution. Since the stable solution is usually desired, the formation of  $A^*A$  must always be avoided, because the useful information can be destroyed by the cross products. The stability problem caused by the cross product  $A^*A$  is illustrated by one example in the following section.

#### Stability of Cross Product of $A^*A$

SVD is widely used for statistical and signal processing computation. It reduces the computational deficiency of ill-conditioned problems and provides a stable solution. The solution is directly obtained from the decomposition of  $A$ , rather than  $A^*A$ .

The common least squares problem often used in system identification and stochastic modeling can be written as

$$\underline{Y} = A \underline{X} + \underline{V}, \quad (3.62)$$

where  $A$  is an  $M \times N$  data matrix,  $\underline{Y}$  is an  $M \times 1$  output vector,  $\underline{X}$  is an  $N \times 1$  parameter vector, and  $\underline{V}$  is an  $M \times 1$  residual vector. The LS solution can be obtained from

$$\underline{X} = [A^*A]^{-1} A^* \underline{Y}. \quad (3.63)$$

Should the data matrix  $A$  be ill-conditioned, the stable solution could never be obtained from (3.63), and the SVD algorithm must be used. Assume that the data matrix is

$$A = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} \quad (3.64)$$

Thus,

$$A^* A = \begin{bmatrix} 1 + \epsilon^2 & 1 \\ 1 & 1 + \epsilon^2 \end{bmatrix}, \quad (3.65)$$

where  $\epsilon$  is the machine precision, the smallest number that can be represented by the machine. Using floating point, fl, arithmetic

$$\begin{aligned} \text{fl}[A^* A] &= \text{fl} \begin{bmatrix} 1 + \epsilon^2 & 1 \\ 1 & 1 + \epsilon^2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{aligned} \quad (3.66)$$

It is obvious that the rank of  $A^* A$  is 1 while the rank of  $A$  is 2.

The theoretical singular values of the rank 2 matrix  $A$  are

$$\sigma_1 = \sqrt{2 + \epsilon^2}, \quad (3.67)$$

$$\sigma_2 = \epsilon \quad (3.68)$$

Using floating point arithmetic with machine precision  $\epsilon$  gives

$$\sigma_1 = \sqrt{2} \quad (3.69)$$

$$\sigma_2 = \epsilon, \quad (3.70)$$

which still corresponds to a rank 2 matrix. But, the singular values of  $A$  computed from  $A^* A$  would be at best

$$\sigma_1 = \sqrt{2} \quad (3.71)$$

$$\sigma_2 = 0, \quad (3.72)$$

which correspond to a rank 1 matrix. This indicates that cross product techniques can destroy useful information. Thus, the direct calculation of  $A^*A$  or  $[A^*A]^{-1}$  should be avoided if the stable solution is desired. The SVD of  $A$  must be calculated from a reliable software package such as IMSL, LINPACK, or EISPACK rather than by direct calculation of the eigenvalue problems,

$$(A^*A)\underline{v}_i = \lambda_i \underline{v}_i \quad (3.73)$$

$$(AA^*)\underline{u}_i = \lambda_i \underline{u}_i. \quad (3.74)$$

This insures accurate results.

#### SVD Identities

The most useful SVD identities often used in system identification and stochastic modeling are as follows,

1. Complex conjugate transpose of  $A$ ,  $A^*$

$$\begin{aligned} A^* &= \left( \sum_{k=1}^N \sigma_k \underline{u}_k \underline{v}_k^* \right)^* \\ &= \sum_{k=1}^N \sigma_k \underline{v}_k \underline{u}_k^* \\ &= V \Sigma U^*. \end{aligned} \quad (3.75)$$

2. Inverse of  $A$ ,  $A^{-1}$ , where  $A$  is a square matrix ( $M = N$ )

$$\begin{aligned}
A^{-1} &= \left( \sum_{k=1}^N \sigma_k \underline{u}_k \underline{v}_k^* \right)^{-1} \\
&= \sum_{k=1}^N \frac{1}{\sigma_k} (\underline{v}_k^*)^{-1} \underline{u}_k^{-1} \\
&= \sum_{k=1}^N \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^* \\
&= V \Sigma^{-1} U^*.
\end{aligned} \tag{3.76}$$

3. Inverse of  $A^*A$ ,  $(A^*A)^{-1}$ , using (3.58) gives

$$\begin{aligned}
(A^*A)^{-1} &= \left( \sum_{k=1}^N \sigma_k^2 \underline{v}_k \underline{v}_k^* \right)^{-1} \\
&= \sum_{k=1}^N \frac{1}{\sigma_k^2} \underline{v}_k \underline{v}_k^* \\
&= V \Sigma^{-2} V^*
\end{aligned} \tag{3.77}$$

4. Inverse of  $AA^*$ ,  $(AA^*)^{-1}$ , using (3.61) gives

$$\begin{aligned}
(AA^*)^{-1} &= \left( \sum_{k=1}^N \sigma_k^2 \underline{u}_k \underline{u}_k^* \right)^{-1} \\
&= \sum_{k=1}^N \frac{1}{\sigma_k^2} \underline{u}_k \underline{u}_k^* \\
&= V \Sigma^{-2} V^*
\end{aligned} \tag{3.78}$$

5. Pseudoinverse of  $A$ ,  $A^+$ , for overdetermined systems

$$A^+ = (A^*A)^{-1}A^*. \tag{3.79}$$

Using (3.75) and (3.77) gives



$$\begin{aligned}
A^+ &= \sum_{k=1}^N \frac{1}{\sigma_k} \underline{v}_k \underline{v}_k^* \sum_{j=1}^N \sigma_j \underline{v}_j \underline{u}_j^* \\
&= \sum_{k=1}^N \sum_{j=1}^N \frac{\sigma_j}{\sigma_k} \underline{v}_k \underline{v}_k^* \underline{v}_j \underline{u}_j^* \\
&= \sum_{k=1}^N \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^* \\
&= V \Sigma^{-1} U^*.
\end{aligned} \tag{3.80}$$

From (3.76) and (3.80) one can conclude that the pseudoinverse of any square matrix  $A$  is equal to its own inverse.

6. Pseudoinverse of  $A$ ,  $A^+$ , for underdetermined systems

$$A^+ = A^* (AA^*)^{-1}. \tag{3.81}$$

Using (3.75) and (3.78) gives

$$\begin{aligned}
A^+ &= \sum_{k=1}^N \sigma_k \underline{v}_k \underline{u}_k^* \sum_{j=1}^N \frac{1}{\sigma_j} \underline{u}_j \underline{u}_j^* \\
&= \sum_{k=1}^N \sum_{j=1}^N \frac{\sigma_k}{\sigma_j} \underline{v}_k \underline{u}_k^* \underline{u}_j \underline{u}_j^* \\
&= \sum_{k=1}^N \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^* \\
&= V \Sigma^{-1} U^*
\end{aligned} \tag{3.82}$$

7. Frobenius norm

The Euclidean matrix norm, or the Frobenius norm, of any arbitrary complex-valued  $M \times N$  matrix  $A$  is a non-negative real scalar given by

$$\begin{aligned} \|A\|_F &= \left( \sum_{i=1}^M \sum_{j=1}^N |a_{ij}|^2 \right)^{1/2} \\ &= [\text{Tr}(A^* A)]^{1/2}, \end{aligned} \quad (3.83)$$

where  $\text{Tr}[A]$  is the trace of the square matrix  $A^* A$ . The trace of any square matrix  $B = A^* A$  is defined by

$$\text{Tr}[B] = \sum_{i=1}^N \lambda_i, \quad (3.84)$$

where  $\lambda_i$  are the eigenvalues of the matrix  $B$ . Using the expression derived in Equations (3.58) and (3.73), the Frobenius norm of  $A$  can be written as

$$\begin{aligned} \|A\|_F &= [\text{Tr}(B)]^{1/2} \\ &= \left( \sum_{i=1}^N \lambda_i \right)^{1/2} \\ &= \left( \sum_{i=1}^N \sigma_1^2 \right)^{1/2} \end{aligned} \quad (3.85)$$

#### Data Matrix Based Identification

In this section, an algebraic characterization of an ARMA model for the ideal noise free case is developed. It is assumed that both input and output measurements are interrelated by a linear, shift invariant ARMA model of order  $(p, q)$ , Equation (3.1). The model is assumed to be irreducible so that the associated transfer function

$$H_{p,q}(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}}, \quad (3.86)$$

has no pole-zero cancellation. The algebraic characterization obtained from the noise-free data matrix, input-output measurements, is extended so as to account for model inaccuracy and additive noise that contaminates both input and output measurements.

Several classical methods of identification were briefly discussed in Chapter II to identify the model's  $a_k$  and  $b_k$  parameters from a finite set of input-output measurements. The model orders  $(p,q)$  are assumed to be known a priori in all these methods. New methods of identification based on the eigencharacterization of the data matrix and correlation matrix are proposed. These methods do not require the knowledge of model orders as long as the extended model order  $p$  and  $q$  are higher than the true model orders  $\bar{p}$  and  $\bar{q}$ .

The AR model order,  $p$ , is obtained from a new nullity algorithm based on the singular values of either data matrix or correlation matrix. The SVD is used for determining the ARMA model's parameters from a set of insensitive features, the singular values, and a set of orthogonal vectors, the right singular vectors, obtained from either data matrix or correlation matrix. It has been shown [45] that the noise invariant right singular vectors corresponding to the  $r$  smallest singular values form a basis whose dimension can be obtained from the nullity algorithm. The linear combination of the basis vectors is used to determine the minimum norm, the Euclidean norm of the parameter vector, solution of the ARMA model's parameters.

The linear relationship in Equation (3.1) can be evaluated for the sample interval  $k=0$  through  $k=N$  where the input-output measurement outside the interval  $[1,N]$  are identically zero. Using the matrix format, Equation (3.1) can be written as

$$\begin{bmatrix} y(1) \\ y(2) \\ \cdot \\ \cdot \\ \cdot \\ y(N) \end{bmatrix} = \begin{bmatrix} -y(0)\dots-y(1-p) & u(1)\dots u(1-q) \\ -y(1)\dots-y(2-p) & u(2)\dots u(2-q) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ -y(N-1)\dots-y(N-p) & u(N)\dots u(N-q) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_p \\ b_0 \\ \cdot \\ \cdot \\ b_q \end{bmatrix}, \quad (3.87)$$

or more compactly

$$\underline{Y} = \begin{bmatrix} -Y_p & | & U_{q+1} \end{bmatrix} \underline{g}, \quad (3.88)$$

where  $Y_p$  is the  $N \times p$  Toeplitz structured output matrix,  $U_{q+1}$  is the  $N \times (q+1)$  Toeplitz structured input matrix,  $\underline{Y}$  is the  $N \times 1$  output measurement vectors, and  $\underline{g}$  is the  $(p+q+1) \times 1$  ARMA model parameter vector. Equation (3.88) can be written as

$$\underline{Y} = D_{p,q+1} \underline{g}, \quad (3.89)$$

where  $D_{p,q+1}$  is called the data matrix.

The system of linear equations, (3.89), is overdetermined with a unique solution if the nullity of  $D_{p,q+1}^* D_{p,q+1}$  is equal to one. Multiplying both sides of (3.89) by  $D_{p,q+1}^*$  gives

$$D_{p,q+1}^* \underline{Y} = D_{p,q+1}^* D_{p,q+1} \underline{g}. \quad (3.90)$$

The least square solution for the parameter vector,  $\underline{g}$ , can be given as

$$\underline{g} = [D_{p,q+1}^* D_{p,q+1}]^{-1} D_{p,q+1}^* \underline{Y}. \quad (3.91)$$

Using the SVD identity, Equation (3.80), the extended model order solution can be obtained as

$$\underline{g} = \sum_{k=1}^{p+q+1} \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^* \underline{Y}. \quad (3.92)$$

The true model order must be known a priori or the nullity algorithm can be used to estimate it. The unbiased, reduced-order solution is expressed as

$$\underline{c} = \begin{bmatrix} 1 \\ \underline{g} \end{bmatrix}, \quad (3.93)$$

where

$$\underline{g} = \sum_{k=1}^{\bar{p}+\bar{q}+1} \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^*. \quad (3.94)$$

Equation (3.87) can be rewritten according to

$$\begin{bmatrix} y(1) & y(0) & \dots & y(1-p) \\ y(2) & y(1) & \dots & y(2-p) \\ \vdots & & & \\ \vdots & & & \\ y(N) & y(N-1) & \dots & y(N-p) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} u(1) \dots u(1-q) \\ u(2) \dots u(2-q) \\ \vdots \\ u(N) \dots u(N-q) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_q \end{bmatrix}, \quad (3.95)$$

or more compactly

$$\underline{Y}_{p+1} \underline{a}_{p+1} = \underline{U}_{q+1} \underline{b}_{q+1}, \quad (3.96)$$

where  $\underline{a}_{p+1}$  and  $\underline{b}_{q+1}$  are the  $(p+1) \times 1$  and  $(q+1) \times 1$  AR and MA parameter vectors, respectively. Combining both sides of Equation (3.96) gives

$$\begin{bmatrix} \underline{Y}_{p+1} & \vdots & -\underline{U}_{q+1} \end{bmatrix} \begin{bmatrix} \underline{a}_{p+1} \\ \underline{b}_{q+1} \end{bmatrix} = 0 \quad (3.97)$$

or

$$D_{p+1,q+1} \underline{C} = \underline{0}, \quad (3.98)$$

where  $D_{p+1,q+1}$  is the  $N \times (p+q+2)$  block Toeplitz structured data matrix and  $\underline{C}$  is the  $(p+q+2) \times 1$  parameter vector. It is clear that  $N > p+q+1$  must be satisfied if the ARMA  $(p,q)$  parameters are desired. This will ensure that the number of model equations will be at least equal to the number of unknown parameters to be identified. The system of linear equations expressed by (3.98) is often overdetermined with  $p+q+1$  unknowns. The number of independent equations must be equal to  $p+q+1$  if a unique set of parameter vector is desired.

This condition will be obtained if and only if the rank of the data matrix,  $D_{p+1,q+1}$ , is equal to  $p+q+1$ , the rank reduced conditions. In this case the unique solution can be achieved from the null space of  $D_{p+1,q+1}$ ,

$$\begin{aligned} \underline{C} &= \begin{bmatrix} a_{-p+1} \\ b_{-q+1} \end{bmatrix} \\ &= \frac{\underline{V}_1}{V_1(1)}, \end{aligned} \quad (3.99)$$

where  $\underline{V}_1$  is the  $(p+q+2) \times 1$  basis vector which spans the one dimensional null space of the data matrix  $D_{p+1,q+1}$ , and  $V_1(1)$  is the normalization factor determined such that the first AR coefficient is equal to one,  $a_0=1$ . The unique solution is obtained from the normalized eigenvector of  $D_{p+1,q+1}$  corresponding to the zero eigenvalue,

$$\begin{aligned} D_{p+1,q+1} \underline{V}_1 &= \lambda_1 \underline{V}_1 \\ &= \underline{0}. \end{aligned} \quad (3.100)$$

Thus, one can conclude that the ARMA model parameter vector is equal to the null space solution of the data matrix. If the data matrix

has nullity equal to zero, the given input-output measurements cannot be perfectly modelled as ARMA (p,q). If the nullity is equal to one, the unique solution is given in (3.99). If the nullity is greater than one then the system of linear equations, (3.89), does not have a unique solution. In this case, the minimum error solution can be obtained from the constraint minimization problem, namely that the first element of the AR parameters is equal to one. It will be shown that the ARMA model's parameters are directly related to the eigenvalue-eigenvector characterization of the matrix product,  $D_{p+1,q+1}^* D_{p+1,q+1}$ .

Because of the orthogonal properties of the right singular vector it follows that

$$\begin{aligned} D_{p+1,q+1} \underline{v}_m &= \sum_{k=1}^{p+q+2} \sigma_k \underline{u}_k \underline{v}_k^* \underline{v}_m \\ &= \sigma_m \underline{u}_m \end{aligned} \quad (3.101)$$

This relationship implies that the data matrix null space will be spanned by those right singular vectors whose associated singular values are zero. Assume that  $r$  singular values are zero for the noise free case. Thus, the basis for the null space is given by

$$[ \underline{v}_1, \underline{v}_2, \dots, \underline{v}_r ]. \quad (3.102)$$

Once this null space is formed, the ARMA model's parameters are obtained from the linear combination of the vectors in this basis.

#### Minimum Error Solution: Data Matrix Approach

In the last section, some restrictions were imposed on the excitation,  $u(k)$ , and response,  $y(k)$ , time series so that an appropriate

ARMA (p,q) model could be developed. Clearly, both input and output measurements are noise contaminated, and they cannot perfectly be related by an ARMA (p,q) model. Assume that noisy input and output measurements can be expressed as follows,

$$\hat{u}(k) = u(k) + w(k) \quad (3.103)$$

$$\hat{y}(k) = y(k) + v(k), \quad (3.104)$$

where  $u(k)$  and  $y(k)$  represent the true value of the input and output measurements,  $w(k)$  and  $v(k)$  represent the input and output additive noise, and  $\hat{u}(k)$  and  $\hat{y}(k)$  represent the noisy input and output measurements, respectively.

To obtain a desired ARMA (p,q) model from these noisy input-output measurements, the same procedure as for the noise free case will be performed. It is assumed that the input-output measurements are related through an ARMA (p,q) process, and a residual term is introduced to account for the model inaccuracy and measurement noise. Thus, the noise contaminated data matrix can be shown as follows,

$$\begin{aligned} \underline{\varepsilon} &= [Y_{p+1} \quad | \quad -U_{q+1}] \begin{bmatrix} \underline{a}_p \\ \underline{b}_q \end{bmatrix} \\ &= D_{p+1,q+1} \underline{C}, \end{aligned} \quad (3.105)$$

where  $D_{p+1,q+1}$  is the  $N \times (p+q+2)$  noisy data matrix,  $\underline{C}$  is the  $(p+q+2) \times 1$  parameter vector, and  $\underline{\varepsilon}$  is the  $N \times 1$  residual vector.

It is desired to select the  $\underline{a}_p$  and  $\underline{b}_q$  parameter vectors to minimize the Euclidean norm of the error vector,  $(\underline{\varepsilon}^* \underline{\varepsilon})^{1/2}$ . To minimize  $\underline{\varepsilon}^* \underline{\varepsilon}$  with the constraint that the first AR parameter,  $a_0$ , is one, the Lagrange



multiplier technique is used. Using the SVD representation of the noisy data matrix, Equation (3.105) can be written as

$$\underline{\varepsilon} = \left( \sum_{k=1}^{p+q+2} \sigma_k \underline{u}_k \underline{v}_k^* \right) \underline{C}. \quad (3.106)$$

Due to noise and possible perturbation of the system itself, the data matrix will be full rank and all  $(p+q+2)$  singular values will be positive. Thus the minimization criterion can be defined as

$$\begin{aligned} \underline{\varepsilon}^* \underline{\varepsilon} &= \left[ \sum_{k=1}^{p+q+2} \sigma_k \underline{u}_k \underline{v}_k^* \underline{C} \right]^* \left[ \sum_{j=1}^{p+q+2} \sigma_j \underline{u}_j \underline{v}_j^* \underline{C} \right] \\ &= \underline{C}^* \left[ \sum_{k=1}^{p+q+2} \sum_{j=1}^{p+q+2} \sigma_k \sigma_j \underline{v}_k \underline{u}_k^* \underline{u}_j \underline{v}_j^* \right] \underline{C}. \end{aligned} \quad (3.107)$$

Using the orthogonal property of  $\underline{u}_j$ , Equation (3.107) can be simplified as

$$\underline{\varepsilon}^* \underline{\varepsilon} = \underline{C}^* \sum_{k=1}^{p+q+2} \sigma_k^2 \underline{v}_k \underline{v}_k^* \underline{C}. \quad (3.108)$$

Recall that parameter vector  $\underline{C}$  is in the space spanned by the vector space  $[\underline{v}_1, \underline{v}_2, \dots, \underline{v}_{p+q+2}]$ , where  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_{p+q+2}$  are the right singular vectors of the noisy data matrix. Thus, the parameter vector  $\underline{C}$  can be written as a linear combination of the  $(p+q+2)$  linearly independent vectors  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_{p+q+2}$ ,

$$\underline{C} = \alpha_1 \underline{v}_1 + \alpha_2 \underline{v}_2 + \dots + \alpha_{p+q+2} \underline{v}_{p+q+2} \quad (3.109)$$

or more compactly

$$\underline{C} = \sum_{k=1}^{p+q+2} \alpha_k \underline{v}_k, \quad (3.110)$$

where  $\alpha_k$  are constant parameters.

Substituting for  $\underline{C}$  in Equation (3.108),  $\underline{\varepsilon}^* \underline{\varepsilon}$  can be written as

$$\begin{aligned} \underline{\varepsilon}^* \underline{\varepsilon} &= \sum_{k=1}^{p+q+2} \alpha_k^* \underline{v}_k^* \sum_{j=1}^{p+q+2} \sigma_j^2 \underline{v}_j \sum_{i=1}^{p+q+2} \alpha_i \underline{v}_i \\ &= \sum_{k=1}^{p+q+2} \sum_{j=1}^{p+q+2} \sum_{i=1}^{p+q+2} \alpha_k^* \alpha_i \sigma_j^2 \underline{v}_k^* \underline{v}_j \underline{v}_i. \end{aligned} \quad (3.111)$$

Using the orthogonal property of  $\underline{v}_k$ , Equation (3.111) can be simplified as

$$\underline{\varepsilon}^* \underline{\varepsilon} = \sum_{k=1}^{p+q+2} \sigma_k^2 \alpha_k^2, \quad (3.112)$$

which is a function of  $\alpha_k$ . Using the Lagrange multiplier technique the minimum error solution will be obtained by minimizing the expression  $\underline{\varepsilon}^* \underline{\varepsilon}$  in (3.112) subject to the constraint

$$\begin{aligned} a_0 &= \underline{C}(1) \\ &= \sum_{k=1}^{p+q+2} \alpha_k \underline{v}_k(1) \\ &= 1. \end{aligned} \quad (3.113)$$

The Lagrangian function,  $g$ , is formed as follows

$$\begin{aligned} g &= \underline{\varepsilon}^* \underline{\varepsilon} + 2\lambda [1 - \sum_{k=1}^{p+q+2} \alpha_k \underline{v}_k(1)] \\ &= \sum_{k=1}^{p+q+2} \sigma_k^2 \alpha_k^2 + 2\lambda [1 - \sum_{k=1}^{p+q+2} \alpha_k \underline{v}_k(1)], \end{aligned} \quad (3.114)$$

where the constant parameter  $\lambda$  is called the Lagrange multiplier.

The gradient of  $g$  with respect to  $\lambda$  and  $\alpha_k$  is taken and has been set to zero,

$$\frac{\partial g}{\partial \lambda} = 1 - \sum_{k=1}^{p+q+2} \alpha_k v_k(1) = 0 \quad (3.115)$$

and

$$\frac{\partial g}{\partial \alpha_k} = 2\alpha_k \sigma_k^2 + 2\lambda v_k(1) = 0. \quad (3.116)$$

Solving for  $\alpha_k$  in (3.116)

$$\alpha_k = \frac{-\lambda v_k(1)}{\sigma_k^2}, \quad (3.117)$$

and substituting in (3.115) gives

$$\sum_{k=1}^{p+q+2} \frac{-\lambda v_k(1)}{\sigma_k^2} v_k(1) = 1. \quad (3.118)$$

Thus

$$\lambda = \frac{-1}{\sum_{k=1}^{p+q+2} \frac{|v_k(1)|^2}{\sigma_k^2}}. \quad (3.119)$$

Upon the substitution of  $\lambda$  in (3.117) and using (3.110), the parameter vector,  $\underline{C}$ , can be calculated as

$$\underline{C} = \left[ \sum_{k=1}^{p+q+2} \frac{|v_k(1)|^2}{\sigma_k^2} \right]^{-1} \sum_{k=1}^{p+q+2} \frac{v_k(1)}{\sigma_k^2} \underline{v}_k. \quad (3.120)$$

If the excitation and response time series are closely related by an ARMA (p,q) model and the additive noise is sufficiently small, the minimum error solution in (3.120) will provide an acceptable parameter

vector. However, in practice these conditions are not met and an alternative method is required to obtain satisfactory results.

### Correlation Matrix Based Identification

Up to this point, it has been shown that the estimated parameters derived from the LS, GLS, and IV methods are biased when both excitation and response time series are corrupted with correlated noise or when the SNR is very low. The modified EEI also required prior knowledge of the noise for an unbiased parameter estimation. Moreover, the minimum error solution based on the SVD of the data matrix is only satisfactory if the additive noise is sufficiently low and the input and output are closely related by an ARMA (p,q) model. A new technique is proposed to alleviate the bias effects due to the correlated noise, low SNR, and the model inaccuracy. This method utilizes the null space of the correlation matrix for determining the smallest Euclidean norm of the parameter vectors.

Assume that the measured input and output are related by an ARMA model of order (p,q) as given by Equation (3.1). Multiply both sides of Equation (3.1) by  $y(k-l)$  and calculate its expected value to get

$$E[y(k)y(k-l)] = E\left[-\sum_{i=1}^p a_i y(k-i)y(k-l)\right] + E\left[\sum_{j=0}^q b_j u(k-j)y(k-l)\right]. \quad (3.121)$$

Using the linear property of the expected value operator and the following input-output autocorrelation and cross correlation lag formulas

$$r_{uu}(n) = E[u(n)u(n+k)] \quad \text{input autocorrelation} \quad (3.122)$$

$$r_{yy}(n) = E[y(n)y(n+k)] \quad \text{output autocorrelation} \quad (3.123)$$

$$r_{uy}(n) = E[u(n+k)y(n)] \quad \text{input-output cross correlation} \quad (3.124)$$

$$r_{yu}(n) = E[y(n+k)u(n)] \quad \text{output-input cross correlation,} \quad (3.125)$$

Equation (3.121) can be written as

$$r_y(\ell) = - \sum_{i=1}^p a_i r_y(\ell-i) + \sum_{j=0}^q b_j r_{uy}(\ell-j) \quad (3.126)$$

$$0 \leq \ell \leq N \quad N \geq p+q+2.$$

Using matrix format, Equation (3.126) can be written as

$$\begin{bmatrix} r_y(0) & \dots & r_y(-p) \\ r_y(1) & \dots & r_y(1-p) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ r_y(N-1) & \dots & r_y(N-p-1) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = \begin{bmatrix} r_{uy}(0) & \dots & r_{uy}(-q) \\ r_{uy}(1) & \dots & r_{uy}(1-q) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ r_{uy}(N-1) & \dots & r_{uy}(N-q-1) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_q \end{bmatrix}, \quad (3.127)$$

or more compactly as

$$R_{yy}\underline{a} = R_{uy}\underline{b}. \quad (3.128)$$

Combining both sides of (3.128) gives

$$\begin{bmatrix} R_{yy} & | & -R_{uy} \end{bmatrix} \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix} = \underline{0}, \quad (3.129)$$

where  $R_{yy}$  is the  $N \times (p+1)$  Toeplitz structured output autocorrelation matrix,  $R_{uy}$  is the  $N \times (q+1)$  Toeplitz structured input-output cross correlation matrix,  $\underline{a}$  is the  $(p+1) \times 1$  AR Parameter vector, and  $\underline{b}$  is the  $(q+1) \times 1$  MA parameter vector. Equation (3.129) can be written as

$$S_{p+1,q+1}^{(1)} \underline{C} = \underline{0}, \quad (3.130)$$

where  $S_{p+1,q+1}^{(1)}$  is the  $N \times (p+q+2)$  correlation matrix, and  $\underline{C}$  is the  $(p+q+2) \times 1$  ARMA parameter vector.

The overdetermined system of linear equations, (3.130), has a unique solution provided that the correlation matrix,  $S_{p+1,q+1}^{(1)}$  has rank  $p+q+1$ . The unique solution can be obtained by a nonzero vector lying in the null space of  $S_{p+1,q+1}^{(1)}$ , i.e.,  $S_{p+1,q+1}^{*(1)} S_{p+1,q+1}^{(1)} \underline{V}_1 = 0$ .

In this case, the unique solution is given as

$$\underline{C} = \frac{\underline{V}_1}{V_1(1)} \quad (3.131)$$

where  $\underline{V}_1$  corresponds to an eigenvector of the matrix  $S_{p+1,q+1}^* S_{p+1,q+1}$  associated with its zero eigenvalue. Similarly, by multiplying both sides of Equation (3.1) by  $u(k+P)$  and taking the expected value gives

$$R_{yu} \underline{a} = R_{uu} \underline{b} \quad (3.132)$$

or

$$\begin{bmatrix} -R_{yu} & R_{uu} \end{bmatrix} \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix} = \underline{0}, \quad (3.133)$$

where  $R_{uu}$  is the  $N \times (q+1)$  Toeplitz structured input autocorrelation matrix. Equation (3.133) can be written as

$$S_{p+1,q+1}^{(2)} \underline{C} = 0. \quad (3.134)$$

The unique solution can be achieved from the null space of  $S_{p+1,q+1}^{(2)}$  the same as Equation (3.131).

To form the correlation matrices, the following equations can be used to estimate the unbiased correlation lags from finite length input-output data

$$r_{yy}(n) = \frac{1}{N-n} \sum_{k=0}^{N-1-n} y(k+n)y(k) \quad 0 \leq n \leq N-1, \quad (3.135)$$

$$r_{uu}(n) = \frac{1}{N-n} \sum_{k=0}^{N-1-n} u(k+n)u(k) \quad 0 \leq n \leq N-1, \quad (3.136)$$

and

$$r_{yu}(n) = \frac{1}{N-n} \sum_{k=0}^{N-1-n} y(k+n)u(k) \quad 0 \leq n \leq N-1, \quad (3.137)$$

$$r_{uy}(n) = \frac{1}{N-n} \sum_{k=0}^{N-1-n} u(k+n)y(k) \quad 0 \leq n \leq N-1. \quad (3.138)$$

Moreover, Equations (3.130), and (3.134) can be used to form the combined correlation based identification as

$$\begin{bmatrix} R_{yy} & -R_{uy} \\ -R_{yu} & R_{uu} \end{bmatrix} \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix} = \underline{0}, \quad (3.139)$$

or more compactly

$$S_{p+1,q+1}^{(c)} \underline{C} = \underline{0}, \quad (3.140)$$

where  $S_{p+1,q+1}^{(c)}$  is the  $2N \times (p+q+2)$  combined correlation matrix. Equation (3.140) has a unique solution if the rank of the combined correlation matrix is  $p+q+1$ . The form of solution is given in Equation (3.131). For noise contaminated data, Equations (3.130), (3.134), and (3.140) need to be modified to match the underlying ARMA  $(p,q)$  model as

$$S_{p+1,q+1} \underline{C} = \underline{\varepsilon}, \quad (3.141)$$

whose solution is discussed in the last section. One may blindly think that these three Equations, (3.130), (3.134), and (3.140), will all result in the same parameter estimates. The empirical results and numerous simulations show that the parameter vectors estimated from these equations are different due to errors in estimating the correlation lags. By intuition one can introduce enough mathematical reasoning to support the simulations, which will be discussed briefly. Since the amount of noise power in the output is much higher than the input for the second order system considered, (3.134) will result in a better parameter estimate than (3.130). Similarly, Equation (3.140), which is the combination of (3.130) and (3.134) with added flexibility of choosing different rows from these two equations, will result in a better parameter estimation than each individual equation. One also may improve the parameter estimates by scaling the excitation-response measurement so the input-output noise power is the same and use either Equations (3.130) or (3.134) for the analysis.

#### Minimum Norm Solution: Correlation Matrix

##### Approach

As mentioned before, the noise contaminated case, both input and output corrupted with additive noise, is only an approximation to the ideal noise free case. Moreover, the correlation lag estimates introduce additional inaccuracy in the model.

Replacing the combined correlation matrix,  $S_{p+1,q+1}^{(c)}$  (S), by its SVD representation, gives

$$S = \sum_{j=1}^{p+q+2} \sigma_j \frac{u}{-j} \frac{v}{-j}, \quad (3.142)$$



where  $\sigma_j$  are the singular values, and  $\underline{v}_j$  and  $\underline{u}_j$  are the right and left singular vectors of  $S$ , respectively. Using the orthogonal property of  $\underline{u}_j$ ,  $S^*S$  can be expressed as

$$\begin{aligned} S^*S &= \sum_{j=1}^{p+q+2} \sigma_j \underline{v}_j \underline{u}_j^* \sum_{k=1}^{p+q+2} \sigma_k \underline{u}_k \underline{v}_k^* \\ &= \sum_{k=1}^{p+q+2} \sigma_k^2 \underline{v}_k \underline{v}_k^*. \end{aligned} \quad (3.143)$$

Upon taking the SVD of the combined correlation matrix  $S$ , (i.e. eigenanalysis of  $S^*S\underline{V}_k = \lambda_k\underline{V}_k$ ), two eigenvalue-eigenvector related properties follow. First, the  $r$  smallest singular values are theoretically zero for the noise free case. But, in the noise corrupted case, these singular values are affected by noise and are relatively small compared to the rest of the singular values. These  $r$  smallest singular values will identify the null space from the  $p+q+2-r$  larger singular values. Second, the right singular vectors associated with the  $p+q+2-r$  smallest singular values are orthogonal to the ARMA parameter vector,  $\underline{C}$ . Assume that the space spanned by the noise-corrupted right singular vectors is approximately equal to the space spanned by the noise-free right singular vectors. Thus,  $\{\bar{\underline{v}}_1, \bar{\underline{v}}_2, \dots, \bar{\underline{v}}_r\}$  will form a basis due to the orthogonality property of the right singular vectors, where  $\bar{\underline{v}}_k$  is the  $\underline{v}_{p+q+2-r+k}$  right singular vectors.

In the ideal noise free case, the extended correlation matrix has a null space of dimension  $r$ . However, in practice, the measurements are corrupted with additive noise and the extended correlation matrix has full rank, i.e., zero nullity. It is desired to approximate the correlation matrix  $S$  with another matrix,  $S^{(r)}$ , whose null space has a dimension of  $r$  and is closest to the correlation matrix in the Frobenius

norm sense (see Equation (3.85)). The matrix  $S^{(r)}$  is called the best rank approximation of  $S$  and can be obtained from

$$S^{(r)} = \sum_{k=1}^{p+q+2-r} \sigma_k \underline{u}_k \underline{v}_k^* , \quad (3.144)$$

which is the same as the SVD representation of  $S$ , Equation (3.142), but its  $r$  smallest singular values are set to zero. The  $S^{(r)}$  is no longer a block Toeplitz structured matrix, but has the nullity characteristic compatible with the underlying model. Thus, the parameter vector will lie in the null space of  $S^{(r)}$ . The candidate ARMA parameter vector is the linear combination of the basis vectors,

$$\underline{C} = \sum_{k=1}^r \sigma_k \underline{\bar{v}}_k . \quad (3.145)$$

It is desired to choose the parameter vector,  $\underline{C}$ , that has the smallest Euclidean norm, with the first component of the AR parameters,  $a_0$ , to be 1. This condition is met if the following relationship holds

$$\sum_{k=1}^r \alpha_k \underline{\bar{v}}_k(1) = 1 . \quad (3.146)$$

By using the Lagrange multiplier technique the desired solution can be obtained. The Lagrangian function is formed as

$$g = \underline{C}^T \underline{C} + 2\lambda [1 - \sum_{k=1}^r \alpha_k \underline{\bar{v}}_k(1)] . \quad (3.147)$$

Substituting for  $\underline{C}$  from Equation (3.145) gives

$$g = \left( \sum_{k=1}^r \alpha_k \underline{\bar{v}}_k \right)^T \left( \sum_{k=1}^r \alpha_j \underline{\bar{v}}_j \right) + 2\lambda \left( 1 - \sum_{k=1}^r \alpha_k \underline{\bar{v}}_k(1) \right) . \quad (3.148)$$

The gradient of  $g$  with respect to  $\lambda$  and  $\alpha_k$  is taken and has been set to zero.

$$\frac{\partial g}{\partial \lambda} = 1 - \sum_{k=1}^r \alpha_k \bar{v}_k(1) = 0 \quad (3.149)$$

$$\begin{aligned} \frac{\partial g}{\partial \alpha_k} &= \bar{v}_k^{-T} \sum_{j=1}^r \alpha_j \bar{v}_j + \sum_{k=1}^r \alpha_k \bar{v}_k^{-T} \bar{v}_j \cdot 2\lambda \bar{v}_k(1) \\ &= 2\alpha_k - 2\lambda \bar{v}_k(1) \\ &= 0 \end{aligned} \quad (3.150)$$

Thus,

$$\alpha_k = \lambda \bar{v}_k(1) . \quad (3.151)$$

Substituting for  $\alpha_k$  in (3.146) gives

$$\sum_{k=1}^r \lambda \bar{v}_k(1) \bar{v}_k(1) = 1 \quad (3.152)$$

or

$$\lambda = \frac{1}{\sum_{k=1}^r |\bar{v}_k(1)|^2} . \quad (3.153)$$

Substituting for  $\lambda$  in (3.151) gives

$$\alpha_k = \frac{\bar{v}_k(1)}{\sum_{k=1}^r |\bar{v}_k(1)|^2} . \quad (3.154)$$

Therefore, the minimum norm solution is obtained as

$$\underline{c} = \left( \sum_{k=1}^r |\bar{v}_k(1)|^2 \right)^{-1} \sum_{k=1}^r \bar{v}_k(1) \bar{v}_k . \quad (3.155)$$

#### Bias Removal in Combined Correlation Method

It has been shown [45] that the parameter vector obtained from Equation (3.134) or (3.140) is biased when both excitation and response are corrupted with additive noise. To overcome this problem, the proper scaling is suggested so that the amount of noise power in both input and

output are the same. Two similar methods of scaling are proposed. These methods are as follows:

Method 1: Scale the input-output data so that their root mean square value is the same,

$$\tilde{u}(n) = \alpha_1 u(n) \quad (3.156)$$

$$\tilde{y}(n) = \alpha_2 y(n) \quad (3.157)$$

where

$$\alpha_1 = \left[ \frac{1}{N} \sum_{n=1}^N u^2(n) \right]^{1/2} \quad (3.158)$$

$$\alpha_2 = \left[ \frac{1}{N} \sum_{n=1}^N y^2(n) \right]^{1/2} \quad (3.159)$$

and N is the input-output data length. The criterion to be minimized is

$$\underline{\underline{\epsilon}}^* \underline{\underline{\epsilon}},$$

$$\begin{aligned} \underline{\underline{\epsilon}}^* \underline{\underline{\epsilon}} &= \begin{bmatrix} \underline{a} & \underline{b} \end{bmatrix} \begin{bmatrix} \tilde{R}_{yy}^* \\ -\tilde{R}_{uy}^* \end{bmatrix} \begin{bmatrix} \tilde{R}_{yy} & -\tilde{R}_{uy} \end{bmatrix} \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix} \\ &= \begin{bmatrix} \underline{a} & \underline{b} \end{bmatrix} \begin{bmatrix} \tilde{R}_{yy}^* \tilde{R}_{yy} & -\tilde{R}_{yy}^* \tilde{R}_{uy} \\ -\tilde{R}_{uy}^* \tilde{R}_{yy} & \tilde{R}_{uy}^* \tilde{R}_{uy} \end{bmatrix} \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix}. \end{aligned} \quad (3.160)$$

Substituting for  $R_{yy}$  and  $R_{uy}$  gives

$$\underline{\underline{\epsilon}}^T \underline{\underline{\epsilon}} = \begin{bmatrix} \underline{a} & \underline{b} \end{bmatrix} \left[ \begin{array}{cc|cc} \alpha_2^4 R_{yy}^* & R_{yy} & -\alpha_2^3 \alpha_1 R_{yy}^* & R_{uy} \\ -\alpha_2^3 \alpha_1 R_{uy}^* & R_{yy} & \alpha_1^2 \alpha_2^2 R_{uy}^* & R_{uy} \end{array} \right] \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix}. \quad (3.161)$$

For the noise free case, Equation (3.161) can be set to zero, but for the noise contaminated data the eigenvalue problem will be formulated as

$$\begin{bmatrix} \alpha_2^4 R_{yy}^* & -\alpha_2^3 \alpha_1 R_{yy}^* R_{uy} \\ -\alpha_2^3 \alpha_1 R_{yy}^* R_{yy} & \alpha_1^2 \alpha_2^2 R_{uy}^* R_{uy} \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha_2} \underline{a} \\ \frac{1}{\alpha_1} \underline{b} \end{bmatrix} = \lambda \begin{bmatrix} \frac{1}{\alpha_2} \underline{a} \\ \frac{1}{\alpha_1} \underline{b} \end{bmatrix} \quad (3.162)$$

$$= \left( \frac{\lambda}{\alpha_2} \right) \begin{bmatrix} \underline{a} \\ \frac{\alpha_2}{\alpha_1} \underline{b} \end{bmatrix} .$$

From Equation (3.162) it is easy to see that the estimated value for the parameter vector  $\underline{b}$  needs to be scaled by a factor of  $\left(\frac{\alpha_1}{\alpha_2}\right)$  in order to satisfy the transformation in Equations (3.156) and (3.157).

Method 2: In this method, only the input data is scaled

$$\tilde{u}(N) = \alpha u(n) \quad (3.163)$$

$$\text{where } \alpha = \left[ \frac{\sum_{n=1}^N y^2(n)}{\sum_{n=1}^N u^2(n)} \right]^{1/2}. \quad (3.164)$$

Similarly, the estimated value of the parameter vector  $\underline{b}$  needs to be scaled by a factor  $\left(\frac{1}{\alpha}\right)$  in order to satisfy Equation (3.163).

In the beginning of Chapter III, it was shown that the MEEI can be used to utilize the prior knowledge of the additive noise power and identify the unbiased parameter estimates of an ARMA (p,q) process. Moreover, assume that noise contaminated data are defined as

$$x(n) = u(n) + \varepsilon_u(n) \quad (3.165)$$

and

$$z(n) = y(n) + \varepsilon_y(n) \quad (3.166)$$

The combined correlation matrix can be written as

$$\begin{aligned}
 S &= \begin{bmatrix} R_{ZZ} & -R_{XZ} \\ -R_{ZX} & R_{XX} \end{bmatrix} \\
 &= \begin{bmatrix} R_{yy} + \sigma_{\epsilon y}^2 I & -R_{uy} \\ -R_{yu} & R_{uu} + \sigma_{\epsilon u}^2 I \end{bmatrix} .
 \end{aligned}
 \tag{3.167}$$

From Equation (3.167) one can see that the variances of the additive noise are only contributed to diagonal elements of the combined correlation matrix. These variances are either known or can be estimated from the input-output observations. Upon removing the variances of input and output additive noise from the diagonal elements, the noiseless combined correlation matrix,  $S$ , results in a significant improvement in parameter estimates when SNR is low. If either input-output data are scaled by the suggested methods, or the SNR is very high, the effect of noise removal is not so significant. It has been shown [45] that scaling has much more effect on parameter improvement than noise power cancellation. This can be true because the correlation lags in the combined correlation matrix have been estimated from finite length data. A very long input-output measurement along with an accurate estimate of the additive noise power can reverse the action and leads to unbiased parameter estimation. The simulated results in Table III shows the effects of scaling and noise cancellation on the combined correlation matrix. The choice of  $T_1$  and  $T_2$ , the number of rows in the upper and lower partitions of  $S$  is discussed in the next section.

TABLE III  
 COMPARISON OF SCALING AND NOISE REMOVAL TECHNIQUE  
 SNR OUTPUT = 3 dB      SNR INPUT = 6 dB

MSE		a	b	c	d
7	29	0.030799240	0.0068713464	0.65789741	0.13861471
28	56	0.032309245	0.0095207654	0.12042093	0.55992073
7	7	0.034265712	0.0076690502	0.67499244	0.02401763
29	29	0.029818557	0.0069434494	0.13159504	0.6360714
7	35	0.030938726	0.0070895329	0.65416169	0.10667670
7	45	0.031071663	0.0073542409	0.65654367	0.1285670

- (a) Scaling with no noise subtraction
- (b). Scaling with noise subtraction
- (c) no scaling, no noise subtraction
- (d) no scaling, noise subtraction.

Another bias removal technique that is proposed is called the row reduction method. Should the first (p+1) rows of the top partition of S, and the first (q+1) rows of the bottom partition of S be deleted, the following noiseless combined correlation matrix will be obtained, provided the additive noise is zero mean and white.

$$S^{(p+1)(q+1)} = \begin{bmatrix} R_{yy}^{(p+1)} & -R_{uy}^{(p+1)} \\ -R_{yu}^{(q+1)} & 0 \end{bmatrix} \quad (3.168)$$

This matrix no longer has the input and output noise power in its diagonal elements. It will provide a better parameter estimate than Equation (3.140) if the impulse response of the system decays slowly, otherwise some information will be lost and a worse parameter estimates than Equation (3.140) will result.

#### Optimum Dimension of the Correlation Matrix

It has been shown that Equations (130), (134), and (140) are the fundamental equations for the correlation based identification. Define the number of rows of  $S_{p+1,q+1}^{(1)}$  and  $S_{p+1,q+1}^{(2)}$  as  $T$  and the number of rows in the top partition and bottom partition of  $S_{p+1,q+1}^{(c)}$  as  $T_1$  and  $T_2$ , respectively. If the number of equations is selected to be equal to the number of unknowns, i.e.,  $T=p+q+2$  for  $S_{p+1,q+1}^{(1)}$  and  $S_{p+1,q+1}^{(2)}$  or  $T_1=T_2=p+q+2$  for  $S_{p+1,q+1}^{(c)}$ , then the estimated parameters are not accurate because they are very sensitive to the noise in a minimal order model. Therefore, an overdetermined system is required for the purpose of noise smoothing.

It is not so clear how large the overdetermined systems should be. For larger  $T_1$  and  $T_2$ , more points are satisfied. But correlation lags with high indices are estimated from a few data points and introduce more inaccuracies. Therefore, an optimum number of rows for forming the combined correlation matrix must exist. The optimum selection of  $T_1$  and  $T_2$  can be obtained from an optimization problem based on the data length,  $N$ , the model orders  $(p,q)$ , and SNR, with the constraint that the first AR parameter be one, along with a criterion, minimum error or minimum norm solution discussed earlier.



An easier approach, which is used in this thesis, is based on the  $T_1$  and  $T_2$  that provide the minimum error between the true spectrum and the model spectrum with prior knowledge of the true spectrum. A standard second order linear shift-invariant system is used for the simulation. The extended model orders are selected to be 5,  $p=q=5$ .

The noisy input-output are recorded after the system transient has vanished. The length of input-output measurements,  $N$ , is 300, and SNR in both input and output are 6 dB. It has been shown [45] that the estimated parameters are improved by increasing the number of equations in the overdetermined system until  $T_1=T_2=29$  beyond which the estimated parameters are not improved further. Upon increasing  $T_1$  and  $T_2$  beyond 29, the estimated parameters start to be worse due to inaccuracies of higher lag indicies. The estimated parameters for three fundamental correlation matrices are calculated, where  $T_1=T_2=T=29$ . The normalized mean square error (MSE) between the true spectrum and the model spectrum is tabulated in Table IV. The results clearly indicate that the combined correlation based identification has a better performance for determining the parameter estimates of the underlying system compared with the correlation based identification introduced in (3.134) and (3.140).

TABLE IV  
COMPARISON OF THE RESOLUTION OF CORRELATION MATRICES  
AND COMBINED CORRELATION MATRIX

p	q	MSE		
		$S_{p+1,q+1}^{(1)}$	$S_{p+1,q+1}^{(2)}$	$S_{p+1,q+1}^{(c)}$
5	5	0.814412	0.0274906160	0.0012503644

Note: The SNR on both input and output = 6 dB and  $T_1=T_2=T=29$

#### Determination of the AR Model Order

After the optimal  $T_1$  and  $T_2$  are selected and  $S_{p+1,q+1}^{(c)}$  is formed, the singular values and the right singular vectors of  $S_{p+1,q+1}^{(c)}$  are calculated. The singular values that convey much information about the characteristics of the systems are used to obtain the rank, the AR model order, and the nullity of the combined correlation matrix. The order of the model is directly calculated from the number of nonzero singular values, if the input-output measurements are noise free, and a linear shift-invariant model can exactly relate these measurements. However, in practice these conditions are not met and the presence of the additive noise on both input and output measurements and model inaccuracy will change the singular values dependent upon the variance of the additive noise. In this case, the recognition of  $p+q+2$  larger singular values from the  $p-\bar{p}+q-\bar{q}$  smaller singular values is not practical and an accurate algorithm is required to separate the AR model order from the nullity of  $S_{p+1,q+1}^{(c)}$ .

In the ideal noise free case the combined correlation matrix has a null space of dimension  $r$ , but in the noise contaminated case the dimension of the null space is equal to zero, full rank condition. The matrix  $S_{p+1,q+1}^{(c)}$  is replaced by its best rank approximation,  $S_{p+1,q+1}^{(r)}$ , and a null space of dimension  $r$  is obtained. The nullity and rank of the combined correlation matrix can be calculated from a method called the effective rank ratio (ERR), based on the Frobenius norm representation given in Equation (3.85). The ERR is formulated as [8]

$$\begin{aligned} \text{ERR} &= v(k) \\ &= \frac{\| \| S_{p+1,q+1}^{(r)} \| \|_F}{\| \| S_{p+1,q+1}^{(c)} \| \|_F} \quad (3.169) \\ &= \frac{\left( \sum_{k=1}^{p+q+2-r} \sigma_k^2 \right)^{1/2}}{\left( \sum_{j=1}^{p+q+2} \sigma_j^2 \right)^{1/2}} \quad 0 < r < p+q+2 \end{aligned}$$

where  $k=p+q+2-r$  identifies the rank of  $S_{p+1,q+1}^{(c)}$  as  $v(k)$  approaches the predetermined value of ERR. Clearly, the effective rank ratio,  $v(k)$ , reaches its maximum value of one as  $k$  approaches  $j$ . If  $v(k)$  is close to one for values of  $k$  significantly smaller than  $j$ , the combined correlation matrix is said to be low effective rank. On the other hand, if  $v(k)$  is close to one for values of  $k$  almost equal to  $j$ , the combined correlation matrix is said to be of high effective rank. Although the ERR provides some important information about the rank of the combined correlation matrix, the predetermined value of the effective rank ratio explicitly must be defined. The value of ERR is dependent upon the system characteristics, the variance of additive noise, and SNR. Thus the effective rank ratio algorithm cannot be used as a reliable tool for

determining the AR model order from the combined correlation matrix because of some uncertainty about the predetermined value of  $v(k)$ .

In most applications, the system characteristics, the variance of additive noise, and the SNR are not known a priori. Thus, a reliable and powerful algorithm is required to identify the dimension of the null space and the rank of the combined correlation matrix. A new nullity algorithm is proposed based on the empirical results, the inherent characteristics of the singular values of the signal subspace and noise subspace, and the shifting property of the singular values in the presence of additive white noise. This algorithm can accurately calculate the AR model order from the singular values according to

$$\rho_1(k) = \frac{\sigma_{k-1}}{\sigma_k} \log \left[ \frac{\sigma_{\max}}{\sigma_k} \right]^2 \quad (3.170)$$

or

$$\rho_2(k) = \frac{\sigma_{k-1}}{\sigma_k} \frac{\log [1+k\sigma_{\min}]}{\log [1+k\sigma_{\max}]} \quad (3.171)$$

The peak value of either method will separate the nullity and rank of the correlation matrix, and the model order can be considered as the rank of  $S_{p+1,q+1}^{(c)}$ .

#### Numerical Example

In order to illustrate the effectiveness of the proposed bias removal techniques, the standard second order difference equation has been simulated [45]

$$y(n) - 1.5y(n-1) + 0.7y(n-2) = u(n) + 0.5u(n-1), \quad (3.172)$$

where  $u(n)$  and  $y(n)$  are the input and output of the noise free ARMA (2,2) process, respectively. A set of input and output measurements are recorded after 100 iterations to make sure that system is in steady state. (See Appendix A). The length of the recorded input and output data vectors is 300. The recorded input and output are corrupted with zero mean additive white noise

$$\tilde{u}(n) = u(n) + \varepsilon_u(n) \quad 1 \leq n \leq 300 \quad (3.173)$$

$$\tilde{y}(n) = y(n) + \varepsilon_y(n) \quad 1 \leq n \leq 300 \quad (3.174)$$

with adjustable variance so that the SNR at input and output is equal to predetermined values, i.e. 6 dB. The input signal is also zero mean white noise with variance equal to one.

This set of noisy input-output measurements is used to compare the magnitude of the transfer functions obtained from MEEI, LS, GLS, IV, and EIGSP methods. The magnitude of each transfer function has been compared with the magnitude of the true transfer function, where the true transfer function is

$$H(e^{j\omega}) = \frac{1 + 0.5e^{-j\omega}}{1 - 1.5e^{-j\omega} + 0.7e^{-j2\omega}} \quad (3.175)$$

Ten sets of statistically independent data, contaminated by uniform noise, are generated in order to provide a basis for comparison. These data are used to estimate the parameters of an ARMA model via various algorithms, LS, GLS, IV, and EIGSP. The magnitudes of the resulting system transfer functions are calculated and plotted, Figure 4 through Figure 12. Furthermore, the mean value of the system transfer functions,

$$\bar{H}(e^{j\omega}) = \frac{1}{10} \sum_{k=1}^{10} \hat{H}_k(e^{j\omega}), \quad (3.176)$$

and standard deviations (SD)

$$SD = \frac{1}{10} \sum_{k=1}^{10} | \hat{H}_k(e^{j\omega}) - H(e^{j\omega}) |^2, \quad (3.177)$$

are calculated and plotted to compare with the magnitude of the true transfer function, Figure 4 through 11. Their normalized mean square errors,

$$MSE = \frac{1}{10} \sum_{k=1}^{10} [ | \hat{H}_k(e^{j\omega}) | - | H(e^{j\omega}) | ]^2, \quad (3.178)$$

are used as a means of comparison and have been printed along with each plot, Figure 4 through Figure 12.

It has been shown that all mean values are close to the actual system function and only differ near the maximum points. The mean values of LS, GLS, IV, and EIGSP methods are plotted along with the actual system function, Figure 13. It has been shown that EIGSP method has the closest mean value to the actual function. The noise removal and scaling have significantly improved the mean value via EIGSP method (see Figure 14 through Figure 19). From these results, it is clear that the proposed correlation based identification has provided a better parameter estimate than LS, GLS, and IV methods. This improvement is due to noise desensitization, and parameter vector optimization via SVD, a numerically stable method.

Thus, the null space solution of the correlation matrix can be considered as a reliable and accurate method of system identification. In Chapters IV-VII, several applications of the proposed method are presented. The empirical results and simulations clearly indicate the

superiority and high resolution capability of the method over the classical linear prediction methods. Since these applications are selected from diverse fields, a brief discussion of each problem is presented. Moreover, remarks at the end of each chapter provide useful information for further investigation.

## CHAPTER IV

### CARMA MODEL METHOD OF TWO-DIMENSIONAL SHAPE

#### CLASSIFICATION: AN EIGENSYSTEM APPROACH

#### VS. THE $L_p$ NORM

#### Motivation

Because of periodicity of the time series derived from the  $N$  angularly equispaced radii of a closed boundary analysis problem, the correlation matrix has an invariant feature under rotation, translation, and scaling. The periodic characteristics possessed by the time series can be utilized to obtain improvement for texture boundary detection. A new circular ARMA (CARMA) model is introduced to represent the time series obtained for shape classification. This model is compared with a regular ARMA model and its high resolution and accuracy is tested for several two dimensional objects. Singular value decomposition is used to calculate the insensitive features for shape classification and boundary reconstruction. The invariant right singular vectors of the correlation matrix are used as an orthogonal basis for the solution space. The dimension of the spanned space is calculated from the nullity algorithm. To show the high resolution of the eigensystem approach,  $L_1$  and classical  $L_2$  solutions are compared.



## Introduction

Several parametric and nonparametric techniques are presently available for shape classification and texture boundary detection. A few notable examples are listed here. Persoon and Fu [55] have used the Fourier descriptor technique to obtain the skeleton of an object. Dubois and Glanz [14] have proposed an autoregressive approach to shape classification based on the least squares error criterion. Kashyap and Chellappa [31] suggested a stochastic model for closed boundary analysis. The proposed method is based upon an eigensystem analysis of the time series of samples obtained from the closed boundary. The periodicity of the derived time series allows one to obtain the same spectral shape for the rotated, or translated, or scaled object. It has been shown [43] that the CARMA model parameters of the time series can be obtained with high resolution by an eigenanalysis approach via SVD.

This method utilizes the correlation matrix and can be viewed as the null space solution which uses only the right singular vectors associated with the smallest singular values. (See minimum norm solution proposed in Chapter III). Once the circular autoregressive (CAR) parameters are obtained from Equation (3.155), the time series is filtered by a  $p$ th-order circular moving average (CMA) filter with parameters  $a_k$ , generating a sequence of white-like residuals. The circular autocorrelation of the residual sequence is windowed and the Blackman-Tukey approach is used to estimate the power spectrum of the CMA process. The  $b_k$  parameters are obtained from the inverse Fourier transform of the causal part of the CMA spectrum. This set of parameters,  $a_k$  and  $b_k$ , along with the size parameter,  $\alpha$ , and  $p$  initial conditions are used to reconstruct the two-dimensional shape boundary.

The power spectrum of the time series, calculated from the ratio of the CMA spectrum to the CAR spectrum, also is compared with other solutions computed based on both  $L_1$  and  $L_2$  error criteria. The SVD approach is shown to provide good resolution for the closed boundary shape spectrum and is less sensitive to the shape reconstruction.

#### Mathematical Formulation

Assume that the time series may be modeled as CARMA of order  $(p,q)$ . A size parameter is included in the model to account for scaling in a manner following Kashyap and Chellappa [31]. A residual error term is included as well. The model can then be formulated as

$$y(n)-\alpha + \sum_{k=1}^p a_k (y((n-k))-\alpha) = \sum_{k=0}^q b_k x((n-k)) + w(n) , \quad (4.1)$$

where  $\alpha$  is the expected value of the time series  $y(n)$ ,  $p$  and  $q$  denote the model orders, and  $w(n)$  is zero mean white noise with unit variance.

Note that

$$y((n-k)) = \begin{cases} y(n-k) & \text{for } n-k > 0 \\ y(n-k+N) & \text{for } n-k < 0 \end{cases} \quad (4.2)$$

Dropping the size parameter from (4.1) by removing the mean from the time series samples results in

$$y(n) + \sum_{k=1}^p a_k y((n-k)) = \sum_{k=0}^q b_k x((n-k)) . \quad (4.3)$$

The corresponding Z-transform is given by

$$Y(z) [1 + \sum_{k=1}^p a_k z^{-k}] = X(z) \sum_{k=0}^q b_k z^{-k} \quad (4.4)$$

It is assumed that the model has a reduced order form. Therefore, the power spectral density function associated with the boundary time series is given by the CARMA (p,q) rational form [8]

$$S_{\text{CARMA}}(e^{j\omega}) = \left| \frac{b_0 + b_1 e^{-j\omega} + \dots + b_q e^{-j\omega q}}{1 + a_1 e^{-j\omega} + \dots + a_p e^{-j\omega p}} \right|^2 \quad (4.5)$$

Upon multiplication of both sides of (4.3) by  $y(n-1)$  and taking the expected value, the extended Yule-Walker equations can be simplified as

$$\begin{bmatrix} r_y(q+1) & r_y(q) & \dots & r_y(q-p+1) \\ r_y(q+2) & r_y(q+1) & \dots & r_y(q-p+2) \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ r_y(L+q) & r_y(L+q-1) & \dots & r_y(L+q-p) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad (4.6)$$

or more compactly

$$\underline{R}\underline{a} = \underline{0}, \quad (4.7)$$

where  $R$  is the  $L \times (p+1)$  Toeplitz structured correlation matrix and  $\underline{a}$  is the  $(p+1) \times 1$  AR parameter vector. The elements of  $R$  are estimated from the circular autocorrelation

$$r_y(n) = \frac{1}{N} \sum_{k=1}^N y(k \oplus n) y(k), \quad (4.8)$$

where  $\oplus$  denotes modulo  $N$  addition,

$$y(k+N) = y(k) \quad \text{for all } k. \quad (4.9)$$

It has been shown [43] that only the first  $\bar{p}$  singular values of  $R$  are large compared with the other  $p+1-\bar{p}$ . This indicates that the boundary information is contained in the first  $\bar{p}$  dominant singular values and the rest,  $r=p+1-\bar{p}$ , are artifacts due to boundary approximation and error from the camera and digitizing device. Since in reality the  $r$  singular values are zero, one can obtain a best rank approximation from Equation (3.144) repeated as

$$R(r) = \sum_{k=1}^{\bar{p}} \sigma_k \underline{u}_k \underline{v}_k^* . \quad (4.10)$$

This will allow for elimination of the existing artifacts and the formulation of an improved model. The proposed solution based on the null space of  $R$  is given in Equation (3.155) repeated as

$$\underline{a} = \beta \sum_{k=1}^r \bar{v}_k(1) \underline{v}_k , \quad (4.11)$$

where  $\beta$  is a constant for normalization.

Once the CAR coefficients have been defined, the residual sequence produced from CMA is

$$e(n) = \sum_{k=0}^p a_k y((n-k)) . \quad (4.12)$$

The circular autocorrelation,  $r_e(n)$ , of the residual sequence can be calculated from Equation (4.8). Using the triangular window,

$$w(n) = 1 - (|n| \mid (q+1)), \quad (4.13)$$

to insure the positiveness of the Fourier transform of the sequence  $r_e(n)$ , the CMA spectrum is estimated as

$$S_{\text{CMA}}(e^{j\omega}) = \sum_{n=-q}^q w(n) r_e(n) e^{-j\omega n} . \quad (4.14)$$

The inverse Fourier transform of the causal part of  $S_{\text{CMA}}(e^{j\omega})$  gives the  $b_k$  coefficients for boundary reconstruction. In order to obtain some basis for the comparison with the  $L_1$  solution proposed by [67], the following circular linear prediction (CLP) model has been used,

$$x(n) = \sum_{k=1}^p a_k x((n-k)) \quad n=1, 2, \dots, N, \quad (4.15)$$

where  $y(n) = x(n+1)$  is called the one step ahead predictor. The mean value of the time series samples,  $x(n)$ , is removed prior to modeling. Using matrix format, (4.15) can be written as

$$\begin{bmatrix} x(1) & x(N) & x(N-1) & \dots & x(p) \\ x(2) & x(1) & x(N) & \dots & x(p+1) \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ x(N-1) & x(N-2) & x(N-3) & \dots & x(p-2) \\ x(N) & x(N-1) & x(N-2) & \dots & x(p-1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = \begin{bmatrix} x(2) \\ x(3) \\ \cdot \\ \cdot \\ \cdot \\ x(N) \\ x(1) \end{bmatrix} \quad (4.16)$$

or more compactly,

$$B \underline{a} = \underline{Y}, \quad (4.17)$$

where  $B$  is the  $N \times p$  CLP data matrix,  $\underline{a}$  is the  $p \times 1$  CLP parameters, and  $\underline{Y}$  is the  $N \times 1$  predictor. The  $L_2$  norm solution to (4.17) is used as a first estimate of the  $L_1$  solution (4.18) and

$$\underline{a} = [B^T B]^{-1} B^T \underline{Y} \quad (4.18)$$

is calculated based on the complex residual steepest descent (RSD) algorithm [67]. The eigenspace solution via SVD has been calculated according to (4.19).

Equation (4.17) can be rewritten as

$$\begin{bmatrix} -\underline{Y} & B \end{bmatrix} \begin{bmatrix} 1 \\ \underline{a} \end{bmatrix} = \underline{0} \quad (4.19)$$

or more compactly,

$$\underline{DC} = 0, \quad (4.20)$$

where

$$\underline{C} = [ 1, a_1, a_2, \dots, a_p ]^T \quad (4.21)$$

The null space solution of (4.20) can be obtained from (4.11).

An important feature of the null space solution via SVD is the information contained in the singular values. These singular values can be used for shape classification and pattern analysis. They also can be used to identify the CARMA model or the dimension of the null space.

The spectrum obtained for each method has been compared for different objects under rotation and scaling including zero mean additive white noise to account for any undesirable disturbance. The result has shown that the null space solution can obtain a high resolution spectrum even at low signal to noise ratio. This method compares favorably with other complex methods of spectral estimation. Since the additive noise and other artifacts are assumed to be zero mean white processes, they only bias the correlation lag  $r(0)$ , of the first  $p+1$  rows of the correlation matrix. Upon deleting the first  $p+1$  rows of the correlation lags, the effect of noise has been removed or greatly reduced. With regular ARMA this may not result in large model improvement, because of trade offs between the noise effect and bias involved in higher lags. In CARMA, since all lags are estimated from

the same data length, the  $p+1$  row elimination will result in excellent model improvement.

### Results and Conclusions

A new ARMA model has been introduced to represent periodic time series. The null space solution of the proposed model has been studied and improvement was shown. The improvement over ARMA obtained from CARMA via SVD is shown in Figures 20 through 25. These plots of power spectra clearly indicate that the regular ARMA model is not optimum here. Additionally, the spectrum of the CARMA and  $L_1$  models have been plotted for comparison, Figures 26 through 27. Further study will be done to apply the CARMA model to multiple periodic signals in additive white noise.

## CHAPTER V

### ESTIMATION OF THE VOCAL TRACT PARAMETERS

#### FROM ARMA MODEL: AN EIGENSYSTEM

#### APPROACH VS. LPC

##### Motivation

An all-pole model is a very good representation of the vocal tract parameters for a majority of speech sounds, but the acoustic theory tells us that nasals and fricatives require both resonances and anti-resonances (poles and zeros). ARMA spectral estimation is used to approximate the spectrum with a filter transfer function containing zeros as well as poles. This model is compared with an LPC model, and its high resolution and accuracy are tested for several male/female speakers.

Assume that the input is a sequence of quasi-periodic impulses or white Gaussian noise for voiced or unvoiced speech, respectively. In order to obtain an optimum set of parameters from short time analysis, the estimator model orders,  $p$  and  $q$ , are set higher than the true model orders,  $\bar{p}$  and  $\bar{q}$ . Singular value decomposition is used to calculate the insensitive feature of the speech signal. The AR parameter vector is obtained from the null space solution of the correlation matrix developed in Chapter III. The MA parameters are calculated from the Blackman-Tukey approach discussed in Chapter IV. To show the capability of the proposed method for resolving the closely spaced formant



frequencies, the spectra obtained from the ARMA and LPC models are compared.

### Introduction

The speech waveform can be modeled as the response of the vocal tract system to a sequence of quasi-periodic impulses or white noise for voiced or unvoiced speech, respectively. The resonances of the vocal tract are called the formants, and usually can be obtained from the spectrum of the vocal system. These formants primarily depend upon the shape of the vocal tract, and change by the position of the articulators. Since the formant frequencies play an important role in characterization of the speech sounds, an accurate and robust method for computing these frequencies would be essential for speech analysis, synthesis, and recognition.

A variety of spectral estimation techniques exist for estimating the vocal tract parameters. Linear predictive coding (LPC) is a powerful and reliable technique and has become the predominant method for estimating the basic speech parameters, i.e., pitch, formants, spectrum, and vocal tract area functions [58]. Several essentially equivalent formulations of LPC are presently available that can be applied for modeling of speech [6, 47, 48]. Wakita inverse filtering is another method for estimating the vocal tract area functions directly from the acoustic speech waveform [77]. In both cases the discrete area functions can easily be obtained from the reflection coefficients. An all-pole model, which has been used in these methods, is a very good representation of the vocal tract for a majority of speech sounds. But, according to the acoustic theory, the effect of anti-resonances (zeros)

needs to be considered as well as resonances, (poles). The pole-zero model can more closely represent the speech system function than can the LPC all-pole model.

Although predictive coding has been used in communication for some time, it was not applied to speech analysis until the early 1970's [46]. It can be used to estimate the frequency and amplitude of the first three formants of all vowel-like segments of the speech signal. The first three peaks of the power spectrum can be considered to be the first three formant frequencies in the ideal case. However, in practice, this condition is rarely satisfied, and peaks may either merge, or spurious peaks may appear due to noise. Thus the identification of the formants from the spectral peaks will be difficult and high resolution spectral estimation using the eigenspace solution via SVD is suggested.

The model that is proposed has both poles and zeros corresponding to the vocal tract characteristics and the effect of the glottal pulse, respectively. An eigensystem analysis of the correlation matrix obtained from the speech samples has been used. It has been shown [44] that the ARMA model parameters of the speech signal can provide a better spectral estimate than AR model parameters. This method utilizes the null space solution proposed in Chapter III.

The artifacts caused by A/D conversion and model inaccuracy only change the singular values of the correlation matrix; however the singular vectors are unchanged. The invariant right singular vectors can be used as an orthogonal basis for the solution space. The space spanned by the right singular vectors corresponding to the smallest

singular values is formed and the AR parameters,  $a_k$ , are obtained from the minimum norm solution proposed in Chapter III.

The speech signal is pre-emphasized and windowed by a Hamming window and a set of AR coefficients is obtained from Equation (4.11). The resulting time series is filtered by a  $p$ th-order moving average filter with  $a_k$  selected as the filter coefficients. The power spectrum of the pre-emphasized speech signal is calculated in the same manner as presented in Chapter IV. The resulting spectrum is compared with the solution computed based on LPC. The ARMA spectral estimation of the speech signal is shown to provide good resolution for the formant frequencies compared with LPC.

#### Pre-processing of the Speech Samples

The analog speech signal sampled at 8000 samples/sec is first pre-emphasized with a first order filter to take into account radiation effects, which appear as a differentiation at low frequencies. The one-zero pre-emphasis filter can be considered as  $1 - \alpha z^{-1}$ , where the parameter  $\alpha$  is in the range of 0.9 to 1.0. Thus, if the vocal tract spectral characteristics are desired, the speech samples should be pre-emphasized according to

$$\tilde{y}(n) = y(n) - \alpha y(n-1) \quad 0.9 \leq \alpha < 1 \quad n = 1, 2, \dots, N. \quad (5.1)$$

The pre-emphasized speech samples are blocked into frames of  $N = 128$  samples (16 msec) with 25 percent overlap ( $L = 4$ ), 32 samples (4 msec). Thus, the  $j$ -th frame of speech can be written as

$$y_j(m) = \tilde{y}(m+(j-1)*32), \quad m = 1, 2, \dots, 128 \quad j = 1, 2, \dots, L+1. \quad (5.2)$$

Each speech sample contributes to L consecutive analysis frames. Each frame is then smoothed by an N-point Hamming window,

$$y_j(n) = w(n)y_j(n) , \quad (5.3)$$

where

$$w(n) = 0.54 - 0.46 \cos \left[ \frac{2\pi n}{N-1} \right] . \quad (5.4)$$

The resulting time series, the smoothed speech data, is then used to estimate the vocal tract parameters.

#### Mathematical Formulation

Assume that the speech samples may be modeled as an ARMA process of order (p,q). A residual error term is included in order to take into account the model inaccuracy and noise introduced by the sampling device. The model can then be formulated as

$$y(n) + \sum_{k=1}^p a_k y(n-k) = \sum_{k=0}^q b_k x(n-k) + e(n) , \quad (5.5)$$

where p and q denote the model orders, e(n) is the uncorrelated residual, and x(n) is zero mean white noise with unit variance. For a perfect model, the corresponding Z-transform is given in (4.4). Thus, the power spectral density function associated with the speech signal is given by the ARMA (p,q) rational form, repeated from (4.5), as

$$S_{\text{ARMA}}(e^{j\omega}) = \left| \frac{b_0 + b_1 e^{-j\omega} + \dots + b_q e^{-j\omega q}}{1 + a_1 e^{-j\omega} + \dots + a_p e^{-j\omega p}} \right|^2 . \quad (5.6)$$

It has been shown that the extended Yule-Walker equations can be simplified as

$$\sum_{k=0}^p a_k r_y(n-k) = 0 \quad \text{for } n \geq q+1, \quad (5.7)$$

where the autocorrelation lags,  $r_y(n)$ , are estimated as

$$r_y(n) = \frac{1}{N} \sum_{k=1}^{N-n} y(k+n)y(k) \quad n=0,1, \dots, N-1. \quad (5.8)$$

Since the correlation lags calculated from (5.8) are biased and less accurate for the higher lag indices, a long correlation method is proposed to estimate the correlation lags from the data of the same length. First,  $N$  samples of the pre-emphasized speech are smoothed by an  $N$ -point Hamming window,  $w_1(n)$ ,

$$\tilde{y}_1(n) = y(n)w_1(n) \quad n = 1, 2, \dots, N. \quad (5.9)$$

Then,  $N+m$  samples of the pre-emphasized speech are smoothed by an  $N+m$ -point Hamming window,  $w_2(n)$ ,

$$\tilde{y}_2(n) = y(n)w_2(n) \quad n = 1, 2, \dots, N+m, \quad (5.10)$$

where  $m$  is the highest correlation lag index required for the modeling. Thus the long correlation lags can be formulated as

$$r_y(n) = \frac{1}{N} \sum_{k=1}^N y_1(k)w_1(k)y_2(k+n)w_2(k+n). \quad (5.11)$$

It has been shown [44] that the long correlation method gives a better spectral estimation as compared to methods typically used. Since the ARMA model order parameters are not known a priori, an extended order ARMA  $(p,q)$  model is assumed, where  $p > \bar{p}$  and  $q > \bar{q}$ .

Equation (5.7) can be put into matrix form, repeated from (4.7) as,

$$R \underline{a} = 0. \quad (5.12)$$

The SVD analysis of the correlation matrix  $R$  shows that the important information of the speech samples is contained in the first  $\bar{p}$  dominant singular values, and the rest of the singular values are considered as the variance of the artifacts due to the sampling process and error from the A/D conversion filter. For practical applications, only  $\bar{p}$  singular values corresponding to  $\bar{p}/2$  formant frequencies are nonzero. Thus, the correlation matrix  $R$  can be replaced by its best rank approximation, Equation (4.10).

The power spectrum of the speech signal is obtained from the null space solution of the correlation matrix  $R$ , the same way as discussed in Chapter IV. This is done by replacing the circular correlation lags with the correlation lags calculated from either Equation (5.8), the regular correlation lags, or Equation (5.11), the long correlation lags.

Once the power spectrum of the speech samples is calculated, the vocal tract formant frequencies can be determined from the definite peaks in the log power spectrum. These formants also can be used to calculate the vocal tract functions. In order to obtain some basis for comparison, the LPC forward-backward linear prediction has been used.

#### Forward-Backward Prediction of Speech

Assume that the speech samples can be predicted from the linear combination of the past speech samples according to

$$e_f(n) = y(n) + \sum_{k=1}^p a_k y(n-k) , \quad (5.13)$$

where the prediction error,  $e_f(n)$ , arises from the fact that the given speech samples may not be perfectly predicted by the weighted sum of  $p$  past speech samples. Equation (5.13) can be put into matrix form as

$$\begin{bmatrix}
 y(1) & 0 & \dots & 0 \\
 y(2) & y(1) & 0 & \dots & 0 \\
 \cdot & & & & \cdot \\
 \cdot & & & & \cdot \\
 y(p+1) & y(p) & \dots & y(1) \\
 \cdot & & & \cdot \\
 \cdot & & & \cdot \\
 y(N) & y(N-1) & \dots & y(N-p)
 \end{bmatrix}
 \begin{bmatrix}
 a_0 \\
 a_1 \\
 \cdot \\
 \cdot \\
 \cdot \\
 a_p
 \end{bmatrix}
 =
 \begin{bmatrix}
 e_f(1) \\
 e_f(2) \\
 \cdot \\
 \cdot \\
 \cdot \\
 e_f(p)
 \end{bmatrix}
 \quad (5.14)$$

or more compactly

$$Y_f \underline{a} = \underline{e}_f, \quad (5.15)$$

where  $Y_f$  is the  $N \times (p+1)$  data matrix,  $\underline{a}$  is the  $(p+1) \times 1$  forward prediction coefficient vector whose first component is constrained to be one, and  $\underline{e}_f$  is the  $(p+1) \times 1$  forward prediction error. If the forward prediction model is compatible with the speech samples, the prediction error,  $e_f(n)$ , will be approximately zero for  $n \geq p+1$ . Upon deleting the first  $p$  rows of the data matrix, Equation (5.15) can be written as

$$Y_f^{(p)} \underline{a} = \underline{\epsilon}, \quad (5.16)$$

where  $Y_f^{(p)}$  is the  $(N-p) \times (p+1)$  Toeplitz structured data matrix. SVD is used to obtain the minimum norm solution as defined in (4.11) and proposed in Chapter III.

Similarly, the backward prediction model of the speech samples can be written as

$$e_b(n) = y(n) + \sum_{k=1}^p a_k y(n+k). \quad (5.17)$$

Equation (5.17) can be put into matrix form as

$$\begin{bmatrix}
 y(1) & y(2) & & \dots & y(p+1) \\
 \cdot & & & & \\
 \cdot & & & & \\
 y(N-p) & y(N-p+1) & & \dots & y(N) \\
 y(N-p+1) & y(N-p+2) & & \dots & 0 \\
 \cdot & & & & \\
 \cdot & & & & \\
 \cdot & & & & \\
 y(N-1) & y(N) & 0 & \dots & 0 \\
 y(N) & 0 & 0 & \dots & 0
 \end{bmatrix}
 \begin{bmatrix}
 a_0 \\
 a_1 \\
 \cdot \\
 \cdot \\
 \cdot \\
 \cdot \\
 a_p
 \end{bmatrix}
 =
 \begin{bmatrix}
 e_b(1) \\
 e_b(2) \\
 \cdot \\
 \cdot \\
 \cdot \\
 \cdot \\
 e_b(p)
 \end{bmatrix}
 \quad (5.18)$$

or more compactly

$$Y_b \underline{a} = \underline{e}_b, \quad (5.19)$$

where  $Y_b$  is the  $N \times (p+1)$  backward data matrix, and  $\underline{e}_b$  is the  $(p+1) \times 1$  backward prediction error vector. The backward prediction error  $e_b(n)$  will be approximately zero for  $n \leq N-p$  if the speech samples are compatible with the backward prediction model. Upon deleting the last  $p$  rows of the backward data matrix, Equation (5.19) can be written as

$$Y_b^{(p)} \underline{a} = \underline{\epsilon}, \quad (5.20)$$

where  $Y_b^{(p)}$  is the  $(N-p) \times (p+1)$  Toeplitz structured data matrix. SVD is also used to obtain the minimum norm solution as defined in (4.11).

Equations (5.16) and (5.20) are combined as

$$\begin{bmatrix}
 Y_f^{(p)} \\
 Y_b^{(p)}
 \end{bmatrix}
 \underline{a} = \underline{e} \quad (5.21)$$

or more compactly,

$$D \underline{a} = \underline{e}. \quad (5.22)$$



It has been shown in Chapter III that the minimum norm solution obtained from (5.22) has better spectral resolution than (5.16) or (5.20).

The spectrum obtained from LPC and forward-backward prediction has been compared for several male/female speakers. The results have shown that the ARMA modeling of the pre-emphasized speech samples can produce a high resolution spectrum, including the case for two formant frequencies which are closely spaced. This method compares favorably with other methods of spectral estimation of speech. Since the modeling error and other artifacts are assumed to be zero mean white processes, they only bias the diagonal elements of the first  $p+1$  rows of the correlation matrix. Upon deleting these rows, the effect of noise has been removed or greatly reduced. Of course, the trade-off between the noise effect and bias involved in higher lags needs to be considered. In long correlation lags, since all lags are estimated from the same data length, the  $p+1$  row elimination will result in excellent model improvement.

#### Results and Conclusions

As ARMA modeling of the speech samples has been used to represent the resonances and anti-resonances (poles and zeros) which exist in the vocal tract system function. The speech samples are pre-emphasized in order to remove the radiation effect and obtain a better representation of the vocal tract system function. Moreover, the long correlation method with SVD improved the modeling over the regular correlation method. The null space solution of the ARMA model has been used and improvement is shown in Figure 28 through Figure 35. The power spectrum of the vowel portion /a/ of "cat" has been plotted for comparison. The

ARMA model via SVD has resolved all four formant frequencies, while LPC and forward-backward prediction failed to resolve those two formant frequencies which are closely spaced. The number of the dominant formant frequencies is calculated from the nullity algorithm developed in Chapter III.

An accurate model of the nasal and fricatives sounds requires ARMA modeling. Since a zero can accurately be represented by infinite poles, an all pole model cannot be used to resolve two closely spaced formant frequencies. Moreover, the moving average process provides additional filtering of the data and extended orders ARMA model facilitates more degrees of freedom than AR and LC.

## CHAPTER VI

### SPEAKER-INDEPENDENT WORD RECOGNITION: AN LPC

#### FEATURE EXTRACTION APPROACH VS NON-LINEAR

#### SPECTRAL MATCHING

##### Motivation

An LPC based word recognition technique is used to extract the feature vectors from an utterance spoken by several different speakers. First, the utterance is divided into L segments with equal samples. Then, the beginning and the endpoint of the utterance are obtained by using short time zero crossing and energy for each segment. The LPC feature vectors are selected as zero crossing (ZC), energy (ENG), normalized residual error (ERRN), LPC coefficients, and the normalized correlation coefficients. Nonlinear spectral matching is used to minimize the inter-speaker variability. This method maps the spectrum of the input speaker into the spectrum of the reference speaker via dynamic time warping (DTW).

The spectrum of the input speaker and reference speakers are obtained using the proposed null space solution of the input and reference correlation matrix, respectively. A new distance measure is developed based on the eigenanalysis of the correlation matrix. The experiment was performed for 25 isolated words uttered by 5 different speakers. The recognition performance of the nonlinear spectral

matching technique based on the proposed eigenspace solution and new distance measure was better than that of the LPC approach.

### Introduction

A variety of approaches have been introduced for speaker-independent isolated word recognition including [19], [24], [54], [59], [79]. These methods utilize the same statistical pattern recognition model as is depicted in Figure 36. The classical pattern matching approach using dynamic time warping is still popular, in spite of new approaches in speech recognition using the Markov model and a phonetic knowledge base. However, the speech signal has a special characteristic that is different from any classical pattern analysis signals. The inherent dialectal and physiological differences among speakers makes the task of word recognition difficult. Each speaker has a different vocal tract resulting in each speaker having his own peculiar speech characteristics. Speakers also emphasize different parts of an utterance and have different accents. This inter-speaker variation causes difficulty for speaker independent word recognition. This problem may be alleviated by the projection of the representation of two speakers under some invariant canonical form. The modified spectrum of an input speaker can be mapped onto a modified reference speaker via DTW so that the high frequency region that depends on the individual speaker becomes narrow compared to those on a linear frequency scale [51].

With the linear predictive coding (LPC) approach, the speech signal is first divided into  $L$  segments, of an equal number of samples. The endpoints of the utterance are detected by using the energy and zero crossing of each segment. Each segment is then weighted by a Hamming

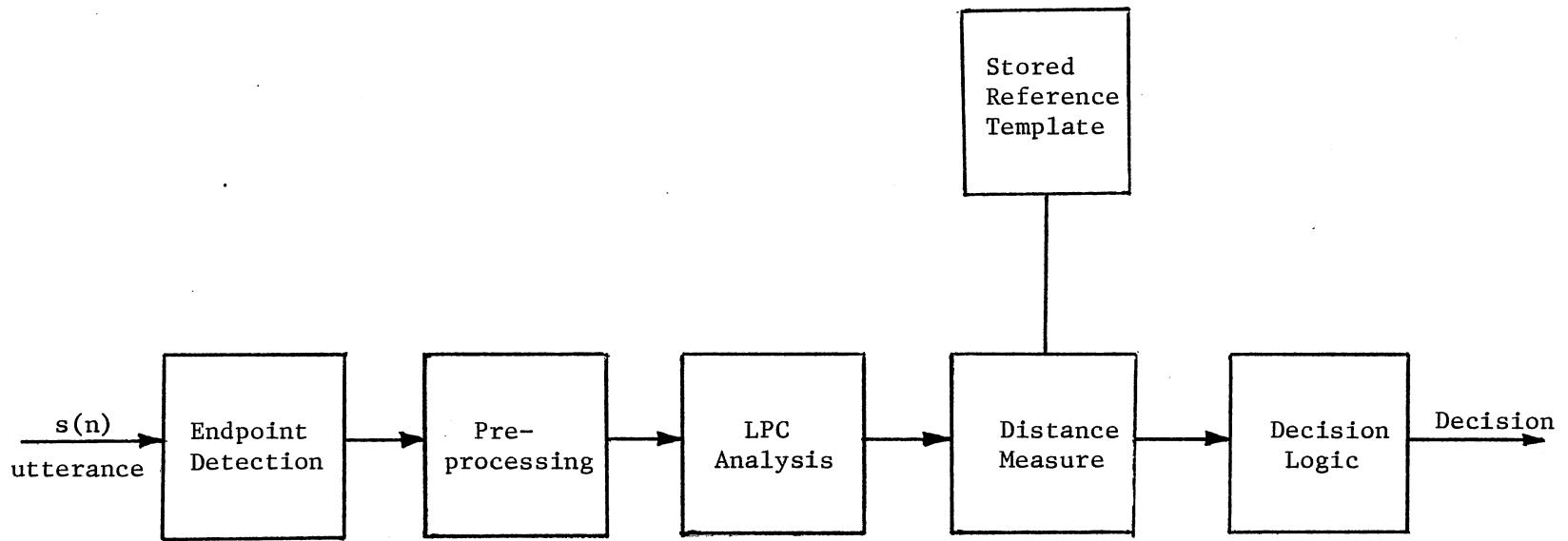


Figure 36. Block Diagram Representation of Word Recognition System.

window and the short time spectrum of the signal is obtained. The resulting feature vectors are compared to the pre-stored feature vectors using some distance criterion. It has been shown that the Itakura distance measure is better for word recognition than the Euclidean distance measure. DTW is used to align the reference word and the input word patterns. The resultant dynamic time aligned reference and input word patterns are compared with some threshold for a word recognition decision [64].

In nonlinear spectral matching the power spectrum of the input and reference word patterns are first calculated from the eigenanalysis of their correlation matrices via SVD (See the proposed spectral analysis of the speech signal in Chapter V). The spectra are then approximated using a least squares fit. The modified spectrum is then computed from the difference between the analyzed spectrum and the least square fit for the voiced region, and as a linear combination of the spectrum in each frame for the unvoiced region. By using the modified spectrum, some of the inter-speaker variabilities in the glottal characteristic can be eliminated. The spectral distance with frequency warping for speaker normalization is first calculated. Then the time warping is carried out via DTW for time normalization. This method, which utilizes the high resolution spectral estimation technique proposed in Chapter V, provides a better word recognition capability than LPC for the spectra with two closely spaced formants.

#### Pre-Processing of the Isolated Words

Analog utterances are first sampled at a rate of 10 KHZ. The zero-mean sampled utterance is obtained by computing the mean over the entire

signal, then subtracting it from the sampled value. The zero-mean signal exhibits greatly reduced artifacts due to dc offset in the A/D converter and 60 HZ hum in the signal. This signal is pre-emphasized with a first order filter, discussed in Chapter V, to account for radiation effects. The pre-emphasized speech samples are blocked into frames of  $N=128$  samples (12.8 ms) with no overlap. Thus the  $j$ -th frame of the utterance can be written as

$$S_j(m) = \tilde{S}(m+(j-1)N) \quad m = 1, 2, \dots, 128 \quad (6.1)$$

where  $N$  is the frame length, and  $\tilde{S}(m+(j-1)N)$  is the pre-emphasized signal. Each frame is then smoothed by an  $N$ -point Hamming window. The smoothed utterances are used as input to the word recognition algorithms.

#### Endpoint Detection

The reference words are divided into  $L$  segments with equal samples. Each segment is then represented by the same size feature vector. If the frame size is selected to be small, it will influence the accuracy of the estimated word boundaries. Experiments have shown that the word recognition endpoint detection algorithm cannot be performed reliably enough. Errors in these regions propagate into all other segments, which leads to high distance measurement between the input word and the reference word. The effect of error caused by unprecisely determined word boundaries can be reduced to some extent by enlarging the segments. The inner region is then divided into non-overlapping segments. Notice that in the process of estimating the word boundaries, not only small uncertainties are encountered, but also

severe distortions are caused by missing final syllables. Additional noise is introduced by sources such as aspiration and lip smacks. These effects may exceed a duration of 180-200 msec and simply cannot be compensated by enlarging the final segments.

Thus, the method of nonlinear spectral matching may alleviate the existing problem. The other methods that require a modified version, i.e., an enlarged or shortened version, of the words in the lexicon occupy a large amount of memory and are not suggested for large vocabulary word recognition. Tables V through IX show the endpoints of the 25 isolated words spoken by 5 different speakers, calculated using the zero crossing and energy in each segment. A segment is removed from the endpoint of the utterance if the energy of the segments at the endpoint is less than the energy threshold but the zero crossing is above the zero crossing threshold. Otherwise, the segment is assumed to be part of the word.

#### Feature Extraction Using LPC

The linear prediction coefficients and the autocorrelation values can be used to obtain the formant frequencies, the spectral envelopes, etc. It has been shown [44] that the power spectrum obtained from the LPC method deteriorates when two peaks either merge, or spurious peaks appear. In this case, the first three peaks of the power spectrum do not correspond to the first three formant frequencies. Thus, the linear prediction coefficients cannot be used very successfully as feature parameters for isolated word recognition without supplemental features. It is known that the zero crossing and the energy of each segment convey much useful information and can be selected as additional



feature parameters. Other useful features are the normalized residual error, and the normalized correlation coefficients.

The choice of distance measure, a measure of distance between the input and reference feature vectors, has significant effect on the performance of the isolated word recognition technique. It has been shown [24] that the LPC feature extraction algorithm is most successful when the Itakura distance measure is used. A new distance measure is proposed based on the eigenanalysis of the input correlation matrix. It has been shown that the proposed distance measure gives a better result than the absolute norm or the Euclidean distance measure.

#### Distance Measure Calculation

In LPC analysis the utterance is divided into L segments of equal samples and smoothed by a Hamming window. The speech samples in each segment are approximated as a linear combination of the past speech samples as

$$\hat{S}(n) = \sum_{k=1}^p a_k S(n-k), \quad (6.2)$$

where the  $a_k$  are called the predictor coefficients and  $p$  is the order of the predictor. The model order is assumed to be  $p=12$ . A set of  $a_k$  parameters is obtained for each segment using the Levinson-Durbin algorithm. The input and reference feature vectors are formed from the LPC coefficients, ZC, ENG, ERRN, and the normalized correlation coefficients.

The distance measure between the input and reference feature vectors are calculated according to

$$d(\underline{C}_i, \underline{C}_r) = \log \left( \frac{\underline{C}_r^T \Lambda \underline{C}_r}{\underline{C}_i^T \Lambda \underline{C}_i} \right) \quad (6.3)$$

where

$$\Lambda = \text{diag} [\sigma_1, \sigma_2, \dots, \sigma_k], \quad (6.4)$$

and  $\sigma_1, \sigma_2, \dots, \sigma_k$  are the singular values of the input correlation matrix, and  $k$  is the dimension of the feature vectors. This new distance measure provides an improved word recognition capability when compared to the absolute norm or Euclidean distance measure.

#### Dynamic Time Warping

Dynamic time warping has been demonstrated to be the most effective method for a speaker-independent isolated word recognition system. It can be used to reduce the speaking rate variation via time normalization. Linear transformations are known to be unable to eliminate the timing difference between speech patterns. The timing difference between two speech patterns is caused by speaker intonation, stress, etc. It can be eliminated by warping the time axis of one pattern so that the maximum coincidence is attained with the other [64]. The time-normalized distance is calculated as the minimized residual distance between these two patterns. This minimization process has been done efficiently by use of dynamic programming (DP). A constraint is applied on the slope of the warping function because too much emphasis on a warping function may result in poor discrimination between the input and reference words. Several DP algorithms were introduced by Sakoe and Chiba [64]. Empirical results and simulations show that no slope constraint will result in better word recognition

when the length of the input word is more than twice the reference word. The optimum DTW algorithm used here is suggested by Sakoe and Chiba [64] as follows

Initialization:

$$D(1,1) = 2d(1,1). \quad (6.5)$$

Dp-equation:

$$D(i,j) = \min \begin{bmatrix} D(i-1,j-2) + 2d(i,j-1) + d(i,j) \\ D(i-1,j-1) + 2d(i,j) \\ D(i-2,j-1) + 2d(i-1,j) + d(i,j) \end{bmatrix} \quad (6.6)$$

Adjustment Window:

$$j-r \leq i \leq j+r \quad (6.7)$$

Time-normalization distance:

$$D(A,B) = \frac{1}{N} D(I,J), \quad \text{where } N = I + J, \quad (6.8)$$

and  $d(i,j)$  is the local distance between input word and reference words,  $D(i,j)$  is the global distance, and  $I, J$  are the number of frames in the input and reference words.

### Nonlinear Spectral Matching

The vocal tract-length varies among speakers and these variations affect the formant frequencies. The inter-speaker variabilities in the glottal characteristics can be eliminated by the unconstrained endpoint dynamic warping algorithm in the frequency region. By using this algorithm, the speech spectrum is optimally shifted in frequency without vocal tract length estimation. The speech spectrum is first obtained by

a high resolution spectral estimation via eigenspace analysis of the correlation matrix, as proposed in Chapter V. The input spectrum,  $x_i$ , is approximated by a least square fit [51].

$$y_i = \alpha x_i + b, \quad (6.9)$$

where  $\alpha$  is the slope of the analyzed spectrum,  $y_i$  is the approximated spectrum, and  $i$  is the channel number. The value of  $\alpha$  is mostly negative for the voiced segment but positive for the fricative segments. The spectrum is then modified as follows:

$$z_i = x_i - y_i \quad \alpha < 0 \text{ voiced}, \quad (6.10)$$

$$z_i = x_i - \frac{1}{N} \sum_{k=1}^N x_k \quad \alpha \geq 0 \text{ fricative}, \quad (6.11)$$

where  $N$  is the number of the frame in the analysis, and  $z_i$  is the modified input spectrum. Similarly, the spectrum of the reference word,  $r_i$ , is modified, and the spectral distance in every frequency range is calculated as follows:

$$q(i,j) = | z_i - r_i |. \quad (6.12)$$

The recursive equation for speaker normalization is calculated according to

$$p(i,j) = \min \begin{bmatrix} p(i-1,j) + q(i,j) \\ p(i-1,j-1) + 2q(i,j) \\ p(i,j-1) + q(i,j) \end{bmatrix}. \quad (6.13)$$

Once dynamic warping in frequency is performed and frequency normalization is obtained; the high frequency region, that depends on the individual, becomes narrow compared to those in the linear frequency

scale. Time warping in word matching is also carried out for time normalization, where the local distance,  $d(i,j)$ , is the spectral distance in the frequency normalization process.

### Results and Conclusions

An LPC based feature extraction is used for speaker-independent isolated word recognition. A new distance measure based on the eigenanalysis of the input correlation matrix is proposed and improvement over the Euclidean distance measure is demonstrated. It is known that the Itakura distance measure requires the Gaussian assumption on the distribution of the LPC coefficients. This assumption is often weak for the distribution of the feature vector, and the proposed distance measure at a cost of eigensystem analysis is suggested. Nonlinear spectral matching is performed, and significant improvement is obtained over the LPC method. Tables X through XII show the results. The spectrum used in the nonlinear spectral matching is obtained from the null space solution of the correlation matrix, introduced in Chapter V. Further research is needed to obtain an optimal dynamic time warping method when the length of the input word is more than twice the reference word.

TABLE V  
 THE LENGTH OF 5 DIFFERENT WORDS SPOKEN  
 BY SPEAKER NO. 1 (MALE)

	No. of frames	Beginning frame	Ending frame
Colt	58	3	56
Moosehead	80	1	80
Star	105	14	103
Drink	68	1	68
Taste	94	21	83

TABLE VI  
 THE LENGTH OF 5 DIFFERENT WORDS SPOKEN  
 BY SPEAKER NO. 2 (MALE)

	No. of frames	Beginning frame	Ending frame
Colt	72	12	71
Moosehead	100	2	98
Star	104	2	102
Drink	74	1	73
Taste	92	1	92

TABLE VII  
 THE LENGTH OF 5 DIFFERENT WORDS SPOKEN  
 BY SPEAKER NO. 3 (MALE)

	No. of frames	Beginning frame	Ending frame
Colt	58	7	55
Moosehead	74	1	74
Star	68	2	68
Drink	48	5	48
Taste	60	5	58

TABLE VIII  
 THE LENGTH OF 5 DIFFERENT WORDS SPOKEN  
 BY SPEAKER NO. 4 (FEMALE)

	No. of frames	Beginning frame	Ending frame
Colt	64	1	62
Moosehead	86	1	81
Star	93	4	93
Drink	50	2	50
Taste	78	2	75

TABLE IX  
 THE LENGTH OF 5 DIFFERENT WORDS SPOKEN  
 BY SPEAKER NO. 5 (FEMALE)

	No. of frames	Beginning frame	Ending frame
Colt	96	2	96
Moosehead	98	5	98
Star	94	1	90
Drink	72	4	71
Taste	80	2	78

TABLE X  
 SPEAKER NO. 2 (MALE): WORD RECOGNITION FAILURE OF  
 1. LPC USING EUCLIDEAN NORM  
 2. ARMA USING NEW DISTANCE MEASURE CRITERION

	Colt	Moosehead	Star	Drink	Taste
Colt	✓	—	—	—	—
Moosehead	—	✓	—	—	—
Star	—	✓ ?	×	—	—
Drink	—	—	—	✓	—
Taste	—	—	—	—	✓



TABLE XI

SPEAKER NO. 4 (FEMALE): WORD RECOGNITION FAILURE  
OF LPC USING CLIDEAN NORM

	Colt	Moosehead	Star	Drink	Taste
Colt	×	-	-	√ ?	-
Moosehead	-	√	-	-	-
Star	-	-	√	-	-
Drink	-	-	-	√	-
Taste	-	-	-	-	√

TABLE XII

SPEAKER NO. 5 (FEMALE): WORD RECOGNITION FAILURE  
OF LPC USING EUCLIDEAN NORM

	Colt	Moosehead	Star	Drink	Taste
Colt	×	-	-	√ ?	-
Moosehead	-	√	-	-	-
Star	-	-	√	-	-
Drink	-	-	-	√	-
Taste	-	-	-	-	√

## CHAPTER VII

### ESTIMATING THE FREQUENCIES OF MULTIPLE SINUSOIDS

#### IN WHITE NOISE: A CARMA MODEL APPROACH

#### VS LINEAR PREDICTION

##### Motivation

Traditional forward-backward linear prediction (FBLP) can resolve the frequencies of multiple sinusoids from the spectral peaks in a relatively high SNR environment. When the observed data is corrupted with strong additive noise, SNR is low, and the data length is short, the regular FBLP is unable to detect the closely spaced sinusoids from the spectral peaks.

A CARMA model based on the circular FBLP (CFBLP) and circular correlation matrix is proposed to improve the spectral resolution of the estimated frequencies. This method utilizes the inherent periodic characteristic possessed by the time series observations. The SVD is used to obtain the proposed null space solution of the circular correlation matrix and CFBLP data matrix. The estimated frequencies obtained from a short data record are compared with the linear prediction and Tuft-Kumaresan method [75]. The periodic characteristic of the underlying time series, along with the null space solution of the CARMA model, is the source of improvement over the traditional methods.

## Introduction

The estimation of frequencies of multiple sinusoids in additive white noise has been studied by many system analysts. The traditional periodogram is often used for a large data record at high SNR. When the sinusoidal signals are closely spaced in frequency, or the data length is short, the periodogram is not able to detect these frequencies. A variety of procedures based on linear prediction models can be employed to increase the spectral resolution of closely spaced sinusoids. Nuttall [52], and Ulrych and Clayton [76] developed the FBLP to obtain the frequencies of closely spaced sinusoids from a short data length when the SNR is sufficiently high. Papoulis [53] proposed the adaptive data extrapolation technique to obtain the high resolution spectrum for closely spaced sinusoidal signals. However, in practice, the observed time series, a sum of sinusoids, is corrupted with additive noise, and these methods are not capable of obtaining the high resolution spectrum.

Tufts and Kumaresan [75] proposed a method based on FBLP which provides a significant improvement in spectral resolution of two closely spaced sinusoids if the SNR is sufficiently high and the number of sinusoids are known. Should the additive noise be strong, low SNR, or the number of sinusoids be unknown, their method will not perform very well.

An alternative method is proposed which is based on the eigensystem analysis of the CFBLP and of the related circular correlation matrix. This method utilizes the null space solution of the circular models, due to inherent periodic characteristics possessed by the time series, to improve the spectral resolution [43]. The empirical studies and simulated results show the high resolution capability of the proposed

method when the data length is short and the additive noise is Gaussian. This new procedure provides a significant improvement in a mean square error sense relative to the Tufts-Kumaresan method at low SNR. Moreover, the number of sinusoids is calculated from an accurate nullity algorithm based on the singular values of either the CFBLP data matrix or the circular correlation matrix, as developed in Chapter III.

The Gaussian distribution assumption for the additive noise is not appropriate for many applications. The additive noise is modeled by a mixture of two Gaussian distributions, one with a long tail, in order to account for any unpredictable outliers in the time series observations, sum of sinusoids in additive noise. The probability density of this contaminated Gaussian or the mixed Gaussian noise,  $f_n(n)$  is given according to [23]

$$f_n(n) = (1-\alpha) f_w(w) + \alpha f_v(v), \quad (7.1)$$

where  $\alpha$  is called the mixture parameter,  $f_w(w)$  is the probability density of a zero mean Gaussian random variable with variance  $\sigma_{ww}^2$ , and  $f_v(v)$  is the probability density of a Gaussian random variable with a non-zero mean,  $\mu$ , and variance  $\sigma_{vv}^2$ ,  $\sigma_{vv}^2 \gg \sigma_{ww}^2$ . The value of  $\alpha$  is often unknown,  $\alpha \ll 1$ . Should the  $\alpha$  be zero, the additive noise is assumed to have an exact Gaussian distribution. In this thesis, the value of  $\alpha$  is assumed to be zero. The mixed Gaussian case is under investigation and a possible solution is proposed in Chapter VIII.

In summary, the following steps are taken to perform the analysis. First, the CFBLP data matrix is formed. Then, the SVD is used to obtain the minimum norm solution based on the linear combination of those right singular vectors that span a basis for the solution

space. The dimension of the subspace is obtained from the proposed nullity algorithm based on the singular values of the CFBLP data matrix. Then, the number of sinusoids is found as equal to the rank of the CFBLP data matrix. Once the number of sinusoids is obtained, the power spectrum of the multiple sinusoids is calculated in the same manner as in Chapter IV, Equation (4.5).

#### Mathematical Model

Assume that a time series (observed data sequence)  $y(n)$  is composed of uniformly spaced samples of  $M$  complex sinusoids in additive white noise,  $\eta(n)$ , such that

$$y(n) = \sum_{k=1}^M a_k e^{j\omega_k n} + \eta(n) \quad 1 \leq n \leq N, \quad (7.2)$$

where  $N$  is the length of the data sequence, and the entities  $a_k$  and  $\omega_k$  represent the sinusoidal amplitudes and radian frequencies, respectively. Moreover, the sinusoidal amplitudes and frequencies are unknown constants and  $\eta(n)$  is a complex white noise process with uncorrelated real and imaginary components such that

$$\eta(n) = w(n) + jv(n). \quad (7.3)$$

We can write

$$E \{w_i w_j\} = \begin{cases} \sigma_{ww}^2 & i=j \\ 0 & i \neq j \end{cases} \quad (7.4)$$

and

$$E \{v_i v_j\} = \begin{cases} \sigma_{vv}^2 & i=j \\ 0 & i \neq j, \end{cases} \quad (7.5)$$

where  $\sigma_{ww}^2$  and  $\sigma_{vv}^2$  are the variance of the real and imaginary components of the additive white noise.

Multiplication of both sides of (7.2) by  $y^*(n+l)$  and taking the expected value yields

$$\begin{aligned} r_y(n) &= E \{y(n)y^*(n+l)\} \\ &= \sigma^2 \delta(n) + \sum_{k=1}^M |a_k|^2 e^{j\omega_k n} . \end{aligned} \quad (7.6)$$

Equation (7.6) can be expressed as

$$R = \sigma^2 I + \sum_{k=1}^M |a_k|^2 \underline{u}_k \underline{u}_k^* , \quad (7.7)$$

where

$$\underline{u}_k = [ 1, e^{j\omega_k}, e^{j2\omega_k}, \dots, e^{jP\omega_k} ] \quad 1 \leq k \leq M \quad (7.8)$$

or more compactly,

$$R = R_n + R_s . \quad (7.9)$$

$R_n$  and  $R_s$  are called the noise and signal covariance matrices, respectively.  $R$  is a  $(p+1) \times (p+1)$  Toeplitz structured correlation matrix. Using the eigenvector characteristic of  $R$ ,

$$R \underline{X}_k = \lambda_k \underline{X}_k , \quad (7.10)$$

the covariance matrix,  $R$ , can be represented by

$$R = \sum_{k=1}^M (\lambda_k - \sigma^2) \underline{X}_k \underline{X}_k^* + \sigma^2 I , \quad (7.11)$$

where  $\sigma^2$  is the variance of the additive noise and  $\lambda_k$  is the eigenvalue of  $R$ . If  $\sigma^2$  is known, then the noise free correlation matrix can be obtained. Upon examining the eigenvalues of  $R$ , one can easily find that

the  $p+1-M$  smallest eigenvalues of the correlation matrix are equal to  $\sigma^2$ . Furthermore, the eigenvectors corresponding to the  $p+1-M$  smallest eigenvalues are orthogonal to the  $M$  sinusoidal vectors and form a basis for the solution space. The form of solution is developed in Chapter III, Equation (3.155).

The exact correlation lags in Equation (7.6) are not available and need to be estimated from the observed data

$$r_y(n) = \frac{1}{N} \sum_{k=1}^N y(k) y^*(k+n), \quad (7.12)$$

where  $+$  denotes modulo  $N$  addition, such that

$$y(k+N) = y(k) \quad \text{for all } k. \quad (7.13)$$

Once the circular correlation lags are estimated from the observed data, the null space solution of the circular correlation can be obtained from Equation (3.155). The inherent periodicity of the multiple sinusoids utilized in (7.12) improves the estimated frequencies of the sinusoids.

Similarly, the circular forward linear prediction (CFLP) model,

$$\epsilon_f^{(c)}(n) = \sum_{k=1}^P a_k y((n-k)), \quad (7.14)$$

and the circular backward linear prediction (CBLP) model,

$$\epsilon_b^{(c)}(n) = \sum_{k=1}^P a_k y(n \oplus k), \quad (7.15)$$

are used to form the proposed CFBLP model. Using the matrix form, Equations (7.14) and (7.15) can be written as

$$Y_f^{(c)} \underline{a} = \underline{\epsilon}_f \quad (7.16)$$

and

$$Y_b^{(c)} \underline{a} = \underline{\varepsilon}_b, \quad (7.17)$$

where  $Y_f^{(c)}$  is the  $N \times (p+1)$  circular forward data matrix,  $Y_b^{(c)}$  is the  $N \times (p+1)$  circular backward data matrix,  $\underline{a}$  is the  $(p+1) \times 1$  forward-backward prediction coefficient vector, and  $\underline{\varepsilon}_f$  and  $\underline{\varepsilon}_b$  are the forward and backward prediction error vectors, respectively. The CFBLP model, given as

$$\begin{bmatrix} Y_f^{(c)} \\ Y_b^{(c)} \end{bmatrix} \underline{a} = \underline{e}, \quad (7.18)$$

is proposed as an alternative method to obtain the high resolution frequency estimates of closely spaced sinusoids. This method will resolve the estimated frequencies of two closely sinusoids when the data length is relative small,  $N=8$ , while the traditional methods of spectral estimation perform poorly.

### Results and Conclusions

A circular ARMA model based on the circular data matrix and the associated circular correlation matrix is proposed to represent the periodic time series, sum of multiple sinusoids in additive white noise. This model estimates the frequencies of multiple sinusoids from a short data length,  $N=8$ . The simulated result is compared with the LP, regular ARMA model, and Tuft-Kumerason method, and improvement is shown, Figures 37 through 44. This model also resolves the frequencies of two closely spaced sinusoids in a low SNR environment. Its high resolution and accuracy over the regular ARMA are shown in Figures 45 and 46.



It is known that the wrong selection of the AR model order may result in spurious peaks in the power spectrum. The AR model order, the number of complex sinusoids, is calculated from an accurate nullity algorithm proposed in Chapter III. Moreover, the periodic characteristic of the time series is used and the null space solution of the circular model is obtained. The frequencies of the multiple sinusoids are obtained from the peaks of the power spectrum. The additive noise is assumed to be Gaussian. The performance of the contaminated or mixed Gaussian case is under investigation and a possible solution is given in Chapter VIII.

## CHAPTER VIII

### SUMMARY AND CONCLUSIONS

New techniques for system identification and ARMA spectral estimation have been proposed. The procedure, which is referred to as modified equation error identification (MEEI), uses a priori knowledge of the additive noise to remove or greatly reduce the bias effect in system modeling. MEEI has been shown to be very effective for estimating the unbiased parameters of an ARMA process based on a set of noisy input-output measurements. In particular, this algorithm is capable of obtaining the unbiased parameter estimates of a time varying system. In addition, this method utilizes on-line identification and does not require a large amount of memory as opposed to off-line identification. Furthermore, new input-output measurements are processed at each sample instant, and the model parameters are updated so as to account for the variations in the system parameters. If the knowledge of the additive noise power is not known or cannot be estimated from a simple algorithm, MEEI does not perform satisfactorily. In this case, unbiased parameter estimates of a model may not be obtained from a set of noisy input-output measurements.

A new procedure based on the eigencharacterization of the input-output correlation matrix has been proposed, and unbiased parameter estimates are obtained without prior knowledge of the additive noise power. This procedure utilizes the singular value decomposition (SVD)

as a powerful and reliable tool for calculating the singular values and the noise invariant characteristics, the right singular vectors, of the correlation matrix. The singular values, which convey useful information about the structure of the correlation matrix and the additive noise power, have been used to improve the stability of ill-conditioned problems by calculating their best rank approximation. Moreover, the right singular vectors corresponding to the smallest singular values, form an orthogonal basis for the solution of the correlation based identification. The dimension of the spanned space, the null space, is obtained from a proposed nullity algorithm based on the singular values of the correlation matrix. The autoregressive (AR) model order, the number of dominant singular values, has been considered as the rank of the correlation matrix, the number of the singular values minus the dimension of the spanned space. The eigenspace (EIGSP) solution or the minimum norm solution has been calculated from the linear combination of the right singular vectors in the null space of the correlation matrix. This solution has been shown to be unbiased without prior knowledge of the additive noise if the proper scaling is used. The results have shown that the proposed null space solution has better performance for estimating the parameters of an autoregressive moving average (ARMA) model than the classical methods of system identification.

A second order ARMA system has been simulated and a set of noisy input-output measurements were used to compare the performance of the MEEI and EIGSP solutions with the solutions obtained from the classical methods of system identification. If the additive noise or disturbance is white, or small, the least square (LS) algorithm will result in a

reasonable parameter estimates and the generalized least square (GLS) algorithm or EIGSP solution can be ignored because of their high computational cost. If the input-output is corrupted with strong additive colored noise, the GLS algorithm can be used to obtain an efficient parameter estimate. The EIGSP solution has been shown to be the most efficient method of parameter estimation when the additive noise is very strong colored noise, i.e. signal-to-noise ratio (SNR) is very low. But, it may result in some extraneous peaks in the spectrum if the nullity of the underlying ARMA model is not known.

The optimum number of rows of the correlation matrix in the over ordered case of ARMA modeling is not known. This optimum number is obtained from the comparison of normalized mean squared error (MSE) of many different trials. More research needs to be done to calculate the optimum number of rows from the statistics of the input-output measurements. The AR model order can be obtained from the nullity criterion algorithms that have been developed to separate the signal subspace and noise subspace. An efficient noise detection algorithm may also be useful to improve the parameter estimation.

A comparison of the spectra and variances of LS, GLS, EIGSP, and instrumental variable (IV) method, Figure 13, has shown that IV is very sensitive to the noise. Although its spectrum is closer to the spectrum of the true system than LS and GLS, the method of IV is not suggested for the noise contaminated input-output measurements. GLS and MEEI can each be used as an effective parameter estimator for the real time process, while the EIGSP solution is an excellent parameter estimator for off-line use. Several applications of the EIGSP solution have been presented in Chapters IV-VIII. The spectra obtained from the proposed

method clearly indicate the high resolution of the EIGSP solution over linear prediction (LP) and other traditional methods of spectral estimation. The additive noise is assumed to be white Gaussian in all analyses. The possible solution for the non-Gaussian case, i.e., impulsive noise or mixed Gaussian case, is suggested in the next section.

A new circular ARMA (CARMA) model has been proposed and its EIGSP solution obtained in the same manner as ARMA. The CARMA model utilizes the periodic characteristics possessed by the time series. It has been shown that the correlation lags estimated from the CARMA model are closer to the true correlation lags due to inherent periodicity of the time series, Figures 48 through 53. This model has been used to represent time series obtained from two dimensional shapes and multiple sinusoids in white Gaussian noise, and improvement over the ARMA model is obtained. The application of the CARMA model for other periodic time series is under investigation.

#### Suggestions for Future Research

Correlation based identification has been studied and the EIGSP solution via SVD has been developed. The performance of the EIGSP solution has been compared with the solutions obtained from several classical methods of system identification. The application of the EIGSP solution for several different time series has been examined. It has been shown that the power spectra obtained from the proposed method compare favorably with the LP and other traditional methods of spectral estimation. The empirical studies and simulated results clearly

indicate the advantages of the proposed method over the classical methods.

Several possibilities exist for expanding the EIGSP solution proposed in this thesis. The following points could be studied in hope of obtaining a more accurate solution with the EIGSP method.

1. The optimum number of rows of the correlation matrix in the correlation based identification is not known. This number has been obtained from the comparison of MSE of many different trials. Should the additive noise power be strong, or the system characteristics be unknown a priori, the optimum number of rows of the correlation matrix cannot be determined via a trial and error procedure. A sub-optimum answer may be obtained. An efficient algorithm based on the statistics of the observations might be used to obtain an estimate of the optimum selection. This algorithm would definitely reduce the time required for calculation, since many different trials are not needed, and the optimum answer can possibly be obtained.

2. The AR model order is obtained from the proposed nullity algorithm based on the singular values of the correlation matrix. The algorithm has obtained the AR model order accurately for several applications at low signal-to-noise-ratio (SNR). Should the SNR be very low or the singular values of the signal subspace and the noise subspace be very close to each other, the nullity algorithm may not be able to determine the AR model order accurately. In these cases, the nullity algorithm needs to be modified by introducing a weight vector based on the arithmetic and the geometric means of the singular values.

3. It has been shown that EIGSP solution provides unbiased parameter estimates for system identification and ARMA spectral

estimation. Should the system parameters be changed or new observations be available, the SVD of the new correlation matrix must be obtained if the estimated model parameters are to follow the variation of the actual system parameters. Since the SVD operation is time consuming, the EIGSP solution of the time varying system is most likely not possible. One possible solution is the eigenanalysis of the perturbed data matrix at every K-sample interval. The correlation lags can be calculated from the proposed recursive relationship

$$r_0^{(k)} = r_0^{(0)} + \sum_{i=1}^k x(N+i)x(N+i) \quad (8.1)$$

$$r_1^{(k)} = r_1^{(0)} + \sum_{i=1}^k x(N-1+i)x(N+i) \quad (8.2)$$

⋮

$$r_L^{(k)} = r_L^{(0)} + \sum_{i=1}^k x(N-L+i)x(N+i), \quad (8.3)$$

where  $r_j^{(k)}$  is the j-th correlation lag at the sample instant k, L is the last correlation lag index, and N is the row dimension of the correlation matrix R. It is easy to see that the new correlation lags at the k-th sample instant are the sum of the old correlation lags and the truncated circular correlation lags. Thus, the correlation matrix at sample instant k,  $R^{(k)}$ , can be written as

$$R^{(k)} = R^{(0)} + R^{(T-c)}, \quad (8.4)$$

where  $R^{(T-c)}$  is the truncated circular correlation matrix. The correlation lags of  $R^{(T-c)}$  are calculated according to

$$r_0^{(T-c)} = \sum_{i=N}^{N+k} x(i)x(i), \quad (8.5)$$

$$r_1^{(T-c)} = \sum_{i=N}^{N+k} x(i-1)x(i), \quad (8.6)$$

⋮

$$r_L^{(T-c)} = \sum_{i=N}^{N+k} x(i-L)x(i). \quad (8.7)$$

The right singular vectors of  $R^{(k)}$  can be obtained from the right singular vector of  $R^{(0)}$  and the right singular vectors of  $R^{(T-c)}$ . More research needs to be done to obtain the recursive EIGSP solution for time varying systems using the special structure of  $R^{(T-c)}$ .

4. In this thesis, the additive noise has been assumed to be a white Gaussian process. The Gaussian distribution assumption may not be appropriate for many applications. The additive noise is modeled by a mixture of two Gaussian processes (Chapter VII), in order to account for any unpredictable outliers in the time series observation. The outliers are referred to as a burst of error of short temporal duration. It is known that an  $L_2$  prediction filter applies equal weights to the observations, and the filter coefficients are very sensitive to the outliers. Yarlagadda, Bednar, and Watt [81] have developed an  $L_p$ ,  $1 < p < 2$ , deconvolution algorithm which is less sensitive to noise bursts. They have shown that the  $L_p$  prediction filters may not be stable in general. One can possibly combine their algorithm with the proposed EIGSP solution and obtain a new procedure for modeling of ARMA processes corrupted with a mixture of two Gaussian noise processes. The possible solution of Equation (4.17), repeated as (8.8),

$$\underline{B}a = \underline{Y}, \quad (8.8)$$

can be obtained from the following algorithm



$$1. \quad \underline{a}(0) = \beta \sum_{k=1}^r \bar{v}_k(1) \bar{v}_k \quad . \quad (\text{repeated from (3.155)}) \quad (8.9)$$

$$2. \quad \underline{r}(k) = \underline{B}\underline{a}(k) - \underline{Y}. \quad (8.10)$$

$$3. \quad \gamma_i(k) = \text{sgn}(r_i(k)) \cdot |r_i(k)|^{p-1} \quad (8.11)$$

4. Minimize  $E(k)$  with respect to  $\Delta_k$  where

$$E(k) = \underline{r}(k) - \Delta_k \begin{pmatrix} p_e+1-s & p_e+1-s \\ \sum_{k=1} \sigma_k \underline{u}_k \underline{v}_k^T & \sum_{k=1} \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^T \end{pmatrix} \underline{\gamma}(k), \quad (8.12)$$

and  $s$  is the nullity of the data matrix  $B$ .

Using iterative least squares (IRLS), gain parameter  $\Delta_k$  can be obtained

$$5. \quad \underline{a}(k+1) = \underline{a}(k) - \Delta_k \begin{pmatrix} p_e+1-s \\ \sum_{k=1} \frac{1}{\sigma_k} \underline{v}_k \underline{u}_k^T \end{pmatrix} \underline{\gamma}(k). \quad (8.13)$$

6. Stop if convergence is achieved, otherwise go to step 2. An alternative selection of  $\Delta_k$  based on the singular values of the data matrix  $B$  is

$$\Delta_k = (1 - \mu \sigma_k)^{k-1} \sigma_{\max}, \quad (8.14)$$

$$\text{where } 0 < \mu < \frac{2}{\sigma_{\max}}. \quad (8.15)$$

5. Although this thesis primarily was concerned with the EIGSP solution of the correlation matrix, a new orthogonal transform, M-transform, is proposed to provide an alternative method for the modeling and identification of an ARMA process. It is known that Fourier series coefficients provide a least square fit to a function approximated as a linear combination of the orthogonal basis, sinusoidal basis functions. These coefficients convey the spectral characterization of the signal and can be used as an alternative method of spectral

estimation. Moreover, the Fourier series coefficients are used to reduce the redundancy in many practical time series. Similarly, the M-transform can be used to approximate the time series signals as a linear combination of non-sinusoidal orthogonal basis functions and provide the M-transform coefficients for the spectral estimation and data compression. The eigenanalysis of the M-transform and determination of the M-transform coefficients are proposed in Appedix B. More research is required to compare the performance of the M-transform with other orthogonal transforms.

## REFERENCES

- [1] Ackenhusen, J. G. and Y. H. Oh, "Single Chip Implementation of Feature Measurement for LPC-Based Speech Recognition", Proc. ICASSP 85, pp. 1445-1448.
- [2] Anderson, T. W., An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.
- [3] Astrom, K. J. and T. Bohlin, "Numerical Identification of Linear Dynamic System from Normal Operating Records," Proc. IFAC Symposium on Self-Adaptive Control Systems, Teddington, England, 1965, pp. 96-110.
- [4] Astrom, K. J. and P. Eykhoff, "System Identification - A Survey", Automatica, 7, pp. 123-162, 1971.
- [5] Astrom, K. J., Introduction to Stochastic Control Theory, Academic Press, New York, 1970.
- [6] Atal, B. S. and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of Speech Wave," J. Acoust. Soc. Am., Vol. 50, No. 2, pp. 637-655, August 1971.
- [7] Blum, J., "Multidimensional Stochastic Approximation Procedures," Ann. Math. Statist., Vol. 25, pp. 737-744, 1954.
- [8] Cadzow, J. A., "Spectral Estimation: An Overdetermined Rational Model Equation Approach," Proc. IEEE, Vol. 70, No. 9, pp. 907-939, September 1982.
- [9] Cadzow, J. A. and O. M. Soloman, Jr., "Algebraic Approach to System Identification," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-34, No. 3, pp. 462-469, June 1986.
- [10] Clarke, D. W., "Generalized Least-Squares Estimation of the Parameters of a Dynamic Model," Proc. IFCA Symposium on Identification in Automatic Control Systems, Prague, Czechoslovakia, Paper 3.17, 1967.
- [11] Cullum, J., "Ill-Posed Deconvolutions: Regularization and Singular Value Decomposition," Proc. of the 19th IEEE Conf. on Decision and Control Including the Symp. on Adaptive Processes, pp. 29-35, 1980.

- [12] Davies, W. D. T., System Identification for Self-adaptive Control, Wiley-Interscience, New York, 1970.
- [13] Deutsch, R., Estimation Theory, Prentice-Hall, Englewood Cliffs, N.J., 1965.
- [14] Dubois, S. R. and F. H. Glanz, "An Autoregressive Model Approach to Two-Dimensional Shape Classification," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 1, pp. 55-66, January 1986.
- [15] Dvoretzky, A., "On Stochastic Approximation," Proc. 3rd Berkeley Symp. Math. Statist. and Prob., pp. 39-55, Univ. of California Press, Berkeley, CA. 1956.
- [16] Eykhoff, P., System Identification and State Estimation, Wiley, London, 1974.
- [17] Eykhoff, P., "Process Parameter and State Estimation," Proc. IFAC Symposium, on Identification in Automatic Control Systems, Prague, Paper 2, 1967.
- [18] Goodwin, G. C. and R. L. Payne, Dynamic System Identification, Academic Press, New York, 1977.
- [19] Gupta, V. N., J. K. Bryan, and J. N. Gowdy, "A Speaker-Independent Speech Recognition System Based on Linear Prediction," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, No. 1, pp. 27-33, February 1978.
- [20] Hastings, J. R. and M. W. Sage, "Recursive Generalized Least Squares Procedures for one-line Identification of Process Parameters," Proc. IEEE, pp. 2057-2062, 1969.
- [21] Henderson, T. L., "Geometric Method for Determining System Poles from Transient Response," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-29, pp. 982-988, October 1981.
- [22] Hsia, T. S., System Identification, Lexington Books, Lexington, MA, 1977.
- [23] Huber, P. J., Robust Statistics, John Wiley & Sons, New York, 1981.
- [24] Itakura, F., "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-23, No. 1, pp. 67-72, February 1975.
- [25] Johnson, C. R. Jr., "Adaptive Parameter Matrix and Output Vector Estimation via an Equation Error Formulation," IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-9, No. 7, pp. 392-397, July 1979.

- [26] Johnson, C. R. Jr., A. L. Hamm, and J. R. Treichler, "Equation Error Frequency Estimation Bias for a White Noise Obscured Sinusoidal," Proc. 12th Asilomar Conf. Circuit, System, and Computers, pp. 132-136, Pacific Grove, CA., 1978.
- [27] Johnson, C. R. Jr., D. A. Lawrence, T. Taylor, and M. V. Malakooti, "Simulated Lumped-Parameter System Reduced-Order Adaptive Control Studies," Department of Electrical Engineering, Virginia Poly. Tech. Instit. and State University, March 1981.
- [28] Kalaba, R. and K. Spigarn, Control, Identification and Input Optimization, Plenum Press, New York, 1982.
- [29] Kalman, R. E., "A new Approach to Linear Filtering and Prediction Problems," Trans. on American Society of Mechanical Engineering, J. Basic Eng., Vol. 820, pp. 35-45, 1960.
- [30] Kalman, R. E. and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," Trans. on American Society of Mechanical Engineering, J. Basic Eng., Vol. 83D, pp. 95-108, 1961.
- [31] Kashyap R. L. and R. Chellappa, "Stochastic Models for Closed Boundary Analysis: Representation and Reconstruction," IEEE Trans. on Information Theory, Vol. IT-27, No. 5, pp. 627-637, September 1981.
- [32] Kashyap R. L. and A. Khotanzad, "A Model-Based Method for Rotation Invariant Texture Classification," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 4, pp. 472-481, July 1986.
- [33] Kendel, M. G. and A. Stuart, "The Advanced Theory of Statistics, Griffin, London, Vol. 2, pp. 398-408, 1961.
- [34] Kiefer, J. and J. Wolfowitz, "Statistical Estimation of the Maximum of a Regression Function," Ann. Math. Statist., Vol. 23, pp. 462-466, 1952.
- [35] Kung, S., "A New and Model Reduction Algorithm via Singular Value Decomposition," Conf. Record of the 12th Asilomar Conf. on Circuit, System, and Computers, pp. 705-712, Pacific Grove, CA., 1978.
- [36] Landau, I. D., "A Survey of Model Reference Adaptive Techniques (theory and applications)," Automatica, Vol. 10, pp. 355-399, July 1974.
- [37] Landau, I. D., "Unbiased Recursive Identification Using Model Reference Adaptive Techniques," IEEE Trans. on Automatic Control, Vol. AC-21, No. 2, pp. 194-202, April 1976.
- [38] Landau, Y. D., Adaptive Control: The Model Reference Approach, Marcel Dekker Inc., New York, 1979.

- [39] Lawson, C. L. and R. J. Hanson, Solving Least-Square Problems, Englewood Cliffs, N.J., Prentice-Hall, 1974.
- [40] Levin, M. J., "Optimum Estimation of Impulse Response in the Presence of Noise," IRE Trans. on Circuit Theory, CT-7, pp. 50-56, 1960.
- [41] Ljung, L., "On Positive Real Transfer Functions and the Convergence of Some Recursive Schemes," IEEE Trans. on Automatic Control. Vol. AC-22, No. 4, pp. 539-551, August 1977.
- [42] Luenberger, D. G., Optimization by Vector Space Methods, John Wiley & Sons, Inc., N.Y., 1969.
- [43] Malakooti, M. V. and K. A. Teague, "CARMA Model Method of Two-Dimensional Shape Classification: An Eigensystem Approach Vs.  $L_p$  Norm," Proc. ICASSP 87, pp. 583-586.
- [44] Malakooti, M. V., "Estimation of the Vocal Tract Parameters from ARMA Model: An Eigensystem Approach vs LPC," Proc. of IEEE Region 5 Conference, pp. 30-35, March 1987.
- [45] Malakooti, M. V. and P. Baltas, "Comparison of the Truncated SVD With Least Squares and Generalized Least Squares," Unpublished Research Project, Department of Electrical Engineering, Arizona State University, May 1984.
- [46] Makhoul, J., "Spectral Analysis of Speech by Linear Prediction," IEEE Trans. on Audio and Electroacoustic, Vol. AU-21, No. 3, June 1973.
- [47] Makhoul, J., "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, pp. 561-580, 1975.
- [48] Markel J. D. and A. H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, New York, 1976.
- [49] Mehra, R. K. and D. G. Lainiotis, System Identification, Academic Press, New York, 1976.
- [50] Mendel, J. M., Discrete Techniques of Parameter Estimation: The Equation Error Formulation, Marcel Dekker, New York, 1973.
- [51] Miwa, J. and K. Kido, "Speaker-Independent Word Recognition for Large Vocabulary Using Pre-selected and Non-linear Spectral Matching," Proc. ICASSP 86, pp. 2695-2698.
- [52] Nuttall, A. H., "Spectral Analysis of a Univariate Process with Bad Data Points Via Maximum Entropy and Linear Predictive Techniques," in Naval Underwater System Center (NUSC) Scientific and Engineering Studies, Spectral Estimation, NUSC, New London, Conn, March 1976.

- [53] Papoulis, A. and C. Chamaza, "Detection of Hidden Periodicities by Adaptive Extrapolation," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 5, pp. 495-500, October 1979.
- [54] Pan, K-C, F. K. Soong, and L. R. Rabiner, "A Vector-Quantization-Based Processor for Speaker-Independent Isolated Word Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-33, No. 3, pp. 546-560, June 1985.
- [55] Persoon, E. and K. S. Fu, "Shape Discrimination Using Fourier Descriptors," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 3, pp. 388-397, May 1986.
- [56] Peterka, V., "Subjective Probability Approach to Real-Time Identification," Proc. IFAC Symp. on Identification and System Parameter Estimation, Tabilisi, Paper 14.1, 1976.
- [57] Peterka, V., "Experience Accumulation for Decision Making in Multivariate Time Series," Problems of Control and Identification Theory, No. 7, pp. 143-159, 1978.
- [58] Rabiner, L. R. and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, N.J., 1978.
- [59] Rabiner, L. R., S. E. Levinson, A. E. Rosenberg, and J. G. Wilpon, "Speaker-Independent Recognition of Isolated Words Using Clustering Techniques," IEEE Trans. On Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 4, pp. 336-349, August 1979.
- [60] Rife, D. C. and R. R. Boorstdn, "Multiple Tone Parameter Estimation From Discrete Time Observations," Bell Syst. Tech. J., Vol. 55, No. 9, pp. 1389-1410, November 1976.
- [61] Robbins, H. and S. Monro, "A Stochastic Approximation Method," Ann. Math. Statist, 22, pp. 400-407, 1951.
- [62] Russell, D., Optimization Theory, W. A. Benjamin, Inc, New York, 1970.
- [63] Sage, A. P. and J. L. Melsa, System Identification, Academic Press, New York, 1971.
- [64] Sakoe, H. and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing-26, No. 1, pp. 43-49, February 1978.
- [65] Sakrison, D. J., "Stochastic Approximation," Advance Communication Systems, No. 2, pp. 51-106, 1966.
- [66] Saridis, G. N., "Comparison of Six On-line Identification Algorithms," Automatica, 10, pp. 69-79, Pergamon Press, 1974.

- [67] Schroeder, J. and R. Yarlagadda, "Linear Predictive Spectral Analysis Via the  $L_p$  Norm," Proc. ICASSP 86, pp. 201-204.
- [68] Sinha, N. K. and B. Kuszta, "Modeling and Identification of Dynamic Systems," Van Nostrand Reinhold Company, New York, 1983.
- [69] Soderstrom, T., "Some Methods for Identification of Linear System With Noisy Input-Output Data," Proc. IFAC Symposium on Identification and System Parameter Estimation, Vol. 1, pp. 357-363, Darmstadt, Federal Republic of Germany, 1979.
- [70] Steiglitz, K. and L. E. McBride, "A Technique for the Identification of Linear System," IEEE Trans. on Automatic Control, AC-10, pp. 461-464, 1965.
- [71] Stewart, G. W., Introduction to Matrix Computations, Academic Press, New York, 1973.
- [72] Strejc, V., "Trends in Identification," Proc. IFAC Symposium on Identification and System Parameter Estimation, Vol. 1, pp. 1-16, Darmstadt Federal Republic of Germany, 1979.
- [73] Taha, H. A., Operations Research: An Introduction, Macmillian Publishing Co., Inc., New York, 1976.
- [74] Toutenberg, H., "Prior Information in Linear Models," John Wiley & Sons, New York, 1982.
- [75] Tufts, D. W. and Kumaresan, "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood," Proc. IEEE, Vol. 70, No. 9, pp. 975-989, September 1982.
- [76] Ulrych, T. J. and R. W. Clayton, "Time Series Modeling and Maximum Entropy," Phys. Earth and Planetary Interiors, Vol. 12, pp. 188-200, August 1976.
- [77] Wakita, H., "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms," IEEE Trans. on Audio and Electroacoustics, Vol. AV-21, No. 5, pp. 417-427, October 1973.
- [78] Wilde, D. J., Optimum Seeking and Method, Prentice-Hall, Englewood Cliffs, N.J., 1964.
- [79] White, G. M. and R. B. Neely, "Speech Recognition Experiments with Linear Prediction, Bandpass Filtering, and Dynamic Programming," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-24, pp. 183-188, April 1976.
- [80] Willsky, A. S., Digital Processing and Control and Estimation Theory, The MIT Press, Cambridge, MA, 1979.



- [81] Yarlagadda, R., J. B. Bednar, and T. L. Watt, "Fast Algorithms for  $L_p$  Deconvolution," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-33, No. 1, pp. 174-182, February 1985.
- [82] Young, P. C., "An Instrumental Variable Method for Real Time Identification of a Boise Process," Automatica, Vol. 6, pp. 271-287, 1970.
- [83] Zadeh, L. A., "From Circuit Theory to System Theory," Proc. IRE, Vol. 50, 856-865, 1962.

## APPENDIXES

## APPENDIX A

### SYSTEM TRANSFER FUNCTION

In Chapter III, it is assumed that the underlying system is characterized by a second order linear shift-invariant difference equation, Equation (3.172), repeated as

$$y(n) - 1.5y(n-1) + 0.7y(n-2) = u(n) + 0.5u(n-1). \quad (\text{A-1})$$

Assume that all initial conditions are equal to zero and take the Z-transform of both sides of (A-1)

$$Y(z) [1 - 1.5z^{-1} + 0.7z^{-2}] = u(z) [1 + 0.5z^{-1}]. \quad (\text{A-2})$$

The associated transfer function

$$\begin{aligned} H(z) &= \frac{Y(z)}{U(z)} \\ &= \frac{1 + 0.5z^{-1}}{1 - 1.5z^{-1} + 0.7z^{-2}} \end{aligned} \quad (\text{A-3})$$

has a pair of conjugate poles

$$z_p = 0.7 \pm j0.3708 \quad (\text{A-4})$$

and two simple zeros

$$z_q(1) = 0.0 \quad (\text{A-5})$$

$$z_q(2) = -0.5. \quad (\text{A-6})$$

Pole-zero analysis indicates that the system is stable and minimum phase. The magnitude squared of the Fourier transform

$$\begin{aligned} H(e^{j\omega}) &= H(z) \Big|_{z=e^{j\omega}} \\ &= \frac{1 + 0.5e^{-j\omega}}{1 - 1.5e^{-j\omega} + 0.7e^{-j2\omega}} \end{aligned} \quad (\text{A-7})$$

is calculated according to

$$\begin{aligned} |H(e^{j\omega})|^2 &= H(e^{j\omega}) H^*(e^{j\omega}) \\ &= \frac{1 + 0.5e^{-j\omega}}{1 - 1.5e^{-j\omega} + 0.7e^{-j2\omega}} \cdot \frac{1 + 0.5e^{j\omega}}{1 - 1.5e^{j\omega} + 0.7e^{j2\omega}} \\ &= \frac{1.25 + \cos \omega}{3.74 - 5.1 \cos \omega + 1.4 \cos 2\omega} \end{aligned} \quad (\text{A-8})$$

The plot of magnitude, or the magnitude squared, of the Fourier transform of the system is used to determine the accuracy of the estimated spectrum using the different methods of system identification.

#### Impulse Response and Time Constant

The system transfer function in Equation (A-3) can be used to calculate the low frequency behavior of the system according to

$$\begin{aligned} H(z) \Big|_{z=1} &= H(e^{j\omega}) \Big|_{\omega=0} \\ &= \frac{1 + 0.5}{1 - 1.5 + 0.7} \\ &= 7.5, \end{aligned} \quad (\text{A-9})$$

where  $H(z) \Big|_{z=1}$  is called the dc gain of the system. Equation (A-3) can be written as

$$H(z) = \frac{z^2 + 0.5z}{z^2 - 1.5z + 0.7} \quad (\text{A-10})$$

Using the partial fraction expansion of  $H(z)$  gives

$$H(z) = \frac{\alpha z}{z - (0.75 + j0.3708)} + \frac{\beta z}{z - (0.75 - j0.3708)} \quad (A-11)$$

The inverse Z-transform of  $H(z)$  can be obtained as

$$h(n) = \alpha [0.75 + j0.3708]^n + \beta [0.75 - j0.3708]^n,$$

where

$$\alpha = 0.5 - j1.6855 \quad (A-12)$$

$$\beta = 0.5 + j1.6855 \quad (A-13)$$

Thus,

$$h(n) = (0.83666)^n [\cos n\theta + 3.371 \sin n\theta] \quad (A-14)$$

The time constant,  $n$ , of the underlying system, the time which the system reaches the steady state condition, is calculated as

$$(0.83466)^n = e^{-1} \quad (A-15)$$

Thus,

$$\begin{aligned} n_0 &= 5n \\ &= 5(0.607346) \\ &\approx 28 \end{aligned} \quad (A-16)$$

is the required time for the steady state condition to be reached. The simulated noise free input-output data, Equation (A-1), are recorded at the sample instant  $n=101$ , where all transients have vanished. The recorded data length is  $N=300$  and the data vector indices are initialized to be one at the sample instant  $n=101$ . This recorded data is then corrupted with a sequence of zero mean additive white noise. Equation (3.173) and (3.174), repeated as

$$\tilde{u}(n) = u(n) + \epsilon u(n) \quad 1 < n < 300 \quad (\text{A-17})$$

$$\tilde{y}(n) = y(n) + \epsilon y(n). \quad 1 < n < 300 \quad (\text{A-18})$$

These noisy input and output data vectors are used to identify the modeling performance of the LS, GLS, IV, MEEI, and EIGSP methods.

#### Calculation of Input-Output Noise Power

The input and output additive noise in (A-17) and (A-18) are generated by a random number generator whose outcomes,  $\text{Rand}(n)$ , are the sequence of random variables uniformly distributed between 0 and 1. Using a linear transformation, a sequence of zero mean white noise with variance equal to one,  $\eta(n)$ , can be generated according to [45]

$$\eta(n) = 2\sqrt{3} [\text{Rand}(n) - 0.5] . \quad (\text{A-19})$$

Assume that the variance of the input signal is known,  $\sigma_u^2 = 1$ , and SNR on both input and output are equal to 6 dB, then

$$\begin{aligned} (\text{SNR})_{\text{input}} &= (\text{SNR})_{\text{output}} \\ &= 6 \text{ dB} . \end{aligned} \quad (\text{A-20})$$

The variance of the additive noise at the input is calculated as

$$\begin{aligned} (\text{SNR})_{\text{input}} &= 6 \text{ dB} \\ &= 10 \log \frac{\sigma_u^2}{\sigma_{\epsilon u}^2} \\ &= 10 \log \frac{1}{\sigma_{\epsilon u}^2} . \end{aligned} \quad (\text{A-21})$$

Thus, the additive noise on the input is generated according to [45]

$$\epsilon_u(n) = 1.736163505 [ \text{Rand}(n) - 0.5 ] . \quad (\text{A-22})$$

Using the linear property of the system, the variance of the output is calculated as

$$\begin{aligned} \sigma_y^2 &= \sigma_u^2 \sum_{k=0}^{\infty} h^2(n) \\ &= \sum_{k=0}^{\infty} h^2(n), \end{aligned} \quad (\text{A-23})$$

where  $h(n)$  is the impulse response of the underlying system calculated in (A-14). One may use Parseval's theorem instead of calculating an infinite summation,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega = \sum_{n=-\infty}^{n=\infty} h^2(n). \quad (\text{A-24})$$

It is assumed that the system is causal, so the impulse response is equal to zero for  $n < 0$ . Thus

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega = \sum_{n=0}^{n=\infty} h^2(n). \quad (\text{A-25})$$

Using Simpson's method of integration, the variance of the output is obtained as [45]

$$\begin{aligned} \sigma_y^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega \\ &= 18.880249, \end{aligned} \quad (\text{A-26})$$

and the variance of the additive noise on output is calculated as

$$\begin{aligned} (\text{SNR})_{\text{output}} &= 6 \text{ dB} \\ &= 10 \log \frac{\sigma_y^2}{\sigma_{\epsilon y}^2} \\ &= 10 \log \frac{18.880249}{\sigma_{\epsilon y}^2}. \end{aligned} \quad (\text{A-27})$$

Thus, the additive noise on output are generated according to [45]

$$\epsilon_y(n) = 7.543874968 [ \text{Rand}(n) - 0.5 ]. \quad (\text{A-28})$$



## APPENDIX B

### M-TRANSFORM ALGORITHM

A new orthogonal transform, the Malakooti transform (M-transform), analogous to the Hadamard transform, has been developed to represent the time series signals with a set of coefficients called the M coefficients. These coefficients contain useful information about the spectral characteristics of the underlying time series and can be used for data transmission and compression. Many time series signals are highly redundant; speech, image, and other periodic signals fall into this category. The M-transform representation enables one to represent the desired signal with fewer coefficients, resulting in a saving of transmission bandwidth and memory.

This transform, like the Hadamard transform, has a complete orthonormal set and has an important role in signal and image processing applications. It has been shown [43] that the time series signals obtained from a two-dimensional shape can be represented with a few coefficients for pattern recognition and shape classification. Similarly, speech signals are represented by a set of coefficients for spectral estimation and word recognition. In all these cases, the right singular vectors of the correlation matrix are used as an orthogonal basis for the solution space. For this reason and many others, unitary transforms or an orthonormal basis, in particular a complete orthonormal

basis, should receive more attention than other transforms which have no unitary property.

### Complete Orthonormal Set

A set of linearly independent vectors  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$  is said to be orthonormal if it is self-reciprocal, i.e., if the vectors are all mutually orthogonal and have unit norm as

$$\underline{u}_i^* \underline{u}_j = \begin{cases} 1 & i=j \\ 0 & i \neq j. \end{cases} \quad (\text{B-1})$$

If time series signals  $\underline{X}$  and  $\underline{Y}$  are represented by a linear combination of a set of orthonormal vectors

$$\underline{X} = \sum_{i=1}^n \alpha_i \underline{v}_i \quad (\text{B-2})$$

and

$$\underline{Y} = \sum_{i=1}^n \beta_i \underline{v}_i, \quad (\text{B-3})$$

then their inner product,  $\langle \underline{X}, \underline{Y} \rangle$ , is easy to find. The inner product of  $\underline{X}$  and  $\underline{Y}$  is obtained as

$$\begin{aligned} \langle \underline{X}, \underline{Y} \rangle &= \left\langle \sum_{i=1}^n \alpha_i \underline{v}_i, \sum_{j=1}^n \beta_j \underline{v}_j \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j^* \underline{u}_i^* \underline{u}_j \\ &= \sum_{i=1}^n \alpha_i \beta_i^* \\ &= \langle \underline{\alpha}, \underline{\beta} \rangle. \end{aligned} \quad (\text{B-4})$$

An orthonormal set is said to be complete if any additional non-zero orthonormal vector is superfluous. If a signal is approximated by

a linear combination of the first  $m$  vectors of a complete orthonormal set with dimension  $n$ , then the norm of the error can be reduced by choosing  $m$  sufficiently large. In the next section, a method for generating the complete orthonormal sets of vectors,  $m$ -transform vectors, with the eigenanalysis of the spanned space is presented.

#### Generation of M-transform Matrix

Assume that the order-1 M-transform matrix,  $M_0$ , is equal to one,

$$M_0 = 1, \quad (\text{B-5})$$

and the order-2 M-transform matrix,  $M_1$ , is formed according to

$$\begin{aligned} M_1 &= \begin{bmatrix} aM_0 & abM_0 \\ -abM_0 & aM_0 \end{bmatrix} \\ &= \begin{bmatrix} a & ab \\ -ab & a \end{bmatrix}, \end{aligned} \quad (\text{B-6})$$

where  $a$  and  $b$  are constant parameters.

The matrix  $M$  is a  $2 \times 2$  anti-symmetric unitary matrix

$$\begin{aligned} M_1^T M_1 &= M_1 M_1^T \\ &= c I, \end{aligned} \quad (\text{B-7})$$

where the matrix  $I$  is a  $2 \times 2$  identity matrix and constant parameter  $c$  is equal to the determinant of  $M_1$ . Thus,

$$c = a^2 (1+b^2) \quad (\text{B-8})$$

and  $M_1$  inverse is given as

$$M_1^{-1} = \frac{M_1^T}{c}. \quad (\text{B-9})$$

Similarly, the order-3 M-transform matrix,  $M_2$ , can be obtained according to

$$M_2 = \begin{bmatrix} aM_1 & abM_1 \\ -abM_1 & aM_1 \end{bmatrix}. \quad (\text{B-10})$$

The matrix  $M_2$  is a 4 x 4 anti-symmetric unitary matrix

$$\begin{aligned} M_2^T M_2 &= M_2 M_2^T \\ &= c I, \end{aligned} \quad (\text{B-11})$$

where the matrix  $I$  is an 4 x 4 identity matrix,  $c$  is given in (B-8), and the inverse of  $M_2$  is calculated according to

$$M_2^{-1} = \frac{M_2^T}{c}. \quad (\text{B-12})$$

Without loss of generality, the  $2^k \times 2^k$  M-transform matrix,  $M_k$  can be obtained from

$$M_k = \begin{bmatrix} aM_{k-1} & abM_{k-1} \\ -abM_{k-1} & aM_{k-1} \end{bmatrix}, \quad (\text{B-13})$$

and  $M_k$  inverse is given according to

$$M_k^{-1} = \frac{M_k^T}{c}. \quad (\text{B-14})$$

Using the Kronecker product notation

$$A \otimes B = \begin{bmatrix} a_{11} B & a_{12} B & \dots & a_{1n} B \\ a_{21} B & a_{22} B & \dots & a_{2n} B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} B & a_{n2} B & \dots & a_{nn} B \end{bmatrix}, \quad (\text{B-15})$$

the M-transform matrices can be written according to

$$\begin{aligned}
 M_1 &= M_1 \otimes M_0 \\
 &= \begin{bmatrix} aM_0 & abM_0 \\ -abM_0 & aM_0 \end{bmatrix},
 \end{aligned} \tag{B-16}$$

and

$$\begin{aligned}
 M_2 &= M_1 \otimes M_1 \\
 &= M_1 \otimes (M_1 \otimes M_0) \\
 &= (M_1 \otimes M_1) \otimes M_0 \\
 &= M_1^{(2)} \otimes M_0 \\
 &= M_1^{(1)} \otimes M_1,
 \end{aligned} \tag{B-17}$$

where  $M_1^{(2)}$  is the Kronecker power 2 of  $M_1$  and the symbol  $\otimes$  denotes the Kronecker product. Similarly,

$$\begin{aligned}
 M_3 &= M_1 \otimes M_2 \\
 &= M_1 \otimes M_1^{(2)} \otimes M_0 \\
 &= M_1^{(3)} \otimes M_0 \\
 &= M_1^{(2)} \otimes M_1
 \end{aligned} \tag{B-18}$$

.

.

.

$$M_k = M_1 \otimes M_{k-1}$$

$$= M_1 \otimes M_1^{(k-1)} \otimes M_0 \quad (\text{B-19})$$

$$= M_1^{(k)} \otimes M_0$$

$$= M_1^{(k-1)} \otimes M_1.$$

It has been shown that, (B.59), (B.63), and (B.65), the eigenvalues of a 4 x 4 matrix D,  $\lambda_i$ ,

$$D = A \otimes B, \quad (\text{B-20})$$

can be calculated from the product of the eigenvalues of B,  $\mu_i$ , and the eigenvalues of A,  $\gamma_i$ , according to

$$\lambda_1 = \mu_1 \gamma_1 \quad (\text{B-21})$$

$$\lambda_2 = \mu_2 \gamma_2 \quad (\text{B-22})$$

$$\lambda_3 = \mu_3 \gamma_3 \quad (\text{B-23})$$

$$\lambda_4 = \mu_4 \gamma_4. \quad (\text{B-24})$$

Thus, the eigenvalues of the M-transform matrices can be obtained from a recursive algorithm proposed in the following section.

#### Eigenvalues-Eigenvectors of M-transform Matrices

Assume that constant parameters a and b are given as a=1 and b=2.

Thus,

$$M_1 = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \quad (\text{B-25})$$

and

$$M_2 = \begin{bmatrix} 1 & 2 & 2 & 4 \\ -2 & 1 & -4 & 2 \\ -2 & -4 & 1 & 2 \\ 4 & -2 & 2 & 1 \end{bmatrix} .$$

The eigenvalues,  $\lambda_i^{(1)}$ , and eigenvectors,  $\underline{x}_i^{(1)}$ , of  $M_1$  are,

$$\lambda_1^{(1)} = 1 + j2 \quad (\text{B-28})$$

$$\lambda_2^{(1)} = 1 - j2$$

$$\underline{x}_1^{(1)} = \begin{cases} j0.7071 \\ -0.7071 \end{cases} \quad (\text{B-29})$$

$$\underline{x}_2^{(1)} = \begin{cases} 0.7071 \\ -j0.7071 \end{cases} , \quad (\text{B-30})$$

where the eigenvalues of  $M_1$  are complex conjugates of each other. Using the Kronecker product relationship between  $M_1$  and  $M_2$ , Equation (B-17), the eigenvalues of  $M_2$ ,  $\lambda_i^{(2)}$ , are calculated according to

$$\begin{aligned} \lambda_1^{(2)} &= \lambda_1^{(1)} \lambda_1^{(1)} \\ &= (1-j2)(1+j2) \end{aligned} \quad (\text{B-31})$$

$$= -3+j4$$

$$\begin{aligned} \lambda_2^{(2)} &= \lambda_2^{(1)} \lambda_1^{(1)} \\ &= (1-j2)(1+j2) \end{aligned} \quad (\text{B-32})$$

$$= 5$$

$$\begin{aligned}
\lambda_3^{(2)} &= \lambda_1^{(1)} \lambda_2^{(1)} \\
&= (1+j2) (1-j2) && \text{(B-33)} \\
&= 5 \\
&= \lambda_2^{*(2)}
\end{aligned}$$

$$\begin{aligned}
\lambda_4^{(2)} &= \lambda_2^{(1)} \lambda_2^{(1)} \\
&= (1-j2) (1-j2) && \text{(B-34)} \\
&= -3 -j4 \\
&= \lambda_1^{*(2)} .
\end{aligned}$$

The matrix  $M_2$  has two complex conjugate eigenvalues. Using the complex conjugate property half of the eigenvalues of  $M_2$  can be obtained without any calculation.

In general the eigenvalues of the  $2^L \times 2^L$  M-transform,  $M_L$ , are calculated recursively from the proposed algorithm as follows

1. Calculate the eigenvalues of  $M_1$

$$\lambda_1^{(1)} = a + jab \quad \text{(B-35)}$$

$$\lambda_2^{(1)} = a - jab \quad \text{(B-36)}$$

2. For  $k=2$  to  $L$  do;

$$N=2^k \quad \text{(B-37)}$$

for  $i=1$  to  $N/2$  do;

$$\lambda_i^{(k)} = \lambda_i^{(k-1)} \lambda_1^{(1)} \quad \text{(B-38)}$$



$$\lambda_{N-i+1}^{(k)} = \lambda_i^{*(k)} \quad (\text{B-39})$$

enddo

enddo.

The eigenvectors of the L-th order M-transform are obtained from a new procedure based on the eigenvectors of the lower order M-transform. The proposed eigenvector algorithm calculates half of the eigenvectors of the  $M_L$  matrix from a simple procedure. This method, which requires few operations, is incomparable with a direct method where the dimension of  $M_L$  is high. To show the effectiveness of the proposed eigenvector algorithm, the eigenvectors of the  $M_2$  matrix are calculated using the eigencharacterization of the  $M_1$  matrix.

The characteristic equation of the  $M_1$  matrix is given as

$$\begin{aligned} f(\lambda) &= \lambda^2 - 2a\lambda + a^2(1+b^2) \\ &= 0 \end{aligned} \quad (\text{B-40})$$

or

$$\lambda^2 = 2a\lambda - a^2(1+b^2) \quad (\text{B-41})$$

Using the Cayley-Hamilton theorem gives

$$\begin{aligned} M_1^2 &= 2aM_1 - a^2(1+b^2)I \\ &= \text{tr}[M_1]M_1 - \det[M_1]I. \end{aligned} \quad (\text{B-42})$$

Assume that  $\lambda$  is the eigenvalue of  $M_2$  corresponding to the eigenvector,

$$\underline{x}, \quad \underline{x} = \begin{bmatrix} \underline{x}_u \\ \underline{x}_l \end{bmatrix}. \quad (\text{B-43})$$

Thus, the eigenvectors of the  $M_2$  matrix are related by the following relationships

$$M_2 \underline{x} = \lambda \underline{x} \quad (\text{B-44})$$

$$(M_2 - \lambda I) \underline{x} = \underline{0}. \quad (\text{B-45})$$

Substituting for  $M_2$  and  $\underline{x}$  into (B-45) gives

$$\begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix} = \begin{bmatrix} aM_1 - \lambda I & abM_1 \\ -abM_1 & aM_1 - \lambda I \end{bmatrix} \begin{bmatrix} \underline{x}_u \\ \underline{x}_\ell \end{bmatrix} \quad (\text{B-46})$$

or

$$(aM_1 - \lambda I) \underline{x}_u + abM_1 \underline{x}_\ell = \underline{0} \quad (\text{B-47})$$

and

$$-abM_1 \underline{x}_u + (aM_1 - \lambda I) \underline{x}_\ell = \underline{0}. \quad (\text{B-48})$$

Since  $ab \neq 0$ ,  $\underline{x}_\ell$  can be obtained from (B-47) as

$$\underline{x}_\ell = -\frac{1}{ab} M_1^{-1} (aM_1 - \lambda I) \underline{x}_u. \quad (\text{B-49})$$

Substituting for  $\underline{x}_\ell$  into Equation (B-48) gives

$$-abM_1 \underline{x}_u - (aM_1 - \lambda I) \frac{1}{ab} M_1^{-1} (aM_1 - \lambda I) \underline{x}_u = \underline{0}$$

or

$$\frac{1}{ab} [2a\lambda M_1 - a^2(1+b^2) M_1^2 - \lambda^2 I] M_1^{-1} \underline{x}_u = \underline{0}. \quad (\text{B-50})$$

Similarly,  $\underline{x}_u$  can be obtained from (B-48)

$$\underline{x}_u = \frac{1}{ab} M_1^{-1} (aM_1 - \lambda I) \underline{x}_\ell. \quad (\text{B-51})$$

Substituting for  $\underline{x}_u$  into Equation (B-47) gives

$$(aM_1 - \lambda I) \left[ \frac{1}{ab} M_1^{-1} (aM_1 - \lambda I) \underline{x}_\ell \right] + abM_1 \underline{x}_\ell = \underline{o} \quad (B-52)$$

or

$$\frac{1}{ab} [a^2 (1+b^2)M_1^2 - 2a\lambda M_1 + \lambda^2 I] M_1^{-1} \underline{x}_\ell = \underline{o} . \quad (B-53)$$

Substituting for  $M_1^2$  from (B-42) into (B-53) gives

$$2a[a^2 (1+b^2) - \lambda] [M_1 - \frac{1}{2a} (a^2 (1+b^2) + \lambda) I] M_1^{-1} \underline{x}_\ell = \underline{o} . \quad (B-54)$$

Similarly, substituting for  $M_1^2$  from (B-42) into (B.50) gives

$$2a[a^2(1+b^2) - \lambda][M_1 - \frac{1}{2a} (a^2(1+b^2) + \lambda) I] M_1^{-1} \underline{x}_u = \underline{o} . \quad (B-55)$$

Two eigenvalues of  $M_2$  are calculated from (B-54) and (B-55) according to

$$\lambda_2^{(2)} = a^2(1+b^2) \quad (B-56)$$

$$= \det[M_1]$$

and

$$\lambda_3^{(2)} = a^2 (1+b^2) \quad (B-57)$$

$$= \det [M_1]$$

or

$$\lambda_2^{(2)} = \lambda_3^{(2)} \quad (B-58)$$

$$= a^2 (1+b^2)$$

$$= (a+jab) (a-jab) ,$$

where  $(a+jab)$  and  $(a-jab)$  are the eigenvalues of the  $M_1$  matrix. Thus,

$$\lambda_2^{(2)} = \lambda_3^{(2)} \quad (\text{B-59})$$

$$\begin{aligned} & \lambda_1^{(1)} \lambda_2^{(1)} \\ & = \det[M_1]. \end{aligned}$$

The remaining two eigenvalues of  $M_2$  are calculated from the following relationship,

$$M_1 - \frac{1}{2a} (a^2 (1+b^2) + \lambda)I = \underline{0} \quad (\text{B-60})$$

and

$$2aM_1 - a^2 (1+b^2)I - \lambda I = \underline{0} . \quad (\text{B-61})$$

Equation (B-61) indicates that the remaining eigenvalues of  $M_2$  are related to the eigenvalues of  $M_1$  according to

$$\begin{aligned} \lambda_1^{(2)} &= 2a\lambda_1^{(1)} - a^2 (1+b^2) \\ &= \text{tr} [M_1] \lambda_1^{(1)} - \det [M_1] \\ &= 2a \lambda_1^{(1)} - \lambda_1^{(1)} \lambda_2^{(1)} \\ &= \lambda_1^{(1)} [2a - \lambda_2^{(1)}]. \end{aligned} \quad (\text{B-62})$$

Substituting for  $\lambda_2^{(1)}$  from (B-36) into (B-62) gives

$$\begin{aligned} \lambda_1^{(2)} &= \lambda_1^{(1)} [2a - (a - jab)] \\ &= \lambda_1^{(1)} (a + jab) \\ &= \lambda_1^{(1)} \lambda_1^{(1)}. \end{aligned} \quad (\text{B-63})$$

Similarly,

$$\begin{aligned}
\lambda_4^{(2)} &= 2a \lambda_2^{(1)} - a^2 (1+b^2) \\
&= \text{tr} [ M_1 ] \lambda_2^{(1)} - \det \{ M_1 \} \\
&= 2a \lambda_2^{(1)} - \lambda_1^{(1)} \lambda_2^{(1)} \\
&= \lambda_2^{(1)} [2a - \lambda_1^{(1)}] \\
&= \lambda_2^{(1)} [2a - (a+jab)] \\
&= \lambda_2^{(1)} (a-jb).
\end{aligned} \tag{B-64}$$

Thus,

$$\lambda_4^{(2)} = \lambda_2^{(1)} \lambda_2^{(1)}. \tag{B-65}$$

Assume that  $\underline{x}_4^{(2)}$  is an eigenvector of  $M_2$ , where

$$\underline{x}_4^{(2)} = \begin{bmatrix} \frac{1}{ab} (aM_1 - \lambda_4^{(2)} I) \underline{x}_2^{(1)} \\ M_1 \underline{x}_2^{(1)} \end{bmatrix}. \tag{B-66}$$

$\underline{x}_2^{(1)}$  is the eigenvector of  $M_1$ , and  $\lambda_4^{(2)}$ , and  $\underline{x}_4^{(2)}$  are the eigenvalues and eigenvector of  $M_2$ , respectively. Using Equation (B-46), the eigenvalues-eigenvector of  $M_2$  can be written according to

$$\begin{aligned}
[M_2 - \lambda_4^{(2)} I] \underline{x}_4^{(2)} &= \underline{0} \\
&= \begin{bmatrix} (aM_1 - \lambda_4^{(2)} I) \frac{1}{ab} (aM_1 - \lambda_4^{(2)} I) \underline{x}_2^{(1)} + abM_1^2 \underline{x}_2^{(1)} \\ -abM_1 \left( \frac{1}{ab} \right) (aM_1 - \lambda_4^{(2)} I) \underline{x}_2^{(1)} + (aM_1 - \lambda_4^{(2)} I) M_1 \underline{x}_2^{(1)} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{ab} (a^2(1+b^2)M_1^2 - 2a\lambda_4^{(2)}M_1 + \lambda_4^{(2)2}I) \underline{x}_2^{(1)} \\ \underline{0} \end{bmatrix}
\end{aligned} \tag{B-67}$$

Substituting for  $M_1^2$  from (B-42) into (B-67) gives

$$\begin{aligned}
 [M_2 - \lambda_4^{(2)} I] \underline{x}_4^{(2)} &= \begin{bmatrix} \frac{1}{ab} (a^2(1+b^2) - \lambda_4^{(2)}) [2aM_1 - (a^2(1+b^2) + \lambda_4^{(2)}) I] \underline{x}_2^{(1)} \\ \underline{0} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{1}{ab} (\det M_1 - \lambda_4^{(2)}) [\text{tr}(M_1) M_1 - \det(M_1) + \lambda_4^{(2)}] I] \underline{x}_2^{(1)} \\ \underline{0} \end{bmatrix}.
 \end{aligned} \tag{B-68}$$

Substituting for  $\lambda_4^{(2)}$  from (B-64) into (B-68) gives

$$\begin{aligned}
 [M_2 - \lambda_4^{(2)} I] \underline{x}_4^{(2)} &= \begin{bmatrix} \frac{1}{ab} [2\det(M_1) - \lambda_2^{(1)} \text{tr}(M_1)] \text{tr}(M_1) [M_1 - \lambda_2^{(1)} I] \underline{x}_2^{(1)} \\ \underline{0} \end{bmatrix} \\
 &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
 \end{aligned} \tag{B-69}$$

Thus,  $\underline{x}_4^{(2)}$  is an eigenvector of  $M_2$  corresponding to  $\lambda_4^{(2)}$ . Similarly,  $\underline{x}_1^{(2)}$  is an eigenvector of  $M_2$  corresponding to  $\lambda_1^{(2)}$ , where

$$\underline{x}_1^{(2)} = \begin{bmatrix} \frac{1}{ab} (aM_1 - \lambda_1^{(2)} I) \underline{x}_1^{(1)} \\ M_1 \end{bmatrix}. \tag{B-70}$$

Since  $M_1$  is nonsingular,  $\underline{x}_1^{(2)}$  and  $\underline{x}_4^{(2)}$  are linearly independent. The other two eigenvectors of  $M_2$ ,  $\underline{x}_2^{(2)}$  and  $\underline{x}_3^{(2)}$  are selected so that

$$T = \begin{bmatrix} \underline{x}_1^{(2)} & \underline{x}_2^{(2)} & \underline{x}_3^{(2)} & \underline{x}_4^{(2)} \end{bmatrix} \tag{B-71}$$

are linearly independent, and

$$\Lambda = T^{-1} M_2 T \tag{B-72}$$

is a diagonal matrix.

This analysis clearly shows that half of the eigenvectors of  $M_L$  can be obtained from a straight-forward procedure and the other half can be selected so that  $T = \underline{x}_1^{(k)}, \underline{x}_2^{(k)}, \dots, \underline{x}_n^{(k)}$  is a linearly independent set. The proposed M-transform, whose eigenvalues are calculated from a simple recursive algorithm and half of its eigenvectors are calculated from a few simple operations, can be used as an orthogonal basis to represent many signal and image processing applications. Moreover, the number of distinct eigenvalues of  $M_L$  is  $L+1$  as opposed to an  $L$ -th order Hadamard transform,  $H_L$ , which only has two distinct eigenvalues. The eigenvalues of the  $M_L$  transform can be used as feature parameters if the elements of the  $M_L$  matrix are the autocorrelation lags of the observation and by proper selection of the  $a$  and  $b$  constants.

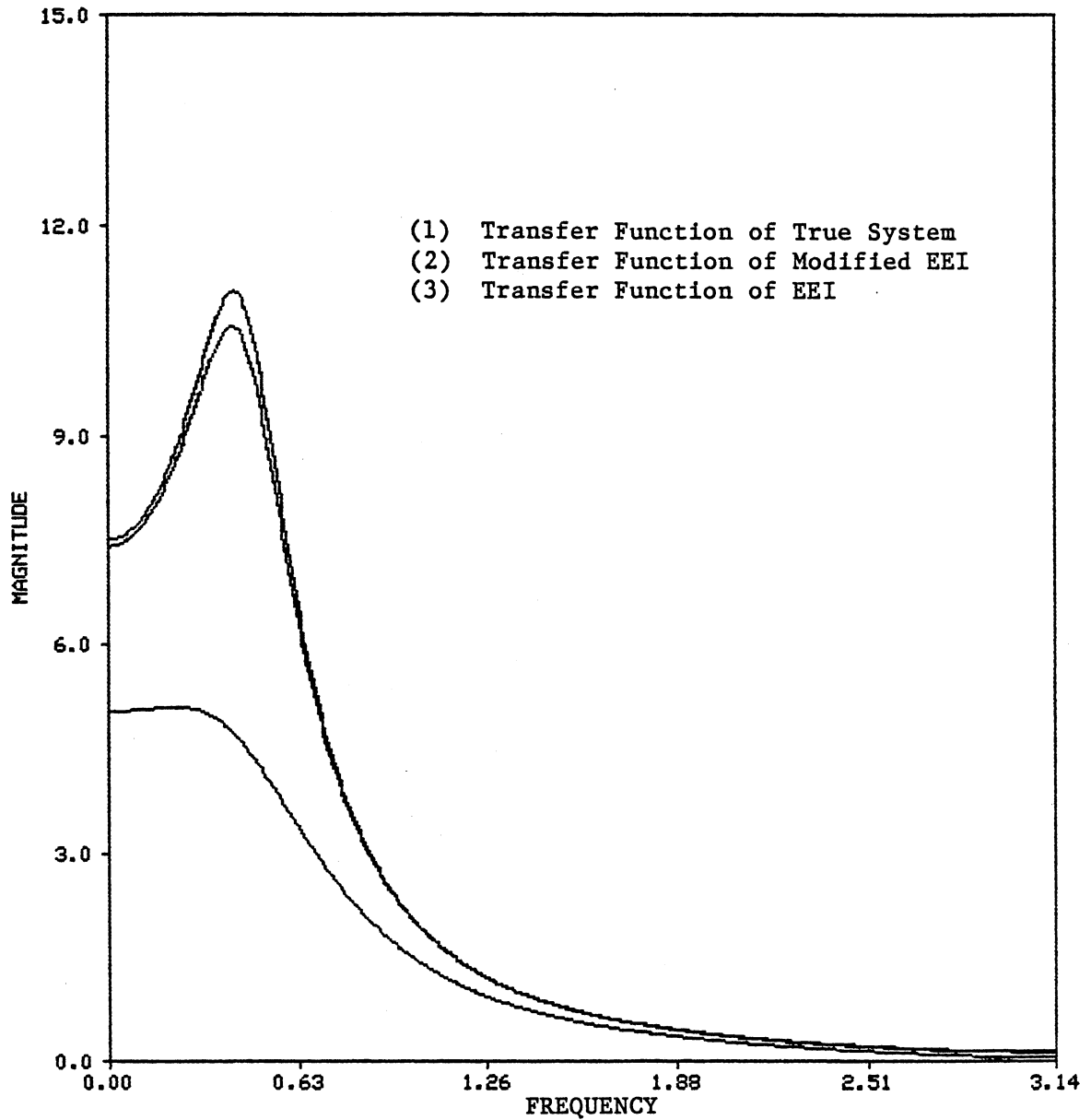


Figure 4. Comparison of Equation Error and Modified Equation Error Identification With True System;  $p=q=2$ , SNR Input = SNR Output = 6 dB.



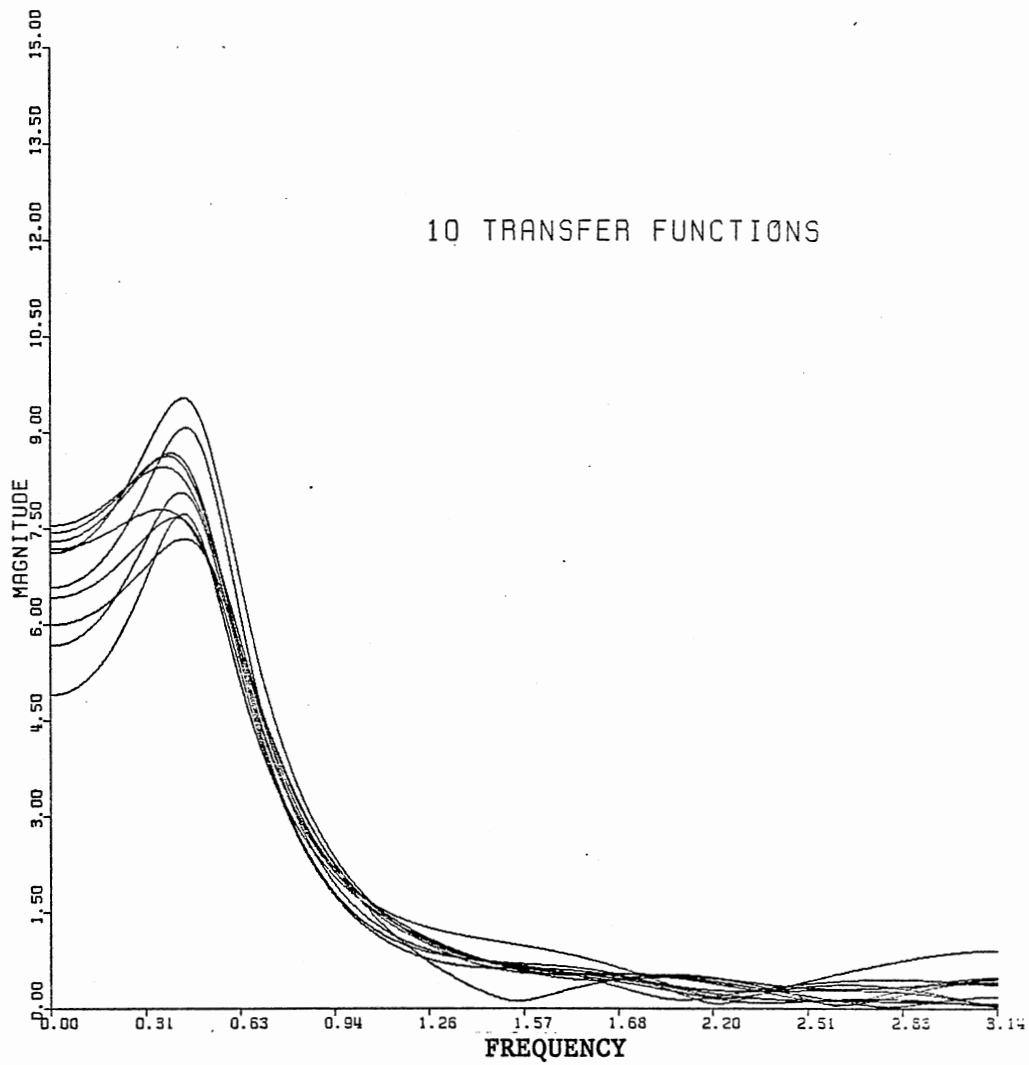


Figure 5. Magnitude of 10 Transfer Functions of the Least Squares Identification;  
 $p=q=5$ , SNR Input=SNR Output = 6 dB.

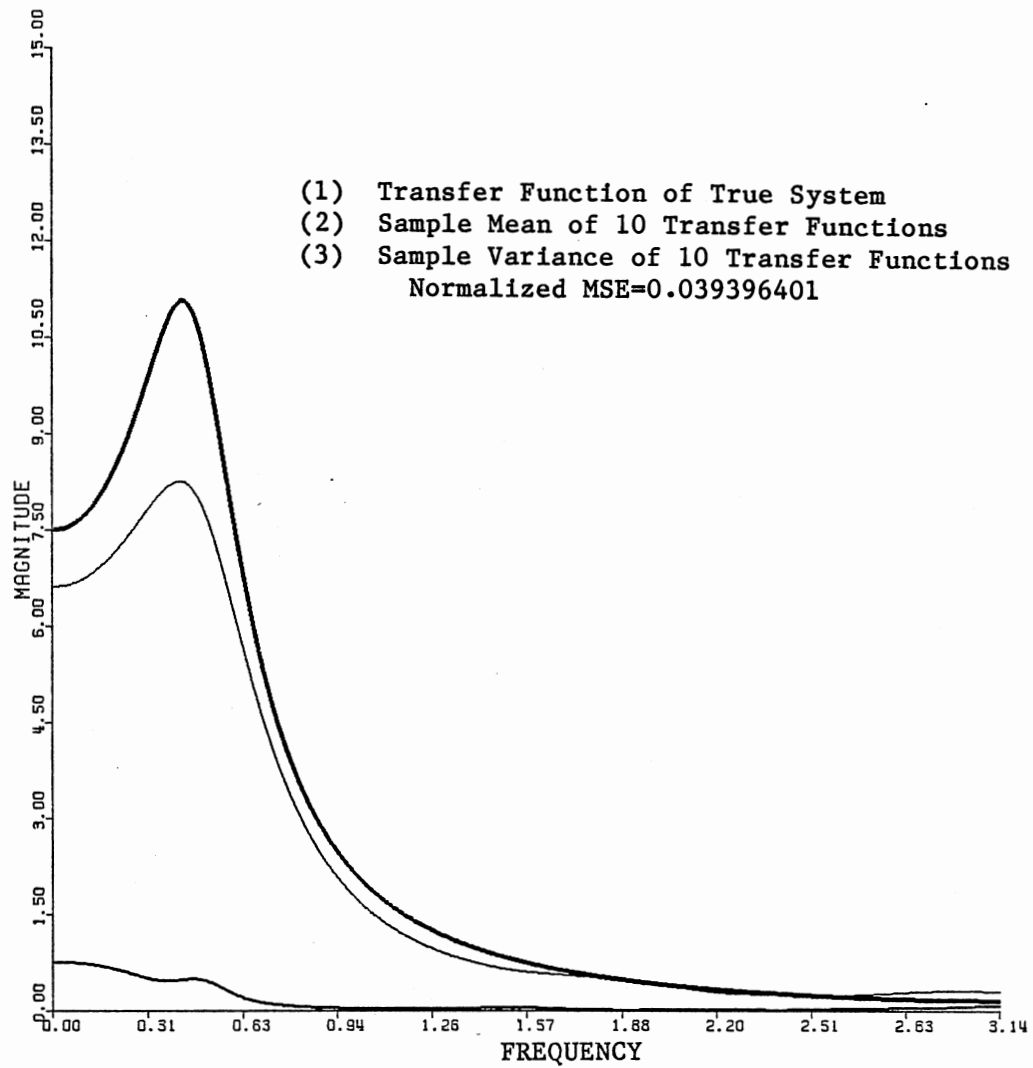


Figure 6. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Least Squares Identification;  $p=q=5$ , SNR Input = SNR Output = 6 dB.

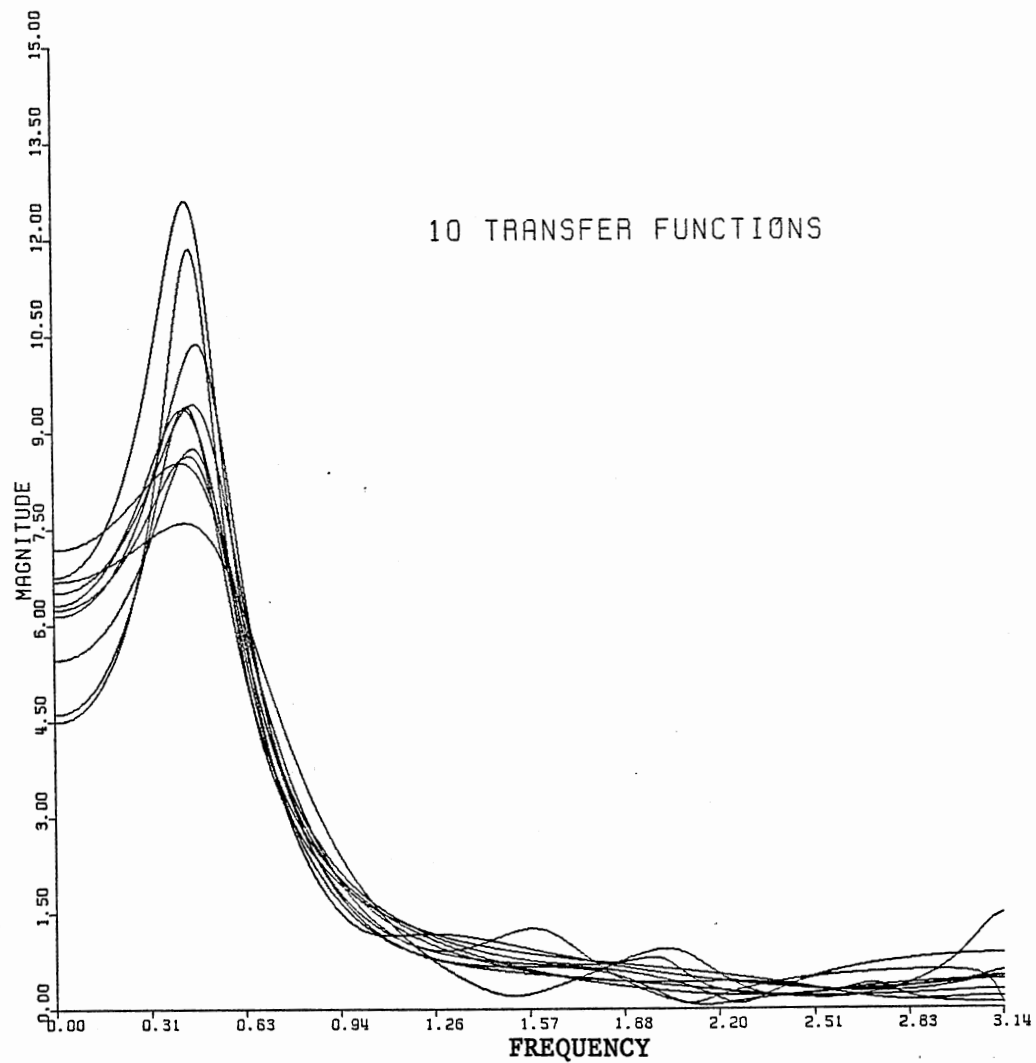


Figure 7. Magnitude of 10 Transfer Functions of the Instrumental Variable Identification;  $p=q=5$ , SNR Input = SNR Output = 6 dB.

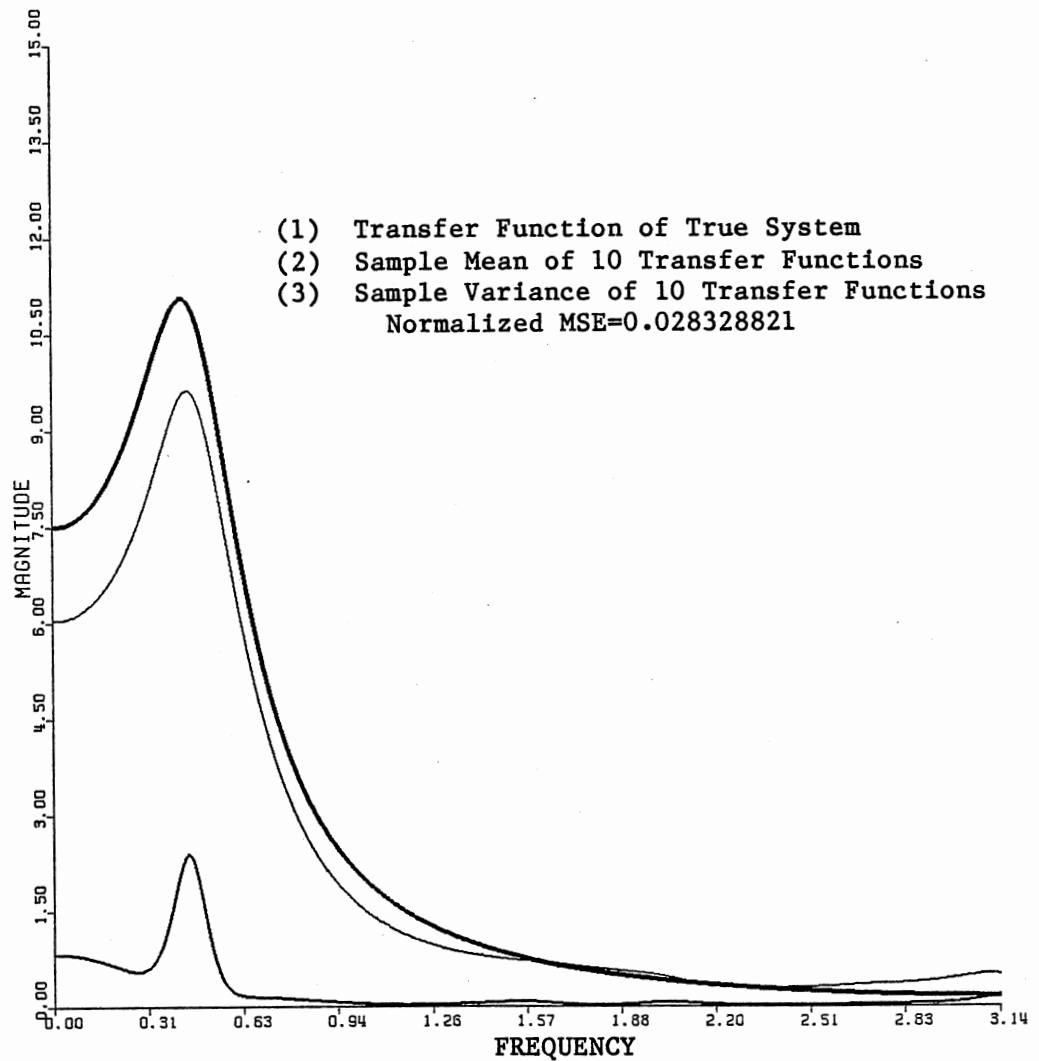


Figure 8. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Instrumental Variable Identification;  $p=q=5$ , SNR Input = SNR Output = 6 dB.

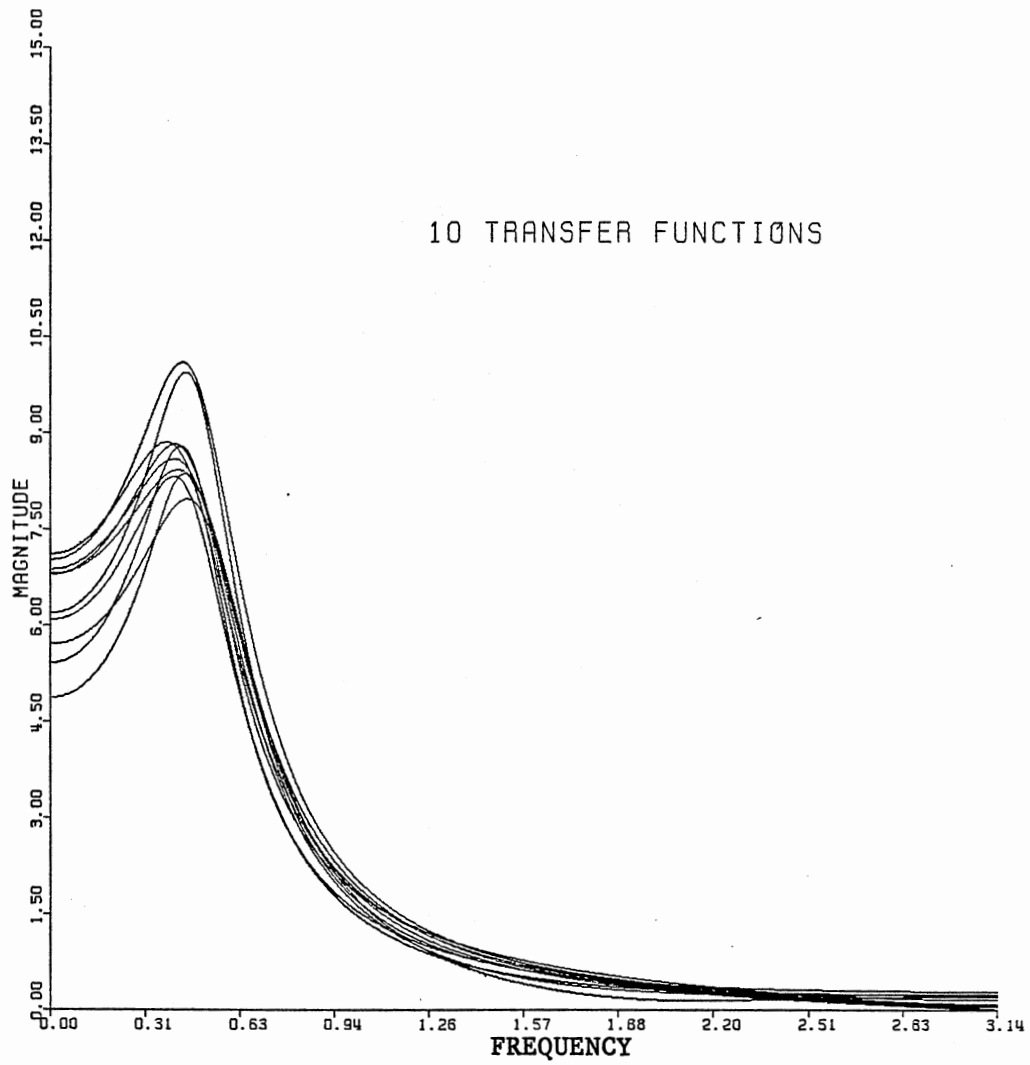


Figure 9. Magnitude of 10 Transfer Functions of the Generalized Least Squares Identification;  $p=q=5$ , SNR Input = SNR Output = 6 dB.

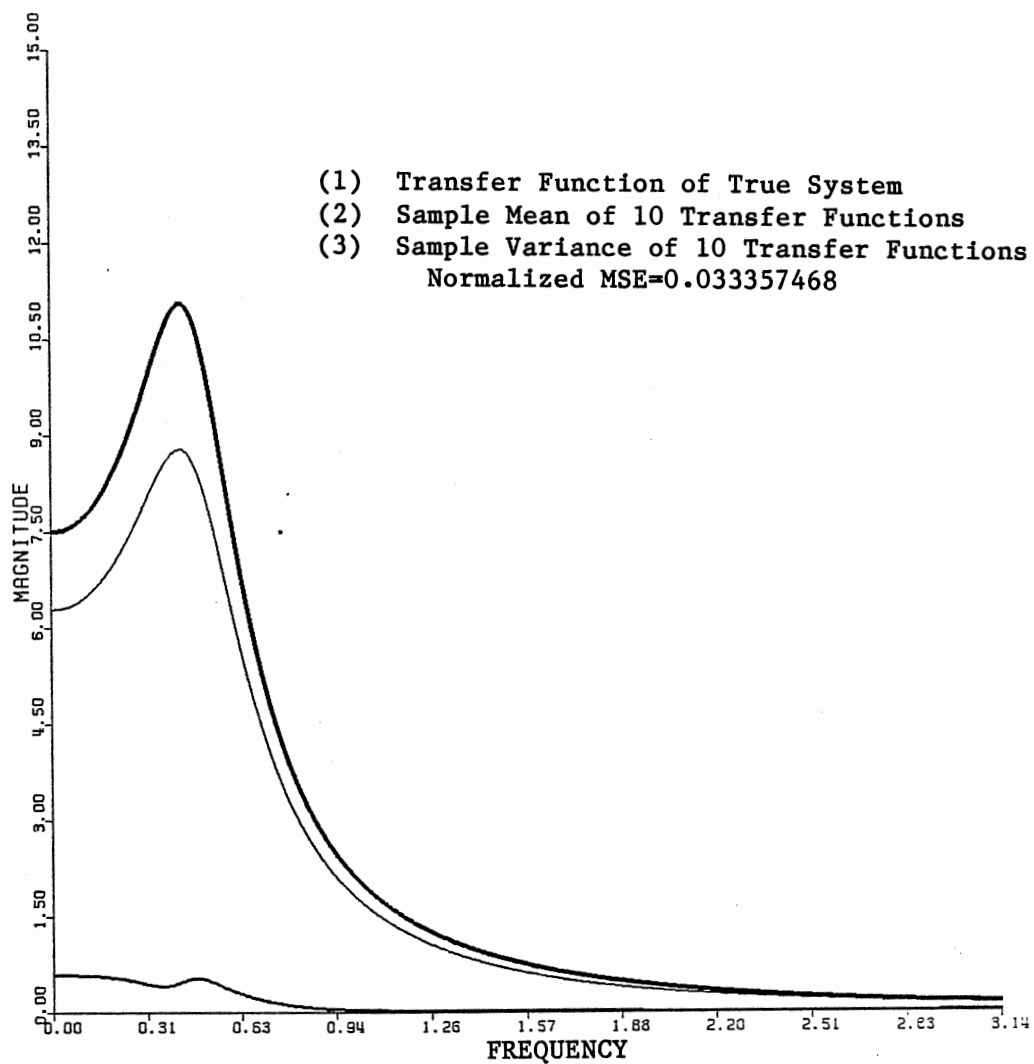


Figure 10. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Generalized Least Squares Identification;  $p=q=5$ , SNR Input = SNR Output = 6 dB.

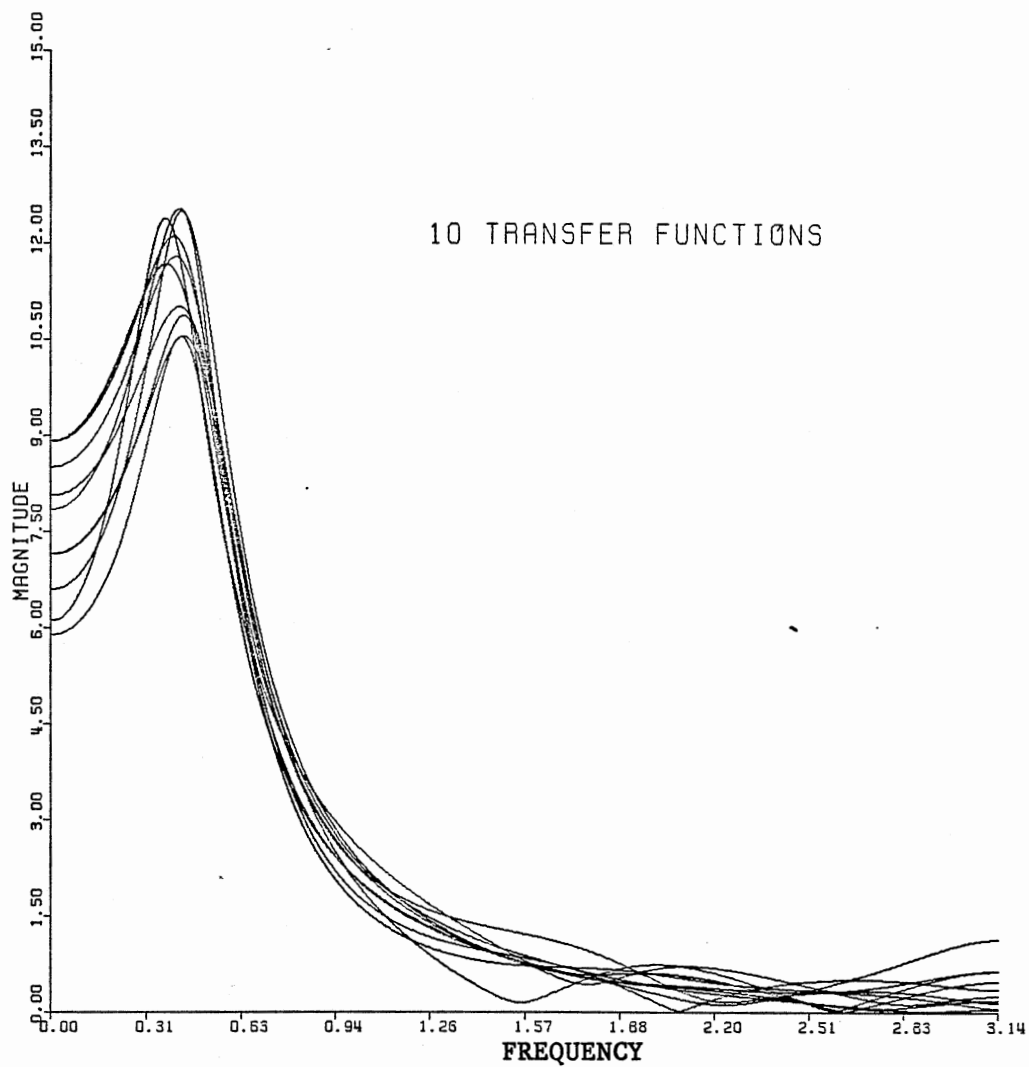


Figure 11. Magnitude of 10 Transfer Functions of the Correlation Based Identification With Scaling and No Noise Subtraction;  $p=q=5$ ,  $T_1 = T_2 = 29$  SNR Input = SNR Output = 6 dB.

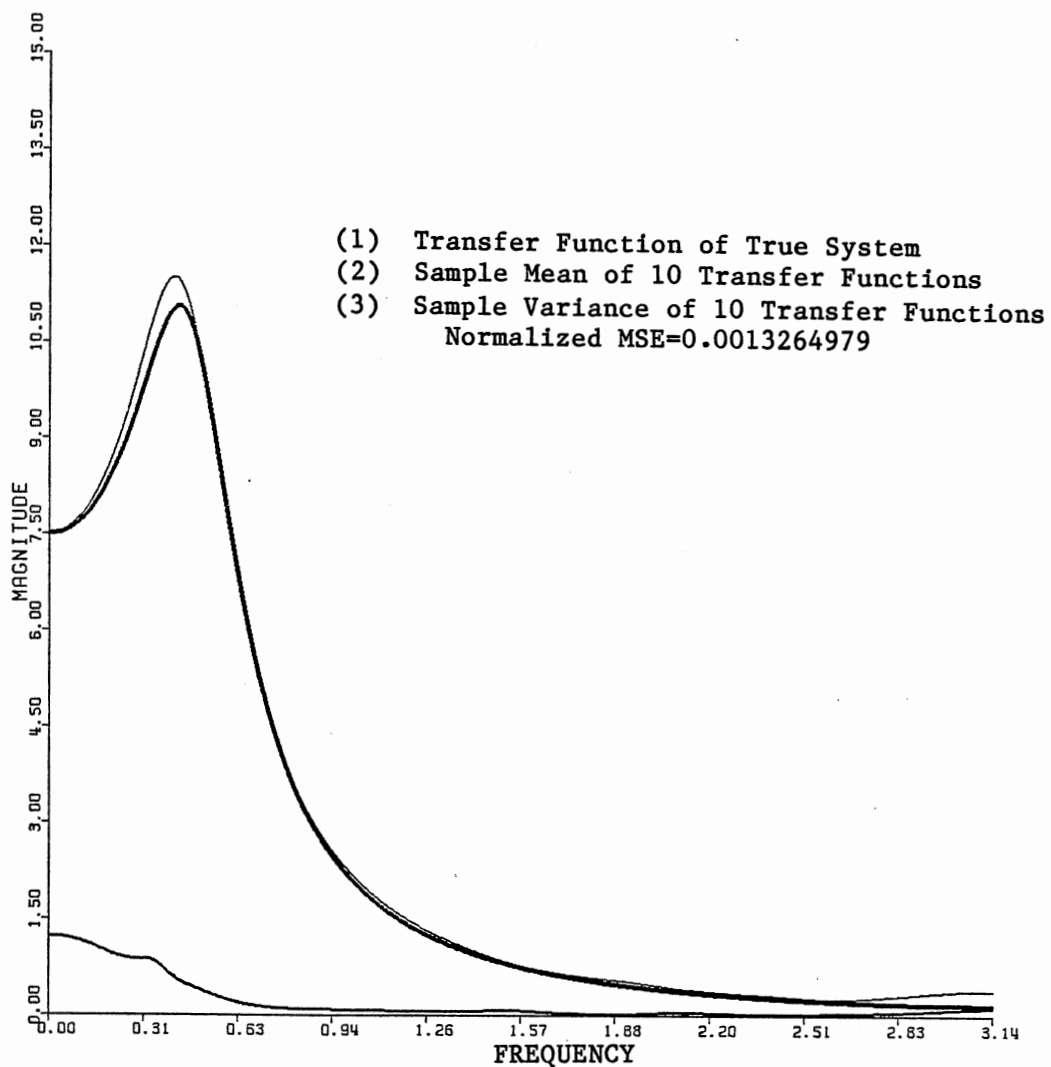


Figure 12. Transfer Function of True System, and Sample Mean and Sample Variance of 10 Transfer Functions of the Correlation Based Identification With Scaling and No Noise Subtraction;  $p=q=5$ ,  $T_1 = T_2 = 29$  SNR Input = SNR Output = 6 dB.



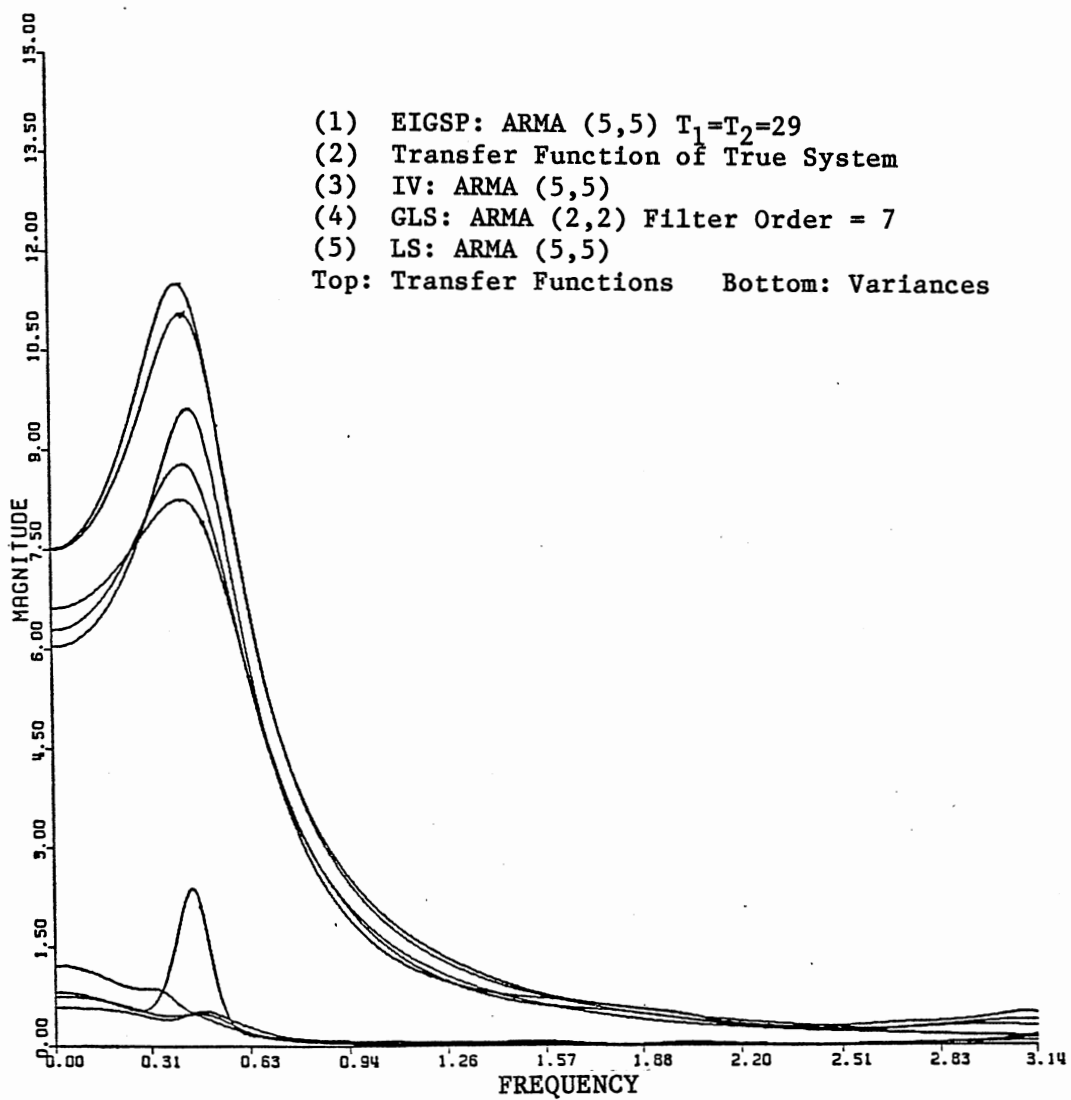


Figure 13. Comparison of the LS, GLS, IV, and EIGSP Solutions With the True System;  
 SNR Input = SNR Output = 6 dB.

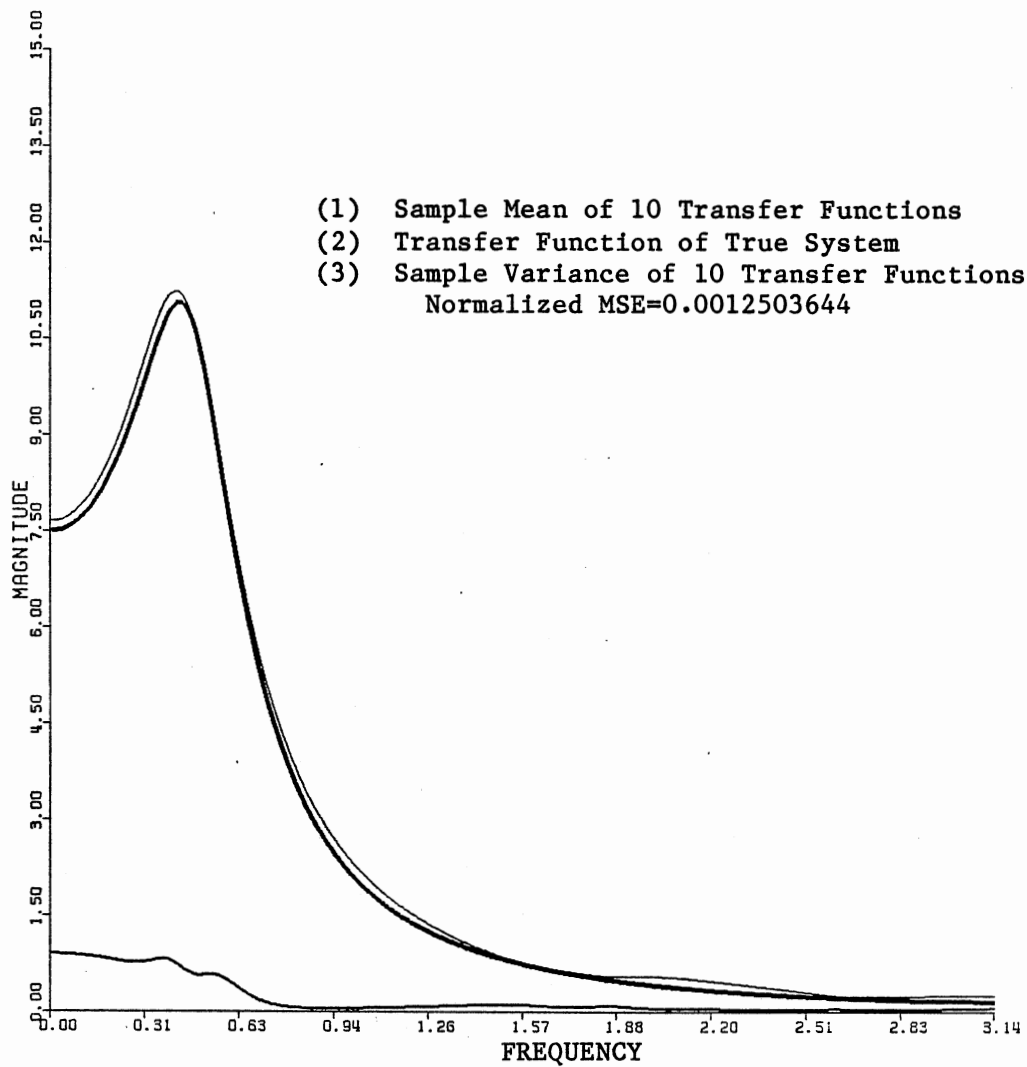


Figure 14. Correlation Based Identification Via SVD With Scaling and Noise Subtraction;  
 $p=q=5$ ,  $T_1 = T_2 = 29$ , SNR Input = SNR  
Output = 6 dB.

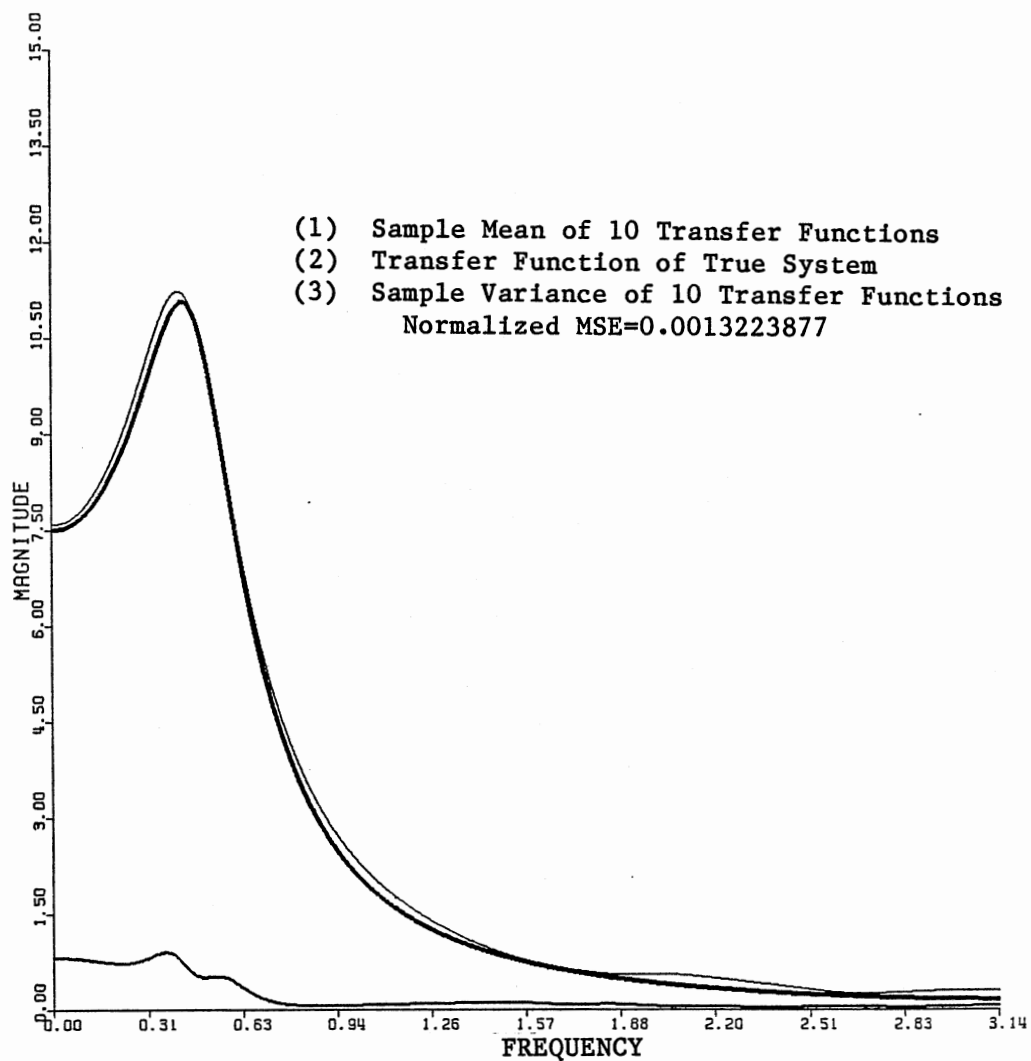


Figure 15. Correlation Based Identification Via SVD With Scaling and No Noise Subtraction;  
 $p=q=5$ ,  $T_1 = T_2 = 29$ , SNR Input = SNR  
Output = 6 dB.

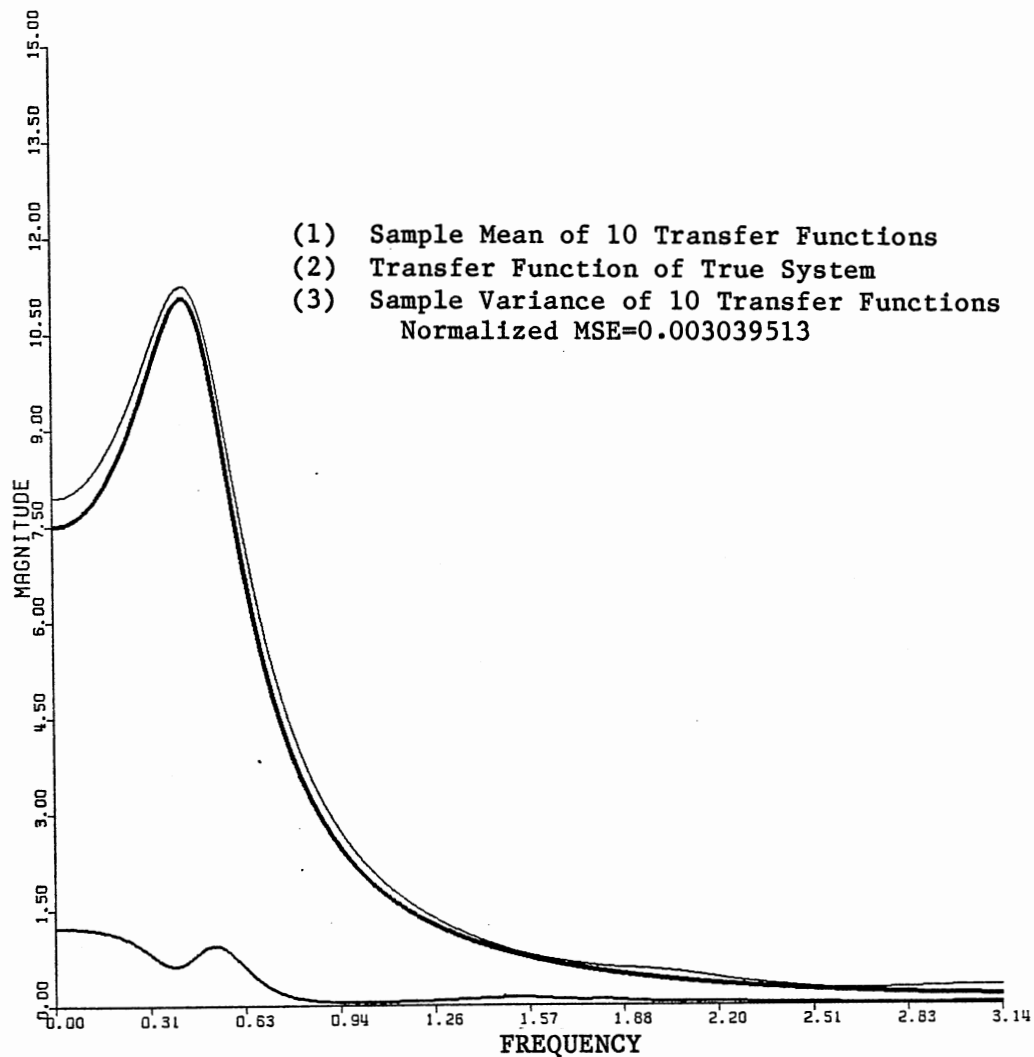


Figure 16. Correlation Based Identification Via SVD With Scaling and Noise Subtraction;  
 $p=q=5$ ,  $T_1 = 8$ ,  $T_2 = 56$ , SNR Input =  
SNR Output = 6 dB.

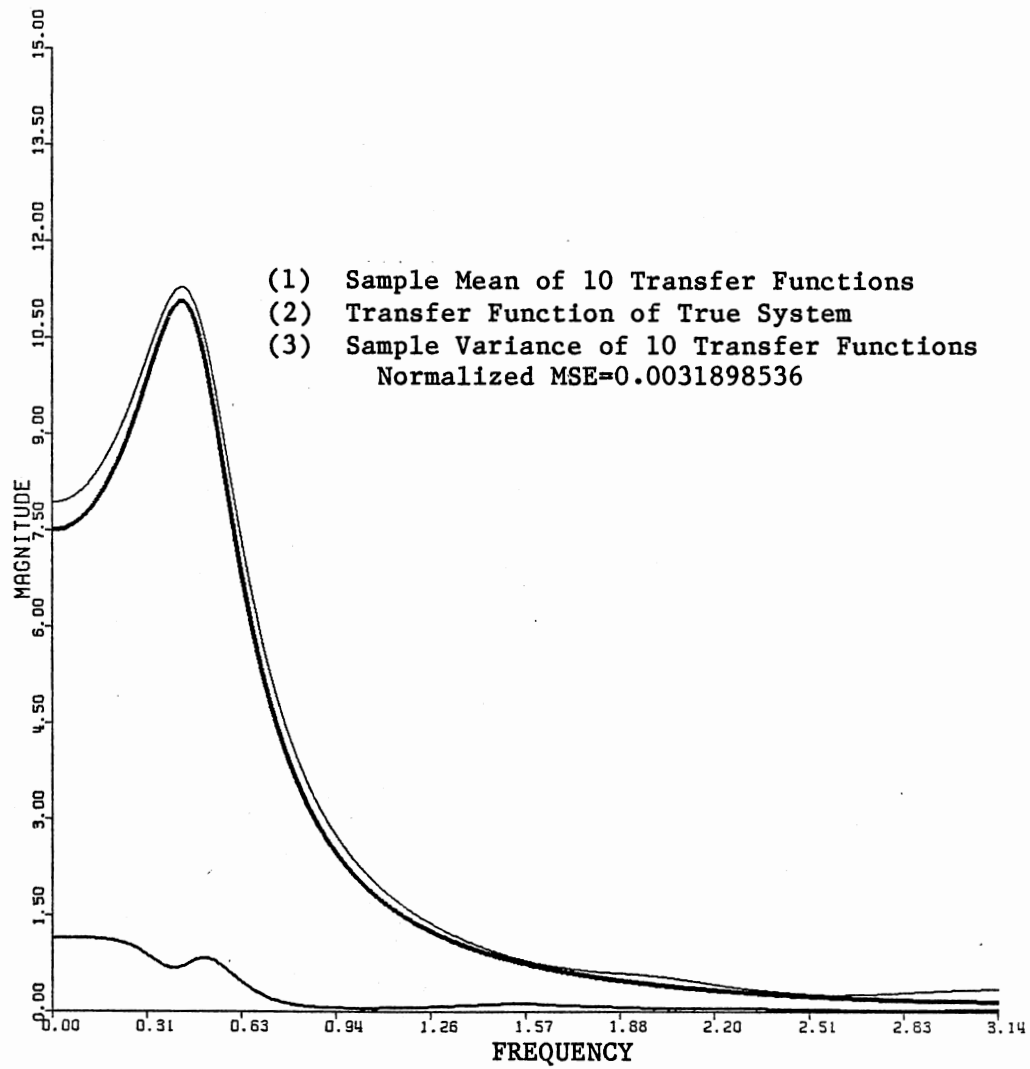


Figure 17. Correlation Based Identification Via SVD With Scaling and No Noise Subtraction;  
 $p=q=5$ ,  $T_1 = 8$ ,  $T_2 = 56$ , SNR Input = SNR Output = 6 dB.

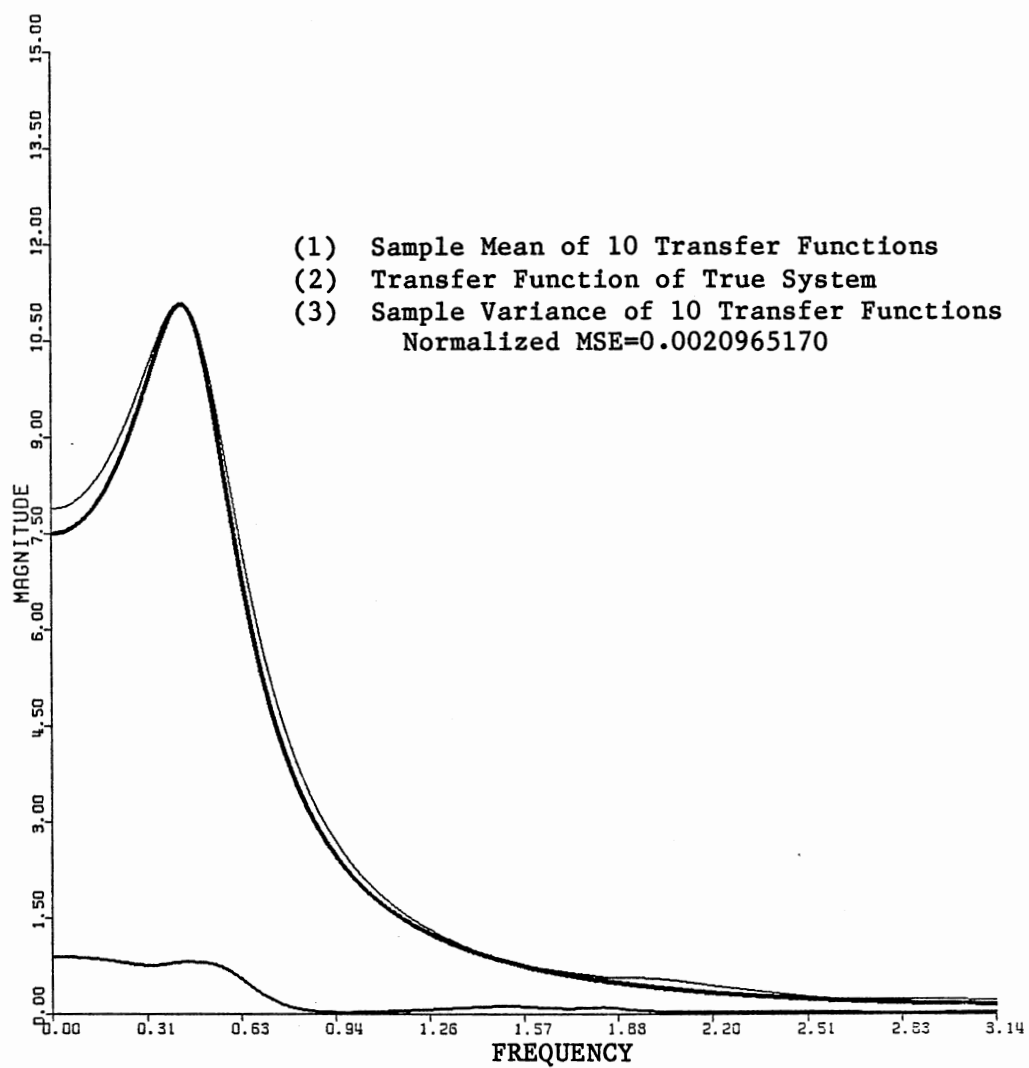


Figure 18. Correlation Based Identification Via SVD With Scaling and Noise Subtraction;  
 $p=q=5$ ,  $T_1 = 6$ ,  $T_2 = 28$ , SNR Input = SNR  
Output = 6 dB.

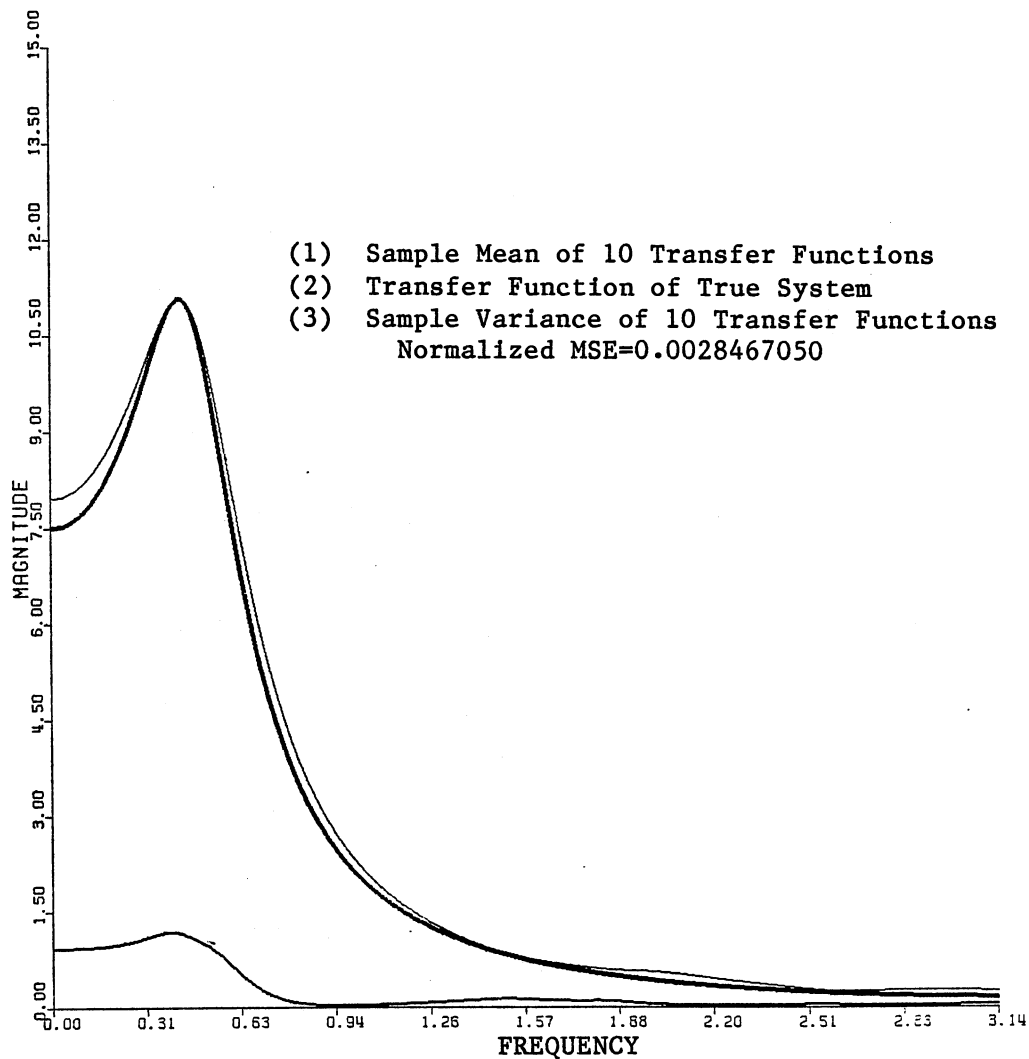


Figure 19. Correlation Based Identification Via SVD With Scaling and No Noise Subtraction;  
 $p=q=5$ ,  $T_1 = 6$ ,  $T_2 = 28$ , SNR Input = SNR Output = 6 dB.

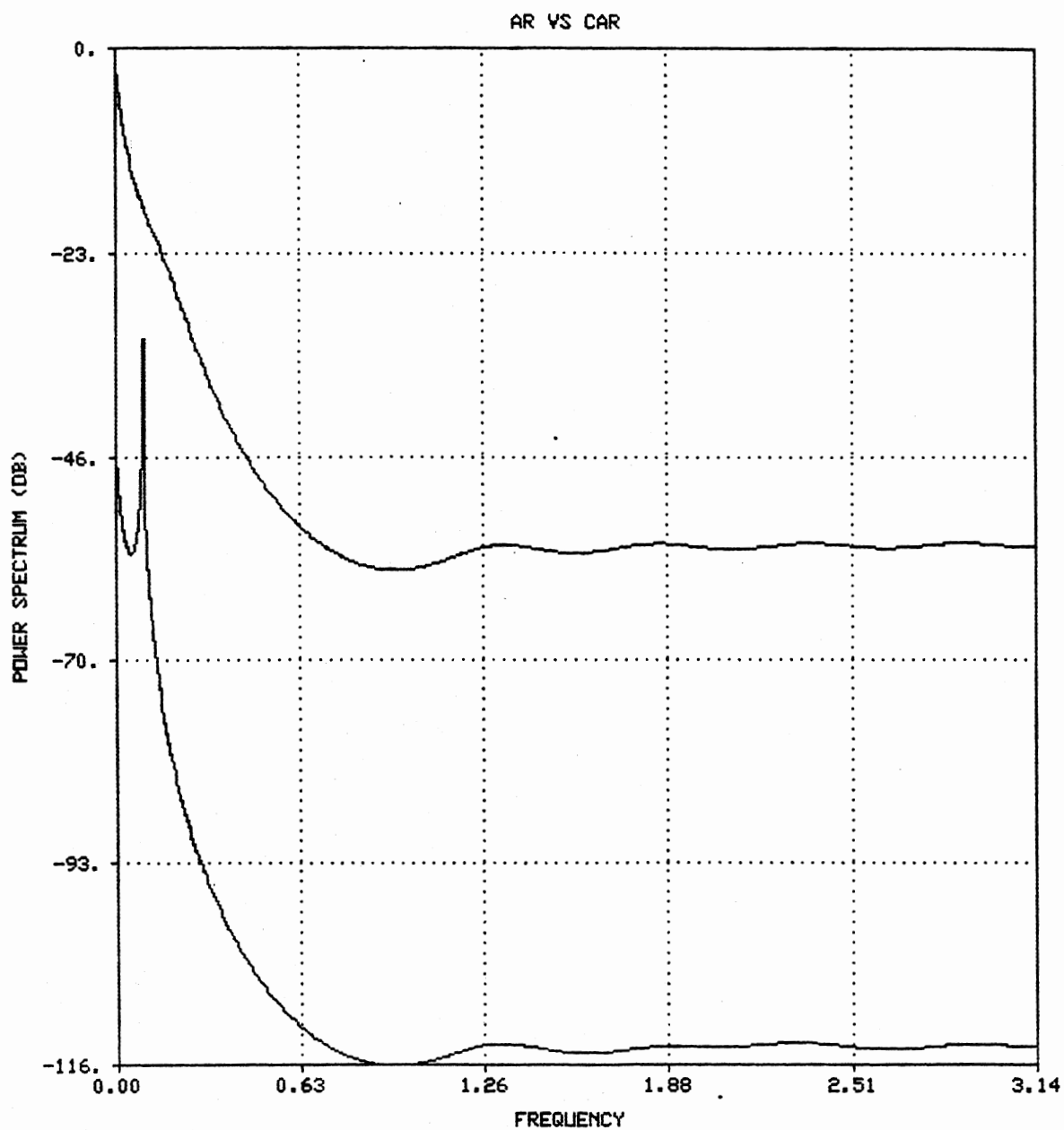


Figure 20. Power Spectrum of AR Vs CAR for a Two Dimensional Shape; Top(AR), Bottom(CAR),  $p=12$ ,  $q=0$ ,  $\bar{p}=4$ ,  $\bar{q}=0$ ,  $N=128$ ,  $L=64$ .



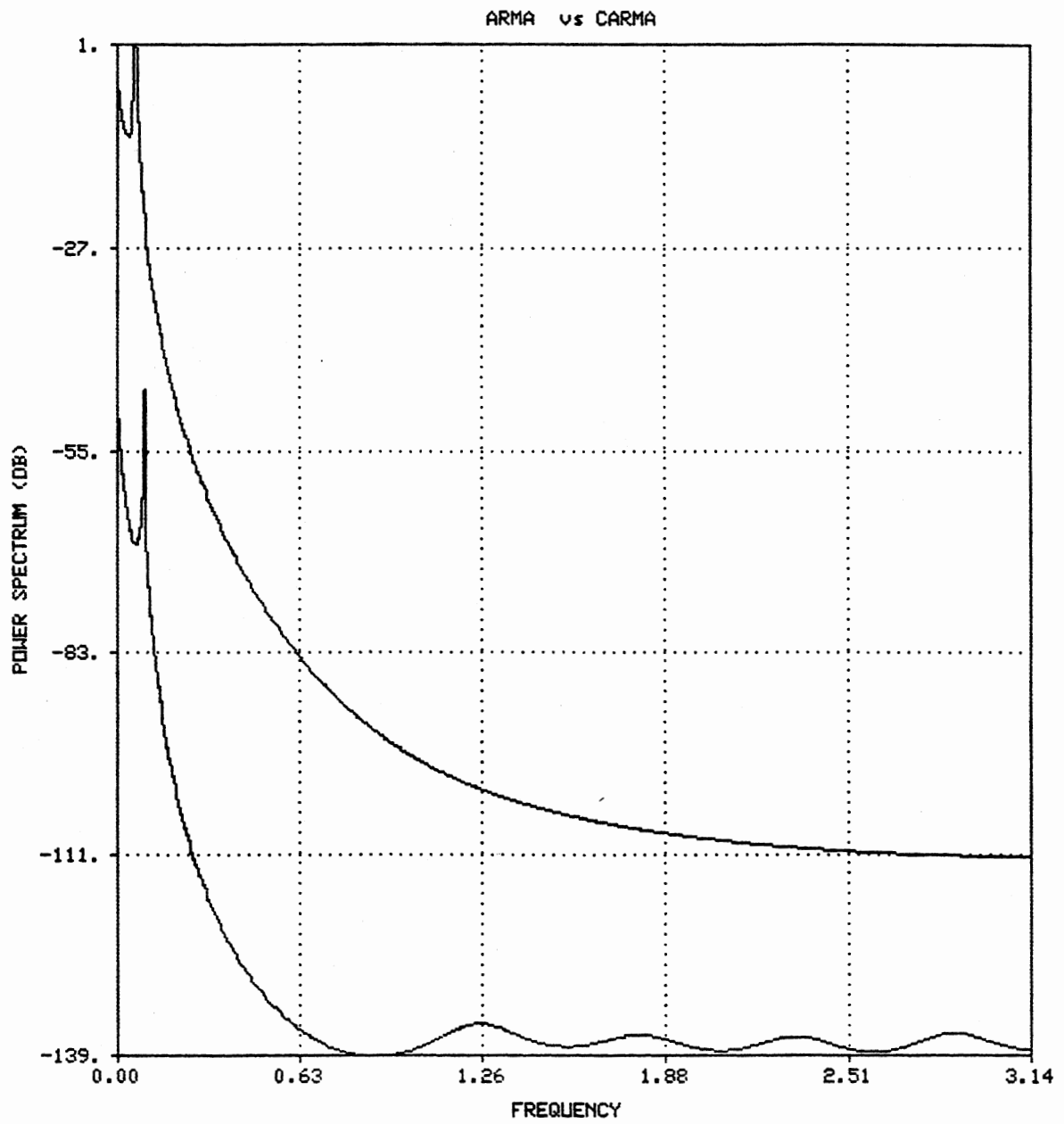


Figure 21. Power Spectrum of ARMA Vs CARMA for a Two Dimensional Shape; Top(ARMA), Bottom(CARMA),  $p=q=12$ ,  $\bar{p}=\bar{q}=4$ ,  $N=128$ ,  $L=64$ .

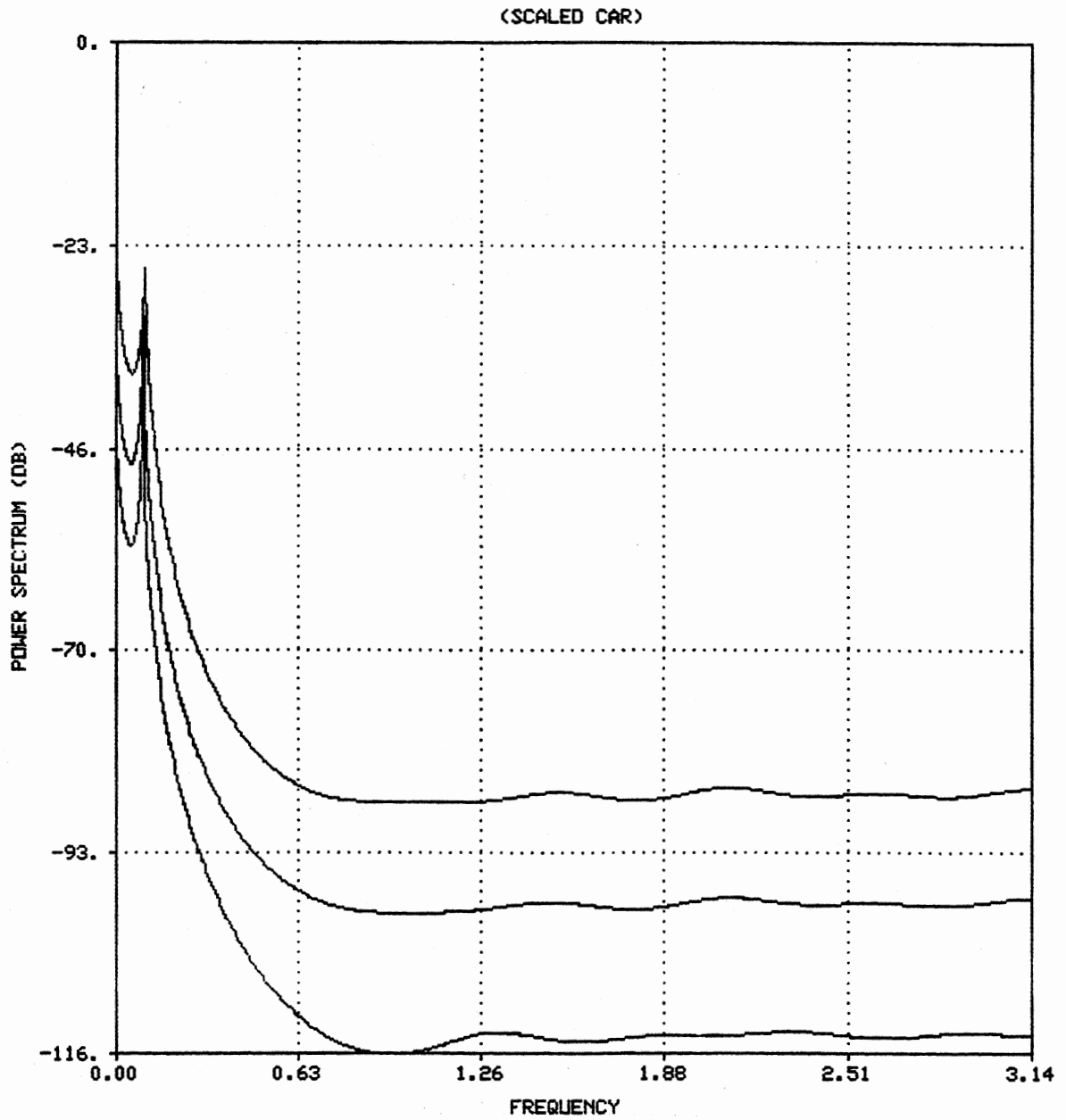


Figure 22. Power Spectrum of Scaled CAR by 1.0, 0.5, and 0.25 for a Two-Dimensional Shape;  $p=12$ ,  $q=0$ ,  $\bar{p}=4$ ,  $\bar{q}=0$ ,  $N=128$ ,  $L=64$ .

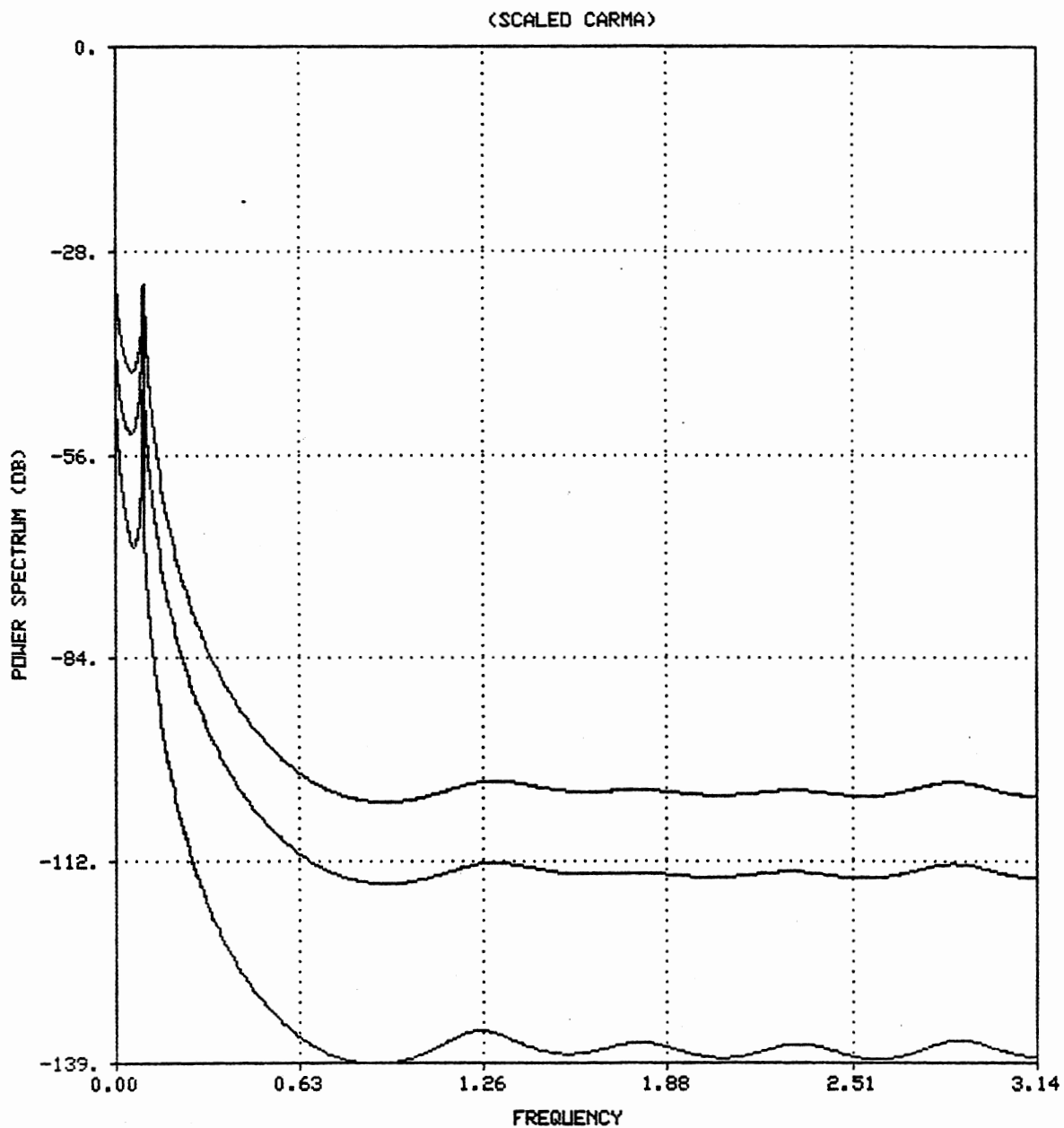


Figure 23. Power Spectrum of Scaled CARMA by 1.0, 0.5, and 0.25  
for a Two-Dimensional Shape;  
 $p=q=12$ ,  $\bar{p}=\bar{q}=4$ ,  $N=124$ ,  $L=64$ .

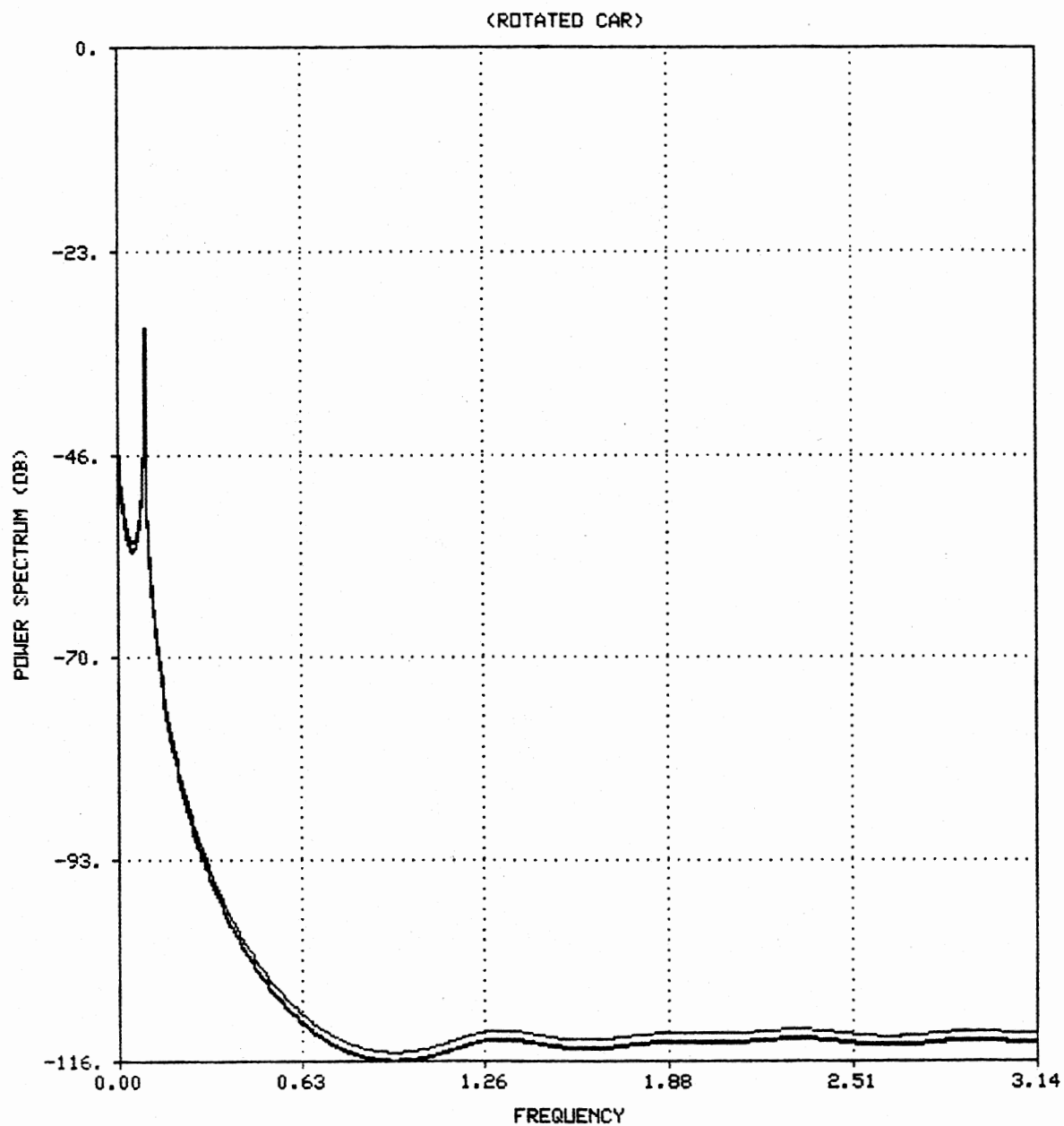


Figure 24. Power Spectrum of Rotated CAR by 0.0,  $\pi/4$ , and  $\pi/2$  for a Two-Dimensional Shape;  $p=12$ ,  $q=0$ ,  $\bar{p}=4$ ,  $\bar{q}=0$ ,  $N=128$ ,  $L=64$ .

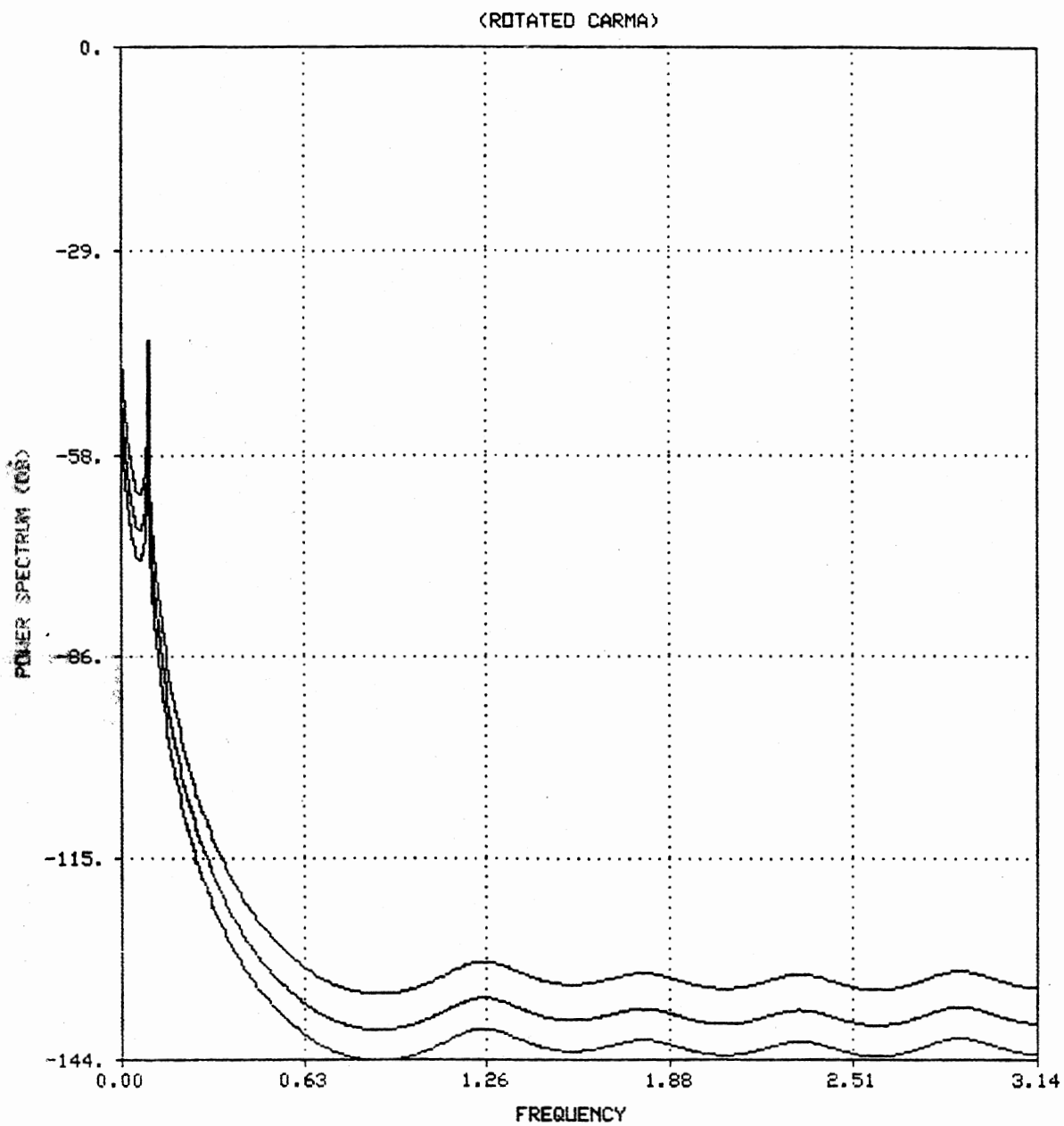


Figure 25. Power Spectrum of Rotated CARMA by  $0.0$ ,  $\pi/4$ , and  $\pi/2$  for a Two-Dimensional Shape;  $p=q=12$ ,  $\bar{p}=\bar{q}=4$ ,  $N=128$ ,  $L=64$ .

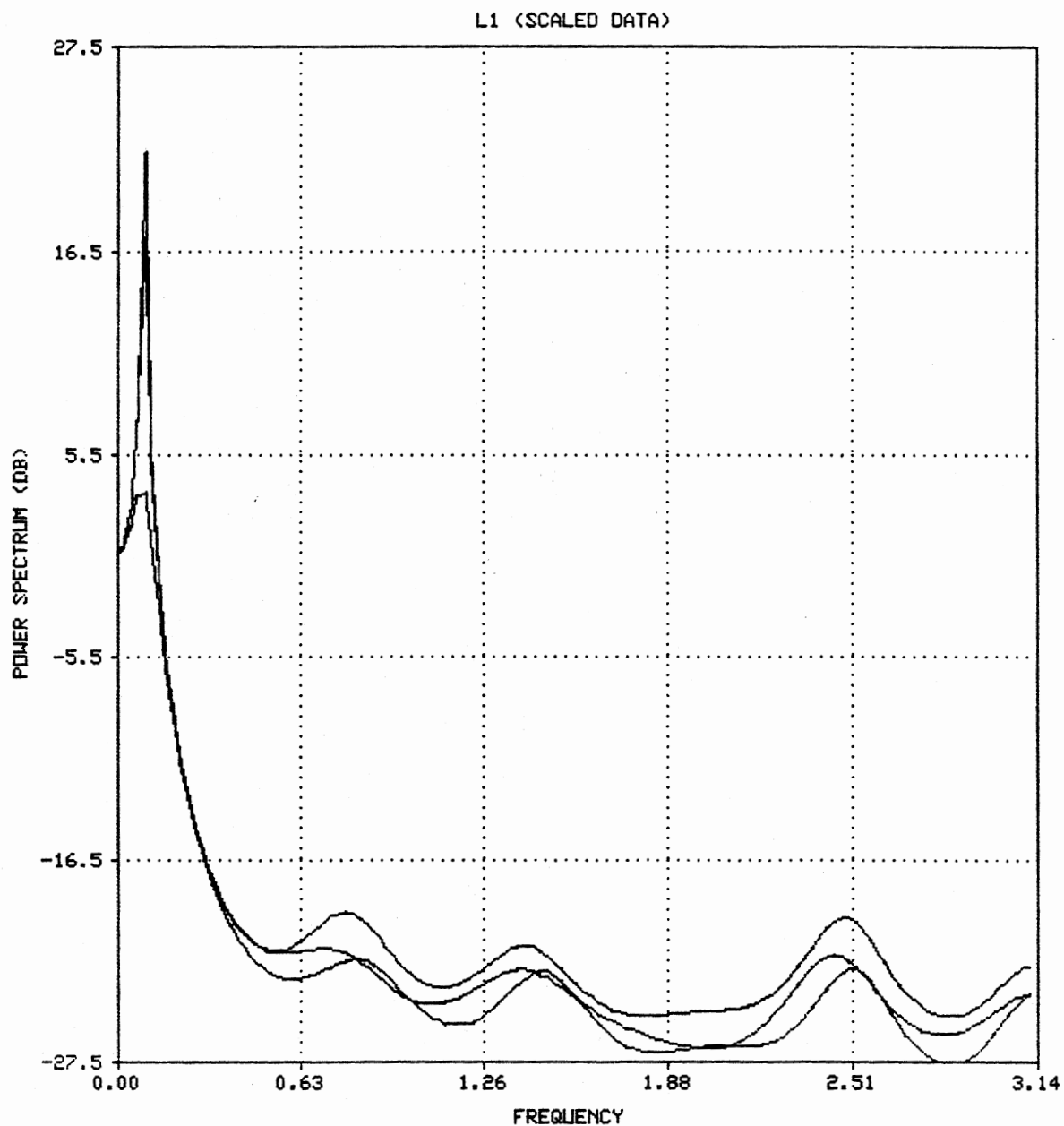


Figure 26. Power Spectrum of Scaled  $L_1$  by 1.0, 0.5, and 0.25 for a Two-Dimensional Shape;  $p=12$ ,  $q=0$ ;  $\bar{p}=4$ ,  $\bar{q}=0$ ,  $N=128$ ,  $L=64$ .

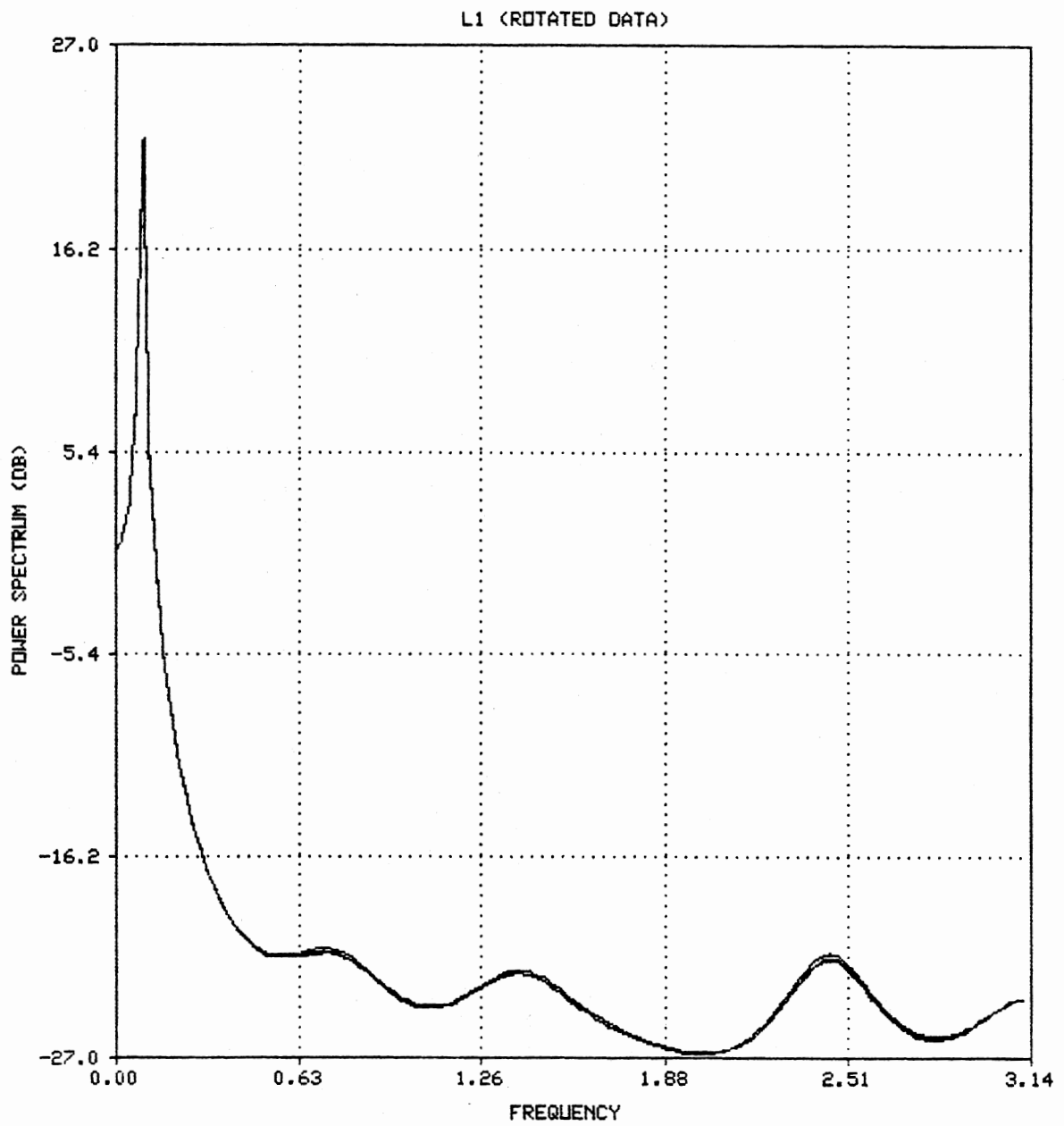


Figure 27. Power Spectrum of Rotated  $L_1$  by  $0.0$ ,  $\pi/4$ , and  $\pi/2$  for a Two-Dimensional Shape;  
 $p=12$ ,  $q=0$ ,  $\bar{p}=4$ ,  $\bar{q}=0$ ,  $N=128$ ,  $L=64$ .

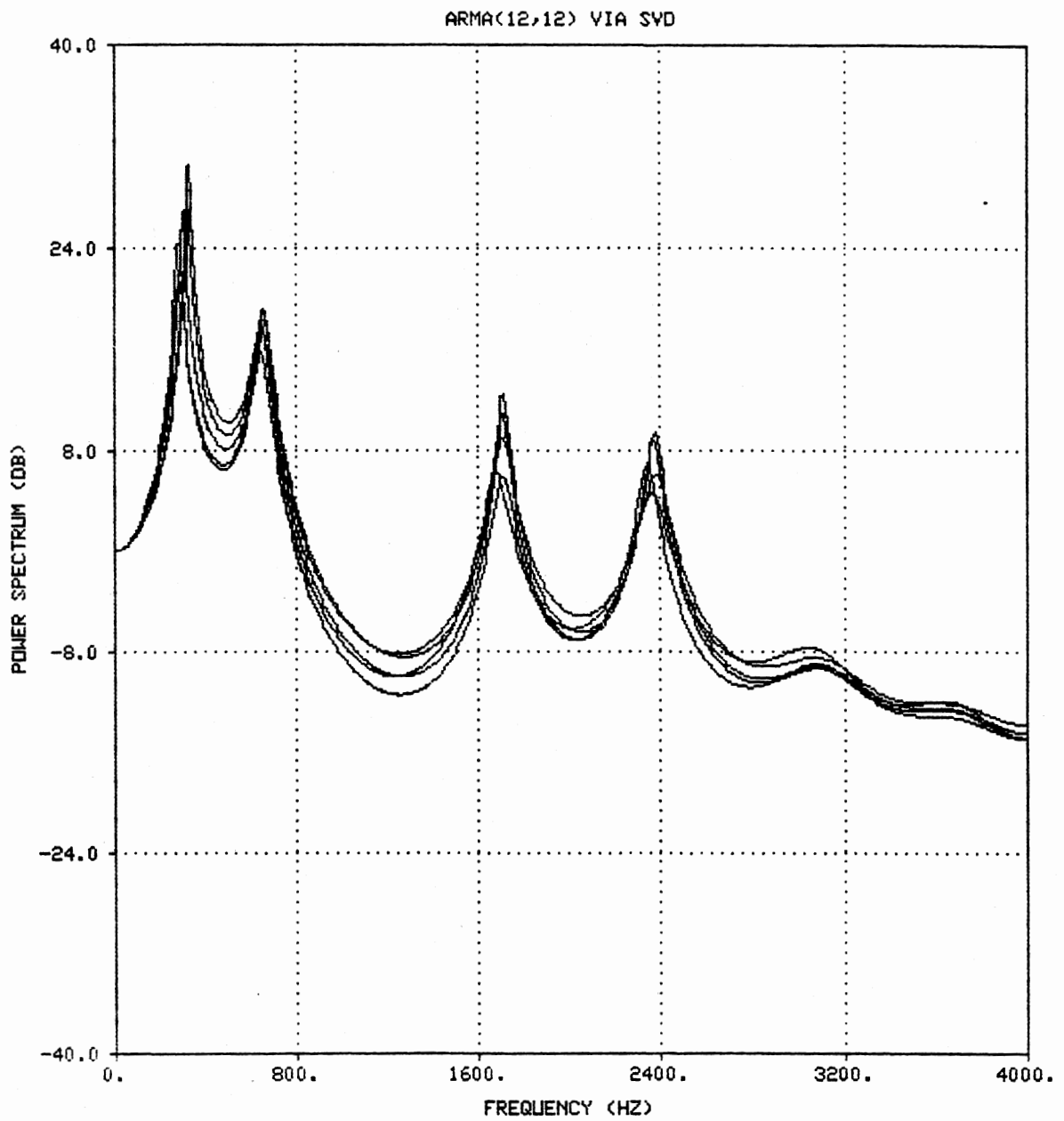


Figure 28. Power Spectrum of Vowel Portion /a/ of "Cat" by a Male Speaker Using EIGSP Solution;  
 $p=q=12$ ,  $\bar{p}=\bar{q}=8$ ,  $N=128$ ,  $L=64$ , 25% Overlap.



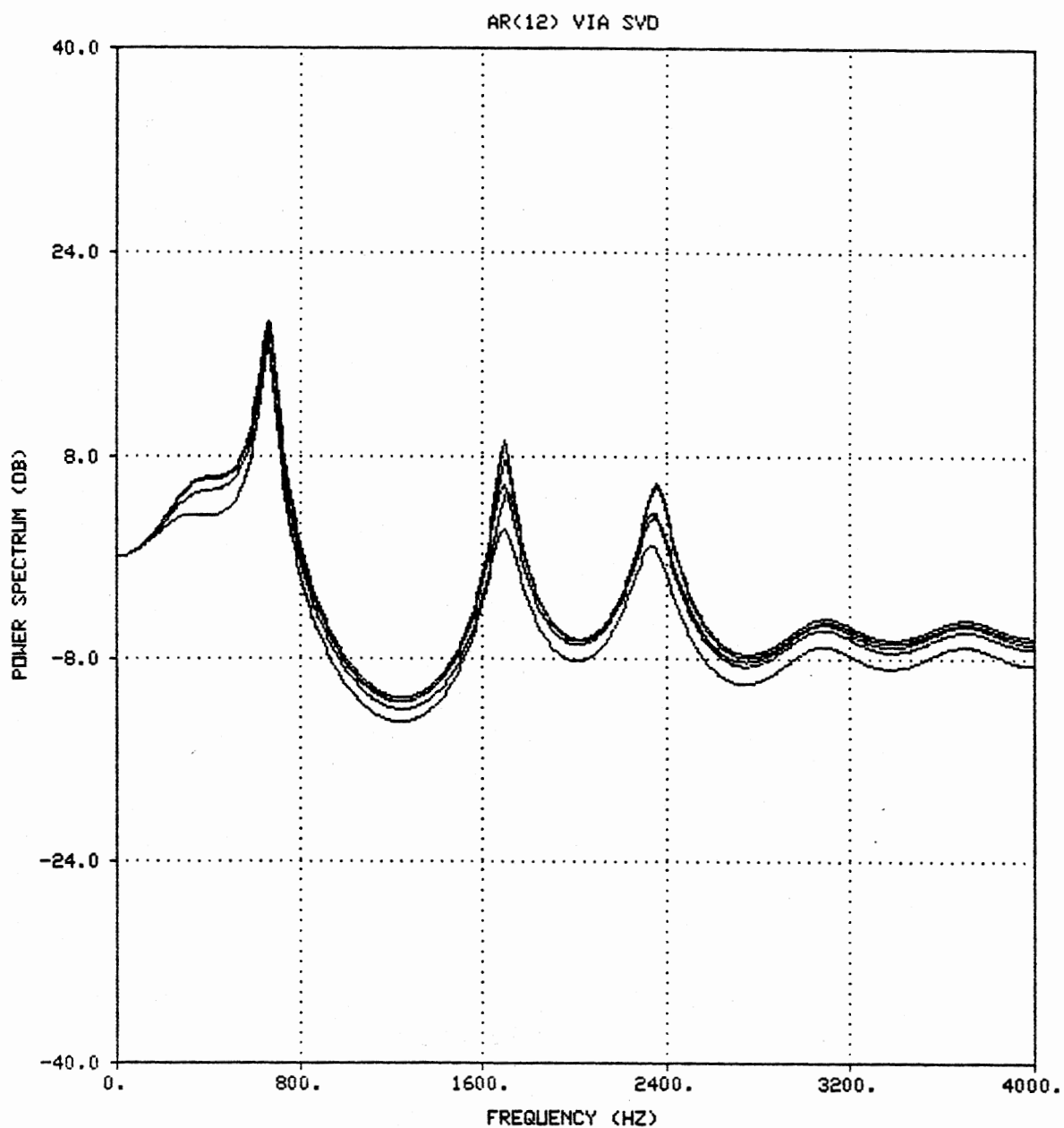


Figure 29. Power Spectrum of Vowel Portion /a/ of "Cat" by a Male Speaker Using EIGSP Solution;  
 $p=12$ ,  $\bar{p}=8$ ,  $q=\bar{q}=0$ ,  $N=128$ ,  $L=64$ , 25% Overlap.

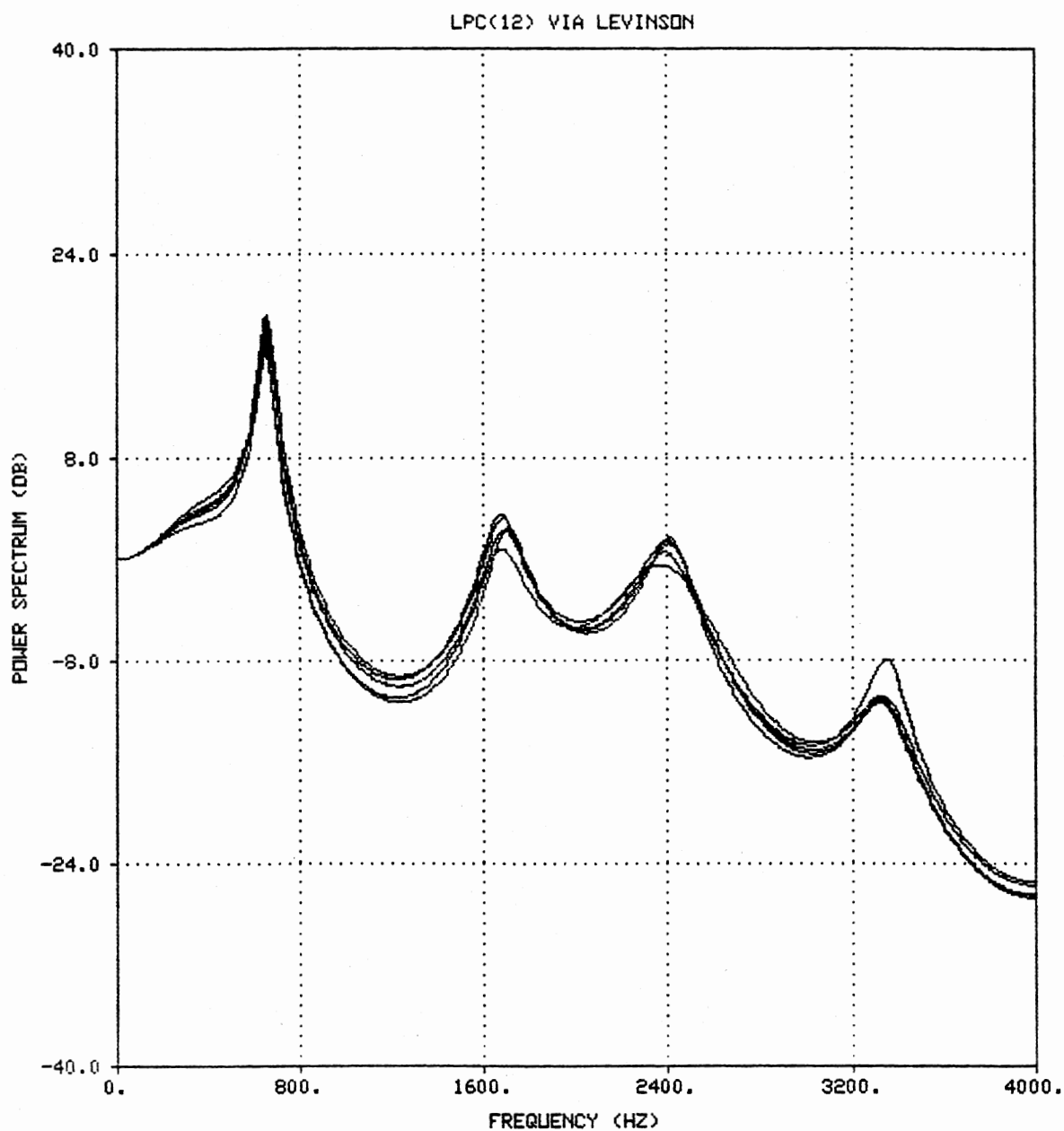


Figure 30. Power Spectrum of Vowel Portion /a/ of "Cat" by a Male Speaker Using LPC;  
p=12, N=128, L=64, 25% Overlap.

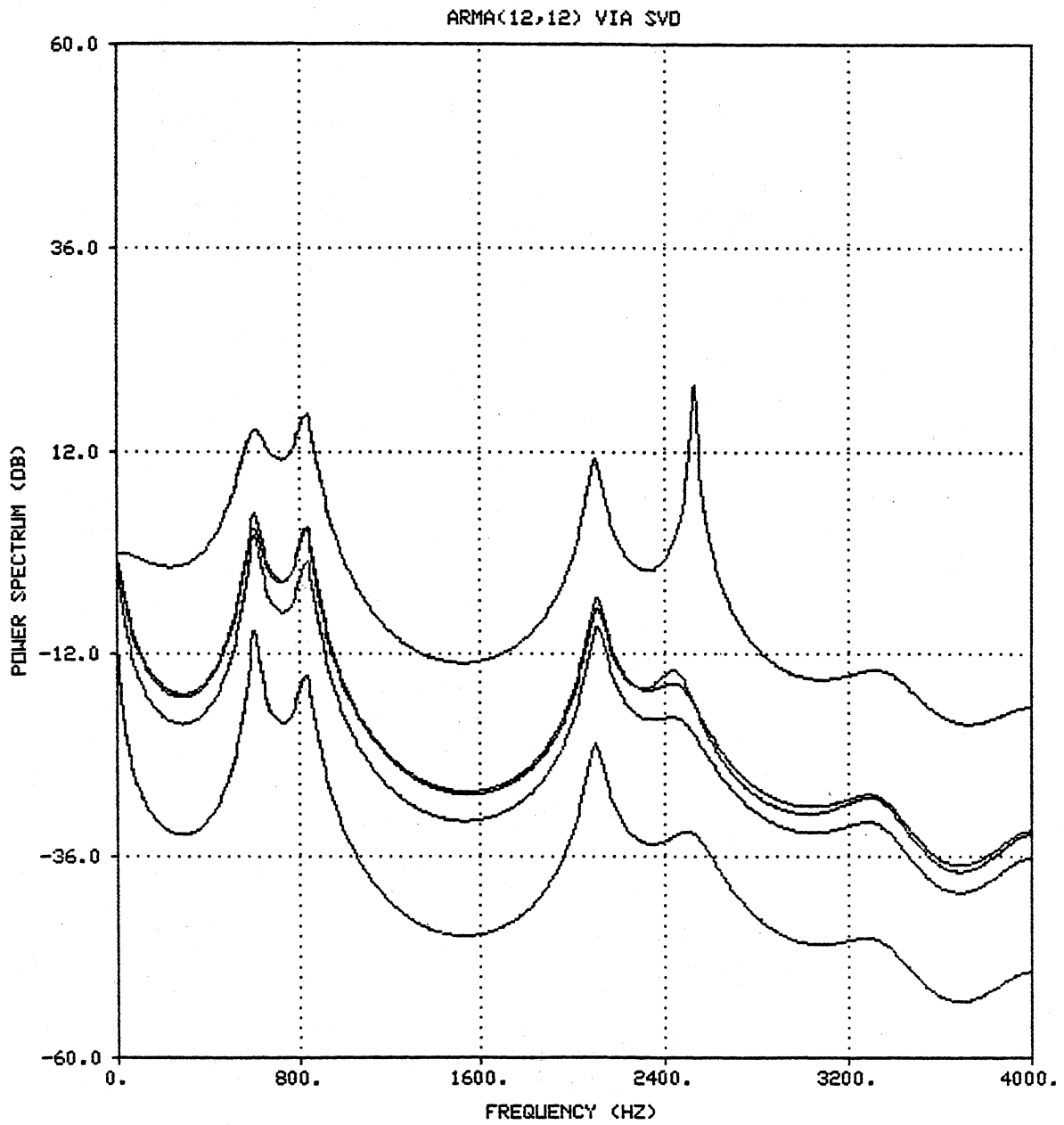


Figure 31. Power Spectrum of Vowel Portion /a/ of "Cat" by a Female Speaker Using EIGSP Solution;  $p=q=12$ ,  $\bar{p}=\bar{q}=8$ ,  $N=128$ ,  $L=64$ , 25% Overlap.

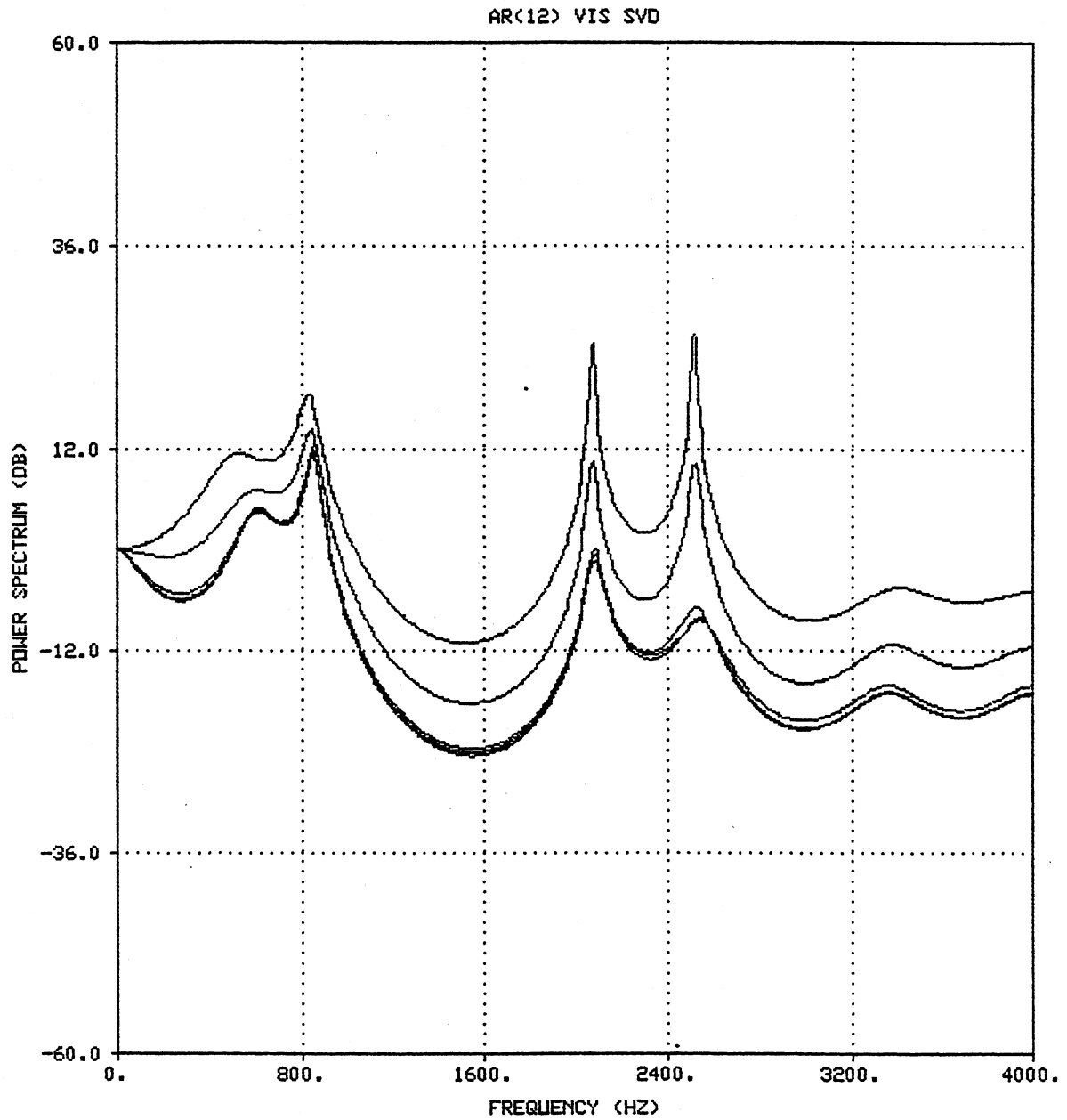


Figure 32. Power Spectrum of Vowel Portion /a/ of "Cat" by a Female Speaker Using EIGSP Solution;  $p=12$ ,  $p=8$ ,  $q=q=0$ ,  $N=128$ ,  $L=64$ , 25% Overlap.

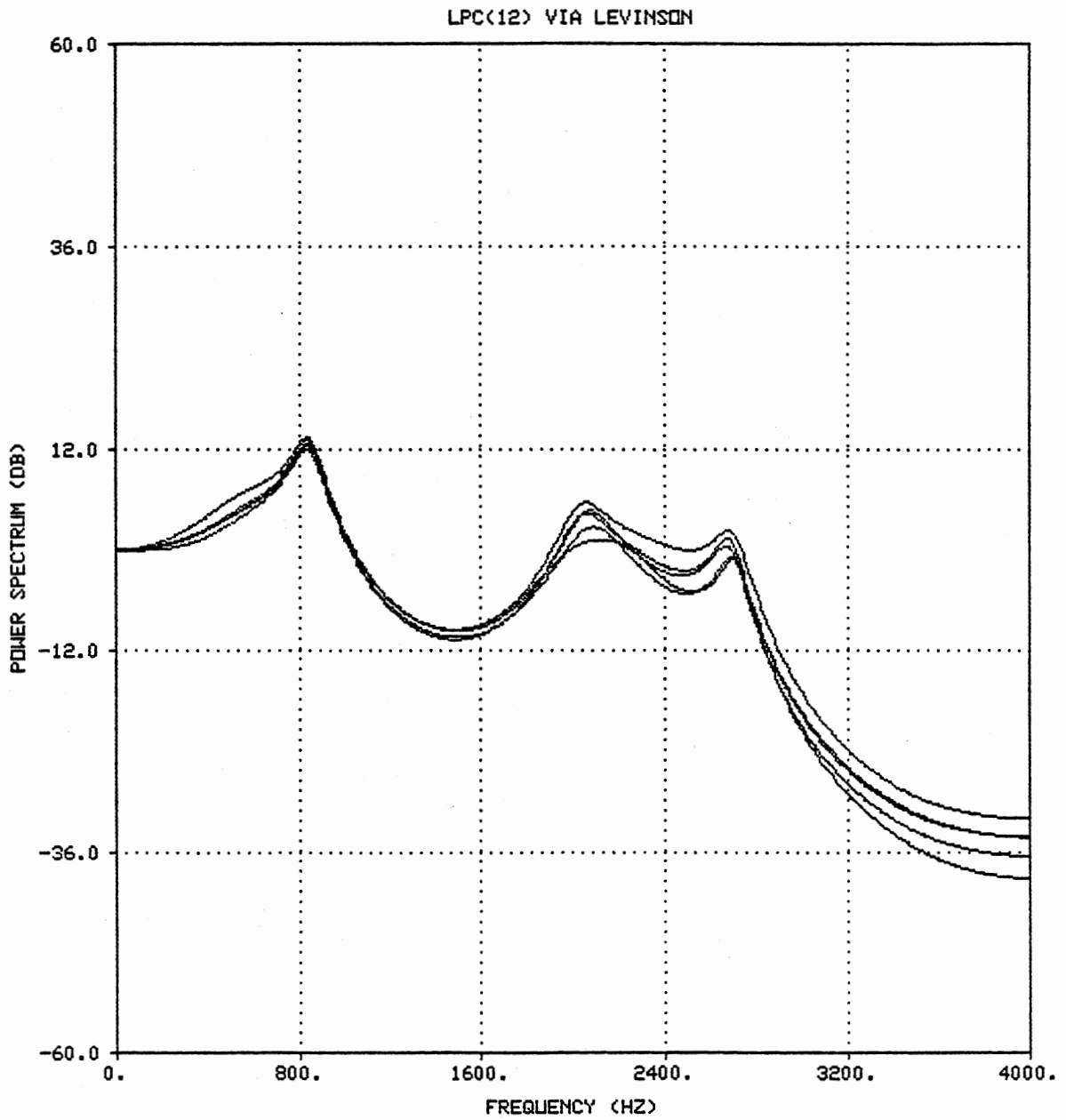


Figure 33. Power Spectrum of Vowel Portion /a/ of "Cat" by a Female Speaker Using LPC;  $p=12$ ,  $N=128$ ,  $L=64$ , 25% Overlap.

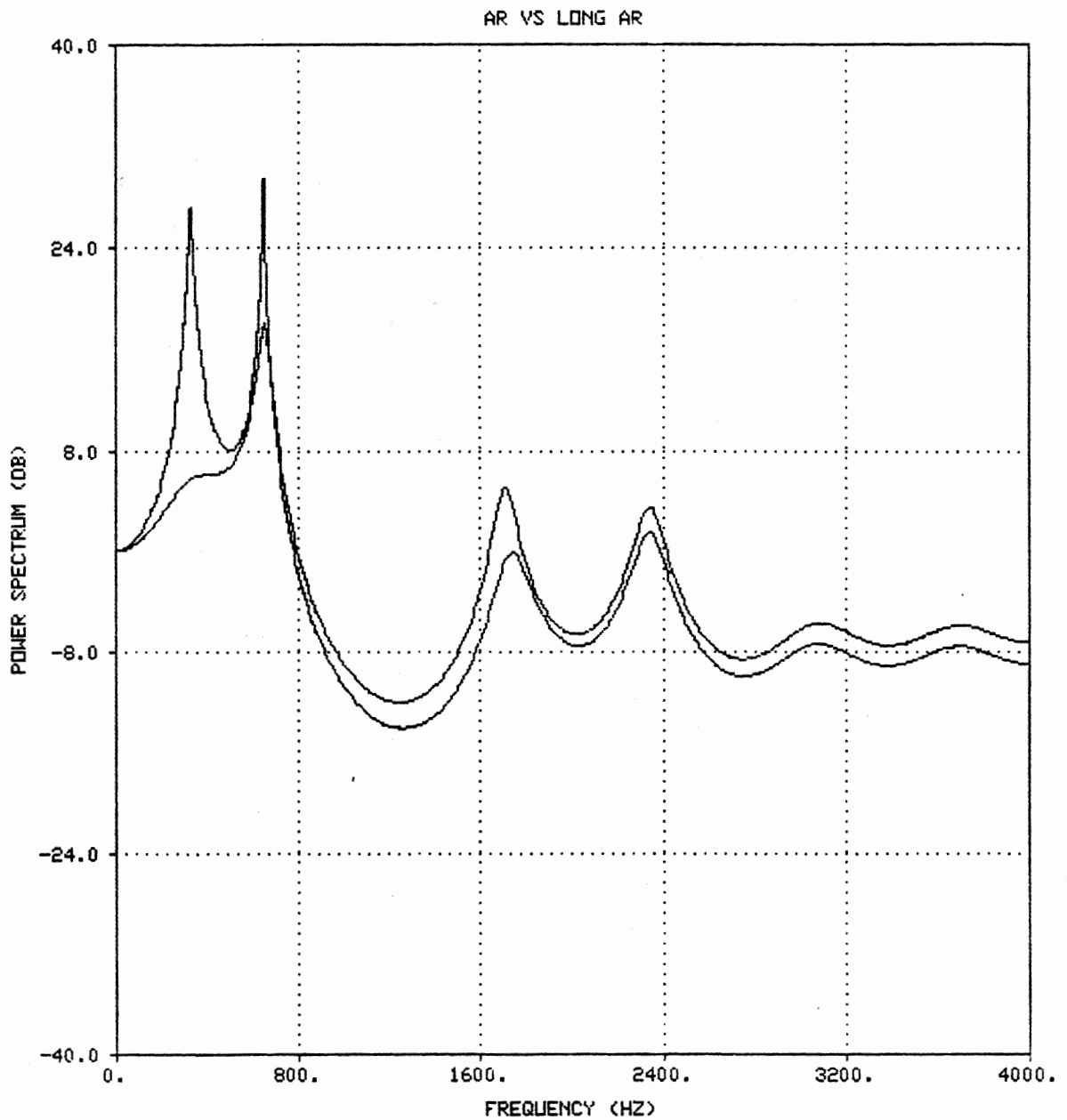


Figure 34. Comparison of the Power Spectrum of AR and Long AR for a Male Speaker;  
 $p=12$ ,  $\bar{p}=8$ ,  $q=\bar{q}=0$ ,  $N=128$ ,  $L=64$ , 25% Overlap.

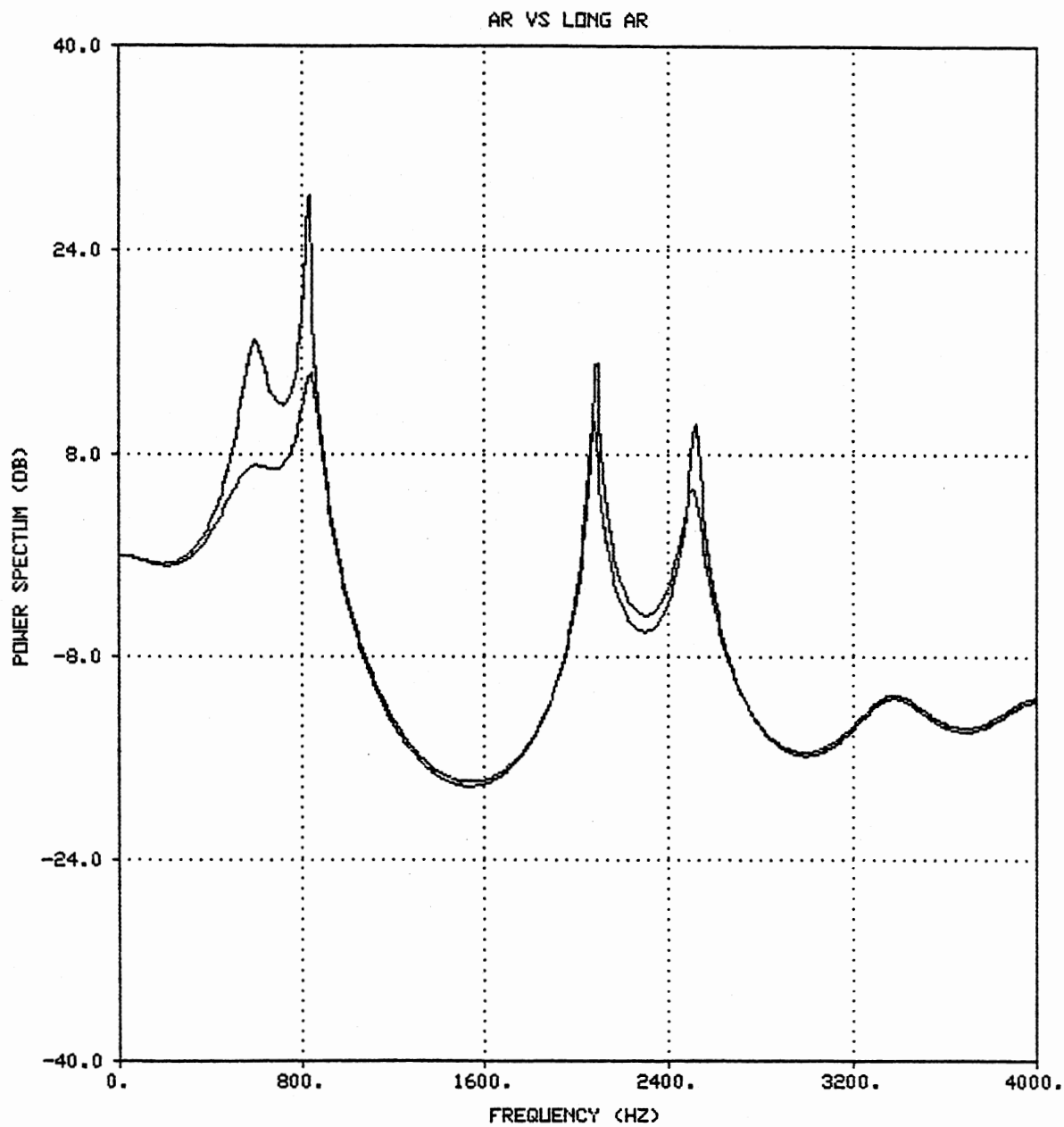


Figure 35. Comparison of the Power Spectrum of AR and Long AR for a Female Speaker;  $p=12$ ,  $\bar{p}=8$ ,  $q=q=0$ ,  $N=128$ ,  $L=64$ , 25% Overlap.

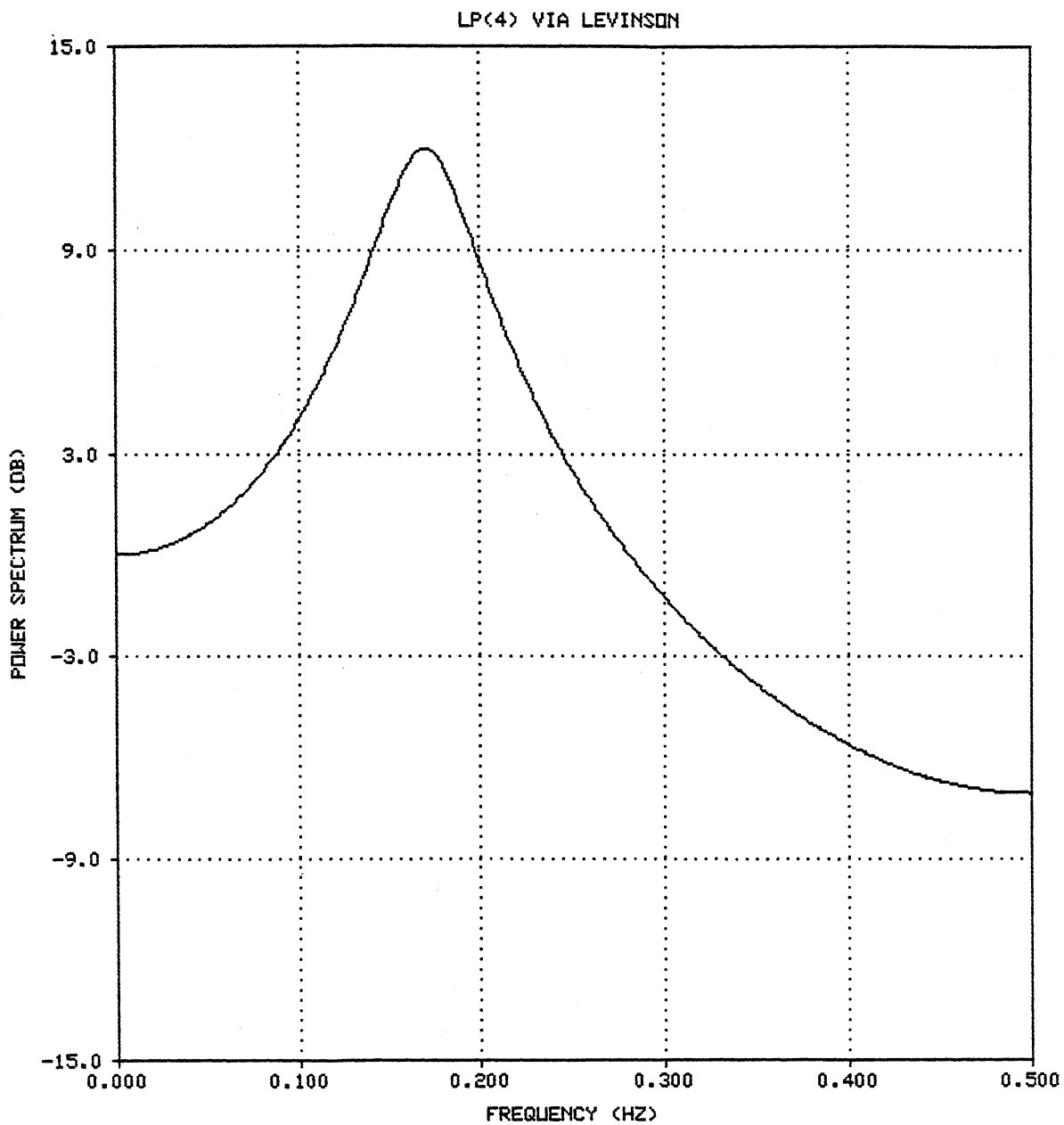


Figure 37. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using LP Solution;  
 $f_1 = 0.15$ ,  $f_2 = 0.185$ , SNR = 30 dB, N=8, p=4.



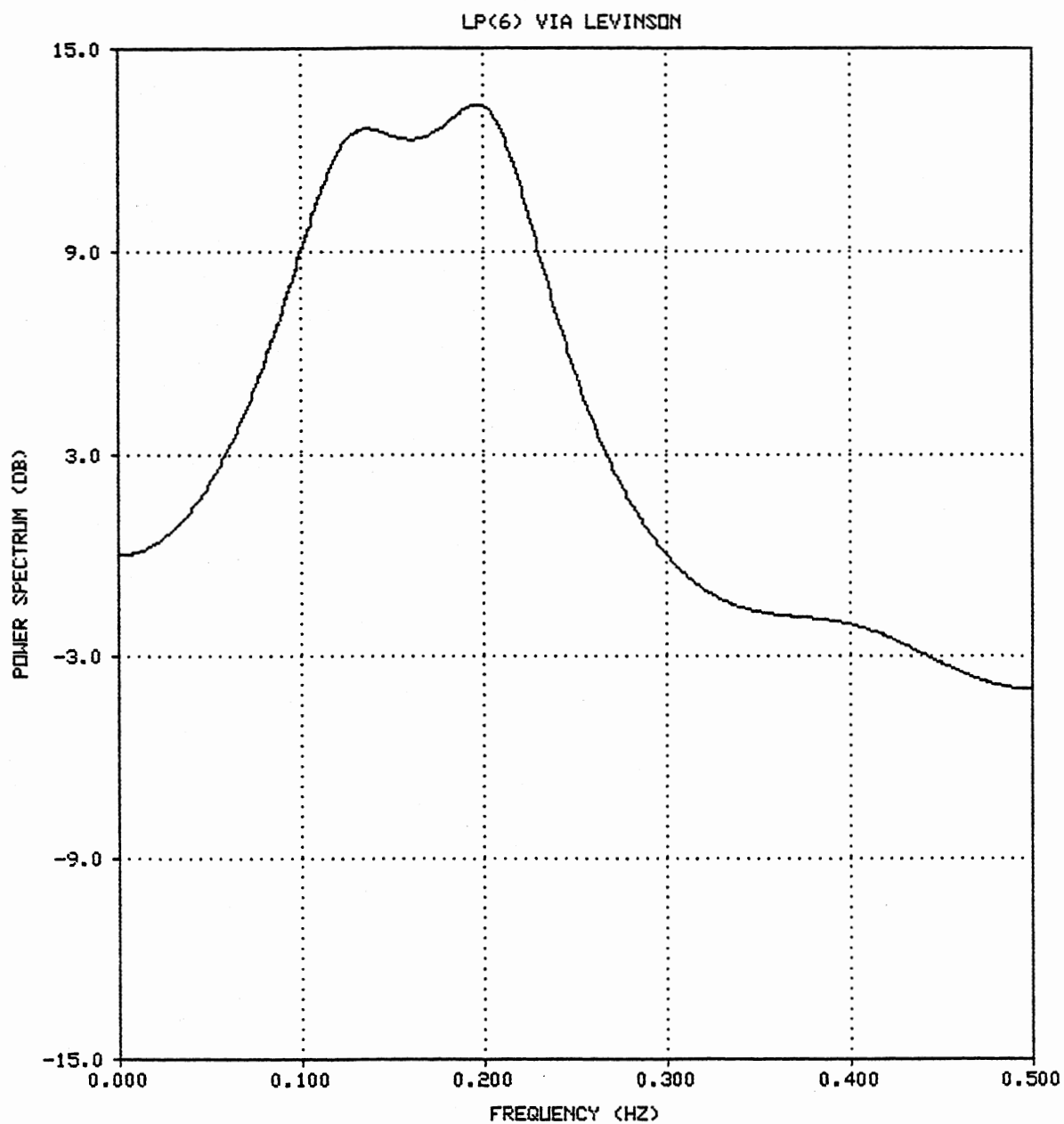


Figure 38. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using LP Solution;  
 $f_1 = 0.15$ ,  $f_2 = 0.185$ , SNR = 30 dB, N=8, p=6.

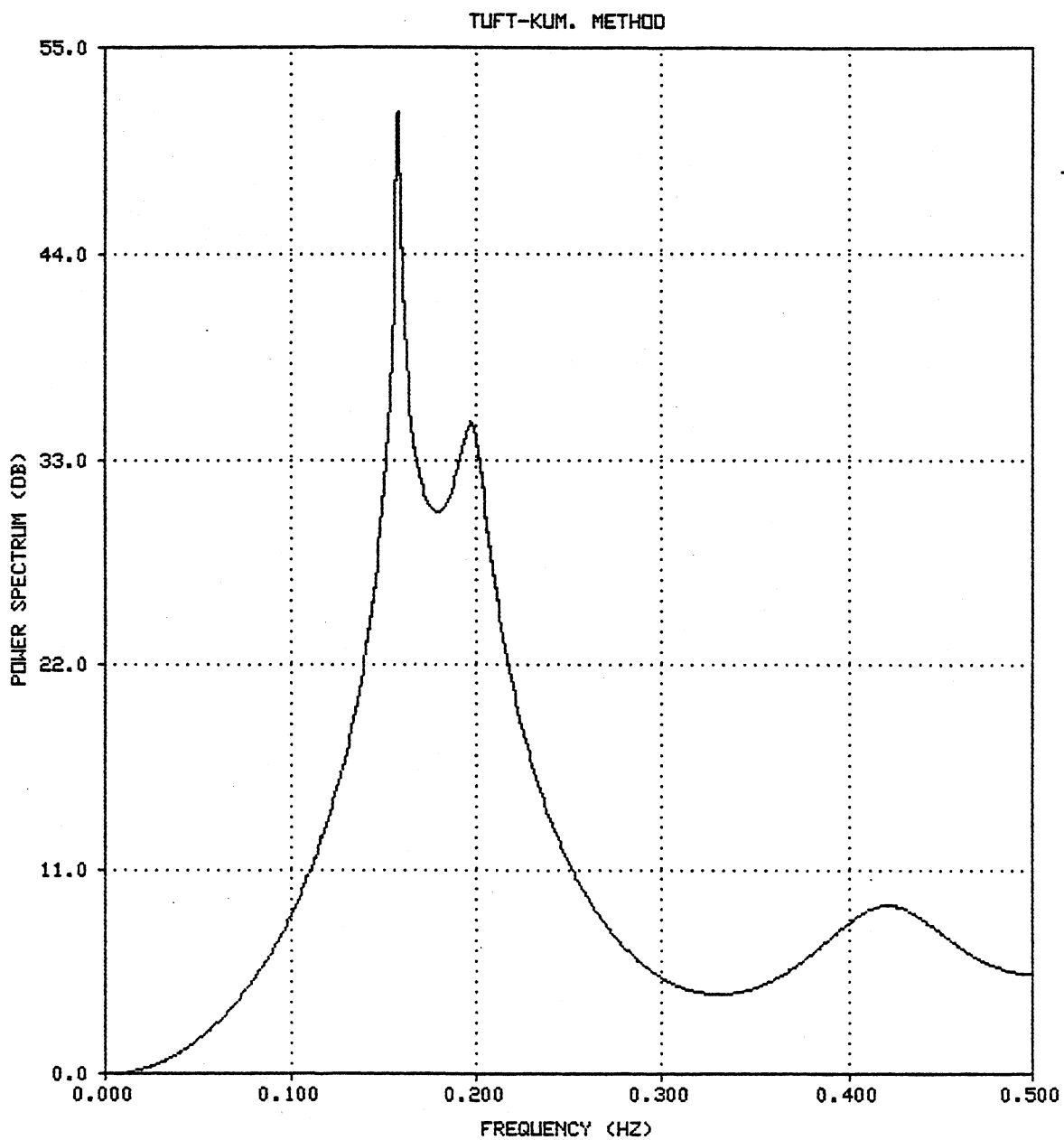


Figure 39. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using Tuft-Kumaresan Method;  
 $f_1 = 0.15$ ,  $f_2 = 0.185$ , SNR = 30 dB, N=8, L=6.

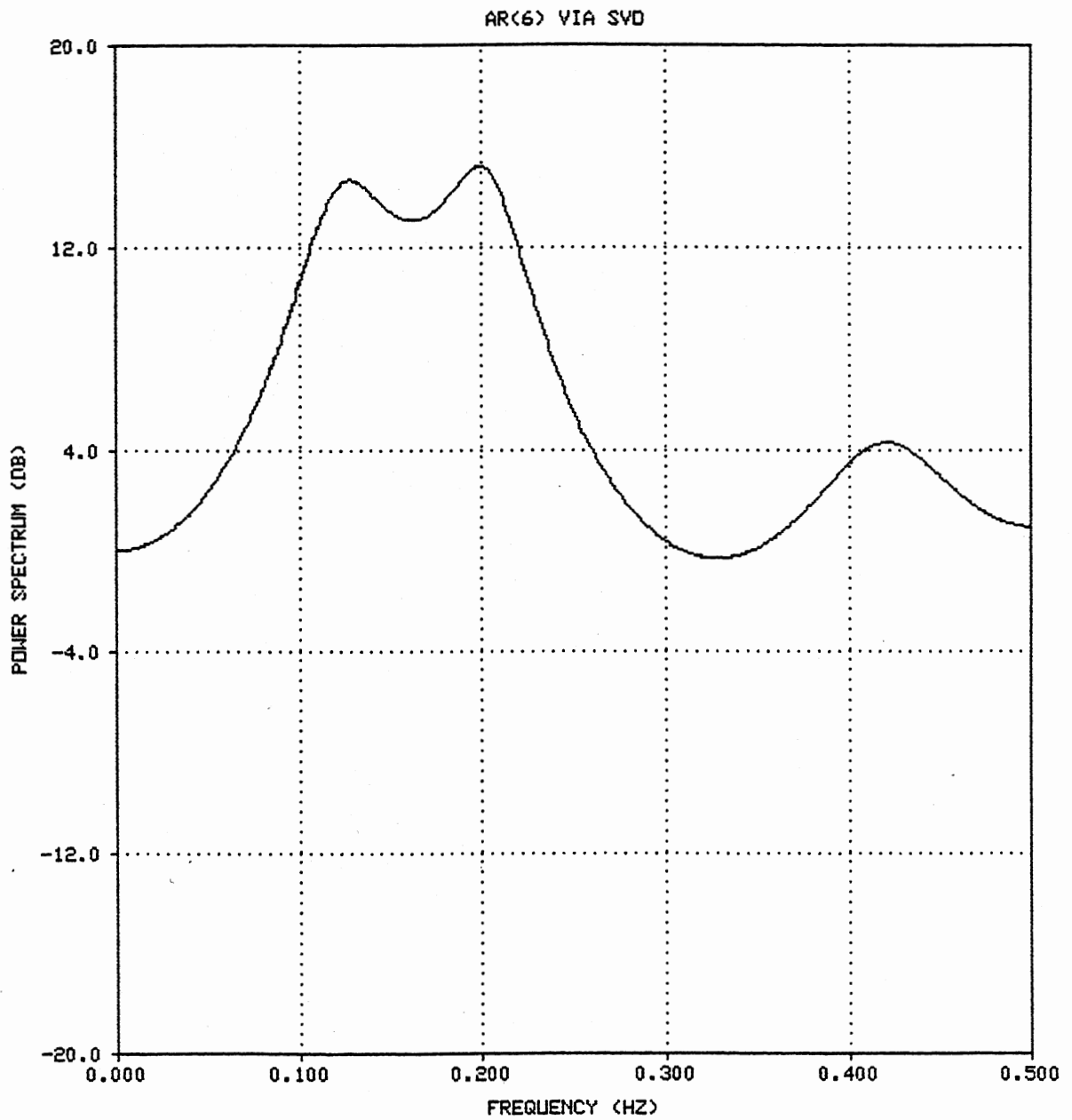


Figure 40. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution;  
 $f_1 = 0.15$ ,  $f_2 = 0.185$ , SNR = 30 dB,  $p=6$ ,  $\bar{p}=4$ ,  
 $q=q=0$ ,  $N=8$ ,  $L=7$ .

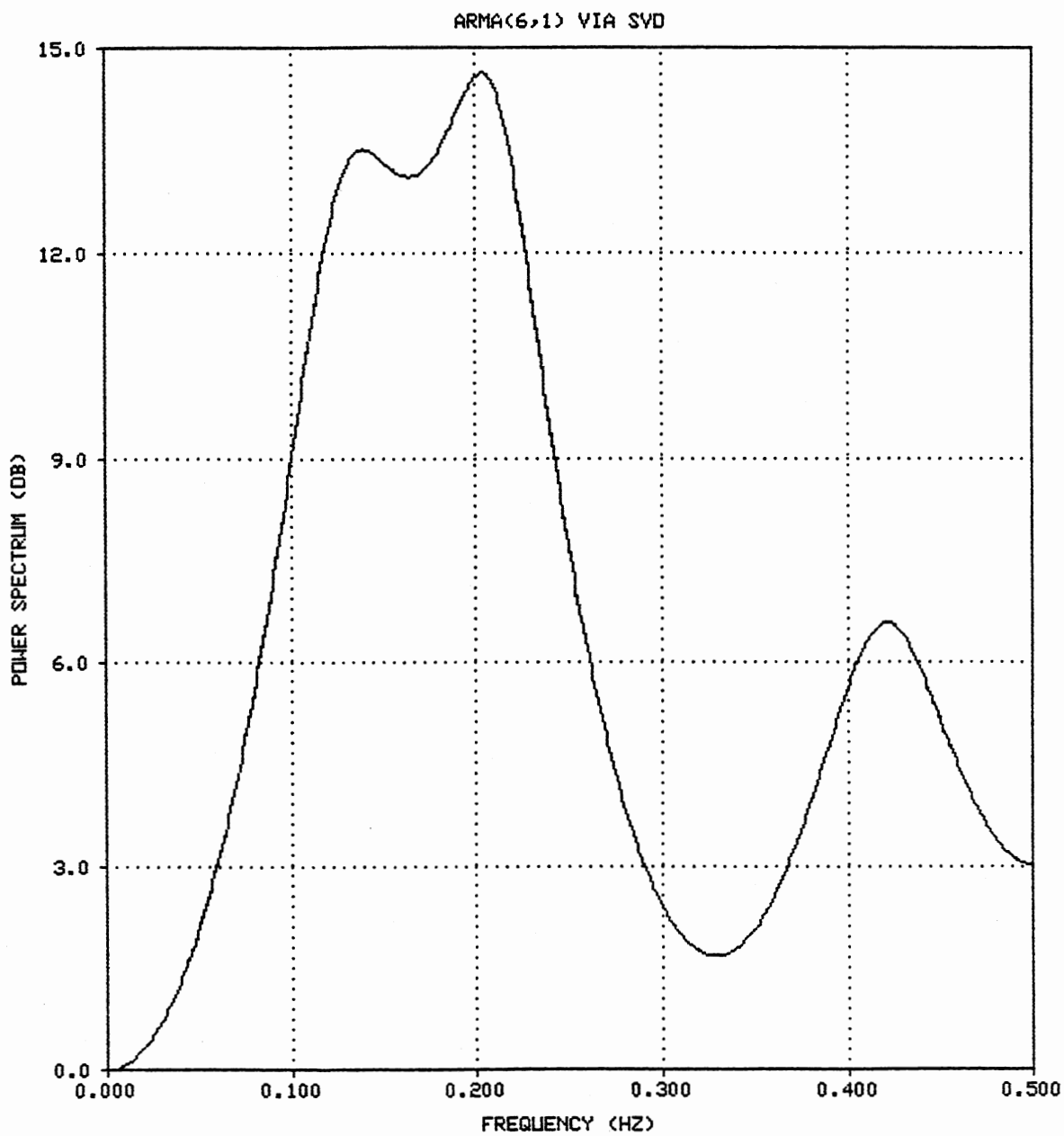


Figure 41. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution;  
 $f_{1_} = 0.15$ ,  $f_2 = 0.185$ , SNR = 30 dB,  $p=6$ ,  $\bar{p}=4$ ,  
 $q=q=1$ ,  $N=8$ ,  $L=7$ .

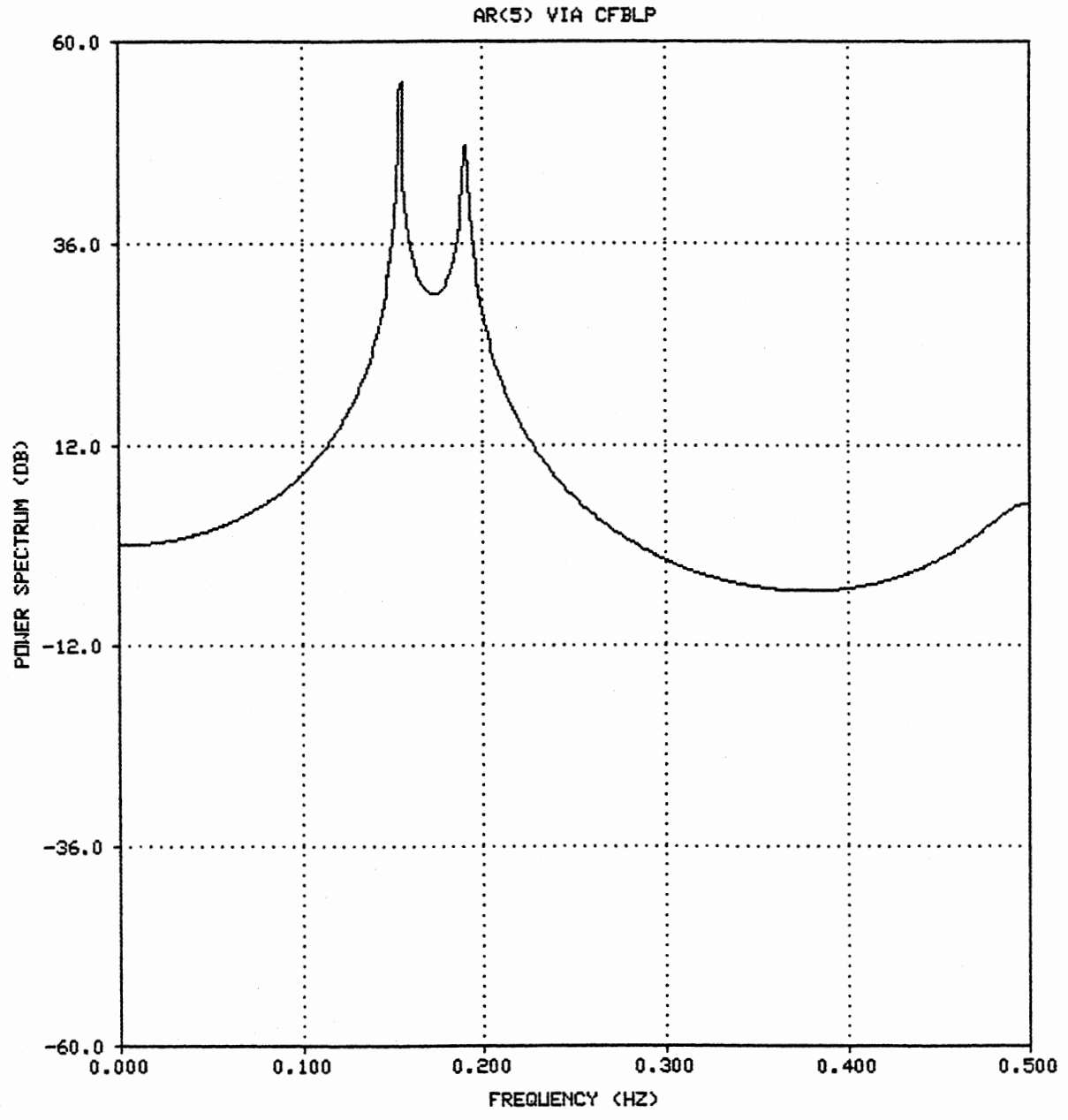


Figure 42. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of CFBLP;  $f_{1\_} = 0.15$ ,  $f_2 = 0.185$ , SNR = 30 dB,  $p=5$ ,  $\bar{p}=4$ ,  $q=q=0$ ,  $N=8$ .

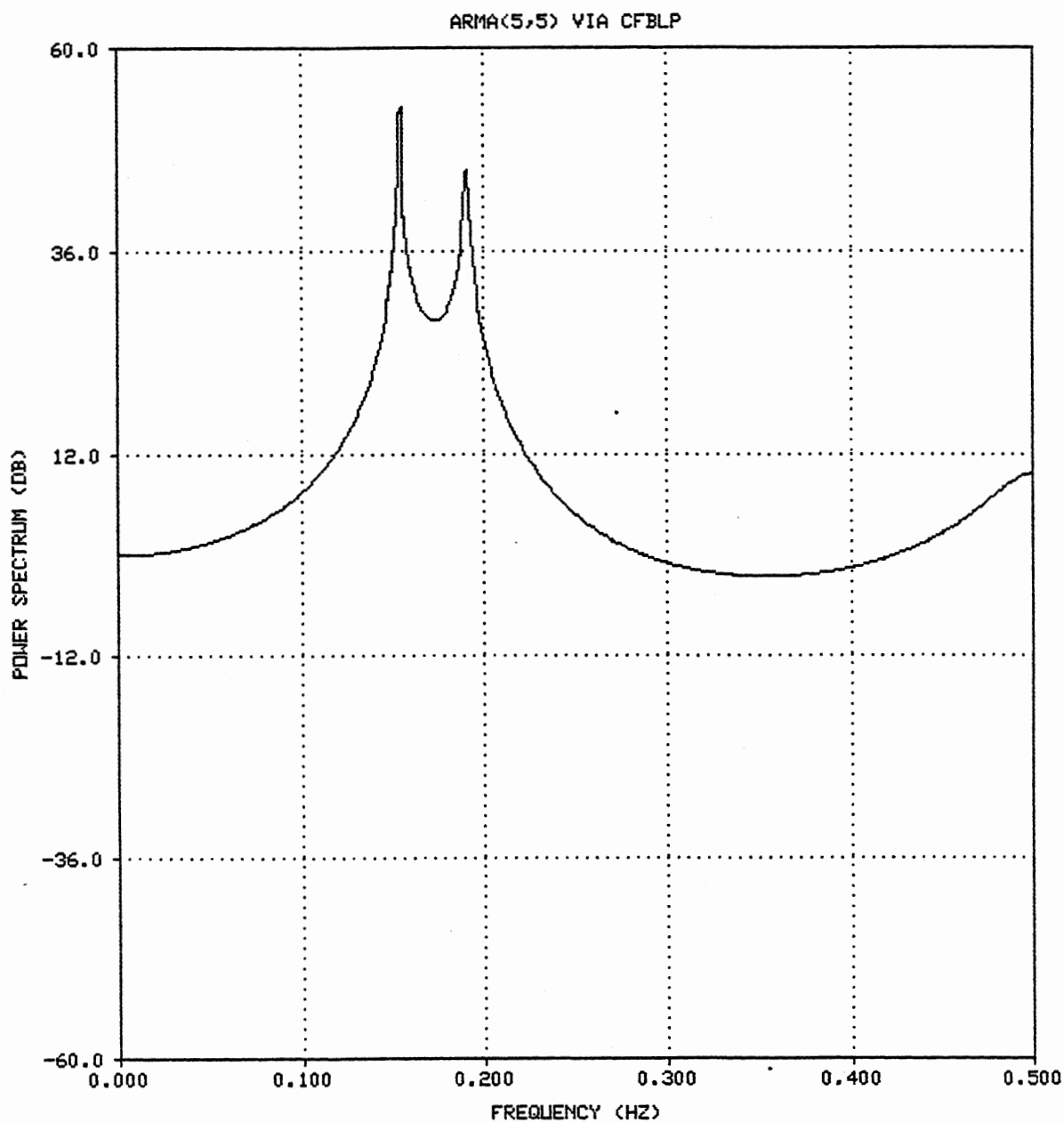


Figure 43. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of CFBLP;  
 $f_1 = 0.15$ ,  $f_2 = 0.185$ , SNR = 30 dB,  $p=q=5$ ,  
 $p=q=4$ ,  $N=8$ .

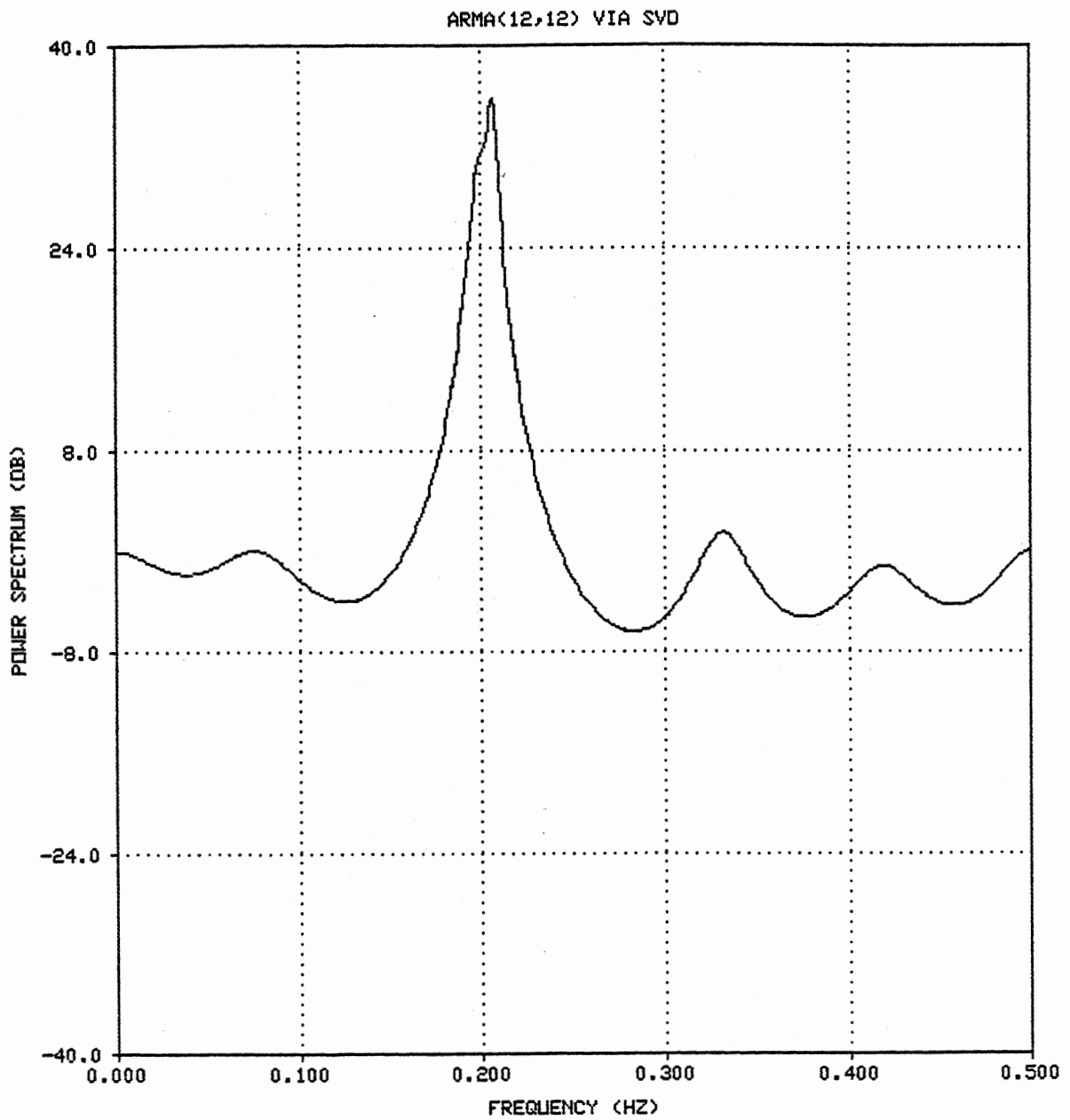


Figure 44. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of Correlation Matrix;  $f_1 = 0.20$ ,  $f_2 = 0.21$ , SNR = 0 dB,  $p=q=12$ ,  $\bar{p}=\bar{q}=4$ ,  $N=128$ ,  $L=64$ .

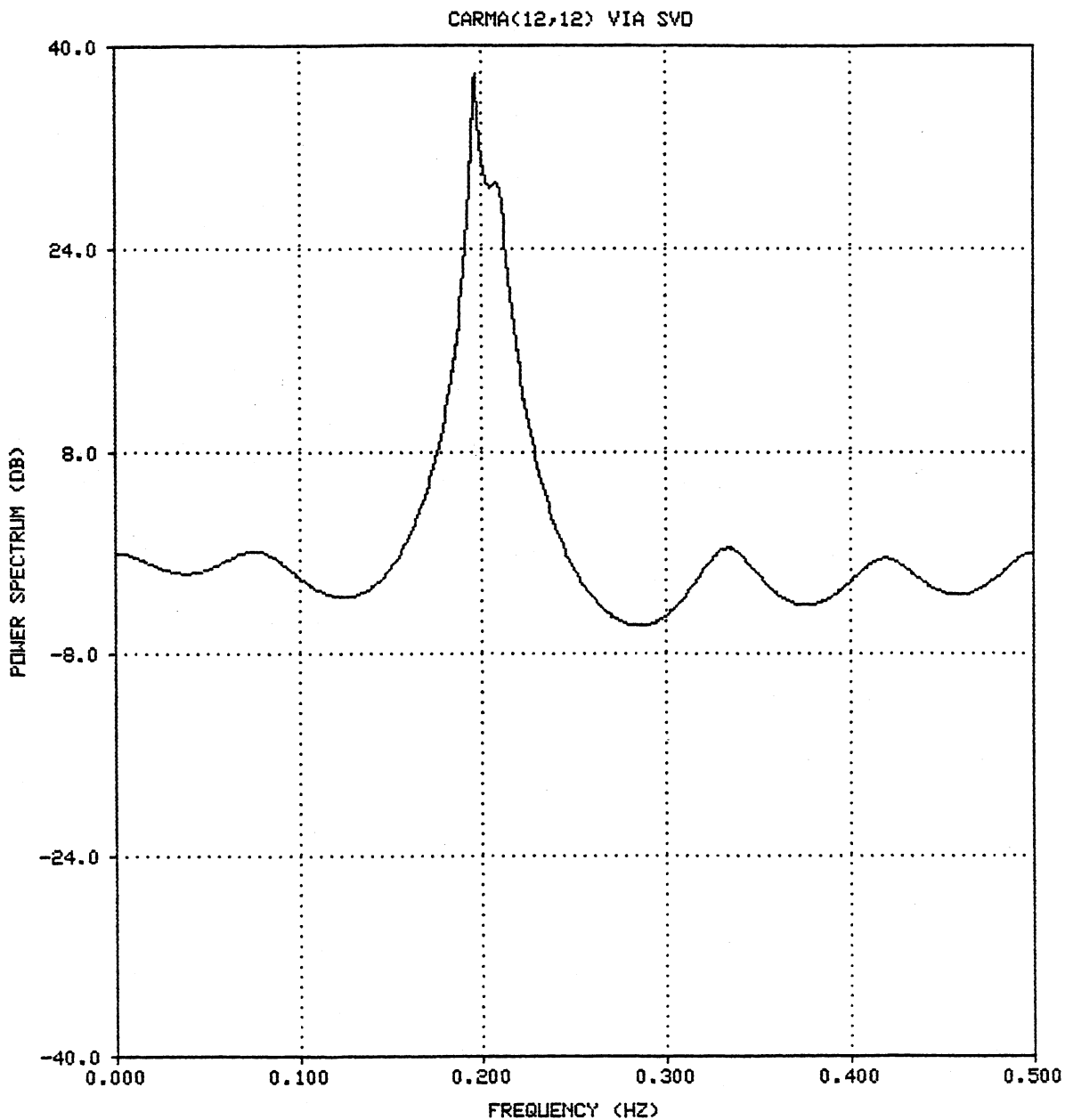


Figure 45. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of Circular Correlation Matrix;  
 $f_1 = 0.20$ ,  $f_2 = 0.21$ , SNR = 0 dB,  $p=q=12$ ,  $\bar{p}=\bar{q}=4$ ,  
 $N=128$ ,  $L=64$ .



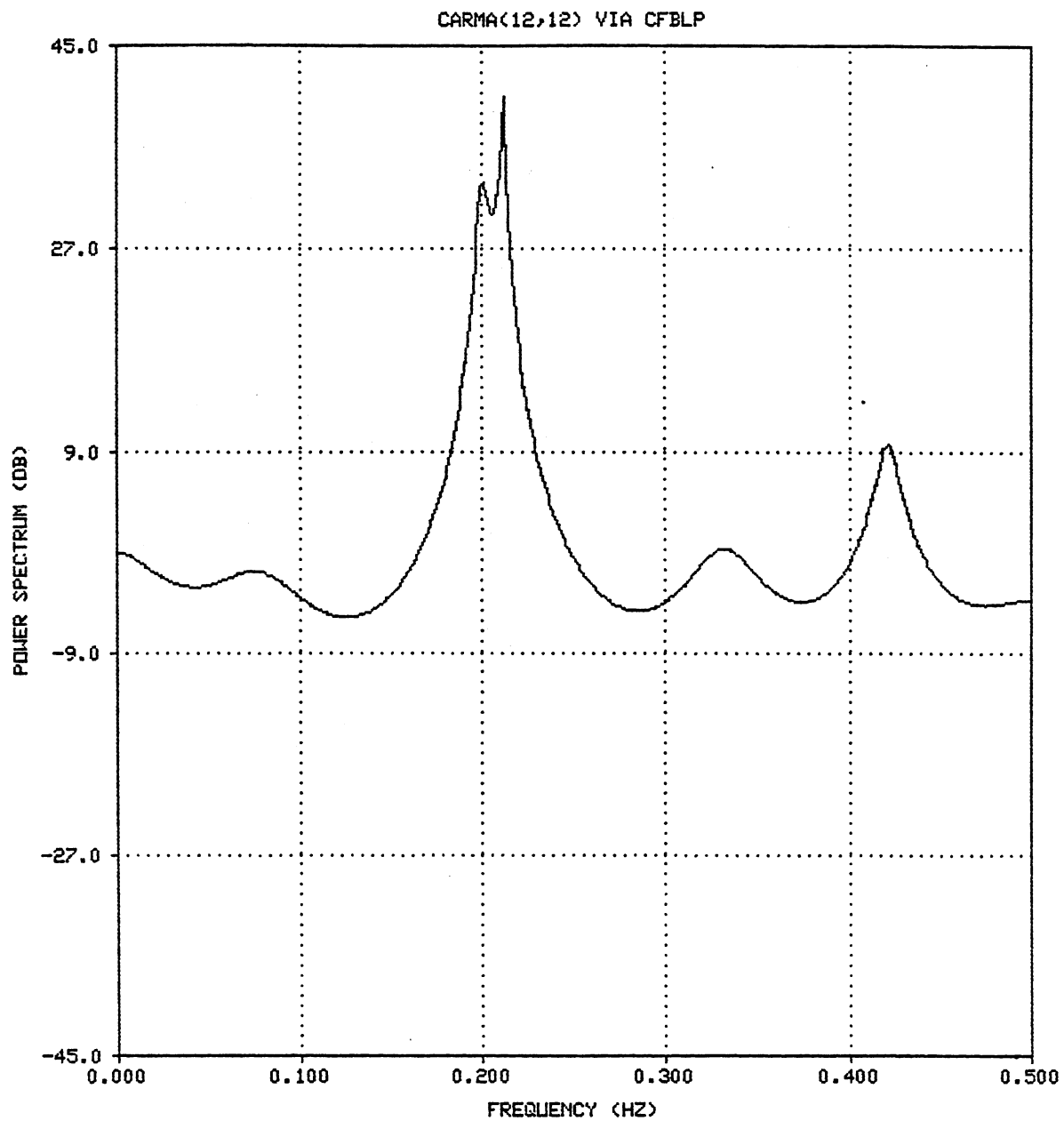


Figure 46. Spectral Estimate for Two Sinewaves in White Gaussian Noise Using EIGSP Solution of CFBLP Data Matrix;  
 $f_1 = 0.20$ ,  $f_2 = 0.21$ , SNR = 0 dB,  $p=q=12$ ,  
 $p=q=4$ ,  $N=128$ .

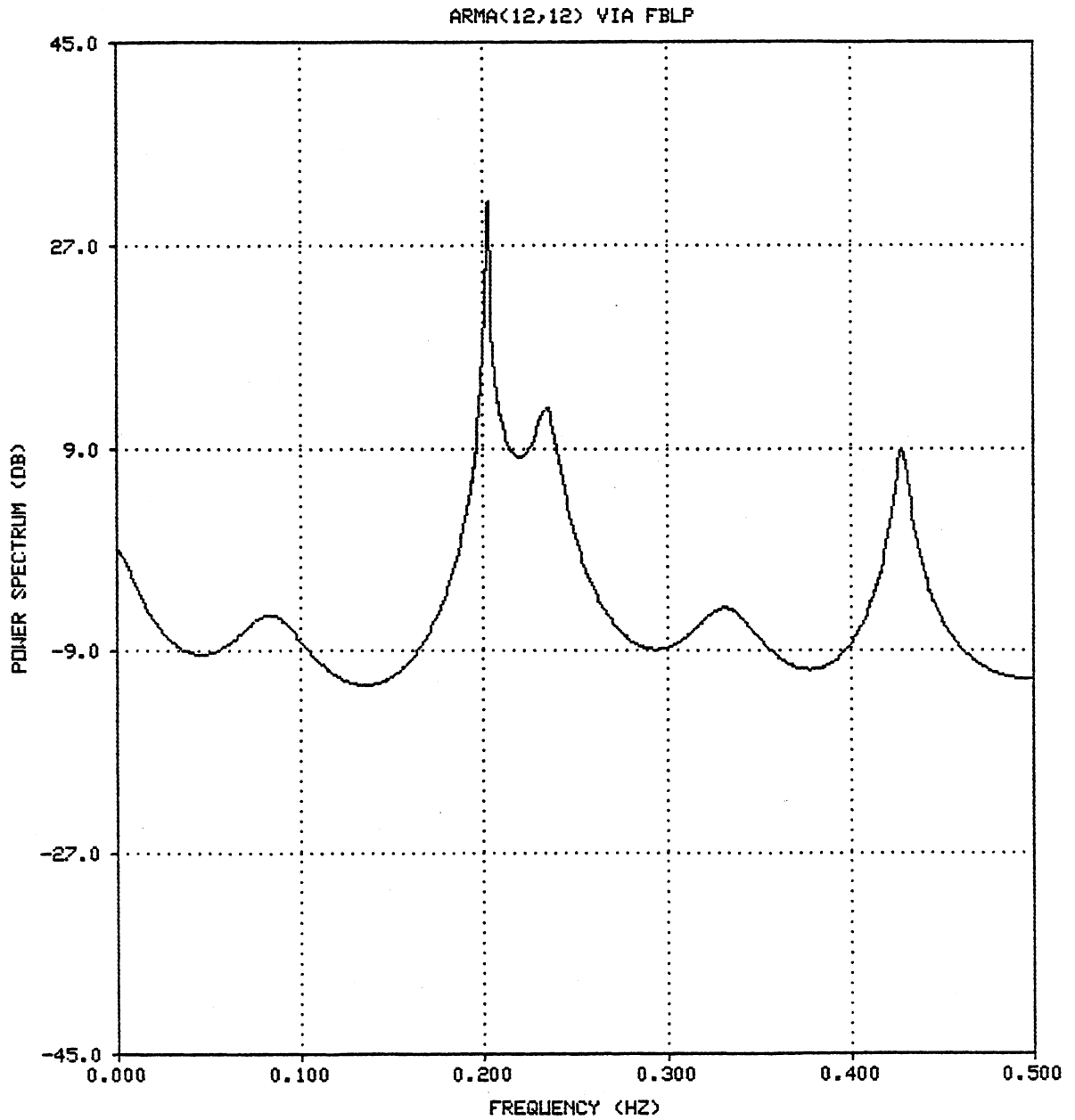


Figure 47. Spectral Estimate of Two Sinewaves in White Gaussian Noise Using EIGSP Solution of FBLP Data Matrix;  
 $f_1 = 0.20$ ,  $f_2 = 0.21$ , SNR = 0 dB,  $p=q=12$ ,  
 $p=q=4$ ,  $N=128$ .

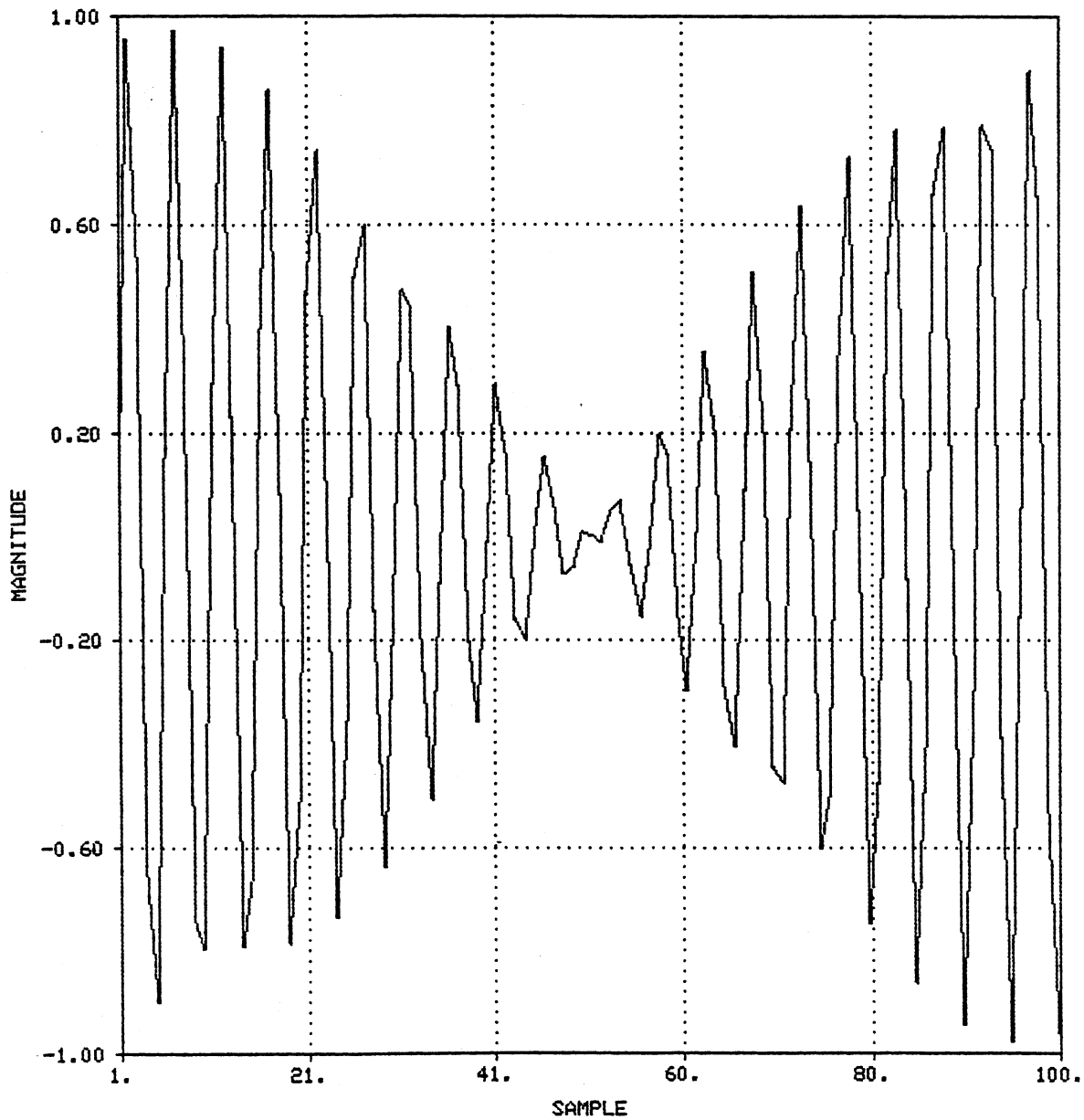


Figure 48. Exact Autocorrelation of Sum of Two Cosinewaves;  
 $f_1=0.2$ ,  $f_2=0.21$ ,  $N=100$ .

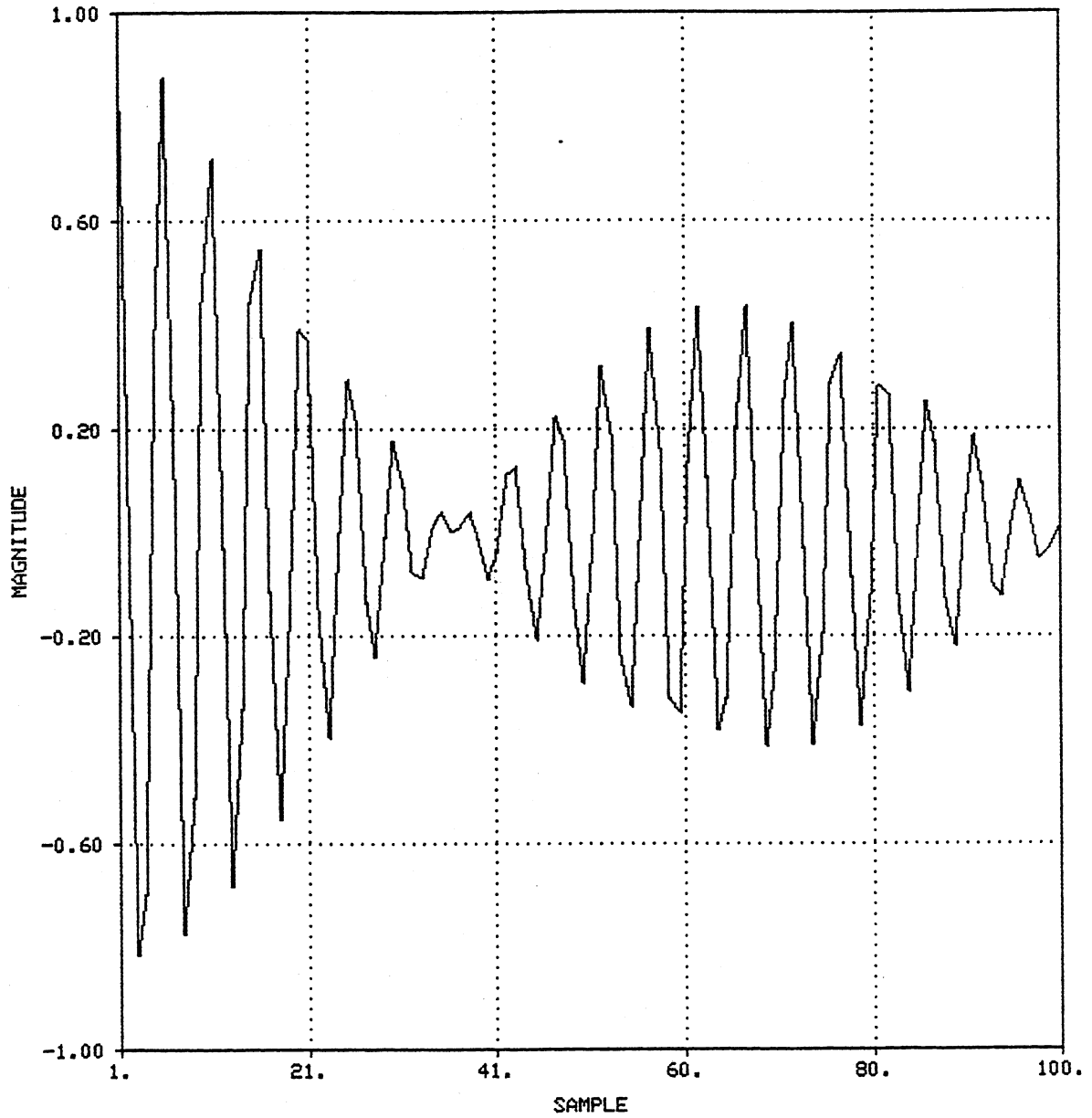


Figure 49. Regular Autocorrelation of Sum of Two Cosinewaves;  
 $f_1=0.2$ ,  $f_2=0.21$ ,  $N=100$ .

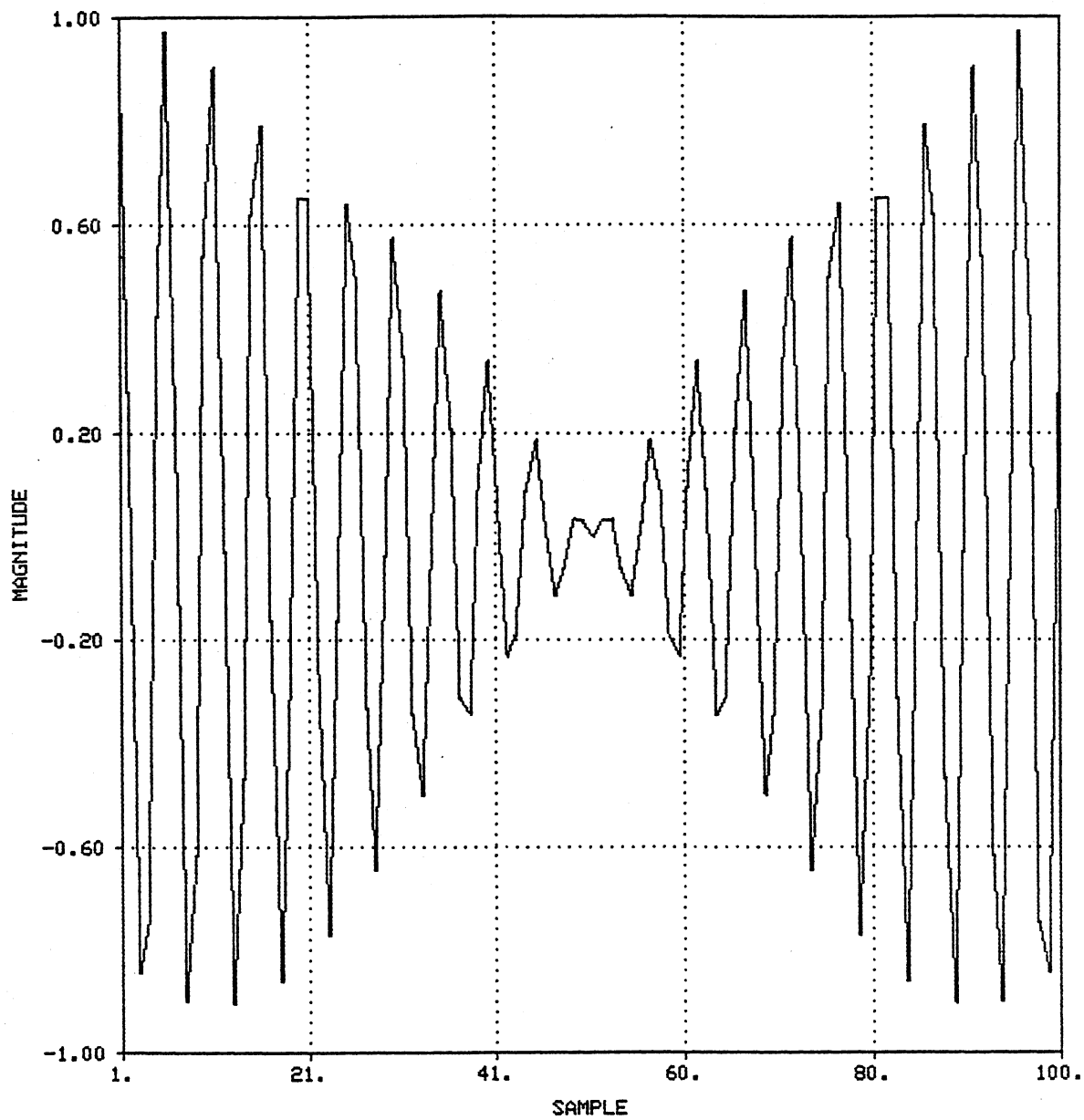


Figure 50. Circular Autocorrelation of Sum of Two Cosinewaves;  
 $f_1=0.2$ ,  $f_2=0.21$ ,  $N=100$ .

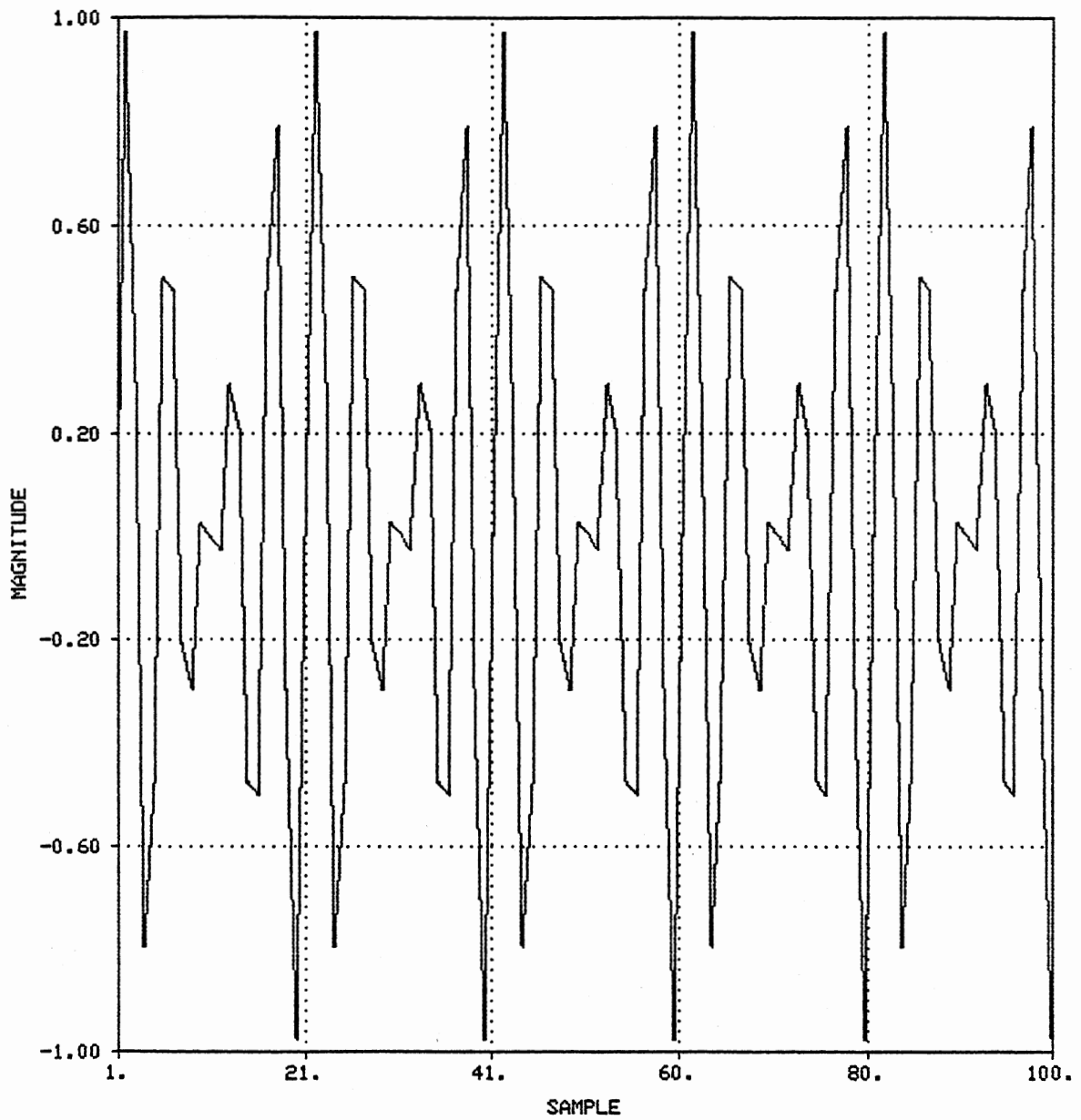


Figure 51. Exact Autocorrelation of Sum of Two Cosinewaves;  
 $f_1=0.2$ ,  $f_2=0.25$ ,  $N=100$ .

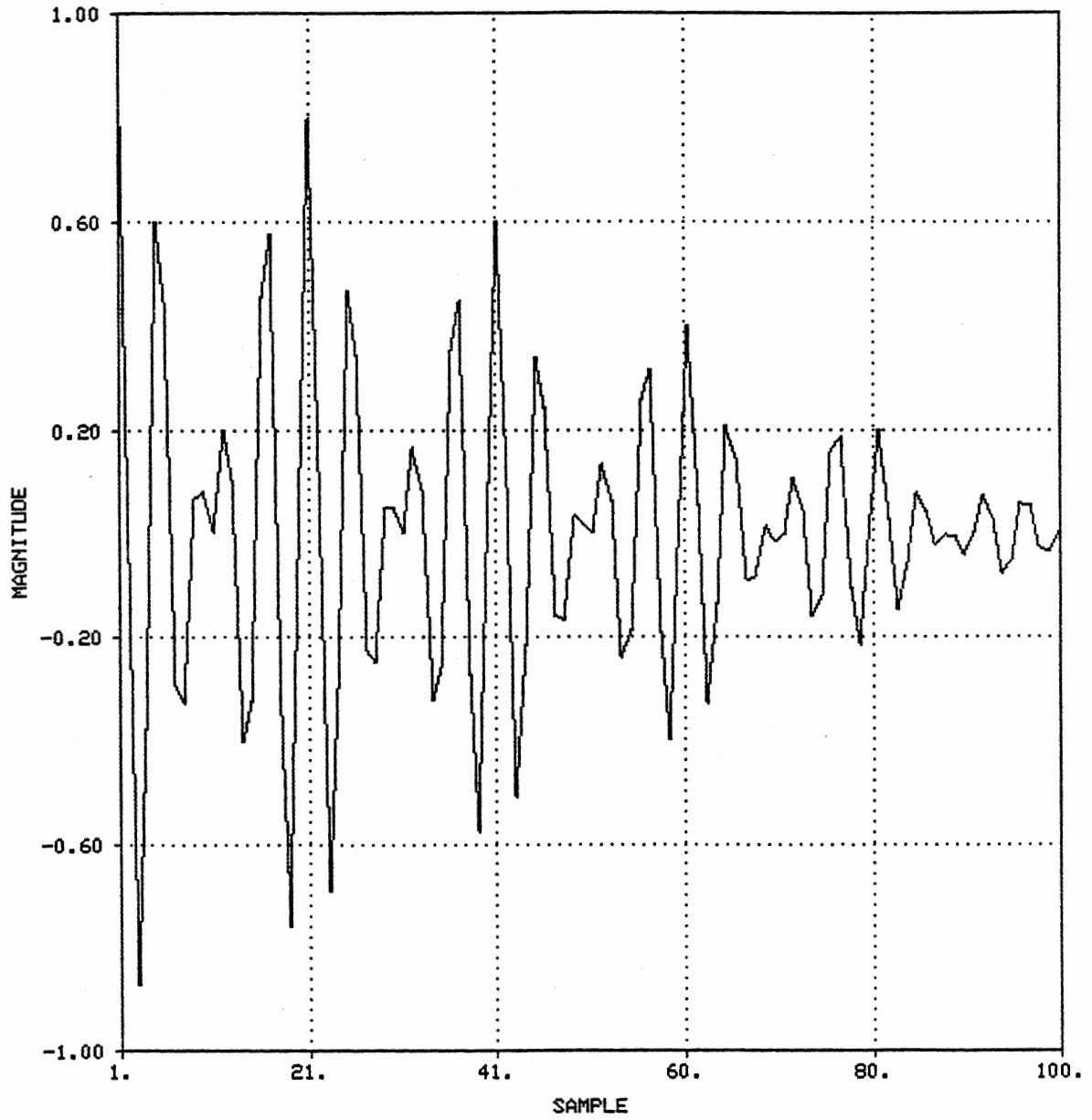


Figure 52. Regular Autocorrelation of Sum of Two Cosinewaves;  
 $f_1=0.2$ ,  $f_2=0.25$ ,  $N=100$ .

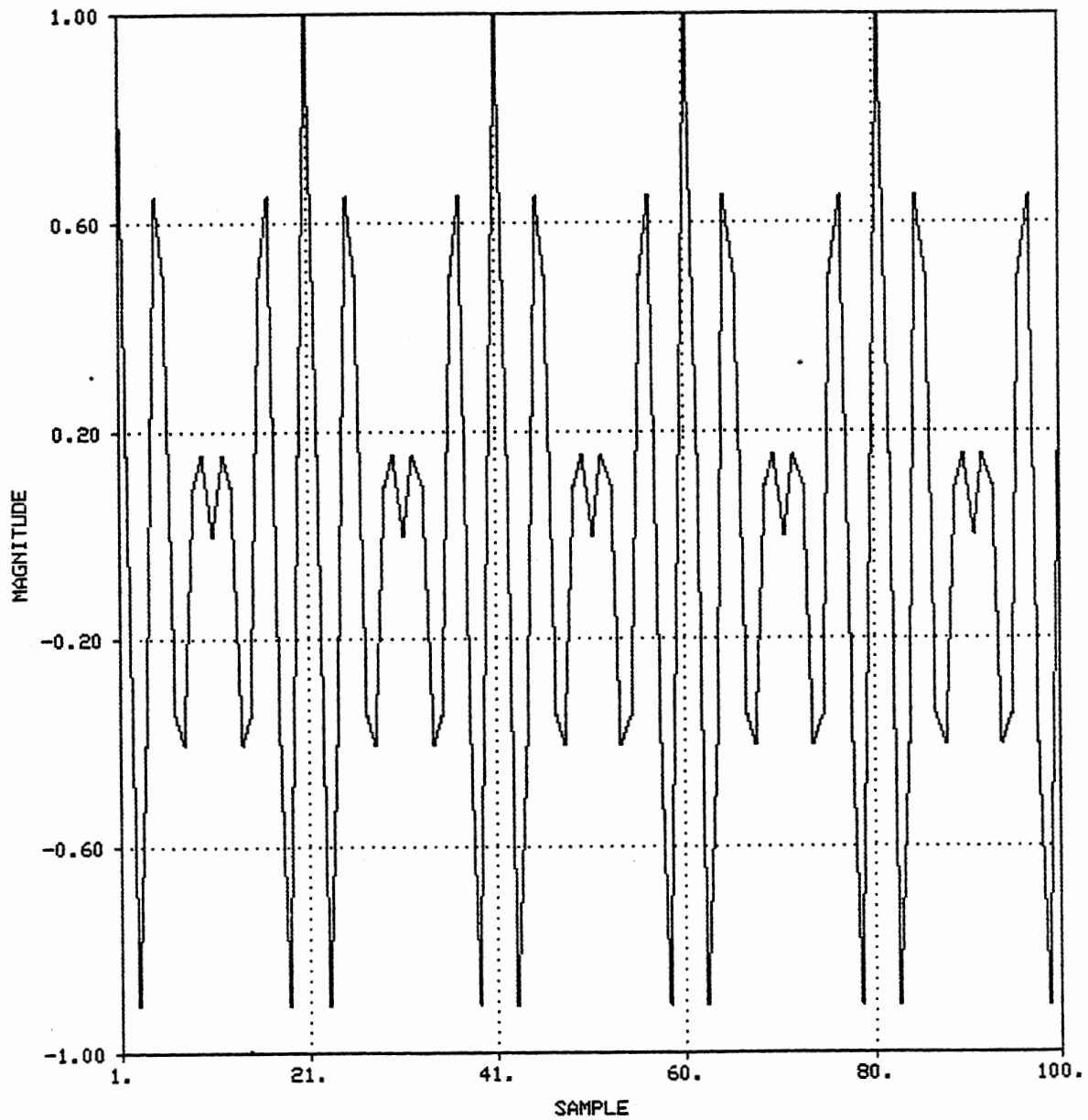


Figure 53. Circular Autocorrelation of Sum of Two Cosinewaves;  
 $f_1=0.2$ ,  $f_2=0.25$ ,  $N=100$ .



VITA

Mohammad Vernosfaderani Malakooti

Candidate for the Degree of

Doctor of Philosophy

Thesis: BIAS REMOVAL APPROACH IN SYSTEM IDENTIFICATION AND ARMA SPECTRAL ESTIMATION

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Isfahan, Iran, November 24, 1953, the son of Haji-Ali V. Malakooti and Hajieh-Gohar V. Kabiri.

Education: Received Diploma of mathematics from Adab High School, Isfahan, Iran, in June, 1972; received the associate's degree in Mechanics from Isfahan Institute of Technology, Isfahan, Iran, in June, 1974; received the Bachelor of Science degree in Electrical Engineering from University of Tulsa, Tulsa, Oklahoma, in December, 1978; received the Master of Science degree in Electrical Engineering from Virginia Polytechnic Institute and State University, Blacksburg, Virginia, in December, 1980; enrolled in doctoral program at Arizona State University, Tempe, Arizona, 1983-1985; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1987.

Professional Experience: Graduate Teaching and Research Assistant, Virginia Polytechnic Institute and State University, 1979-1980; Principal Engineer and Software designer of the Aircraft Simulator, George Moody Incorporated, GMI, Broken Arrow, Oklahoma, May 1981-December 1982; Graduate Teaching and Research Associate, Arizona State University, January 1983-August 1985, Recitation Instructor, Arizona State University, September 1985-December 1985; Graduate Teaching and Research Associate, Oklahoma State University, 1986-1987.

Awards/Affiliations: George Moody Incorporated, GMI, and U.S. Airline for the design of fuel and fire trainer of Boeing 737. Member of Institute of Electrical and Electronic Engineers (IEEE), societies of Automatic Control; Acoustics, Speech, and Signal Processing; Pattern Analysis and Machine

Intelligence; Medical Imaging; Aerospace and Electronic System; Communications; and Geoscience and Remote sensing.