

THE UNIVERSITY OF CALGARY

MULTI-EXPONENTIAL SIGNAL ANALYSIS USING
THE TOTAL LEAST SQUARES METHOD

by

Jasbir Alam

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL ENGINEERING

CALGARY, ALBERTA

June, 1990

© Jasbir Alam, 1990



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-61954-6

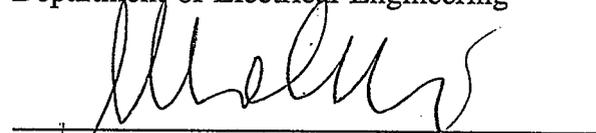
Canada

THE UNIVERSITY OF CALGARY
FACULTY OF GRADUATE STUDIES

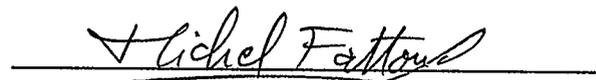
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "Multi-Exponential Signal Analysis Using the Total Least Squares Method", submitted by Jasbir Alam in partial fulfillment of the requirements for the degree of Master of Science.



Dr. S.T. Nichols, Supervisor
Department of Electrical Engineering



Dr. U. Hahn
Department of Radiology



Dr. M. Fattouche
Department of Electrical Engineering



Dr. O.P. Malik
Department of Electrical Engineering

Date 90-09-12

ABSTRACT

The problem of decomposing a multi-exponential signal formed by the linear superposition of exponentials having the same location but different widths and amplitudes is addressed in this dissertation. The problem is of great practical significance and occurs in many fields of applied sciences.

In this thesis several classical iterative search techniques, namely the nonlinear least squares and the maximum likelihood methods, are implemented. A well known approach based on the Gardner transformation has been extended and improved. Most importantly, however, is the introduction of a novel, efficient and robust approach based on the idea of total least squares.

The nonlinear least squares problem is solved using the modified Powell's method. This numerical method iteratively searches for the minimum of a given function in a N-dimensional space. The nonlinear set of equations encountered in the maximum likelihood method are solved using Newton-Raphson's iterative method. Both these methods require a priori knowledge of the number of signal components and the initial estimate of the signal parameters.

A frequency-domain technique based on the Gardner transformation is also implemented. Improvement in the resolution of time-domain impulse train is achieved by modeling the deconvolved frequency data as an autoregressive moving average (ARMA) process using the transient error technique. The autoregressive (AR)

parameters are derived using the singular value decomposition (SVD) of the data matrix and the moving average (MA) portion corresponds to the transient error terms in the unit-step forward prediction. Furthermore, cubic splines are used to obtain the uniform samples that are required to perform the discrete Fourier transform (DFT) of the transformed data. The results are found to be highly sensitive to the choice of nodes at low signal to noise ratio (SNR).

A novel and efficient method called the Total Least Squares method is introduced. The strength of this technique lies in its ability to enhance the signal to noise ratio of noise corrupted data using the singular value decomposition (SVD) method. The data here is modeled as an AR process and the AR-coefficients are obtained using what has recently been referred in the scientific literature as a total least squares approach. The number of roots on the real axis between 0 and 1 of the polynomial subsequently formed with these coefficients give the number of signal components and their locations are related to the widths of the individual components. Knowing the number of signal components and the widths, the amplitudes are estimated using the least squares fit of the data.

The superior accuracy and the robustness of this method is demonstrated by applying this technique to the decomposition of Wood-NMR data to determine the moisture contents of wood during the drying process.

ACKNOWLEDGEMENTS

The author is deeply indebted to Dr. S.T. Nichols for his guidance, patience and support throughout the course of this work.

The financial support received from the Department of Electrical Engineering is gratefully acknowledged.

Special thanks are due to my friend Dr. Lagu and his wife Pushpa whose love and guidance helped me stay on course.

Finally, I would like to thank my wife, Neelam, and my sons, Navdeep and James, for their patience and understanding, without which I could hardly have completed this work.

Dedicated with love to
my wife, Neelam,
and
my children, Navdeep and James.

TABLE OF CONTENTS

APPROVAL PAGE	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
DEDICATION	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF FIGURES	xi
LIST OF SYMBOLS	xiv
1 INTRODUCTION	
1.1 STATEMENT AND THE SIGNIFICANCE OF THE PROBLEM	1
1.2 PREVIOUS METHODS OF ANALYSIS	2
1.3 THESIS OBJECTIVES	4
1.4 THESIS ORGANIZATION	6
2 TIME DOMAIN ANALYSIS OF "MULTICOMPONENT SIGNALS"	
2.1 INTRODUCTION	7
2.2 GRAPHICAL METHODS	8
2.3 ITERATIVE METHODS	9
2.3.1 The Nonlinear Least Squares Method	10
2.3.2 The Maximum Likelihood Method	13
2.3.2.1 Newton-Raphson's Method	15
2.4 THE TOTAL LEAST SQUARES METHOD	20
2.5 THE WEIGHTED TOTAL LEAST SQUARES METHOD	28
3 FREQUENCY DOMAIN ANALYSIS OF "MULTICOMPONENT SIGNALS"	
3.1 INTRODUCTION	34
3.2 THE TRANSFORM TECHNIQUE	35
3.3 DISCRETE REPRESENTATION OF MULTICOMPONENT SIGNAL	43
3.3.1 Cubic Spline Interpolation	44
3.3.2 Sampling Constraints	45

3.3.3	Selecting Good Portion of the Deconvolved Data	46
3.4	PARAMETRIC METHODS	47
3.4.1	SVD-based Transient Error Method	50
3.4.1.1	Modeling of the Deconvolved Data	52
3.4.1.2	Determination of AR Coefficients Using SVD	54
4	RESULTS	
4.1	INTRODUCTION	59
4.2	SIMULATION RESULTS	60
4.2.1	The Transform Method	61
4.2.1.1	Selection of the Nodes for Cubic-Spline Fit	70
4.2.1.2	Selection of Predictor Order	73
4.2.1.3	Selection of Dominant Singular Values	74
4.2.1.4	Effect of Noise	78
4.2.1.5	Effect of Alpha and Sampling Frequency	80
4.2.2	The Total Least Squares (TLS) Method (An Algebraic Method)	81
4.2.2.1	Effect of Noise	90
4.2.2.2	Comments on the Total Least Squares Method	94
4.2.3	The Weighted Total Least Squares Method (An Algebraic Method)	94
4.2.4	Estimating the Amplitudes Using a Least Squares (LS) Method	97
4.2.5	The Nonlinear Least Squares (NLS) Method	99
4.2.6	The Maximum Likelihood (ML) Method	105
4.3	DECOMPOSITION OF WOOD-NMR DATA	111
4.3.1	NMR Data Acquisition	111
4.3.2	Data Processing and Decimation	112
4.3.3	Parameter Estimation	113
4.3.4	An Alternate Approach to Selecting the Dominant Singular Values	119
4.3.5	Estimating the Amplitudes of the Exponentials	121
5	CONCLUSIONS AND RECOMMENDATIONS	126
	REFERENCES	129

LIST OF TABLES

Table 4.1 Comparison of predictor coefficients derived using the Total Least Squares and the Weighted Total Least Squares ($P = I$) methods.	96
Table 4.2 Least Squares estimates of the amplitudes for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-2}$	97
Table 4.3 Least Squares Estimates of the amplitudes for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	98
Table 4.4 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-4}$	100
Table 4.5 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	101
Table 4.6 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$	102
Table 4.7 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	103
Table 4.8 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-2}$	104
Table 4.9 Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-4}$	106
Table 4.10.a Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	107
Table 4.10.b Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	108

Table 4.11 Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$	109
Table 4.12 Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	110
Table 4.13 Effect of zero mean Gaussian noise on the singular values.	121
Table 4.14 Amplitude estimates obtained using LS fit (dope1lms).	122
Table 4.15 Amplitude estimates obtained using LS fit (cpmg10).	122
Table 4.16 Parameters of synthetic data fit.	122

LIST OF FIGURES

Figure 4.1 A sample multi-exponential signal with $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 0.005$	61
Figure 4.2 Cubic spline fit of the transformed data with 30 splines.	63
Figure 4.3 Estimate of the parameter distribution function for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.1, 0.2]^T$ and $\sigma_n = 10^{-6}$	66
Figure 4.4 Estimate of the parameter distribution function for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [1.0, 2.0]^T$ and $\sigma_n = 10^{-6}$	67
Figure 4.5 Estimate of the parameter distribution function for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-5}$	68
Figure 4.6 Estimate of the parameter distribution function for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$	69
Figure 4.7 Spline fit of the transformed data using 20, 30 and 40 splines. $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$	71
Figure 4.8 Distribution function estimates obtained using 20, 30 and 40 splines. $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$	72
Figure 4.9 Distribution function estimates obtained for different predictor orders. $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$	74
Figure 4.10 Effect of noise on the spread of singular values.	75
Figure 4.11 Effect of the effective rank of the data matrix on the distribution for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$	77

Figure 4.12 Estimate of the distribution function for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$	78
Figure 4.13 Estimate of the distribution function for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	79
Figure 4.14 Effect of alpha and sampling on the distribution function.	81
Figure 4.15 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [0.1, 0.2]^T$ and $\sigma_n = 10^{-4}$	84
Figure 4.16 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [1.0, 2.0]^T$ and $\sigma_n = 10^{-4}$	85
Figure 4.17 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-4}$	87
Figure 4.18 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$	89
Figure 4.19 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$	91
Figure 4.20 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-2}$	92
Figure 4.21 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-2}$	93
Figure 4.22 Deconvolved data obtained using file 'dope1lms'.	114
Figure 4.23 Estimate of the parameter distribution function for 'dope1lms'.	115
Figure 4.24 Distribution of the zeroes of the AR polynomial for 'dope1lms'.	117
Figure 4.25 Estimate of the parameter distribution function for 'cpmg10'.	118

Figure 4.26 Distribution of the zeroes of the AR polynomial for 'cpmg10'.	119
Figure 4.27 Synthetic data fit of 'dope1lms'.	123
Figure 4.28 Synthetic data fit of 'cpmg10'.	124

LIST OF SYMBOLS

\underline{A}	amplitude vector
$\underline{\lambda}$	exponent vector
σ_n^2	noise variance
*	convolution operation
AR	autoregressive
ASSP	acoustic, speech and signal processing
DFT	discrete Fourier transform
FFT	fast Fourier transform
LS	least squares
MA	moving average
ML	maximum likelihood
NLS	nonlinear least squares
SNR	signal-to-noise ratio
SVD	singular value decomposition
TLS	total least squares

CHAPTER 1

INTRODUCTION

1.1 STATEMENT AND THE SIGNIFICANCE OF THE PROBLEM

The nontrivial problem of decomposing a multicomponent signal formed by the linear superposition of the basic functions having the same shape and location but different widths and amplitudes arises in many fields of applied sciences. Mathematically, a multicomponent signal is expressed as

$$x(\tau) = \sum_{i=1}^M A_i p(\lambda_i, \tau) + n(\tau); \quad 0 \leq \tau \leq \infty. \quad (1.1)$$

The pulse shape $p(\tau)$ is known while the estimates for the unknown parameters M , A_i and λ_i for $i = 1, 2, \dots, M$ are desired from the knowledge of $x(\tau)$. The additive random noise, $n(\tau)$, is assumed to be Gaussian. In general, the basic function, $p(\tau)$, may be Lorentzian, sinc, Gaussian, exponential or any other complex signal. However, because of practical significance only those signals formed by the linear superposition of real decaying exponentials are studied in this thesis.

The practical significance of this problem finds its roots in such diverse applications as overlapping excitation in fluorescent decay [1], compartmental analysis in physiology [2], pharmacokinetics [3], sedimentation equilibrium [4], speech processing and Electromagnetic wave problems [5,6], imaging systems [7,8], radio

astronomy [9], electronic components reliability study [10], nuclear magnetic resonance (NMR) [11,12] and determination of important material properties in Deep-Level Transient Spectroscopy [13,14,15], to mention a few. Depending upon the area of application the number of components, M , amplitudes, A_i , and widths, λ_i , have different biological or physiological meanings.

1.2 PREVIOUS METHODS OF ANALYSIS

Peeling or the graphical method [16,17,18] is one of the oldest and crudest time-domain analysis techniques. More recently, other time-domain analysis techniques such as nonlinear least squares methods [19], linear least squares or the method of moments [1] and the maximum likelihood methods [11] have been applied to the problem of multicomponent signal analysis. The graphical method is applicable only if the number of signal components is small, i.e. $M \leq 2$, and the additive noise level is low. Furthermore, this method of analysis suffers from a cumulative error build-up. Other time-domain analysis methods, on the other hand, require a priori knowledge of the number of signal components, M , as well as the initial estimates of amplitudes and widths.

Gardner et al [20], in 1959, introduced a frequency domain or spectral analysis method that neither requires a priori knowledge of the number of signal components nor the initial estimates of the amplitudes and widths. In this method, equation (1:1) is transformed into a convolution integral by the use of nonlinear change of variables, $\tau = e^t$ and $\lambda_i = e^{-\gamma_i}$ so that the data is modeled as a

convolution of the basic signal with an impulse train in t-space. The impulse train is recovered by the process of deconvolution or inverse filtering yielding the number of components, amplitudes and widths simultaneously. Gardner et al used this technique in the analysis of multicomponent exponential decays in tracer data. However, due to the difficulties involved in performing the numerical Fourier transform at that time, they met with limited success only.

Schlesinger [21], in 1973, solved the problem of performing numerical Fourier transform by using the discrete Fourier transform (DFT) and its associated fast Fourier transform (FFT) algorithm. However, the problem of high error ripples associated with inverse filtering remained unsolved. Lin and Dutt [3] used numerical Fourier transform and Bessel functions to reduce the magnitude of the error ripples. However, this approach is difficult to implement. Cohn-Sfetcu et al [22], in 1975, used the FFT and the Gaussian filtering to reduce the high frequency noise so that the signal to noise ratio (SNR) of the deconvolved data could be improved. The frequency-domain filtering, however, resulted in a poor resolution in time-domain.

Arunachalam [23], in 1979, devised a data extension technique so as to reduce the error ripple and to improve the resolution of the deconvolved impulse train. Here the "Good portion" of the deconvolved data is extended beyond the optimal passband by the recursive application of a unit step forward predictor determined by Burg's maximum-entropy method. Though this procedure works well for noise-free exponential data, it gives a poor estimate of the number of signal components.

In a further development, Salami [24] and Nichols et al [25] used the autoregressive moving average (ARMA) model of the deconvolved data to enhance the resolution of the deconvolved impulse train and the accuracy of the number of signal components. Their motivation for using the ARMA model was based on the fact that it often yields a model with the fewest number of coefficients and Pagana's work [26] who showed that an AR(p) process in white additive noise can be best modeled as an ARMA(p,p) process. Though their results are quite encouraging, they obtained the uniform samples of the nonlinearly transformed data, required to perform the DFT, from the direct knowledge of the signal parameters themselves, thus avoiding the complex problem of interpolation.

In view of the practical significance of this problem coupled with the severe limitations of the aforementioned methods, new simpler, more efficient and practical means of analysis are investigated.

1.3 THESIS OBJECTIVES

- i. Initially the objective of this thesis was to develop and implement a singular value decomposition (SVD)-based transient error method based on Nichols et al [27] with the added complexity of obtaining the uniform samples of the nonlinearly transformed data using the least squares cubic spline method.

This technique and the results obtained using this technique are presented in details in Chapter 2 and Chapter 4 respectively. It was found that the spline fitting and noise severely reduced the robustness of this technique. Therefore,

it was decided to investigate other techniques including the well established methods of nonlinear least squares and maximum likelihood. The consequence of this investigation was the development of a novel method which was later found to be based upon the total least squares concept. A survey of the scientific literature from the last few years has revealed a growing interest among many researchers in the total least squares methods. However, the results presented in section 2.4 were arrived at independently and those in section 2.5, to the best of our knowledge, constitute an original contribution.

This technique extends the signal to noise ratio (SNR) enhancement concept of noisy data first purposed by Tufts and Kumaresan [40] using the SVD method. A still better improvement in the SNR is achieved using the new Total Least Squares method by removing noise both from the data matrix and the observation vector as opposed to removing it only from the data matrix in the case of Tufts and Kumaresan [40] method.

This Total Least Squares method does not involve nonlinear change of variable, hence alleviating the need for interpolation. The method is based upon modeling the multi-exponential data as an autoregressive (AR) process. The AR coefficients are derived by solving a total least squares problem. The number of roots on the real axis between 0 and 1 of the polynomial subsequently formed with these coefficients give the number of signal components and their locations are related to the widths of the components. Once the number of signal components and the widths are known, the amplitudes are

estimated using a least squares fit of the original data.

The Total Least Squares method is applied to the Wood-NMR data in Chapter 4. The object here is to estimate the moisture content during Kiln-drying process of wood, defect identification and species determination.

1.4 THESIS ORGANIZATION

This dissertation is organized into five Chapters. Chapter 2 entitled "TIME-DOMAIN ANALYSIS OF MULTICOMPONENT SIGNALS" lays the mathematical ground work for the Total Least Squares and its variation the Weighted Total Least Squares method, the nonlinear least squares method and the maximum likelihood method. The closed form solutions in the final implementation states are presented.

In Chapter 3 the frequency-domain method called the SVD-based transient error method is presented. This technique of analysis is based upon the Gardner's transformation and models the deconvolved data as an ARMA process. Cubic splines are used to obtain the uniform samples of the transformed data.

In Chapter 4 the results are presented in two separate sections. In the first section simulation results compare the performance of all the four methods. In the second section, the new Total Least Squares method and the Gardner's transform based frequency-domain method are applied to the Wood-NMR data to estimate the moisture contents in the different anatomical regions of the wood. Finally, the main conclusions are drawn in Chapter 5.

CHAPTER 2

TIME DOMAIN ANALYSIS OF " MULTICOMPONENT SIGNALS "

2.1 INTRODUCTION

The nonparametric time-domain analysis techniques for decomposing multicomponent signals are the oldest and generally more common outside the field of engineering.

Their popularity lends itself to the readily available algorithms that perform reasonably well provided the number of signal components is known and some initial estimates of the amplitudes and widths are available. The nonlinear least squares and the maximum likelihood are two such well known techniques.

However, in the last decade or so the parametric or modeling techniques have been successfully employed in decomposing multi-sinusoidal [28], exponentially damped multisinusoidal [29] and most recently real, decaying multi-exponential [30] signals.

This Chapter briefly discusses peeling or graphical methods and lays the mathematical ground work for the nonlinear least squares and the maximum likelihood methods. Finally, a novel and efficient parametric technique based on the

total least squares approach is introduced.

2.2 GRAPHICAL METHOD

This method belongs to the class of nonparametric techniques and is the oldest and crudest technique for decomposing a multi-exponential signal. The mathematical justification for this technique is as follows.

Consider the multicomponent signal

$$x(\tau) = \sum_{i=1}^M A_i e^{-\lambda_i \tau} \quad (2.1)$$

where, it is assumed, for convenience, that

$$\lambda_1 < \lambda_2 < \dots < \lambda_M.$$

Taking the natural logarithm of both sides of equation (2.1), yields

$$\ln[x(\tau)] = \ln \left[\sum_{i=1}^M A_i e^{-\lambda_i \tau} \right].$$

For

$$\tau \gg \frac{1}{\lambda_2}$$

$$\ln[x(\tau)] \approx \ln[A_1 e^{-\lambda_1 \tau}] = \ln A_1 - \lambda_1 \tau. \quad (2.2)$$

Equation (2.2) is an equation of a straight line with y-intercept as $\ln A_1$ and slope $-\lambda_1$. Thus fitting the tail-end of the natural logarithm of the data with a straight line gives A_1 and λ_1 . Next $A_1 e^{-\lambda_1 \tau}$ is subtracted from (2.1) and the procedure is repeated to determine A_2 and λ_2 . This sequence of steps is repeated until the remaining data is in the range of some error. The major drawbacks of this method are that:

1. The decay rates, λ_i , must be widely separated, which may not always be the case. Consequently, this method has poor resolving capability.
2. There is an accumulative error buildup.
3. For large ($M > 2$) number of components it is not a very practical way of parameter estimation.

Because of these shortcomings, this method is only useful in obtaining a first order estimate of the signal parameters. Parameters obtained using this method may subsequently be used as seed values for the more sophisticated techniques such as the nonlinear least squares and the maximum likelihood methods.

2.3 ITERATIVE METHODS

Given the rough estimates of A_i and λ_i , for $1 \leq i \leq M$, these methods iteratively update the parameter estimates. Convergence of these methods is guaranteed only by the accuracy of the initial estimates and the numerical stability of the algorithms. Thomasson and Clark [19], in 1974, successfully applied the Marquardt's

nonlinear least squares algorithm to the analysis of exponential decay curves.

The maximum likelihood method is another extremely powerful iterative technique. Sandor et. al. [11] have demonstrated the use of this method in estimating the parameters of exponential data with Poisson distribution. The maximum likelihood method, like the nonlinear least square method, also requires a priori knowledge of the number of signal components, M , and the initial estimates of A_i and λ_i and can converge to different solutions depending on the initial estimates.

2.3.1 The Nonlinear Least Square Method

The nonlinear least squares method converts the parameter estimation problem to a multidimensional minimization of an objective function which is nonlinear in some or all of the parameters. For the multi-exponential problem with the time series $x(n)$; $1 \leq n \leq N$ the objective function is formed as

$$f(A_1, \dots, A_M, \lambda_1, \dots, \lambda_M) = \sum_{j=1}^N [x_j - \sum_{i=1}^M A_i e^{-\lambda_i t_j}]^2 \quad (2.3)$$

Starting from the initial estimates of A_i and λ_i the minimum of the objective function (2.3) is searched in $2M$ dimensions. If the initial estimates are not accurate enough the algorithm may not converge or converge to a local minimum.

In this thesis modified Powell's method [31] is used to search for the minimum of the objective function in N dimensions. The original Powell's method generates N mutually conjugate directions and proceeds in a following

manner:

Initialize the set of directions \underline{u}_i along which the minimum of the given function is to be searched, to the basis vectors, i.e.

$$\underline{u}_i = \underline{e}_i, \quad i = 1, 2, \dots, N$$

where the vector \underline{e}_i has all its elements zero except the i^{th} element, which is unity.

Then the following basic procedure is repeated until the function stops decreasing:

1. Save starting position as \underline{p}_0 .
2. For $i = 1, 2, \dots, N$, move \underline{p}_{i-1} to a minimum along \underline{u}_i and call this point \underline{p}_i .
3. For $i = 1, 2, \dots, N-1$, set $\underline{u}_i \leftarrow \underline{u}_{i+1}$.
4. Set $\underline{u}_N \leftarrow \underline{p}_N - \underline{p}_0$.
5. Move \underline{p}_N to minimum along \underline{u}_N and call this point \underline{p}_0 .

Powell, in 1964, showed that, for a function with quadratic form, k iterations of the above basic procedure produce a set of directions \underline{u}_i whose last k members are mutually conjugate. Therefore, N iterations of the basic procedure, amounting to $N(N+1)$ line minimizations in all, will exactly minimize a quadratic form. But the practice of discarding \underline{u}_1 in favor of $\underline{p}_N - \underline{p}_0$ tends to produce directions set

that are linearly dependent. Once, this happens, the Powell's method tends to minimize the given function over a subspace of the full N -dimensional space.

In the modified Powell's method $\mathbf{p}_N - \mathbf{p}_0$ is still taken as a new direction. However, the old direction along which the function at hand made its largest decrease is discarded. Since, the old direction is still the major component of the new direction we are adding, so dropping the old direction gives us the best chance of avoiding a buildup of linear dependence. However, under the following two conditions it is sometimes even better to retain the direction along which maximum decrease occurred. Define:

$$f_0 \equiv f(\mathbf{p}_0), \quad f_N \equiv f(\mathbf{p}_N), \quad f_E \equiv f(2\mathbf{p}_N - \mathbf{p}_0).$$

Let Δf be the magnitude of the largest decrease along one particular direction. Then:

1. If $f_E \geq f_0$, then keep the old set of directions for the next basic procedure, because the average direction $\mathbf{p}_N - \mathbf{p}_0$ is all played out.
2. If $2(f_0 - 2f_N + f_E)[f_0 - f_N - \Delta f]^2 \geq (f_0 - f_E)^2 \Delta f$, then keep the old set of directions for the next basic procedure, because either (i) the decrease along the average direction was not primarily due to any single direction's decrease or (ii) there is a substantial second derivative along the average direction and we seem to be near the bottom of its minimum.

An algorithm based on this technique is implemented in this thesis and its

performance is compared to the maximum likelihood method.

2.3.2 The Maximum Likelihood Method

The maximum likelihood method is based on a relatively simple idea: different parameter populations generate different data samples and any given data sample is more likely to have come from some population than from others, Kmenta [44]. In this context, the problem of parameter estimation can be restated as follows: Given the time series $x(n)$; $1 \leq n \leq N$ estimate the parameter vector \mathbf{q} by its most plausible values. In other words, maximize the conditional joint probability density function of the random vector \mathbf{x} represented by the observed sample vector \mathbf{u} with respect to the parameter vector \mathbf{q} .

The vectors are given by

$$\mathbf{u} = [x_1, x_2, \dots, x_N]^T$$

and

$$\mathbf{q} = [A_1, \dots, A_M, \lambda_1, \dots, \lambda_M]^T.$$

The conditional joint probability function, in this context, is more commonly known as the likelihood function. It is generally a common practice to maximize the logarithm of the likelihood function, known as the log likelihood function, rather than the likelihood function itself with respect to the parameter vector \mathbf{q} .

For the multi-exponential signal, if the total time interval T_0 of the measurement is subdivided into subintervals, $(\Delta t)_i$, $i = 1, 2, \dots, N$ such that for all i $(\Delta t)_i \ll T_0$, then it may be assumed that the measured signal in each interval follows a Poisson distribution. With this assumption, the likelihood function can be expressed as

$$l(\underline{\mathbf{u}}|\mathbf{q}) = \exp\left[-\sum_{j=1}^N \langle x_j \rangle\right] \prod_{j=1}^N \frac{\langle x_j \rangle^{x_j}}{x_j!} \quad (2.4)$$

where $\langle x_j \rangle$ is the expected value of the signal in the j^{th} interval and $x_j!$ is factorial x_j . Furthermore, the expected value $\langle x_j \rangle$ is given by

$$\sum_{i=1}^M A_i e^{-\lambda_i t_j}.$$

Using (2.4), the natural logarithm of the likelihood function, the log likelihood function, is

$$L(\underline{\mathbf{u}}|\mathbf{q}) = -\sum_{j=1}^N \langle x_j \rangle + \sum_{j=1}^N \left[x_j \ln \langle x_j \rangle - \ln x_j! \right]. \quad (2.5)$$

Differentiating the log likelihood function with respect to A_l and equating it to zero gives

$$-\sum_{j=1}^N e^{-\lambda_l t_j} + \sum_{j=1}^N \frac{x_j}{\langle x_j \rangle} e^{-\lambda_l t_j} = 0$$

$$\sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] e^{-\lambda_l t_j} = 0, \quad l = 1, 2, \dots, M. \quad (2.6)$$

Similarly, differentiating (2.5) with respect to λ_l and equating it to zero gives

$$\sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] A_l t_j e^{-\lambda_l t_j} = 0, \quad l = 1, 2, \dots, M. \quad (2.7)$$

Equations (2.6) and (2.7) can be combined into a single equation as

$$\sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial \langle x_j \rangle}{\partial q_i} = 0, \quad i = 1, 2, \dots, 2M. \quad (2.8)$$

From (2.8), we have a set of $2M$ nonlinear equations that are to be solved for $2M$ independent variables, namely the signal parameters. In this thesis Newton-Raphson's iterative method is used to solve these equations.

2.3.2.1 Newton-Raphson's Method

A typical problem gives N functional relations, in N variables, to be zeroed, i.e.

$$f_i(\underline{x}) = 0, \quad i = 1, 2, \dots, N \quad (2.9)$$

where

$$\underline{\mathbf{x}} = [x_1, x_2, \dots, x_N].$$

Then in the neighborhood of $\underline{\mathbf{x}}$ each of the function, f_i , can be expanded in a Taylor's series as

$$f_i(\underline{\mathbf{x}} + \Delta \underline{\mathbf{x}}) = f_i(\underline{\mathbf{x}}) + \sum_{j=1}^N \frac{\partial f_i}{\partial x_j} \delta x_j + O(\delta x^2) = 0.$$

By neglecting terms of the order δx^2 and higher gives

$$f_i(\underline{\mathbf{x}}) + \sum_{j=1}^N \frac{\partial f_i}{\partial x_j} \delta x_j = 0$$

or

$$\sum_{j=1}^N \frac{\partial f_i}{\partial x_j} \delta x_j = -f_i(\underline{\mathbf{x}}), \quad i = 1, 2, \dots, N.$$

The above equation can be expressed in the matrix form as

$$\mathbf{A} \Delta \underline{\mathbf{x}} = \underline{\mathbf{f}}(\underline{\mathbf{x}}) \quad (2.10)$$

where

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial f_N}{\partial x_1} & \frac{\partial f_N}{\partial x_2} & \cdots & \frac{\partial f_N}{\partial x_N} \end{bmatrix},$$

$$\Delta \underline{\mathbf{x}} = [\delta x_1 \delta x_2 \cdots \delta x_N]^T,$$

and

$$\underline{\mathbf{f}}(\underline{\mathbf{x}}) = [-f_1(\underline{\mathbf{x}}) -f_2(\underline{\mathbf{x}}) \cdots -f_N(\underline{\mathbf{x}})]^T.$$

The error in the estimate of the solution to the homogeneous set of equations (2.9) is denoted by the error vector $\Delta \underline{\mathbf{x}}$. An initial estimate of the solution vector, $\underline{\mathbf{x}}$, is required to start the iterative process. Equation (2.10) is solved for the error vector and the solution vector is updated at the end of each iteration as

$$\underline{\mathbf{x}}^{new} = \underline{\mathbf{x}}^{old} + \Delta \underline{\mathbf{x}}.$$

The iterative process stops if either the sum of the magnitudes of the functions, f_i , is less than some tolerance, or the sum of the absolute values of the corrections, δx_i , is less than some tolerance.

The solution of (2.10) for $\Delta \underline{x}$ proceeds through LU decomposition of the matrix \mathbf{A} using Crout's method which employs partial pivoting, i.e. row interchanges only, to attain stability. The LU decomposition of (2.10) gives

$$(\mathbf{LU}) \Delta \underline{x} = \underline{f}(\underline{x}).$$

Now let

$$\mathbf{U} \Delta \underline{x} = \underline{y}. \quad (2.11)$$

Therefore,

$$\mathbf{L} \underline{y} = \underline{f}(\underline{x}). \quad (2.12)$$

The advantage of breaking up one set of linear equations into two successive ones is that the solution of the triangular set of equations is quite trivial. Equation (2.12) is solved for the elements of the vector \underline{y} using forward substitution and (2.11) is solved for the elements of the vector $\Delta \underline{x}$, the corrections, using backward substitution.

Using the Taylor's expansion of (2.8) in the neighborhood of \underline{q} gives

$$\sum_{h=1}^{2M} \frac{\partial}{\partial q_h} \left[\sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial \langle x_j \rangle}{\partial q_i} \right] \delta q_h = - \sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial \langle x_j \rangle}{\partial q_i}$$

$$\sum_{h=1}^{2M} \left\{ \sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial^2 \langle x_j \rangle}{\partial q_h \partial q_i} - \sum_{j=1}^N \frac{x_j}{\langle x_j \rangle^2} \frac{\partial \langle x_j \rangle}{\partial q_h} \frac{\partial \langle x_j \rangle}{\partial q_i} \right\} \delta q_h$$

$$= - \sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial \langle x_j \rangle}{\partial q_i}, \quad i = 1, 2, \dots, 2M.$$

$$\sum_{j=1}^N \sum_{h=1}^{2M} \left\{ \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial^2 \langle x_j \rangle}{\partial q_h \partial q_i} - \frac{x_j}{\langle x_j \rangle^2} \frac{\partial \langle x_j \rangle}{\partial q_h} \frac{\partial \langle x_j \rangle}{\partial q_i} \right\} \delta q_h$$

$$= - \sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial \langle x_j \rangle}{\partial q_i}, \quad i = 1, 2, \dots, 2M. \quad (2.13)$$

Equation (2.13) has the same form as (2.10), where the elements of the matrix \mathbf{A} , $a_{m,n}$, for $1 \leq m, n \leq 2M$, are given as

$$a_{m,n} = \sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial^2 \langle x_j \rangle}{\partial q_n \partial q_m} - \frac{x_j}{\langle x_j \rangle^2} \frac{\partial \langle x_j \rangle}{\partial q_n} \frac{\partial \langle x_j \rangle}{\partial q_m}$$

and the elements of the vector $\underline{f}(\underline{\mathbf{x}})$, $-f_m(\underline{\mathbf{x}})$, are given as

$$f_m(\underline{\mathbf{x}}) = \sum_{j=1}^N \left[\frac{x_j}{\langle x_j \rangle} - 1 \right] \frac{\partial \langle x_j \rangle}{\partial q_m}.$$

Equation (2.13) is solved for the error vector using Newton-Raphson's method. The estimates of the parameter vectors are updated iteratively until the

desired convergence is achieved.

2.4 THE TOTAL LEAST SQUARES METHOD

The underlying principle of this method is that there is a polynomial

$$A(z) = 1 + \sum_{i=1}^M a_i z^{-i}$$

of degree M whose roots correspond to the M components of the multi-exponential signal. The roots are given by $e^{-\lambda_i \Delta t}$, where Δt is the sampling interval. Thus, for any choice of Δt , $0 \leq \Delta t \leq \infty$, the roots corresponding to the rate constants, λ_i , of the exponentials that constitute the data must be real and lie between the interval zero and one on the real axis.

The polynomial $A(z)$ is obtained from the one step forward predictor and is referred to as the prediction error filter. A prediction error filter of degree $pe > M$ has M roots on the real axis between the interval zero and one and the remaining $pe - M$ roots will be off this interval but inside the unit circle in the complex plane. Furthermore, since all the polynomial coefficients are real for real data, the complex roots occur in conjugate pairs. As all the roots lie inside the unit circle, the prediction error filter is minimum phase and hence the predictor is stable.

The choice of the forward predictor order $pe > M$ ensures lower prediction errors in the presence of noise. However, if the predictor coefficients a_i 's are derived using the classical linear least squares method or any other methods which

do not improve the SNR of the data then the zeros of the polynomial $A(z)$ that do not correspond to the signal parameters become spuriously distributed in the complex plane [36] and may even be located in the interval zero and one on the real axis, thus resulting in false detection. The robustness of this technique lies in its ability to improve the SNR of the data by removing noise both from the data matrix and the observation vector by using the SVD method. It is this SNR enhancement property of the Total Least Squares method that enables the detection of the signal component at high noise levels. Using this method the predictor coefficients are derived as follows.

For a multi-exponential time series $x(n)$, the forward predictor coefficients, in the absence of noise, must satisfy the following equation:

$$\mathbf{X}\underline{a}=\underline{0} \quad (2.14)$$

because the real decaying exponentials are perfectly predictable. The data matrix \mathbf{X} given as

$$\mathbf{X} = \begin{bmatrix} x_{pe} & x_{pe-1} & \cdots & x_0 \\ x_{pe+1} & x_{pe} & \cdots & x_1 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_{N-1} & x_{N-2} & \cdots & x_{N-pe-1} \end{bmatrix}$$

is a $(N-pe) \times (pe+1)$ matrix and the integer variable pe signifies the predictor

order.

The vector

$$\underline{\mathbf{a}} = \left[1 \quad a_1 \quad a_2 \quad \dots \quad a_{pe} \right]^T$$

determines the predictor coefficients.

It can be easily shown that the rank of the data matrix \mathbf{X} is M , the number of exponentials in the signal. Since the data matrix is rank deficient if $pe > M$, which is usually the case, there are infinitely many solution vectors, $\underline{\mathbf{a}}$, satisfying equation (2.14). One method of finding a meaningful solution from all the possible solutions is to select the one that assigns the minimum energy to the polynomial $A(e^{j\omega})$. The energy, I , of the polynomial $A(e^{j\omega})$ is defined as

$$I = \frac{1}{2\pi} \int_{-\pi}^{\pi} |A(e^{j\omega})|^2 d\omega. \quad (2.15)$$

The above condition is equivalent to minimizing the Euclidean norm of the vector $\underline{\mathbf{a}}$. The significance of the minimum norm lies in the fact that it minimizes the variance of prediction errors.

Thus,

$$I = \underline{\mathbf{a}}^T \underline{\mathbf{a}}. \quad (2.16)$$

Hence, the linear prediction problem is transformed here into a norm minimization problem. The problem is redefined as: Minimize the norm of the vector $\underline{\mathbf{a}}$

subject to the following constraints:

1. The vector $\underline{\mathbf{a}}$ lies in the null space of \mathbf{X} .
2. $a_0 = 1$.

The problem is solved by resorting to the singular value decomposition (SVD) of the data matrix \mathbf{X} . The following three properties of the SVD are of fundamental significance to the solution of this problem:

- (i) In the absence of noise, the number of nonzero singular values of \mathbf{X} is equal to its rank which in turn equals the number of signal components, M .
- (ii) The right singular vectors, $\underline{\mathbf{v}}_j$, of \mathbf{X} corresponding to the zero singular values form the orthonormal basis for the null space of \mathbf{X} .
- (iii) A $(pe+1) \times (pe+1)$ matrix \mathbf{V} with right singular vectors of \mathbf{X} as its columns is orthonormal. Therefore,

$$\mathbf{V} \mathbf{V}^H = \mathbf{I}$$

where superscript H is the Hermitian operator indicating transpose conjugation and \mathbf{I} is an identity matrix. Furthermore, from the orthonormality of \mathbf{V} we have

$$\sum_{i=0}^{pe} v_i^2(0) = 1 \quad (2.17)$$

and

$$\sum_{i=0}^{pe} (v_i(0)) \underline{v}_i = \underline{e}_0$$

or

$$\sum_{i=0}^{pe} (\underline{e}_0^T \underline{v}_i) \underline{v}_i = \underline{e}_0 \quad (2.18)$$

where $v_i(0)$ is the first element of the i^{th} . singular vector and

$$\underline{e}_0 = [1 \ 0 \ 0 \ 0 \ \dots]^T.$$

Using properties (i) and (ii), the vector \underline{a} is constructed in the null space of \mathbf{X} as

$$\underline{a} = \sum_{i=M}^{pe} C_i \underline{v}_i \quad (2.19)$$

where C_i 's are the unknown constants and \underline{v}_i 's are the right singular vectors of \mathbf{X} corresponding to the zero singular values. The second constraint can be expressed as

$$\underline{e}_0^T \underline{a} - 1 = 0. \quad (2.20)$$

Forming the Lagrangian, L , gives

$$L = \underline{a}^T \underline{a} - \mu(\underline{e}_0^T \underline{a} - 1) \quad (2.21)$$

where μ is called the Lagrange multiplier.

Minimizing the norm of the vector \underline{a} subject to the constraint that $a_0 = 1$ is then equivalent to minimizing the Lagrangian, L , without any constraint. The constraint requiring \underline{a} to be in the null space of \mathbf{X} is implicitly taken care of in (2.19).

Substituting (2.19) in (2.21), gives

$$L = \sum_{i=M}^{pe} \sum_{j=M}^{pe} C_i \underline{v}_i^T C_j \underline{v}_j - \mu \left(\underline{e}_0^T \sum_{i=M}^{pe} C_i \underline{v}_i - 1 \right). \quad (2.22)$$

Noting that \underline{v}_i are orthonormal vectors, (2.22) can be simplified to

$$L = \sum_{i=M}^{pe} C_i^2 - \mu \left(\underline{e}_0^T \sum_{i=M}^{pe} C_i \underline{v}_i - 1 \right). \quad (2.23)$$

It is now desired to minimize L with respect to C_i 's. For real data C_i 's are also real and differentiating (2.23) with respect to C_k , $M \leq k \leq pe$, and equating it to zero yields expression (2.24).

$$\frac{\partial L}{\partial C_k} = 2C_k - \mu \left(\underline{e}_0^T \underline{v}_k \right) = 0 \quad (2.24)$$

Solving for C_k gives

$$C_k = \frac{\mu}{2} \left(\underline{e}_0^T \underline{v}_k \right). \quad (2.25)$$

Substituting (2.25) into (2.19), yields

$$\underline{\mathbf{a}} = \frac{\mu}{2} \sum_{i=M}^{pe} (\underline{\mathbf{e}}_0^T \underline{\mathbf{v}}_i) \underline{\mathbf{v}}_i. \quad (2.26)$$

Equation (2.18) can be expressed as

$$\underline{\mathbf{e}}_0 = \sum_{i=0}^{M-1} (\underline{\mathbf{e}}_0^T \underline{\mathbf{v}}_i) \underline{\mathbf{v}}_i + \sum_{i=M}^{pe} (\underline{\mathbf{e}}_0^T \underline{\mathbf{v}}_i) \underline{\mathbf{v}}_i. \quad (2.27)$$

Substituting (2.26) into (2.27), we obtain

$$\underline{\mathbf{e}}_0 = \sum_{i=0}^{M-1} (\underline{\mathbf{e}}_0^T \underline{\mathbf{v}}_i) \underline{\mathbf{v}}_i + \underline{\mathbf{a}} / \left(\frac{\mu}{2} \right)$$

giving

$$\underline{\mathbf{a}} = \frac{\mu}{2} \left[\underline{\mathbf{e}}_0 - \sum_{i=0}^{M-1} (\underline{\mathbf{e}}_0^T \underline{\mathbf{v}}_i) \underline{\mathbf{v}}_i \right]. \quad (2.28)$$

From (2.28) the constraint $a_0 = 1$ is equivalent to

$$a_0 = \frac{\mu}{2} \left[1 - \sum_{i=0}^{M-1} v_i^2(0) \right] = 1$$

or

$$\frac{\mu}{2} = \frac{1}{1 - \sum_{i=0}^{M-1} v_i^2(0)}. \quad (2.29)$$

Substituting (2.29) back into (2.28), we obtain the desired result

$$\underline{\mathbf{a}} = \frac{\mathbf{e}_0 - \sum_{i=0}^{M-1} (\mathbf{e}_0^T \mathbf{v}_i) \mathbf{v}_i}{1 - \sum_{i=0}^{M-1} v_i^2(0)}. \quad (2.30)$$

The resulting polynomial $A(z) = \sum_{i=0}^{pe} a_i z^{-i}$, where a_i are given by (2.30),

will have M real roots, $e^{-\lambda_i \Delta t}$, in the interval 0 to 1 on the real axis and $(pe-M)$ roots off this interval and inside the unit circle in the complex plane. From the roots the parameters M and λ_i can be determined, using this method. Once the parameters M and λ_i are known, the amplitudes, A_i , can easily be determined using linear least squares methods, such as Prony's method.

Using equations (2.17) and (2.27), the results given by (2.30) can alternately be expressed as

$$\underline{\mathbf{a}} = \frac{\sum_{i=M}^{pe} (\mathbf{e}_0^T) \mathbf{v}_i}{\sum_{i=M}^{pe} v_i^2(0)}. \quad (2.31)$$

The two solutions obtained in (2.30) and (2.31) to the linear prediction problem posed by (2.14) are identical. However, the solution presented by (2.30) is computationally more efficient if $M \ll pe$, which generally is the case. For this reason (2.30) is used in this thesis to obtain the solution vector $\underline{\mathbf{a}}$.

The solution to the linear prediction problem in the form of (2.30) and (2.31) is referred in the scientific literature as a Total Least Squares solution. In a one-dimensional case this method amounts to fitting the best line through the noisy data points that minimizes the sum of the squares of the normal distances as opposed to the vertical distances in the case of linear least squares.

The concept of total least squares or the constrained norm minimization is further extended, in this thesis, to weighted total least squares or the constrained weighted norm minimization. The objective here is to map the roots of the polynomial $A(z)$ in such a fashion so as to achieve greater resolvability and accuracy of the parameters. The solution for the coefficient vector, $\underline{\mathbf{a}}$, in this case, is derived in the next section.

2.5 THE WEIGHTED TOTAL LEAST SQUARES METHOD

It is sometimes desirable to map the roots of the polynomial $A(z)$ in some region to another region in the complex plane in some optimum fashion. For example, if, in the present case, the roots in the interval zero to one can be mapped on to the unit circle, then the complex task of polynomial root finding can be reduced to discrete Fourier transform to estimate the signal parameters. Under these circumstances the problem transforms to a weighted norm minimization. The weighted norm of a vector $\underline{\mathbf{a}}$ is defined as

$$I_w = \underline{\mathbf{a}}^T \mathbf{P} \underline{\mathbf{a}}, \quad (2.32)$$

where \mathbf{P} is a $(pe+1) \times (pe+1)$ weighting matrix. It is this matrix that determines the mapping to be achieved. In the case of a conventional norm, this matrix is simply an identity matrix. Determination of \mathbf{P} is quite complicated and is presently being investigated by researchers in this field. No attempt is made in this thesis to define any optimal mapping or to determine the \mathbf{P} matrix if one is defined. However, a solution for vector $\underline{\mathbf{a}}$ is presented if \mathbf{P} is known and is a symmetric matrix.

The problem at hand then is to minimize the weighted norm, given in (2.32), subject to the same two constraints we had for conventional norm minimization.

As before, $\underline{\mathbf{a}}$ is constructed in the null space of \mathbf{X} as

$$\underline{\mathbf{a}} = \sum_{i=M}^{pe} C_i \underline{\mathbf{v}}_i \quad (2.33)$$

where $\underline{\mathbf{v}}_i$'s are the right singular vectors of \mathbf{X} corresponding to the zero singular values and C_i 's are the arbitrary constants with respect to which I_w in (2.32) is to be minimized subject to the aforementioned constraints.

With (2.33), the Lagrangian L to be minimized is

$$\begin{aligned} L &= \underline{\mathbf{a}}^T \mathbf{P} \underline{\mathbf{a}} - \mu (\underline{\mathbf{e}}_0^T \underline{\mathbf{a}} - 1) \\ &= \sum_{i=M}^{pe} \sum_{j=M}^{pe} C_i C_j \underline{\mathbf{v}}_i^T \mathbf{P} \underline{\mathbf{v}}_j - \mu \left[\underline{\mathbf{e}}_0^T \sum_{i=M}^{pe} C_i \underline{\mathbf{v}}_i - 1 \right]. \end{aligned} \quad (2.34)$$

Minimization of (2.34) with respect to C_k requires

$$\begin{aligned} \frac{\partial L}{\partial C_k} &= \sum_{\substack{j=M \\ j \neq k}}^{pe} C_j \underline{v}_k^T \mathbf{P} \underline{v}_j + 2C_k \underline{v}_k^T \mathbf{P} \underline{v}_k + \sum_{\substack{j=M \\ j \neq k}}^{pe} C_j \underline{v}_j^T \mathbf{P} \underline{v}_k - \mu \underline{e}_0^T \underline{v}_k \\ &= 0. \end{aligned} \quad (2.35)$$

In equation (2.35), the range of k is from M to pe . Since, \mathbf{P} is a symmetric matrix, we have

$$\underline{v}_j^T \mathbf{P} \underline{v}_k = \underline{v}_k^T \mathbf{P} \underline{v}_j.$$

Therefore, (2.35) simplifies to

$$2 \sum_{\substack{j=M \\ j \neq k}}^{pe} C_j \underline{v}_k^T \mathbf{P} \underline{v}_j + 2C_k \underline{v}_k^T \mathbf{P} \underline{v}_k = \mu \underline{e}_0^T \underline{v}_k$$

$$2 \sum_{j=M}^{pe} C_j \underline{v}_k^T \mathbf{P} \underline{v}_j = \mu \underline{e}_0^T \underline{v}_k$$

$$2\mathbf{A}\underline{c} = \mu\underline{b}. \quad (2.36)$$

In equation (2.36) the symmetric matrix \mathbf{A} is given by

$$\mathbf{A} = [\underline{v}_k^T \mathbf{P} \underline{v}_j]_{k,j=M}^{pe}$$

the vector

$$\underline{\mathbf{c}} = [c_M \ c_{M+1} \ \dots \ c_{pe}]^T$$

and

$$\begin{aligned} \underline{\mathbf{b}} &= \begin{bmatrix} \mathbf{e}_0^T \mathbf{v}_M & \mathbf{e}_0^T \mathbf{v}_{M+1} & \dots & \mathbf{e}_0^T \mathbf{v}_{pe} \end{bmatrix}^T \\ &= [v_M^{(0)} \ v_{M+1}^{(0)} \ \dots \ v_{pe}^{(0)}]^T. \end{aligned}$$

From (2.36), $\underline{\mathbf{c}}$ can be written as

$$\underline{\mathbf{c}} = \frac{\mu}{2} \mathbf{A}^{-1} \underline{\mathbf{b}}. \quad (2.37)$$

Expression (2.33) can alternately be expressed as

$$\underline{\mathbf{a}} = \mathbf{V} \underline{\mathbf{c}} \quad (2.38)$$

where the matrix \mathbf{V} has vectors \mathbf{v}_i , for $M \leq i \leq pe$, as its columns. Substituting (2.37) in (2.38) gives

$$\underline{\mathbf{a}} = \frac{\mu}{2} \mathbf{V} \mathbf{A}^{-1} \underline{\mathbf{b}}.$$

Therefore, noting that \mathbf{V} is a unitary matrix, we get

$$\underline{\mathbf{b}} = \mathbf{A} \mathbf{V}^H \underline{\mathbf{a}} / \left(\frac{\mu}{2} \right). \quad (2.39)$$

From (2.27), we have

$$\begin{aligned}
\underline{\mathbf{e}}_0 &= \sum_{i=0}^{M-1} (\underline{\mathbf{e}}_0^T \underline{\mathbf{v}}_i) \underline{\mathbf{v}}_i + \sum_{i=M}^{pe} (\underline{\mathbf{e}}_0^T \underline{\mathbf{v}}_i) \underline{\mathbf{v}}_i \\
&= \sum_{i=0}^{M-1} v_i(0) \underline{\mathbf{v}}_i + \sum_{i=M}^{pe} v_i(0) \underline{\mathbf{v}}_i \\
&= \sum_{i=0}^{M-1} v_i(0) \underline{\mathbf{v}}_i + \mathbf{V} \underline{\mathbf{b}}.
\end{aligned} \tag{2.40}$$

Substituting for $\underline{\mathbf{b}}$, from (2.39) in (2.40)

$$\underline{\mathbf{e}}_0 = \sum_{i=0}^{M-1} v_i(0) \underline{\mathbf{v}}_i + \mathbf{V} \mathbf{A} \mathbf{V}^H \underline{\mathbf{a}} / \left(\frac{\mu}{2} \right).$$

Therefore,

$$\mathbf{V} \mathbf{A} \mathbf{V}^H \underline{\mathbf{a}} = \frac{\mu}{2} \left[\underline{\mathbf{e}}_0 - \sum_{i=0}^{M-1} v_i(0) \underline{\mathbf{v}}_i \right]$$

gives

$$\underline{\mathbf{a}} = \frac{\mu}{2} \mathbf{V} \mathbf{A}^{-1} \mathbf{V}^H \left[\underline{\mathbf{e}}_0 - \sum_{i=0}^{M-1} v_i(0) \underline{\mathbf{v}}_i \right]. \tag{2.41}$$

Now define a $(pe+1) \times (pe+1)$ matrix \mathbf{D} and a vector $\underline{\mathbf{w}}$ as follows:

$$\mathbf{D} = \mathbf{V} \mathbf{A}^{-1} \mathbf{V}^H,$$

$$\underline{\mathbf{w}} = \underline{\mathbf{e}}_0 - \sum_{i=0}^{M-1} v_i^{(0)} \underline{\mathbf{v}}_i.$$

Using this, equation (2.41) can be expressed as

$$\underline{\mathbf{a}} = \frac{\mu}{2} \mathbf{D} \underline{\mathbf{w}} \quad (2.42)$$

and constraint $a_0 = 1$ reduces to

$$a_0 = \frac{\mu}{2} \sum_{i=1}^{pe+1} \mathbf{D}_{1,i} \underline{\mathbf{w}} = 1.$$

Therefore,

$$\frac{\mu}{2} = \frac{1}{\sum_{i=1}^{pe+1} \mathbf{D}_{1,i} \underline{\mathbf{w}}}. \quad (2.43)$$

Substituting (2.43) in (2.42), we obtain the coefficient vector

$$\underline{\mathbf{a}} = \frac{\mathbf{D} \underline{\mathbf{w}}}{\sum_{i=1}^{pe+1} \mathbf{D}_{1,i} \underline{\mathbf{w}}}. \quad (2.44)$$

In calculating the matrix \mathbf{D} , the inverse of the matrix \mathbf{A} is calculated using the singular value decomposition.

CHAPTER 3

FREQUENCY DOMAIN ANALYSIS OF " MULTICOMPONENT SIGNALS "

3.1 INTRODUCTION

Gardner et al [20], in 1959, proposed a transform method which unlike non-parametric time-domain techniques neither requires prior knowledge of the number of components nor of the initial estimates of the signal parameters. This technique was first used by the authors in the analysis of multicomponent exponential decays in tracer data but met with limited success due to the difficulties of performing numerical Fourier transform at that time. Schlesinger [21] solved this problem using the discrete Fourier transform (DFT) and its associated fast Fourier transform (FFT) algorithm. Since then, this technique has been used by many researchers, Lin and Dutt [3] in 1974, Cohn-Sfetcu et al [22] in 1975, Nichols S.T. et al [25] in 1985, etc., with improvements, in the analysis of multicomponent signals.

The technique introduced in this Chapter is based upon the work by Salami et al [27] with the added complexity of cubic spline fitting the nonlinearly transformed data to obtain uniformly spaced samples. The performance of this method is directly compared to that of the Total Least Squares method introduced

in the previous Chapter.

3.2 THE TRANSFORM TECHNIQUE

A multicomponent signal has the general form

$$x(\tau) = \sum_{i=1}^M A_i p(\lambda_i \tau) + n(\tau) \quad (3.1)$$

where $\tau \in R^+$ and R^+ is the positive real line. The pulse shape $p(\tau)$ is known while the parameters M and A_i, λ_i , for $i = 1, 2, \dots, M$, are to be estimated from the observed signal $x(\tau)$. The additive noise is denoted by $n(\tau)$.

Equation (3.1) can be equivalently rewritten as

$$x(\tau) = \sum_{i=1}^M \int_0^{\infty} A_i \delta(\lambda - \lambda_i) p(\lambda \tau) d\lambda + n(\tau) \quad (3.2)$$

where $\delta(\lambda)$ is a dirac delta function defined by

$$\int_{-\infty}^{\infty} \delta(x - x_0) \phi(x) dx = \phi(x_0)$$

if $\phi(x)$ is continuous at x_0 .

Hence, the equation (3.1) can be considered as a particular case of

$$x(\tau) = \int_0^{\infty} q(\lambda) p(\lambda\tau) d\lambda + n(\tau) \quad (3.3)$$

where the unknown distribution function $q(\lambda)$ is given by

$$q(\lambda) = \sum_{i=1}^M A_i \delta(\lambda - \lambda_i).$$

It has been shown by Smith and Cohn-Sfetcu [32] and Smith et al [22] that using Gardner's transform the integral in equation (3.3) can be mapped to a convolution integral. Nichols et al [25] proposed a further refinement to this transformation by multiplying equation (3.3) by a factor τ^α , $0 < \alpha \leq 1$, instead of τ to vary the A_i/λ_i ratio as well as to enhance the signal to noise ratio (SNR) of the deconvolved data.

Multiplying both sides of the equation (3.3) by τ^α gives

$$\tau^\alpha x(\tau) = \int_0^{\infty} \tau^\alpha q(\lambda) p(\lambda\tau) d\lambda + \tau^\alpha n(\tau).$$

Now introducing the nonlinear transformation, $\tau = e^t$ and $\lambda = e^{-\gamma}$ produces

$$e^{\alpha t} x(e^t) = \int_{-\infty}^{\infty} e^{\alpha t} q(e^{-\gamma}) p(e^{t-\gamma}) e^{-\gamma} d\gamma + e^{\alpha t} n(e^t).$$

The above equation can alternately be expressed as

$$e^{\alpha t} x(e^t) = \int_{-\infty}^{\infty} e^{\gamma(\alpha-1)} q(e^{-\gamma}) e^{\alpha(t-\gamma)} p(e^{t-\gamma}) d\gamma + e^{\alpha t} n(e^t). \quad (3.4)$$

Letting

$$y(t) = e^{\alpha t} x(e^t), \quad (3.4a)$$

$$g(\gamma) = e^{\gamma(\alpha-1)} q(e^{-\gamma}), \quad (3.4b)$$

$$k(t) = e^{\alpha t} n(e^t), \quad (3.4c)$$

$$h(t) = e^{\alpha t} p(e^t) \quad (3.4d)$$

in equation (3.4) gives

$$y(t) = \int_{-\infty}^{\infty} g(\gamma) h(t-\gamma) d\gamma + k(t); \quad -\infty \leq t \leq \infty. \quad (3.5)$$

The above equation is a convolution integral and is conventionally written as

$$y(t) = g(t) * h(t) + k(t) \quad (3.6)$$

where the symbol * denotes convolution of $g(t)$ with $h(t)$. Here $y(t)$ and $h(t)$ are known and the parameter distribution function $g(t)$ is to be estimated. Hence the parameters estimation problem is reduced to a deconvolution problem.

The results from equation (3.5) can easily be modified for the case when the basic pulse is an exponential. In this case, the multicomponent signal in equation (3.1) becomes the multi-exponential signal

$$x(\tau) = \sum_{i=1}^M A_i e^{-\lambda_i \tau} + n(\tau). \quad (3.7)$$

Furthermore, the parameters λ_i may be assumed, without the loss of generality, to be ordered so that $\lambda_1 < \lambda_2 < \dots < \lambda_M$.

In this case, using substitutions (3.4a) to (3.4d), we get

$$y(t) = \sum_{i=1}^M \frac{A_i}{\lambda_i^\alpha} e^{\alpha(t-\gamma_i)} e^{-e^{(t-\gamma_i)}} + k(t). \quad (3.8)$$

Expression (3.8) can be written as

$$y(t) = \sum_{i=1}^M B_i h(t-\gamma_i) + k(t)$$

or

$$y(t) = g(t) * h(t) + k(t) \quad (3.9)$$

where

$$g(t) = \sum_{i=1}^M B_i \delta(t-\gamma_i),$$

$$B_i = \frac{A_i}{\lambda_i \alpha}$$

and the pulse $h(t)$ is as previously defined.

Taking the Fourier transform of both sides of (3.9) gives

$$Y(\omega) = G(\omega) H(\omega) + K(\omega) \quad (3.10)$$

where

$$G(\omega) = \sum_{i=1}^M B_i e^{-j\omega\gamma_i}$$

is the transform of the desired distribution and

$$Y(\omega) = \mathbf{F} [y(t)], \quad H(\omega) = \mathbf{F} [h(t)], \quad K(\omega) = \mathbf{F} [k(t)].$$

In the above expressions \mathbf{F} denotes the Fourier transform operator.

Dividing both sides of equation (3.10) by $H(\omega)$ gives

$$\frac{Y(\omega)}{H(\omega)} = G(\omega) + \frac{K(\omega)}{H(\omega)}. \quad (3.11)$$

The inverse Fourier transform of $G(\omega)$ is given by

$$g(t) = \sum_{i=1}^M B_i \delta(t-\gamma_i)$$

$$= \sum_{i=1}^M \frac{A_i}{\lambda_i^\alpha} \delta(t+\ln\lambda_i).$$

The function $g(t)$ is called the parameter distribution function and contains the desired information about the parameters of the multicomponent signal. It is composed of M impulses each having an area A_i/λ_i^α and occupying locations $\ln(1/\lambda_i)$; $i=1,2,\dots,M$.

It is, however, customary to order impulses as $\ln\lambda_1 < \ln\lambda_2 < \dots < \ln\lambda_M$. This can easily be accommodated by taking the direct Fourier transform of $G(\omega)$ in place of the inverse Fourier transform.

Since $y(t)$ and $h(t)$ are both known, $Y(\omega)$ and $H(\omega)$ can both be computed. This implies that the estimate of the parameter distribution function, $g(t)$, can be obtained from equation (3.11) as

$$\hat{g}(t) = \mathbf{F}^{-1} [\hat{G}(\omega)] \quad (3.12)$$

where \mathbf{F}^{-1} denotes inverse Fourier transform operation, and

$$\hat{G}(\omega) = \frac{Y(\omega)}{H(\omega)} = G(\omega) + \frac{K(\omega)}{H(\omega)}. \quad (3.13)$$

The direct implementation of equation (3.12) via (3.13) poses some serious problems. The most serious of these problems is the problem of error ripples

present in the estimate of the parameter distribution function. This is primarily the consequence of division by $H(\omega)$ in equation (3.13). Since $Y(\omega)$ and $H(\omega)$ approach zero simultaneously, $\hat{G}(\omega)$ is well behaved provided $K(\omega)$ equals zero. However, $K(\omega)$ is in no way related to $H(\omega)$ and does not approach zero as $H(\omega)$ does. Consequently, the term $\frac{K(\omega)}{H(\omega)}$ tends to make $\hat{G}(\omega)$ unbounded. The inverse Fourier transform of such a function can produce error ripples that can conceal the peaks corresponding to signal parameters, thus rendering the results meaningless.

One way to reduce the error ripples is to cutoff the value of $\hat{G}(\omega)$ at some frequency ω_c , before the divergence occurs. Now the solution to the problem at hand poses conflicting requirements. On one hand it is required to reduce the bandwidth of $\hat{G}(\omega)$ to reduce the error ripples but on the other hand the smaller the value of the cutoff frequency ω_c the poorer the resolution, limiting the resolvability of closely related exponents.

Callahan and Pizer [33] and Pizer et al [2] introduced error peak averaging and error ripples subtraction techniques to reduce the error ripples. These techniques are difficult to implement and are not computationally efficient when M is large, (>2), and the exponents are close to each other.

Smith and Cohn-Sfetcu [22] demonstrated the use of Gaussian filters to achieve reduction of error ripples in the parameter distribution function. Provencher [34] used convergence parameters along with Gaussian filtering to improve the parameter distribution. But the use of filters to improve the signal to noise ratio

(SNR) of the deconvolved signal results in broadening the peaks in the distribution function. Hence limiting the resolvability.

Arunachalam [23] used one-step forward predictor, determined by Burg's maximum entropy method, to extend the frequency range of the deconvolved data beyond the optimal passband, ω_{nc} . The justification of this extension method lies in the fact that the frequency spectrum of the signal formed by superimposing delayed dirac delta functions exists for all n and can be extended beyond ω_{nc} by the recursive application of one-step forward predictor. The results produced by this extension method showed marked improvement over the previous methods but produced poor estimates at low SNR.

Salami [24] and Nichols et al [25] modeled the "Good portion" of the deconvolved data as an autoregressive moving average (ARMA) process and determined the parameters of the multicomponent signals from the spectral estimates of the ARMA model. Though they obtained good estimates of M , A_i , and λ_i , the autoregressive (AR) model order required is usually high and the uniform samples of the nonlinearly transformed data, required to perform the DFT, are obtained from the direct knowledge of the signal parameters themselves.

In this dissertation a singular value decomposition (SVD) based transient error method, proposed by Nichols et al [27], is used to obtain high resolution estimates of the exponents of the multi-exponential signal. The "Good portion" of the deconvolved data is modeled using a special ARMA process. The AR coefficients of the

process are determined directly from the data using a SVD-based linear least square technique. The MA coefficients are derived from the residual error sequence so as to account for the nonstationary noise present in the deconvolved data.

3.3 DISCRETE REPRESENTATION OF MULTICOMPONENT SIGNAL

Though all the signals considered so far have been assumed to be continuous in nature, in actual practice only the discrete samples of $x(\tau)$, in equation (3.7), are available. Furthermore, these samples are usually not uniformly spaced. This is so because the signal to noise ratio of the decaying exponentials is monotonically decreasing. It is then prudent to sample these signal densely when the SNR is high and sparsely when the SNR is low. The sampled values of $x(\tau)$ are multiplied by τ_n^α and the Gardner's transformation, $\tau_n = e^{n\Delta t}$, is used to obtain the discrete values of $y(t)$ in equation (3.5). Equation (3.5) can be expressed in the discrete form as

$$y[n] = \Delta t \sum_{m=-\infty}^{\infty} g[m] h[n-m] + k[n] \quad (3.14)$$

where, for convenience, the sample values are denoted using square brackets, i.e. $y[n] = y(n \Delta t)$.

Estimation of the distribution function $g[n]$ from equation (3.14) is a discrete time deconvolution or inverse filtering problem. This problem is solved by performing the discrete Fourier transform (DFT) of equation (3.14) and then dividing

both sides by the DFT of $h[n]$, $H(k\Delta\omega)$. This requires uniformly spaced samples of $y[n]$. In this thesis, uniform samples of $y[n]$ are obtained by using cubic spline interpolation technique.

3.3.1 Cubic Spline Interpolation

A program called FITLOS [35] is used to fit the log transformed data with cubic splines. The spline functions are defined as piecewise polynomials of some degree "q", joining in the so called "nodes" and satisfying the continuity constraints for the function itself and its first (q-1) derivatives. Thus, a cubic spline is a polynomial of degree three and is continuous along with its first two derivatives at the nodes, and has the general form

$$S(x) = a_0 + a_1x + a_2x^2 + a_3x^3. \quad (3.15)$$

The cubic spline interpolator algorithm used combines some of the advantages of least squares polynomial method with the cubic spline technique. In other words it computes the polynomial coefficients a_0 , a_1 , a_2 and a_3 such that the quantity

$$I = \sum_{i=1}^N (y_i - S(x_i))^2$$

is minimized subject to the continuity constraints. In the above expression the polynomial $S(x)$ is fitted over the ordered pairs of points (x_1, y_1) , (x_2, y_2) , \dots , (x_N, y_N) . The constraints are imposed by the method of Lagrangian multipliers.

The choice of nodes itself in this algorithm is in no way optimal. However, the

user has a choice of selecting nodes and/or the number of segments.

Once the log transformed data has been fitted with cubic splines the next step is to obtain the uniform samples. The sampling constraints are discussed in the next section.

3.3.2 Sampling Constraints

In equation (3.13) the Fourier transform of the estimate of the parameter distribution function, $\hat{g}(t)$, is obtained by dividing the Fourier transform of the log transformed data, $y(t)$, by the Fourier transform of the pulse $h(t)$, where

$$h(t) = e^{\alpha t} e^{-t}. \quad (3.16)$$

To perform the DFTs, $y(t)$ and $h(t)$ need to be uniformly sampled and the maximum value of the sampling interval Δt needs to be established so as to minimize aliasing. Salami [24] has shown that for $0 < \alpha \leq 1$, the pulse $h(t)$ is essentially band limited to 2 Hz. Therefore aliasing is avoided for

$$\Delta t \leq 0.25 \text{sec.}$$

Furthermore, $g(t)$ is strictly time limited with time width $\ln(\lambda_M/\lambda_1)$ for $\lambda_1 < \lambda_2 < \dots < \lambda_M$. Then to avoid aliasing of $g(t)$ in the time domain, we need

$$N \Delta t \geq \ln[\lambda_M/\lambda_1]$$

or

$$N \geq \frac{1}{\Delta t} \ln[\lambda_M/\lambda_1] \quad (3.17)$$

where N is the number of samples of $y[n]$ and $h[n]$ in equation (3.14). If the above condition is not satisfied then the data values of $y[n]$ and $h[n]$ must be prealiased before performing the DFTs.

3.3.3 Selecting Good Portion of the Deconvolved Data

Performing the DFT of equation (3.14), we have

$$Y_k = G_k H_k + K_k; \quad -N/2 < k \leq N/2, \quad (3.18)$$

where Y_k , G_k , H_k , and K_k are the N -point DFTs of $y[n]$, $g[n]$, $h[n]$ and $k[n]$ respectively. From equation (3.18), we have

$$\hat{G}_k = \frac{Y_k}{H_k} = G_k + N_k \quad (3.19)$$

where

$$N_k = \frac{K_k}{H_k}$$

Since the noise K_k is wide band and H_k is band limited, dividing K_k by H_k tends to amplify the high frequency noise. Thus \hat{G}_k becomes unbounded for large k and any attempt to model this becomes meaningless. In order to proceed with the modeling a "Good portion" of the deconvolved data \hat{G}_k is selected over which \hat{G}_k

is locally stationary. This is done by plotting $|\hat{G}_k|$ versus k and selecting L such that for $|k| > L$, $|\hat{G}_k|$ begins to diverge. Mathematically this can be expressed as

$$\tilde{G}_k = W_k \hat{G}_k \quad (3.20)$$

where the rectangular window W_k is defined as

$$W_k = \begin{cases} 1; & |k| \leq L \\ 0; & \text{otherwise.} \end{cases}$$

Performing the DFT of \tilde{G}_k in equation (3.20) produces poor spectral resolution due to short, $2L+1$, data length. Using diverging values of \hat{G}_k to obtain more points results in high error ripples. If this happens the detection of false peaks is always a possibility. Filtering \hat{G}_k on the other hand reduces the error ripples but severely limits the resolution due to the finite width of the filter. Under these circumstances the parametric modeling techniques become viable alternatives and are introduced in the next section.

3.4 PARAMETRIC METHODS

Arunachalam [23] used one step forward predictor to extend \tilde{G}_k beyond $2L+1$ points. The results showed a significant improvement over nonparametric methods at high SNR but produced poor results at low SNR. The extended Prony's method gives good estimates of $\ln\lambda_i$ and B_i provided the SNR is high and M is known. This method is based upon polynomial rooting technique. The forward predictor

coefficients, also the polynomial coefficients, are obtained in this thesis from the "Good portion" of the deconvolved data. Once λ_i are determined, B_i can then be found using the linear least squares method. A similar technique is discussed in [36] for decomposing complex sinusoids in the additive noise and is referred to as the modified forward-backward linear predictor (FBLP) method. The name is derived from the fact that the prediction filter coefficients minimize the energy in the error signal both in the forward and backward directions. Pisarenko's harmonic decomposition (PHD) method is another polynomial rooting method. The coefficients of the polynomial in "z" are obtained as follows.

The data sequence, y_n , is a sum of complex sinusoids in an independent noise sequence, w_n , given by

$$y_n = x_n + w_n$$

where

$$x_n = \sum_{i=1}^p b_i e^{j(\omega_i n + \phi_i)}$$

Introducing the vectors

$$\underline{y}^T = [y_n, y_{n-1}, \dots, y_{n-2p}]$$

$$\underline{a}^T = [1, a_1, a_2, \dots, a_{2p}]$$

and

$$\underline{\mathbf{w}}^T = [w_n, w_{n-1}, \dots, w_{n-2p}]$$

the problem can be expressed as the solution to the matrix equation

$$\mathbf{R}_p \underline{\mathbf{a}} = \sigma_w^2 \underline{\mathbf{a}} \quad (3.21)$$

where

$$\mathbf{R}_p = E[\underline{\mathbf{y}} \underline{\mathbf{y}}^T]$$

and

$$\sigma_w^2 = E[\underline{\mathbf{w}} \underline{\mathbf{w}}^T].$$

The sequence w_n is assumed to be white noise sequence with zero mean and variance σ_w^2 . The symbols E and T in the above equations denote expectation and transpose operators respectively. Furthermore, the harmonic signal x_n is assumed to be uncorrelated with the white noise sequence. From equation (3.21) it is evident that the polynomial coefficient vector $\underline{\mathbf{a}}$ is the eigenvector of the auto-correlation matrix \mathbf{R}_p corresponding to the minimum eigenvalue σ_w^2 . Again, the PHD method requires prior knowledge of p , the number of components, and the determination of minimum eigenvalue depends on the accuracy of the estimated autocorrelation functions $R_y(k)$. A biased correlation function ensures a positive

definite Toeplitz matrix \mathbf{R}_p but its use in (3.21) produces inaccurate frequency estimates. An unbiased correlation function on the other hand does not guarantee a positive definite matrix and may have negative eigenvalues.

The above mentioned techniques suffer from the following drawbacks:

1. Inaccurate parameter estimates are obtained for low SNR.
2. The number of components, M , must be known for both Prony's and PHD methods.
3. Whenever the angular frequencies, ω_i , are more closely spaced than the reciprocal of the observation interval or, more generally, when the sinusoids are not orthogonal to each other over the observation interval the spectral resolutions are poor.

A SVD-based transient error method does not require prior knowledge of M and produces high resolution parameter distribution function both at high and relatively low SNR.

3.4.1 SVD-based Transient Error Method

In this method the "Good portion" of the deconvolved data \tilde{G}_k is obtained as outlined in section 3.3.3 of this Chapter. The selected portion of the data is then modeled as a special nonstationary autoregressive moving average (ARMA) process. The AR parameters of this ARMA model are obtained directly from the data using SVD-based linear least squares technique. The MA coefficients correspond to

the residual error sequence so as to account for the nonstationary noise in the deconvolved data.

From equation (3.19), the deconvolved data is given as

$$\hat{G}_k = \frac{Y_k}{H_k} = G_k + N_k$$

where

$$G_k = \sum_{i=1}^M B_i e^{jk\Delta\omega \ln \lambda_i},$$

$$B_i = \frac{A_i}{\lambda_i^\alpha},$$

$$\Delta\omega = 2\pi\Delta f,$$

$$\Delta f = \frac{1}{N\Delta t}$$

and N is the number of DFT points used. The length L of the "Good portion" of the deconvolved data, \tilde{G}_k in equation (3.20), is chosen such that $|\hat{G}_k|$ attains its minimum for $k = L$. The resolvability of λ_i can be further improved by including the negative frequencies of the deconvolved data to account for the arithmetic errors. This is accomplished by right shifting \tilde{G}_k by an amount L , i.e.

$$\tilde{G}_k = \tilde{G}_{k-L}.$$

It can be easily shown that the magnitude of the DFT of \tilde{G}_k is invariant under this transformation.

3.4.1.1 Modeling of the Deconvolved Data

Realizing that \tilde{G}_k is a sequence of complex points and is stationary in the interval $0 \leq k \leq 2L$, one can fit a forward predictor over this sequence. Furthermore, since this sequence is generated by superimposing M complex sinusoids, the sequence, in the absence of noise, is perfectly predicted by an M^{th} . order predictor.

Therefore

$$\tilde{G}_k = -\sum_{i=1}^M a_i \tilde{G}_{k-i}; \quad M \leq k \leq 2L$$

where a_i is the AR coefficients. In actual practice though the perfect prediction is never possible because of the following reasons:

1. Since \tilde{G}_k is zero for $k < 0$, therefore the output transient values e_0, e_1, \dots, e_{M-1} will be nonzero.
2. The data \tilde{G}_k is noisy because of experimental noise and arithmetic errors.

Therefore, the output error sequence is

$$e_k = \sum_{i=0}^M a_i \tilde{G}_{k-i}; \quad a_0 = 1 \quad \text{and} \quad 0 \leq k \leq 2L. \quad (3.22)$$

Taking the z -transform of equation (3.22), we have

$$E(z) = A(z) \tilde{G}(z).$$

Hence

$$\hat{g}(n) = \tilde{G}(z) = \left. \frac{E(z)}{A(z)} \right|_{z=e^{j2\pi \frac{n}{N}}} = \sum_{i=1}^M B_i \delta(n - \ln \lambda_i); \quad 0 \leq n < N \quad (3.23)$$

where

$$E(z) = \sum_{i=0}^{2L} e_i z^{-i}$$

and $A(z)$ is the well known prediction error filter, given as

$$A(z) = \sum_{i=0}^M a_i z^{-i}; \quad a_0 = 1.$$

As seen from equation (3.23), the accuracy of the estimate of the parameter distribution function $\hat{g}(n)$ has a direct dependence on the accuracy of the estimated AR coefficients, which in turn is determined by the algorithm used for its computation. Nichols et al [25] have reported that Burg's [37] and Marple's [38] algorithms produce poor estimates of A_i and λ_i . Burg's algorithm is known to produce biased estimates of λ_i for short data records and Marple's algorithm is quite sensitive to

arithmetic errors and does not ensure a minimum phase AR filter. Barrodale and Erickson's [39] algorithm alleviates the numerical problems associated with Burg's and Marple's algorithms by using Cholesky's decomposition of the covariance matrix to arrive at the AR coefficients. However, this technique also produces poor estimates of A_i and λ_i for low SNR and becomes unstable for large values of M . Consequently, Nichols et al [27] proposed a SVD-based method based on Kumaresan and Tufts [40] work which produces high resolution estimates of A_i and λ_i even at relatively low SNR.

3.4.1.2 Determination of AR Coefficients Using SVD

Autoregressive (AR) coefficients in this method are derived directly from the data using the forward prediction error equation, given as

$$\tilde{G}_k + \sum_{i=1}^p a_i \tilde{G}_{k-i} = e_k; \quad p \leq k \leq 2L \quad (3.24)$$

where p is the predictor order. Equation (3.24) can be expressed in matrix notation as

$$\underline{\mathbf{g}} + \mathbf{G} \underline{\mathbf{a}} = \underline{\mathbf{e}}$$

where

$$\underline{\mathbf{g}} = [\tilde{G}_p, \tilde{G}_{p+1}, \dots, \tilde{G}_{2L}]^T$$

$$\underline{\mathbf{a}} = [a_1, a_2, \dots, a_p]^T,$$

$$\underline{\mathbf{e}} = [e_p, e_{p+1}, \dots, e_{2L}]^T$$

and

$$\mathbf{G} = \begin{bmatrix} \tilde{G}_{p-1} & \tilde{G}_{p-2} & \dots & \tilde{G}_0 \\ \tilde{G}_p & \tilde{G}_{p-1} & \dots & \tilde{G}_1 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \tilde{G}_{2L-1} & \tilde{G}_{2L-2} & \dots & \tilde{G}_{2L-p} \end{bmatrix}.$$

It is then desired to find the prediction error coefficient vector $\underline{\mathbf{a}}$ such that the square of the norm of the error vector $\underline{\mathbf{e}}$ is minimized. Where the square of the norm of the vector $\underline{\mathbf{e}}$ is given as $\underline{\mathbf{e}}^H \underline{\mathbf{e}}$ and the Hermitian operator, H , implies transpose conjugation.

The solution to this well known least squares problem is given as

$$\underline{\mathbf{a}} = -[\mathbf{G}^H \mathbf{G}]^{-1} \mathbf{G}^H \underline{\mathbf{g}}. \quad (3.25)$$

In equation (3.25), solution for the unknown autoregressive coefficient vector $\underline{\mathbf{a}}$ is obtained by using singular value decomposition of the data matrix \mathbf{G} . From the singular value decomposition theorem, a $(2L-p+1) \times p$ matrix \mathbf{G} can be decomposed as

$$\mathbf{G} = \mathbf{U} \mathbf{D} \mathbf{V}^H \quad (3.26)$$

where \mathbf{U} and \mathbf{V} are unitary matrices such that

$$\mathbf{U} \mathbf{U}^H = \mathbf{I},$$

$$\mathbf{V} \mathbf{V}^H = \mathbf{I},$$

$$\mathbf{D} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_w, 0, 0, \dots, 0)$$

and

$$\mathbf{I} = \text{diag}(1, 1, 1, \dots, 1).$$

Here w denotes the rank of the matrix \mathbf{G} such that $w \leq \min(2L-p+1, p)$. The columns of \mathbf{U} are the left singular vectors of \mathbf{G} i.e. the eigenvectors of $\mathbf{G} \mathbf{G}^H$ and the columns of \mathbf{V} are the right singular vector of \mathbf{G} i.e. the eigenvectors of $\mathbf{G}^H \mathbf{G}$. The diagonal elements of \mathbf{D} are called the singular values of \mathbf{G} and are in fact the square roots of the eigenvalues of $\mathbf{G}^H \mathbf{G}$. Let these eigenvalues be denoted by $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$ where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_w > 0$ and $\sigma_{w+1} = \sigma_{w+2} = \dots = \sigma_p = 0$. Furthermore, since $\mathbf{G}^H \mathbf{G}$ is a Hermitian matrix its eigenvalues must be real and positive.

Substituting (3.26) into (3.25) gives

$$\begin{aligned}
\underline{\mathbf{a}} &= -[\mathbf{VDU}^H \mathbf{UDV}^H]^{-1} \mathbf{VDU}^H \cdot \underline{\mathbf{g}} \\
&= -[\mathbf{VD}^2 \mathbf{V}^H]^{-1} \mathbf{VDU}^H \cdot \underline{\mathbf{g}} \\
&= -[\mathbf{VD}^{-2} \mathbf{V}^H \cdot \mathbf{VDU}^H] \cdot \underline{\mathbf{g}}
\end{aligned}$$

or

$$\underline{\mathbf{a}} = -[\mathbf{VD}^{-1} \mathbf{U}^H] \cdot \underline{\mathbf{g}} \quad (3.27)$$

where \mathbf{D}^{-1} is obtained from \mathbf{D} by replacing each nonzero diagonal entry by its reciprocal.

Ideally, the rank w of the matrix \mathbf{G} is equal to the number of signal components M and as well as the number of nonzero singular values of \mathbf{D} . Wilkinson [41] showed that the perturbation, due to noise, of the w originally nonzero singular values of \mathbf{D} and the corresponding first w columns of \mathbf{U} and \mathbf{V} are relatively small. However, in the presence of noise the matrix \mathbf{G} tends to become a full rank matrix and the perturbations in the directions of the rest of the columns of \mathbf{U} and \mathbf{V} are quite large. This necessitates the approximation of \mathbf{G} , in a Frobenius norm sense, by a lower rank matrix $\hat{\mathbf{G}}$. Eckart and Young, 1936, developed the following procedure, based on SVD, to find the best lower rank approximation to a given matrix.

Let the rank of \mathbf{G} be w , and let $s(W)$ be the set of all $(2L-p+1) \times p$ matrices of rank $W < w$. Then for all matrices B in $s(W)$

$$\|G - \hat{G}\| \leq \|G - B\|$$

where

$$\hat{G} = U\hat{D}V^H$$

and $\| \cdot \|$ denotes F-norm (Frobenius norm) defined as

$$\|A\| = \left[\sum_{i=1}^N \sum_{j=1}^N |a_{ij}|^2 \right]^{\frac{1}{2}}.$$

The matrix \hat{D} is obtained from D in equation (3.26) by setting to zero all but its w largest singular values. This is equivalent to retaining only first w columns of U and V , and only first w rows and columns of the matrix D . Thus an improvement in the SNR is achieved by discarding the noise space.

With this, equation (3.27) reduces to

$$\underline{a} = -[V\hat{D}^{-1}U^H] \cdot \underline{g} \quad (3.28)$$

Once the AR parameters are computed using equation (3.28), the residual sequence e_i can be determined using equation (3.22). The distribution function $\hat{g}(n)$ is then estimated using equation (3.23)

CHAPTER 4

RESULTS

4.1 INTRODUCTION

In this Chapter the performance of the transform method, the Total Least Squares method, the nonlinear least squares method and the maximum likelihood method is compared. The theory of these techniques is presented in Chapters 2 and 3. The performance of each method is based on the following criteria:

1. Accuracy of the estimated signal parameters in the absence of the noise. This is an indicator of the robustness of the algorithm to arithmetic errors and the user selected parameters.
2. Effect of the additive noise on the estimates.
3. Resolvability of closely-related exponents, i.e. for $\frac{\lambda_{i+1}}{\lambda_i} \leq 2$, at both low and high noise levels.
4. Accuracy of estimates when the number of signal components are greater than two.
5. Relative numerical stability of the algorithm.

The results are organized into two sections. In the first section simulated data is used to assess the performance of the different methods and in the second section experimentally obtained WOOD-NMR data is used to obtain the estimate of the unknown signal parameters.

4.2 SIMULATION RESULTS

The purpose of this section is to demonstrate the capabilities and the limitations of the four aforementioned algorithms in estimating the parameters of multi-exponential signals, both in the absence and the presence of the additive noise. A typical signal under study is plotted in figure 4.1 using a logarithmic scale to emphasize the effect of noise at low signal levels. The signal is observed in the Gaussian additive noise with a zero mean and a standard deviation (σ_n) of 0.005. It is apparent, both from figure 4.1 and the equation (3.1), that the SNR of these types of signals decreases rapidly with increasing time.

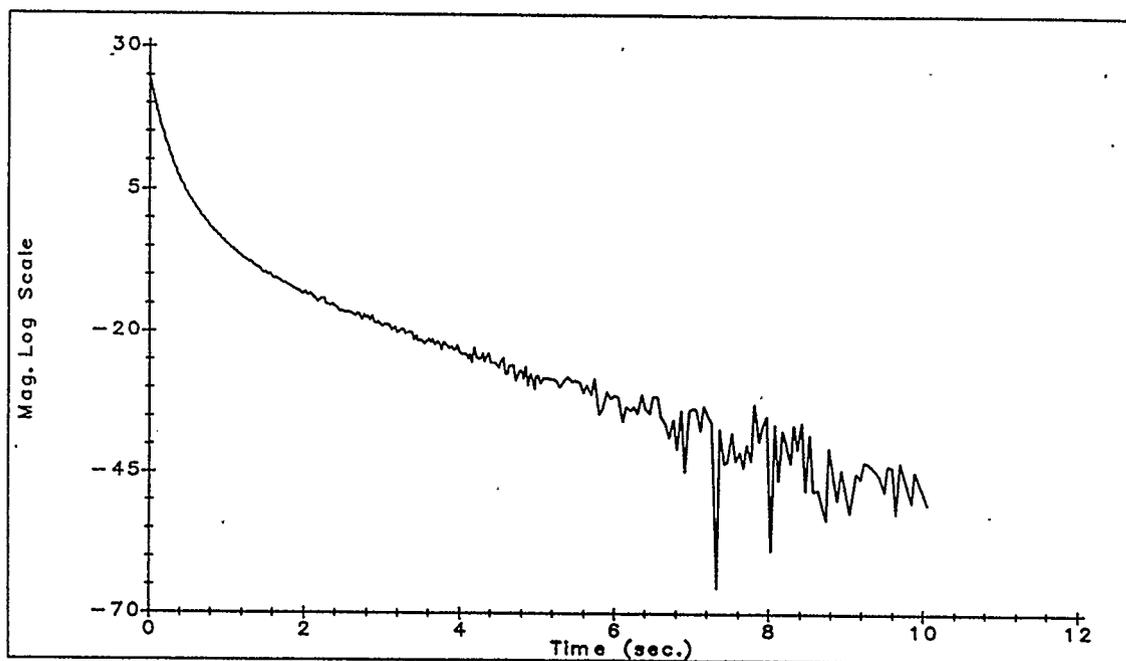


Figure 4.1 A sample multi-exponential signal with $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 0.005$.

4.2.1 The Transform Method

The theoretical aspect of this method is treated in Chapter 3. Following the Gardner's transformation, the estimate of the Fourier transform of the distribution function is obtained via the deconvolution process. This necessitates acquiring equi-spaced samples of the nonlinearly transformed data. These samples are obtained by cubic-spline fitting the transformed data and resampling it at the desired frequency. Nichols et al [25] have shown that the pulse $h(t) = e^{(\alpha t)} e^{-e^t}$

is essentially bandlimited to 2 Hz. Therefore, the sampling frequency needed to avoid aliasing must be at least 4 Hz.

The spline fitting is performed using the program FITLOS [35], capable of fitting up to 500 points of data with up to 50 cubic splines by the method of least squares. The user has the options of either specifying the node points or the number of segments. If the nodes are explicitly specified then the data is spline fitted using these nodes only. If, on the other hand, only the number of segments are specified then the data is divided, as evenly as possible, among these segments to obtain the node points. The functionality of this program is demonstrated by fitting the transformed data of figure 4.1 with 30 cubic splines and the results are shown in figure 4.2. The fit of the resampled data to the transformed data is extremely good and the two curves are hardly distinguishable.

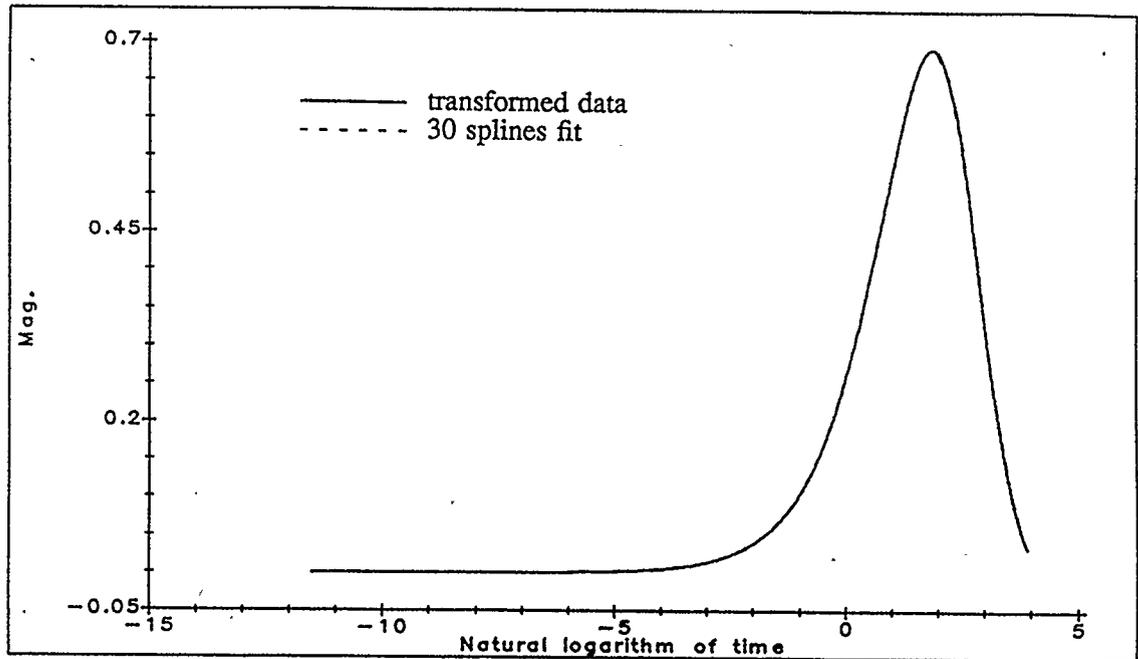


Figure 4.2 Cubic spline fit of the transformed data with 30 splines.

Discrete samples of the multi-exponential signal are obtained by sampling the signal at a much higher rate for small values of time and at a relatively slower rate for large values of time. This is done primarily for two reasons:

1. To obtain a sufficient number of samples of the signal components with large decay rates.
2. To obtain more samples of the signal in the time region where the signal is decaying rapidly.

The time interval is usually selected from the initial time of 0.00001 sec. to ten times the largest time constant.

The parameter distribution function is obtained, as outlined in section 3.5.1, by fitting the "Good portion" of the deconvolved with a special autoregressive moving average (ARMA) process.

As shown in Chapter 3 equation (3.23), the parameter distribution function is given as

$$\hat{g}(n) = \sum_{i=1}^M B_i \delta(n - \ln \lambda_i)$$

where $B_i = \frac{A_i}{\lambda_i \alpha}$. Hence, the locations of the impulses give λ_i 's and the areas under the impulses give A_i 's, implicitly. To obtain a better estimate of the locations of the peaks of the distribution function, the z-transform in equation (3.23) is evaluated on a circle with radius $\frac{1}{r}$, for $0.9 \leq r < 1.0$. That is, the distribution function in equation (3.23) are calculated for $z = \frac{1}{r} \exp(j2\pi \frac{n}{N})$ with $r = 0.99$.

Estimating the A_i 's from the distribution function by calculating the areas under the impulses is a cumbersome and error prone technique, though the peaks of the distribution do give a relative estimate of the amplitudes of $\frac{A_i}{\lambda_i \alpha}$. In actual practice, once the exponents have been estimated, it is easy to find A_i 's using a

least squares technique, such as Prony's extended LS method.

The four cases studied in this section are the ones that have been reported in the literature to be difficult to decompose. The ratios $\frac{A_i}{\lambda_i}$ have been kept constant for no reason other than to keep the relative magnitudes of the peaks of the estimated parameter distribution functions the same.

Case study 1: The signal under study here has two closely-related exponents and has the form

$$x_1(\tau) = 0.1 e^{-0.1 \tau} + 0.2 e^{-0.2 \tau} + n(\tau) \quad (4.1)$$

where $n(\tau)$ is additive Gaussian noise with a zero mean and a standard deviation of 10^{-6} . As seen from figure 4.3, both the exponents are clearly resolved and accurately estimated, though the amplitudes are slightly inaccurate as indicated by the heights of the two peaks. The estimated exponents λ_1 and λ_2 are 0.1030 and 0.2080 respectively.

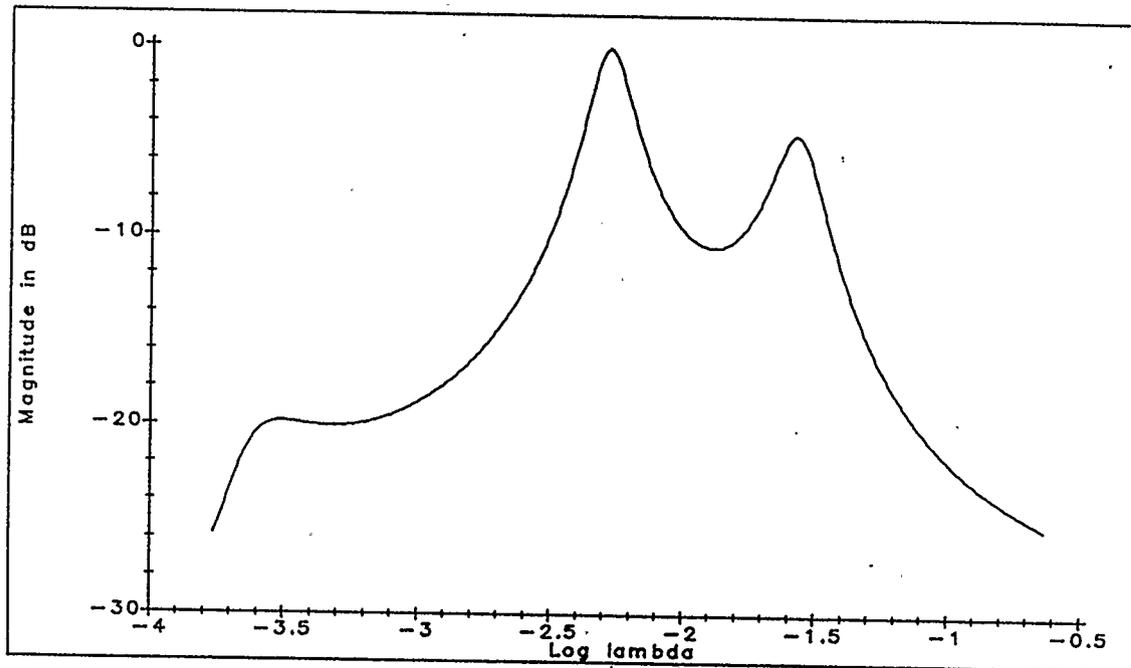


Figure 4.3 Estimate of the parameter distribution function for $\underline{A} = \underline{\lambda} = [0.1, 0.2]^T$ and $\sigma_n = 10^{-6}$.

Case study 2: Callahan and Pizer [33] observed that the presence of component $e^{-1.0t}$ introduced large error ripples in the estimate of the parameter distribution function obtained using Gardner's method. Consequently, the multi-component signal considered in this case study includes an exponential with exponent 1.0 and has the form

$$x_2(\tau) = 1.0 e^{-1.0 \tau} + 2.0 e^{-2.0 \tau} + n(\tau). \quad (4.2)$$

The additive noise is Gaussian with a zero mean and a standard deviation of 10^{-6} . As seen from figure 4.4, both the components are resolved and the exponents λ_1 and λ_2 are estimated as 0.9692 and 2.0200 respectively.

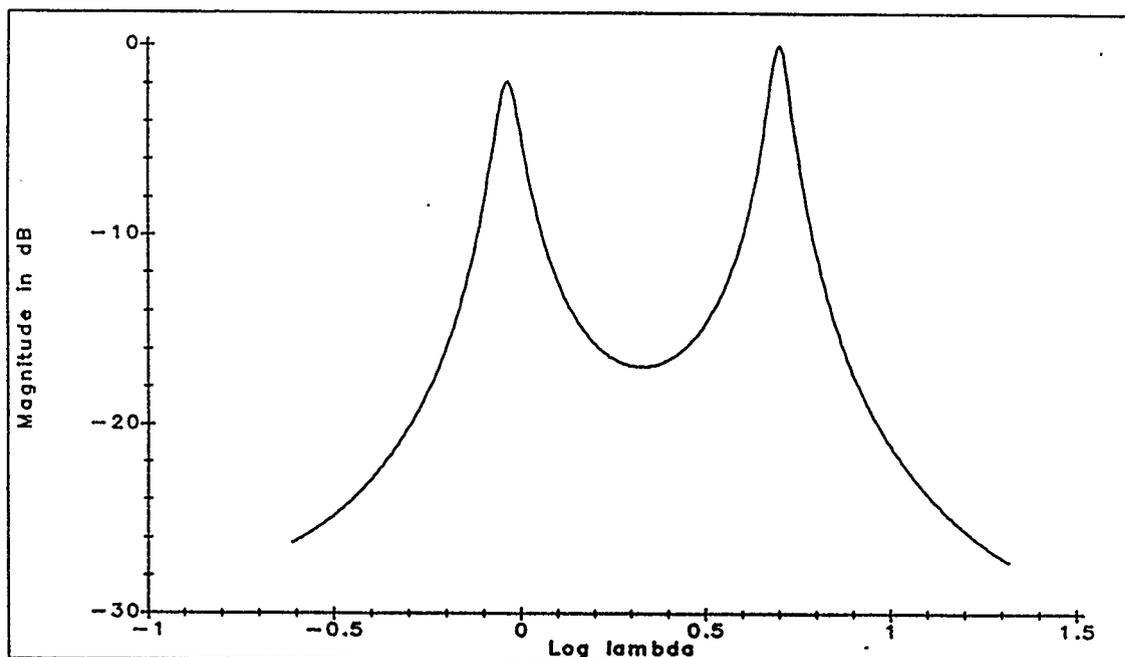


Figure 4.4 Estimate of the parameter distribution function for $\underline{\Lambda} = \underline{\lambda} = [1.0, 2.0]^T$ and $\sigma_n = 10^{-6}$.

Case study 3: The signal considered here has three components with widely-spaced exponents, i.e. $\frac{\lambda_3}{\lambda_1} = 10.0$. The signal also includes a component with exponent 1.0 and is of the form

$$x_3(\tau) = 1.0 e^{-1.0 \tau} + 3.0 e^{-3.0 \tau} + 10.0 e^{-10.0 \tau} + n(\tau). \quad (4.3)$$

The standard deviation of the noise is taken to be 10^{-5} . The results are shown in figure 4.5 with all the components being accurately resolved and the amplitudes of the peaks being the same.

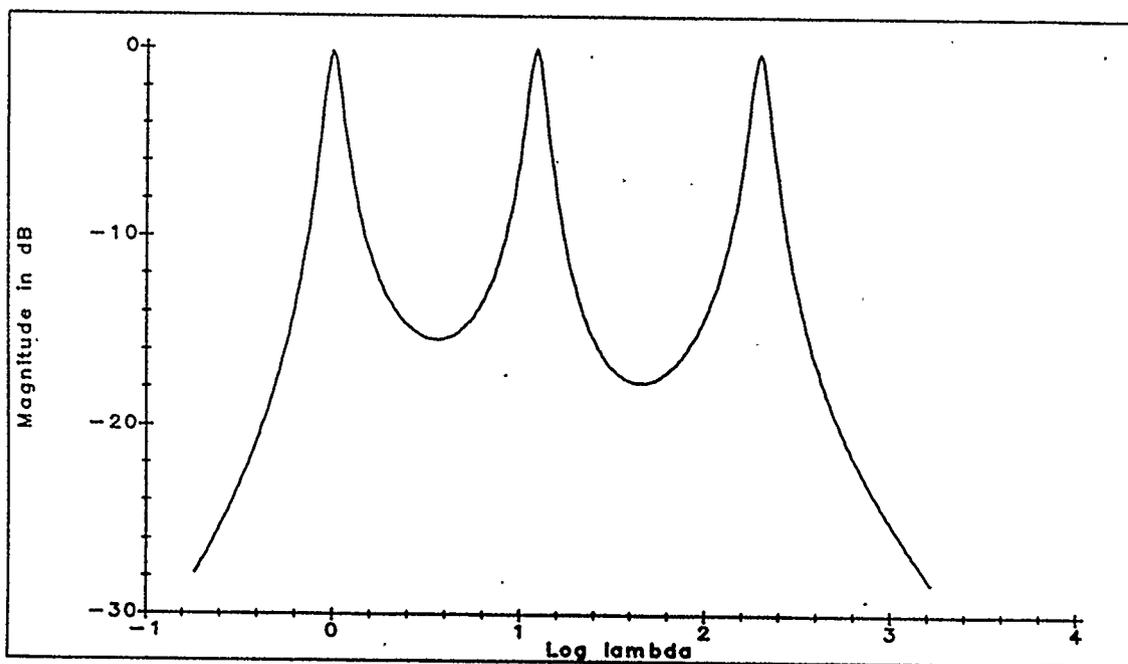


Figure 4.5 Estimate of the parameter distribution function for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-5}$.

The exponents λ_1 , λ_2 and λ_3 are estimated from the distribution function in figure 4.5 as 1.000, 2.985 and 9.943 respectively.

Case study 4: The test signal of the form

$$x_4(\tau) = 0.5 e^{-0.5 \tau} + 2.0 e^{-2.0 \tau} + 5.0 e^{-5.0 \tau} + 10.0 e^{-10.0 \tau} + n(\tau) \quad (4.4)$$

is a typical multicomponent signal problem that is difficult to solve as the number

of signal components is large ($M = 4$) and the ratio of $\frac{\lambda_4}{\lambda_3} = 2$. The estimate of

the parameter distribution function obtained is shown in figure 4.6.

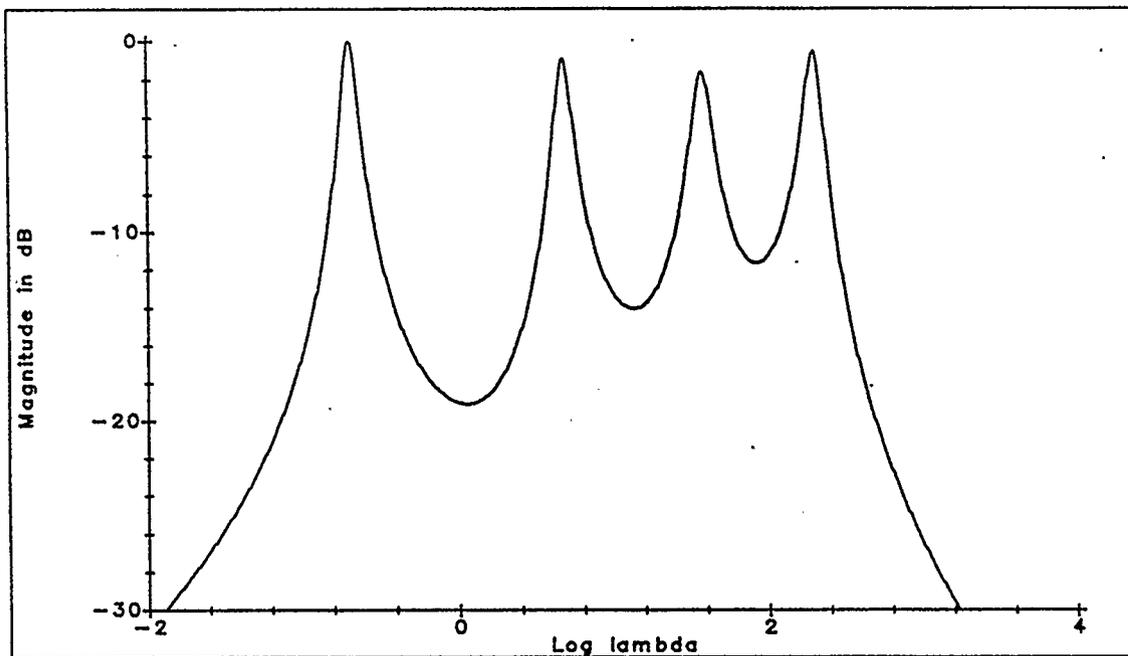


Figure 4.6 Estimate of the parameter distribution function for $\underline{\Lambda} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$.

The additive noise has a standard deviation of 10^{-5} and a zero mean. The exponents λ_1 , λ_2 , λ_3 and λ_4 are estimated as 0.499, 1.958, 4.808 and 9.866 respectively.

During the course of obtaining the estimate of the parameter distribution function, a careful attention has to be paid to the selection of certain parameters. The selection of these parameters is not always obvious and affects the accuracy of the distribution function to a varying degree. It is, therefore, only prudent to study the effect of these parameters on the estimate of parameter distribution function.

4.2.1.1 Selection of the Nodes for Cubic-Spline Fit

A noisy estimate of the Fourier transform of the parameter distribution function is obtained by deconvolving the transformed data with the pulse

$$h(t) = e^{(\alpha t)} e^{-e^t}. \quad (4.5)$$

Since the Gardner's transformation involves a nonlinear change of the independent variable (time), the uniformly spaced samples of the transformed data, required for deconvolution, are obtained by cubic-spline fitting the data. Though the splines are fitted using least squares (LS) technique, the choice of nodes itself is not optimal in the LS sense. Thus, for a different choice of nodes the sample values will be slightly different. In order to investigate the sensitivity of the distribution function to a slight perturbation in the interpolated values, the signal in equation (4.4) is considered here again.

The signal is transformed and resampled at 4.0 Hz. using 20, 30 and 40 splines. The spline fit in all three cases is shown in figure 4.7.

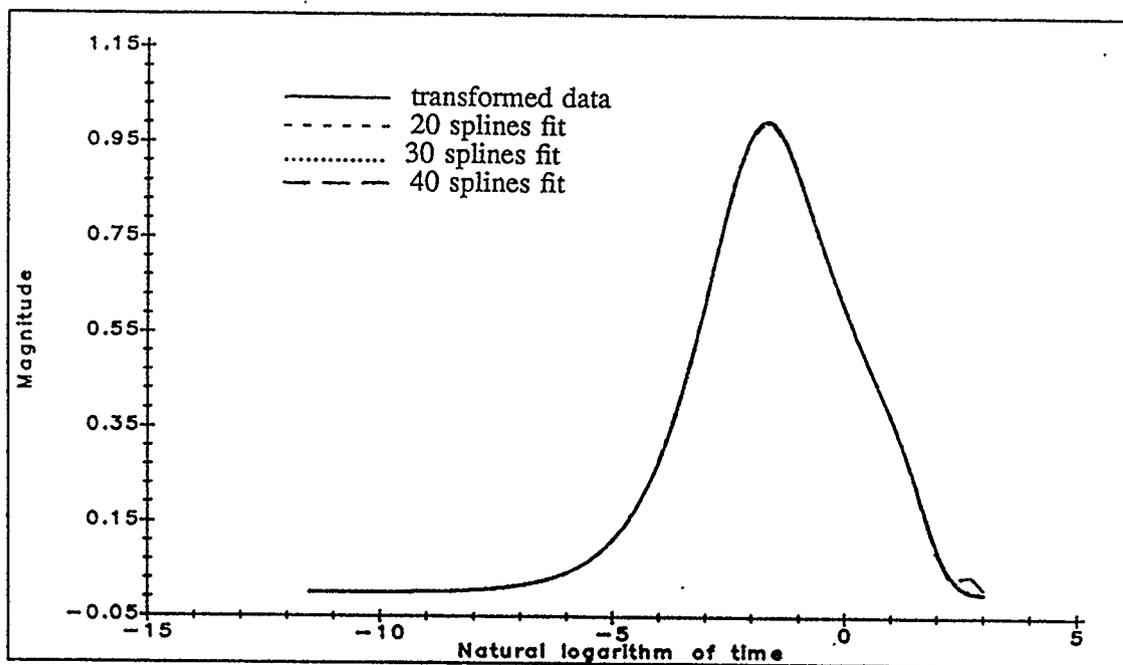


Figure 4.7 Spline fit of the transformed data using 20, 30 and 40 splines.

$$\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T \text{ and } \sigma_n = 10^{-5}.$$

The interpolated values in all cases are found to be very close except at the tail end of the data. In the case of 40 splines last 5 points out of 59 calculated points were notably different. The simple explanation for this deviation in sample values is that with 40 splines the number of points in each segment are comparatively small. Consequently, the spline fitted on the tail section of the data has more

points with high noise levels causing the fit to be quite different as compared to other two cases. The distribution functions obtained in all three cases are shown in figure 4.8.

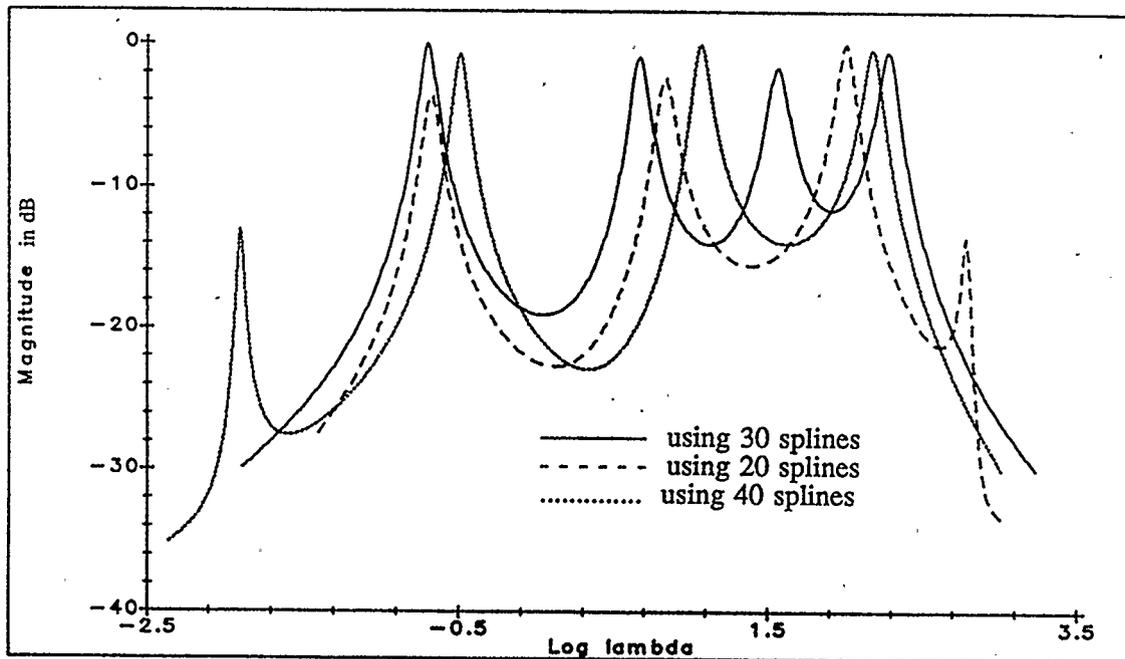


Figure 4.8 Distribution function estimates obtained using 20, 30 and 40 splines. $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$.

Figure 4.8 shows a marked difference in the three distribution functions. In the case of 20 and 40 splines only three signal components are estimated. Furthermore, the decay rates estimates obtained using 20 and 40 splines are grossly inaccurate. Hence, we see that the choice of nodes is extremely critical. However, to

the best of the author's knowledge there exists no least squares technique that is capable of finding the node points in some optimum sense.

4.2.1.2 Selection of Predictor Order

The deconvolution process used to obtain the estimate of the Fourier transform of the distribution function tends to amplify the high frequency noise. Therefore, directly performing the inverse Fourier transform of the deconvolved data to obtain the estimate of the distribution function produces very high noise ripples. If this happens, not only does the possibility of detecting the false components become very real but also the peaks relating to the signal components get buried in noise and may not be detected at all. To avoid this possibility only the "Good portion" of the deconvolved data is selected as outlined in Chapter 3 and is modeled as an ARMA process. The process of predicting the future data values based upon the past values requires finding the best predictor order.

Tufts and Kumaresan [40] have reported that in the case of Principal- Eigenvector method and white stationary noise, the predictor order of $\frac{3}{4}N$, where N is the number of points, practically achieves the Cramer-Rao bound. Since this condition of stationarity is not fulfilled for the case at hand, the predictor orders of $\frac{1}{2}N$, $\frac{2}{3}N$ and $\frac{3}{4}N$ are tried. The resulting distribution estimates are shown in figure 4.9.

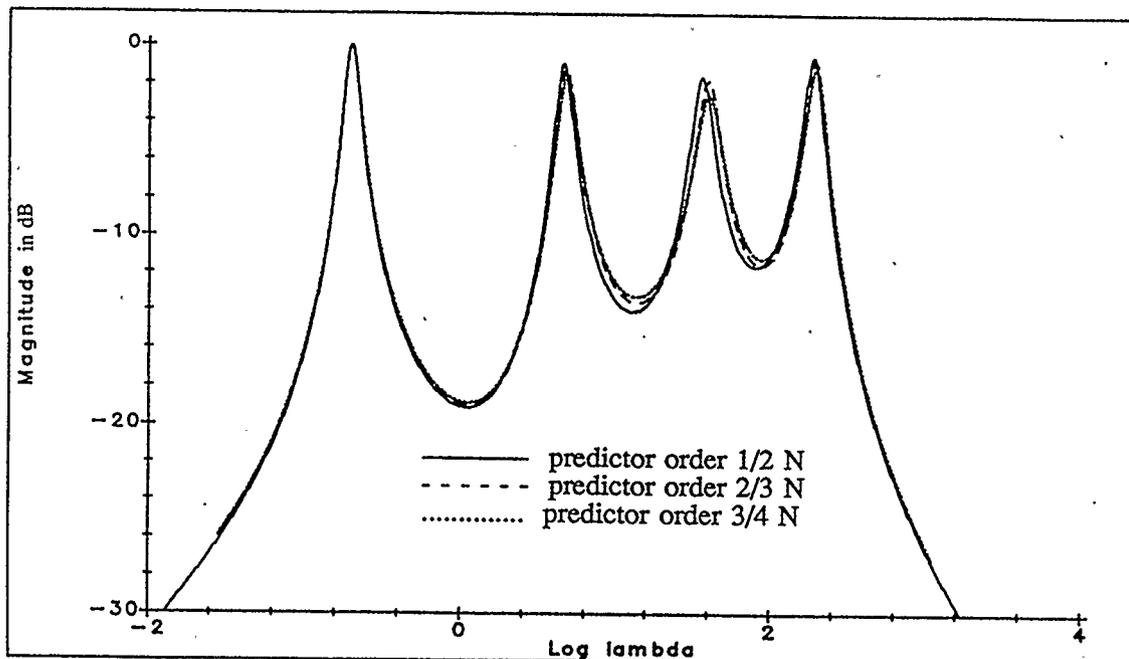


Figure 4.9 Distribution function estimates obtained for different predictor orders. $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$.

Figure 4.9 shows that in all cases the results obtained are accurate and very similar.

4.2.1.3 Selection of Dominant Singular Values

Once the predictor order has been decided upon the predictor coefficients are derived using a SVD-based least squares technique. To ensure stable AR coefficients, a lower rank approximation to the data matrix, in a Frobenius norm

sense [42], is desired. This requires the knowledge of the number of dominant singular values of the data matrix. In the presence of noise this is not always so apparent. Figure 4.10 shows the effect of noise on the spread of singular values of the data matrix generated using 4 complex sinusoids.

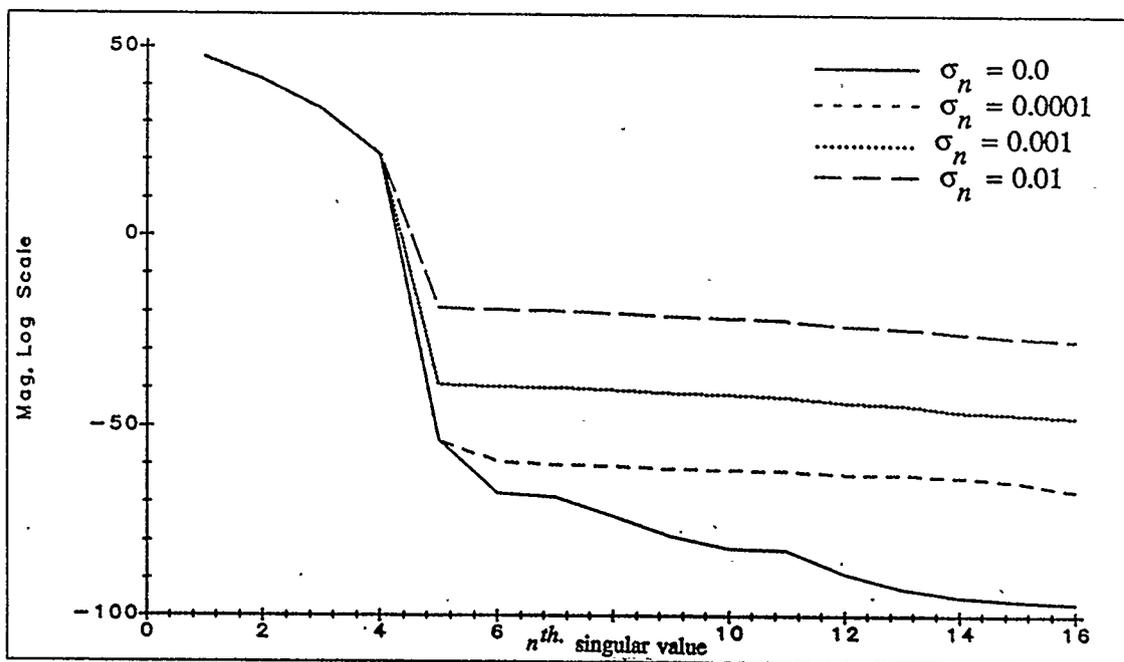


Figure 4.10 Effect of noise on the spread of singular values.

This spread further deteriorates in the case of Gardner's transform method due to the errors introduced by the cubic-spline fit. In such a case Cadzow [43] suggests using a l_2 norm criterion. The n^{th} . l_2 matrix norm is defined as

$$l_2(n) = \left[\frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2}{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2} \right]^{\frac{1}{2}} \quad (4.6)$$

where p is the number of columns of the data matrix and $1 \leq n \leq p$. Then the effective rank of the data matrix is that value of n for which l_2 becomes close to one. The data matrix is said to have low effective rank if the l_2 matrix norm nears one for small values of n relative to p . Otherwise, it is said to have a high effective rank.

The number of dominant singular values are always chosen to be that value of n in equation (4.6) for which l_2 matrix norm approached the value one. Figure 4.11 shows the effect of choosing 4, 6 and 9 dominant singular values when the multicomponent signal is that given in equation (4.4) with $\sigma_n = 10^{-5}$.

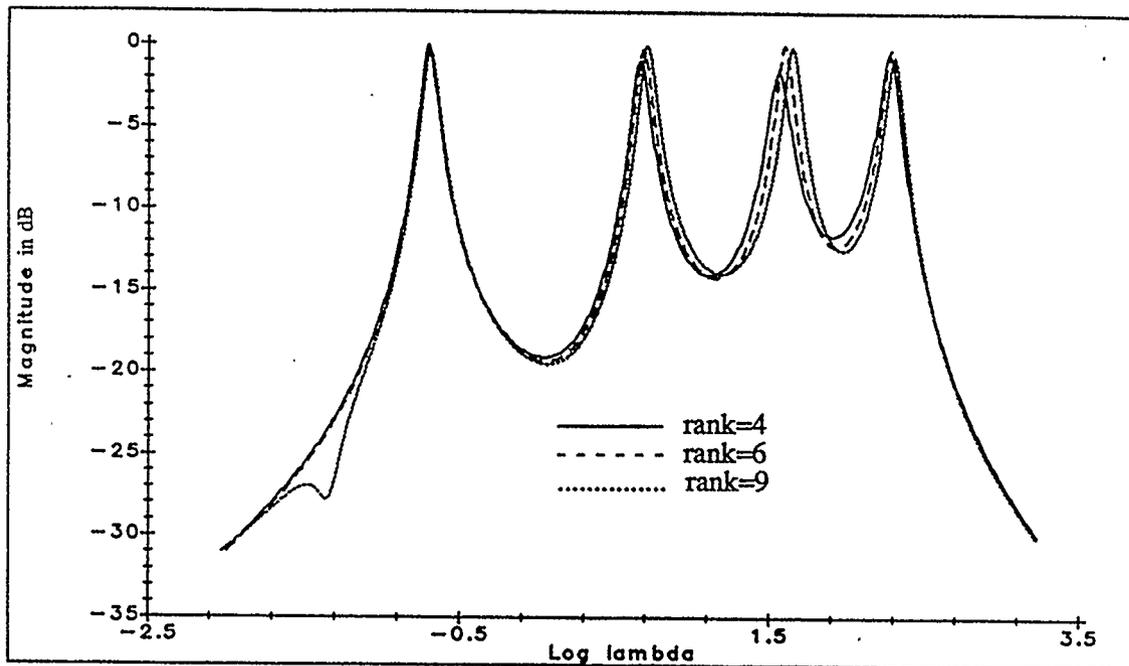


Figure 4.11 Effect of the effective rank of the data matrix on the distribution for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-5}$.

It is seen from figure 4.11 that as long as the effective rank chosen is slightly greater than the number of signal components and the noise level is low, its effect on the estimates is minimal. However, if the effective rank is chosen to be less than the number of components present, which usually is the case with high noise

levels, the results are grossly in error.

4.2.1.4 Effect of Noise

Again, the test signal considered here is the one given in equation (4.4) with the noise standard deviation increased from 10^{-5} to 10^{-4} . The four exponents become barely resolvable with the estimated values of 0.5068, 2.1336, 5.2807 and 9.7889. These results are shown in figure 4.12.

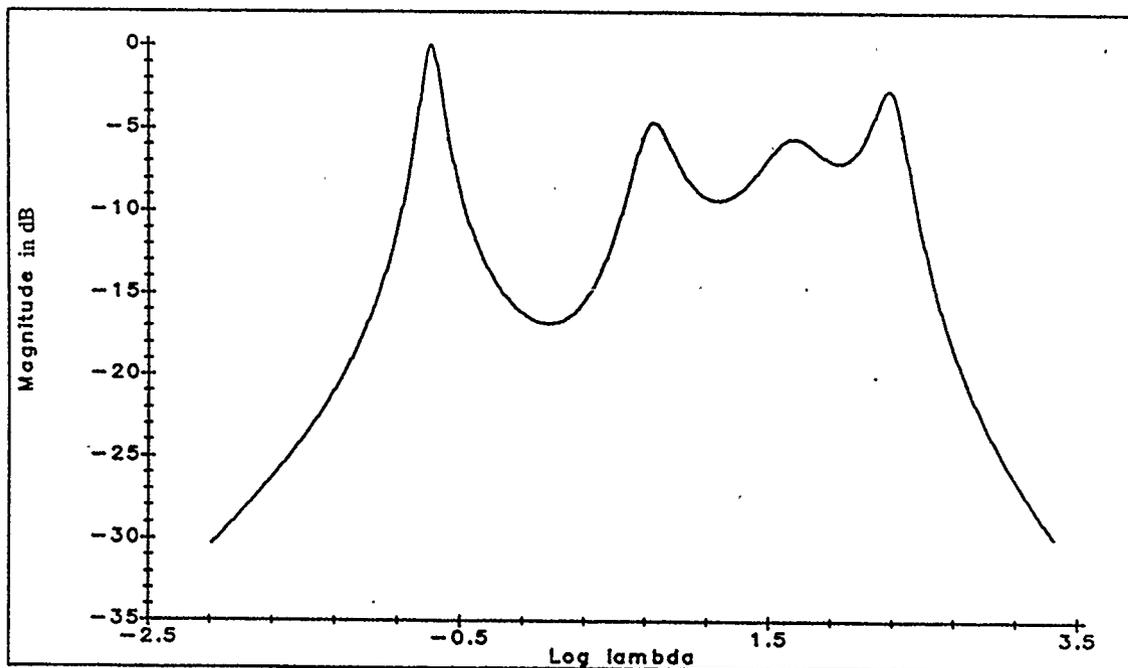


Figure 4.12 Estimate of the distribution function for $\underline{\Lambda} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$.

Next the noise standard deviation is increased to 10^{-3} . As seen from figure 4.13, with the transformed data fitted with 30 splines, only three signal components are detected with exponents 0.5609, 2.7396 and 8.7064. Interestingly enough, when the data is fitted with 25 splines the four components became resolvable with exponents as 0.4724, 1.7010, 4.2432 and 9.6371.

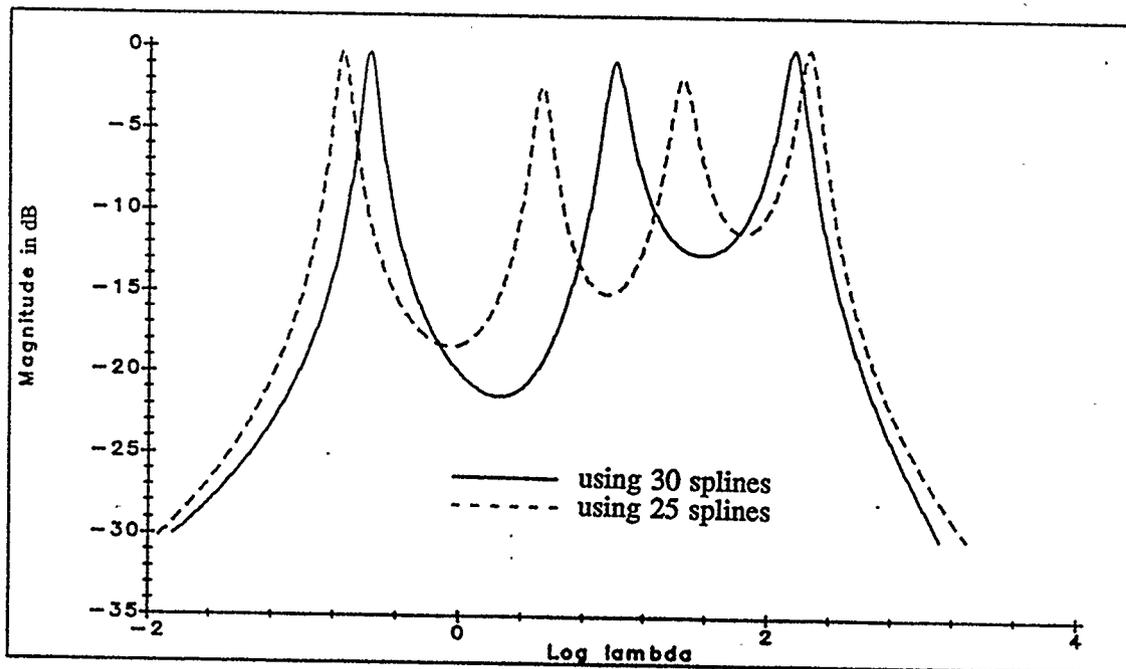


Figure 4.13 Estimate of the distribution function for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

The results again emphasize the significance of the optimum choice of the nodes. It is quite possible that under some other better choice of nodes all four

components can be more accurately resolved even at this noise level.

4.2.1.5 Effect of Alpha and Sampling Frequency

Salami [24] has reported that using $\alpha = 0.5$ and the sampling frequency of 2 Hz. produced best results for the parameter estimates. To test the hypothesis the test signal in Equation (4.4) is again considered with a noise standard deviation of 10^{-3} . The transformed data was fitted with 30 splines. The method failed to estimate both the number of components and the exponents correctly.

In a separate study α was maintained as 0.5. The transformed data was fitted with 25 splines and resampled at 4.0 Hz. The distribution functions obtained in both cases are shown in figure 4.14. Both of these methods failed to estimate both the number of components and the exponents.

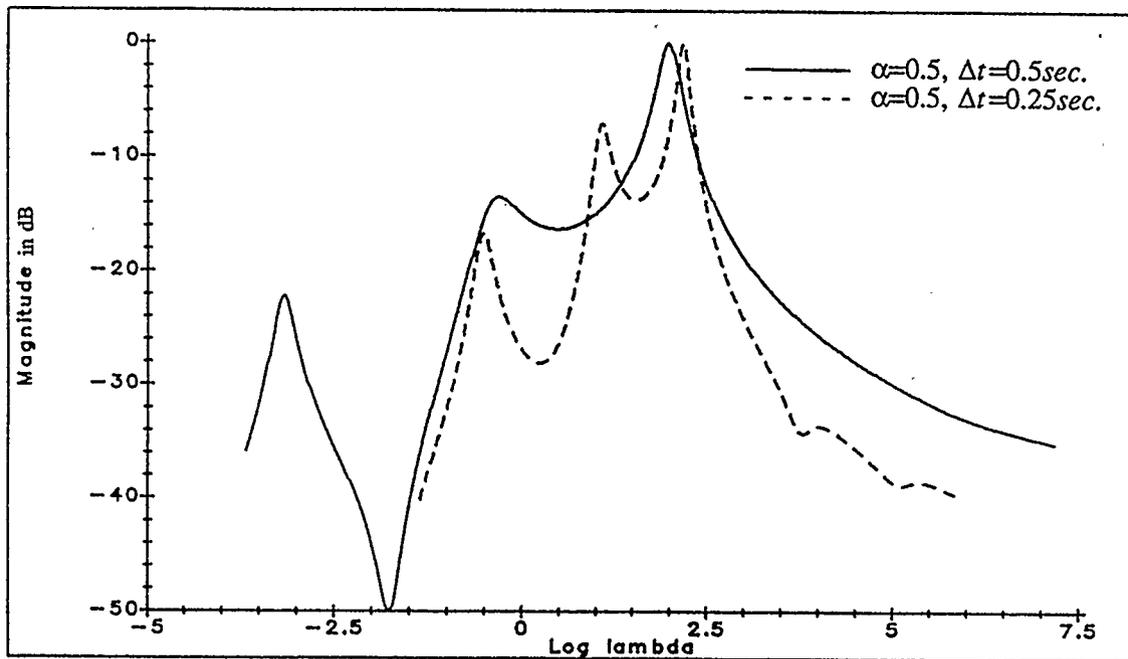


Figure 4.14 Effect of alpha and sampling on the distribution function.

4.2.2 The Total Least Squares (TLS) Method (An Algebraic Method)

This technique of analysis belongs to the class of time-domain methods and models the data, the discrete samples of the multi-exponential signal, as an AR process. In this respect this method can also be classified as a parametric method.

It is then easy to show, [30], that the polynomial

$$A(z) = 1 + \sum_{i=1}^{pe} a_i z^{-i} \quad (4.7)$$

where pe is the predictor order and a_i 's are the predictor coefficients, will have M , the number of components, real roots given by $e^{-\lambda_i \Delta t}$ in the interval 0 to 1 on the real axis and $(pe - M)$ roots off this interval and inside the unit circle in the complex plane. Thus from the location of the roots the parameters M and λ_i can be estimated using this method. Once these parameters have been estimated, the magnitudes of the exponentials, A_i 's, can be estimated easily using any least squares method.

Though it has been reported [30] that for real exponentials the predictor order of $\approx \frac{N}{3}$, where N is the number of data points, produces best results, it is found that using predictor order of $\frac{N}{6}$ in no way affects the detection of the signal components. Consequently, through out this section the data is modeled as a $\frac{N}{6}$ order AR process. This not only speeds up the computation of the AR coefficients using the SVD but also reduces the computation load for the root finding routine.

The AR model coefficients, a_i 's, are derived using equation (2.30) and the roots of the polynomial in (4.7) are then found using a system routine "zcpoly" capable of finding complex roots.

The case studies undertaken have the ratio of $\frac{\lambda_{i+1}}{\lambda_i} = 2$ and includes up to

four exponentials. This encompasses the possibilities of closely-related exponents and also where the number of signal components are large. In all the cases the time interval is chosen from zero to five times the longest time-constant.

Case study 1: The signal considered here is shown in equation (4.1). One hundred data samples are obtained by uniformly sampling the signal between the interval 0.0 to 50.0 seconds. Giving the sampling interval

$$\Delta t \approx 0.50 \text{ sec.}$$

Gaussian noise with a zero mean and a standard deviation (σ_n) of 10^{-4} is then added to the data to simulate additive noise. The data is modeled as a 15th. order AR process. The AR coefficients are derived using equation (2.30) and the roots of the 15th. order polynomial, formed according to the equation (4.7), are plotted in the complex plane in figure 4.15.

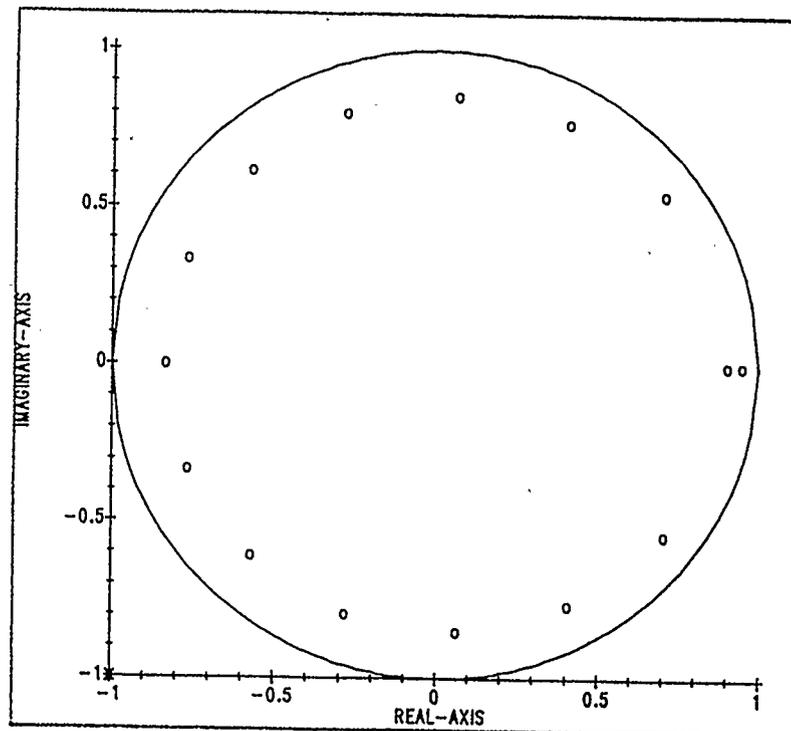


Figure 4.15 Distribution of the zeroes of the AR polynomial for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.1, 0.2]^T$ and $\sigma_n = 10^{-4}$.

Only two roots corresponding to the two exponentials lie on the real-axis between the interval zero and one. From the location of these roots the exponents λ_1 and λ_2 are estimated as 0.1005 and 0.1998 respectively. Furthermore, it is interesting to note that one real root lies on the negative real-axis while the remaining roots are conjugate pairs and are uniformly distributed inside the unit circle. This situation confirms the theory [36] here and in all the subsequent cases studied in this section.

Case study 2: The test signal is shown in equation (4.2) and is considered here for the reasons outlined in the previous section. Again, 100 data samples are obtained by uniformly sampling the signal between the interval 0 and 5.0 secs. Additive noise is Gaussian with a zero mean and 10^{-4} standard deviation.

The data is fitted with a 15th. order AR model and the roots of the polynomial with AR coefficients as its coefficients are plotted in the complex plane in figure 4.16.

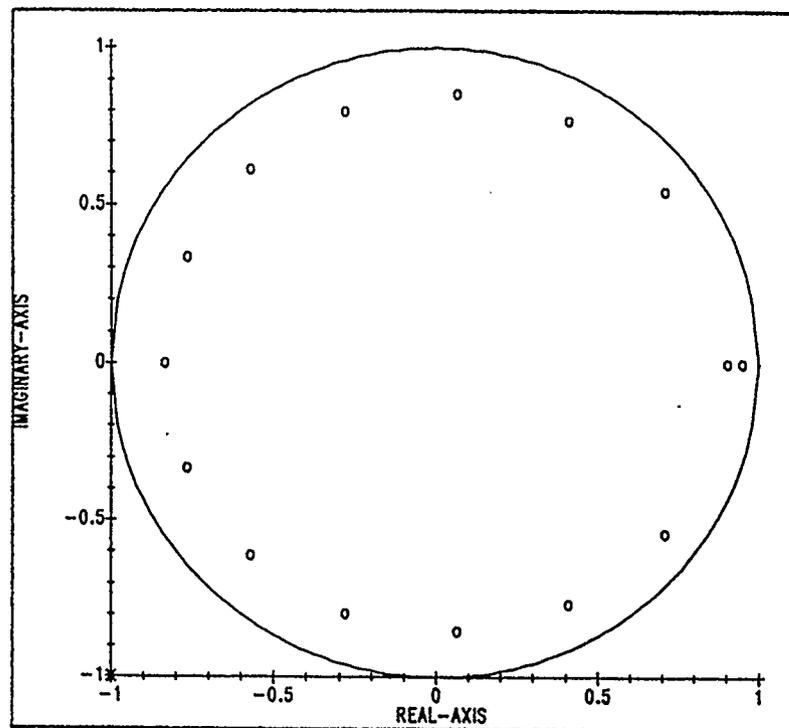


Figure 4.16 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [1.0, 2.0]^T$ and $\sigma_n = 10^{-4}$.

The two roots on the real-axis between the interval zero and one correspond to the two exponentials. The location of the roots gives λ_1 and λ_2 as 1.0004 and 1.9997 respectively.

Case study 3: The test signal considered here is composed of three components and is given in equation (4.3). The object in this and the next case study is to show that the Total Least Squares is a powerful technique for decomposing multi-exponential signals even when the number of components are large ($M > 2$) and the exponents are closely-related. One hundred data samples are acquired with a sampling interval of 0.0505 seconds. Standard deviation of the noise and the sampling time of the signal is maintained as in case study 2. The distribution of the zeroes of the 15th order polynomial in z is shown in the figure 4.17.

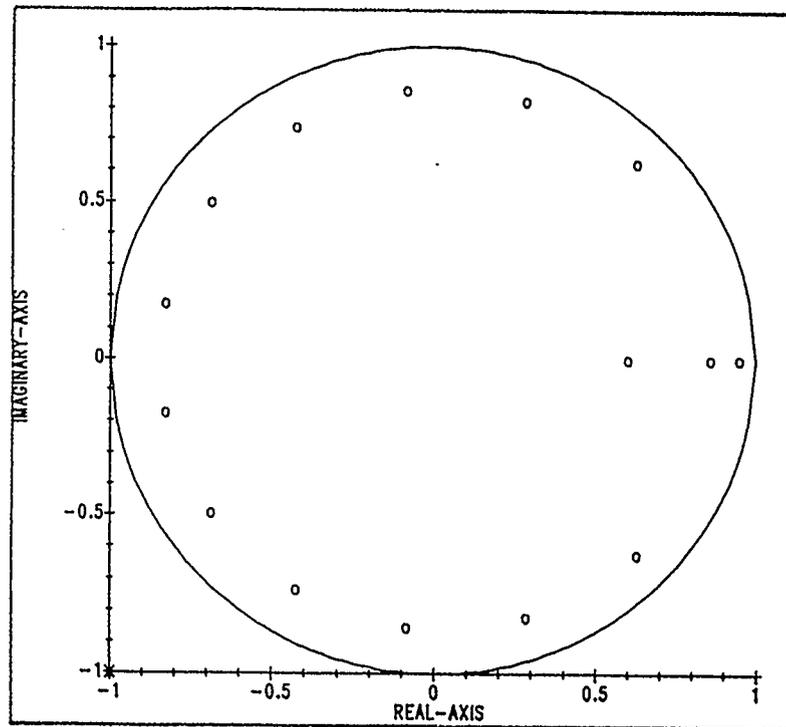


Figure 4.17 Distribution of the zeroes of the AR polynomial for $\underline{\mathbf{A}} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-4}$.

Three zeroes corresponding to the three signal components are located on the real-axis between the interval zero and one, again. The remaining twelve zeroes are distributed uniformly in the complex plane inside the unit circle as conjugate pairs. From the location of the zeroes on the positive real-axis between zero and one the exponents λ_1 , λ_2 and λ_3 are estimated as 1.0000, 2.9988 and 9.9993 respectively.

Case study 4: The multi-component signal considered here is shown in equation (4.4) and presents all the worst case scenarios, namely

1. The number of signal components is large, i.e. $M = 4$.
2. Some of the exponents are closely-related, i.e. $\frac{\lambda_4}{\lambda_3} = 2$.
3. The ratio of the largest to the smallest exponent is large, i.e. $\frac{\lambda_4}{\lambda_1} = 20.0$.

Again, one hundred uniform samples are obtained by sampling the signal between the interval zero and ten seconds. Gaussian noise with a standard deviation of 10^{-4} and zero mean is then added and the process is fitted with a 15th order AR model. The distribution of the zeroes of the polynomial with AR model parameters as its coefficients is shown in figure 4.18.

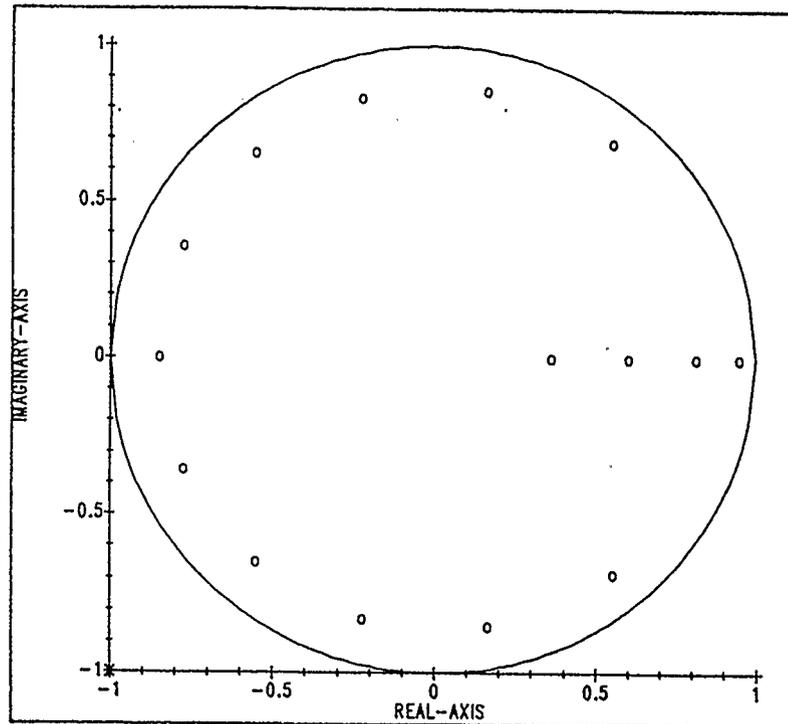


Figure 4.18 Distribution of the zeroes of the AR polynomial for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$.

The four zeroes laying on the real-axis, between the interval zero and one, correctly suggest four signal components and the locations of the zeroes give λ_1 , λ_2 , λ_3 and λ_4 as 0.4999, 1.9938, 4.9833 and 9.9867. The ratio of fourth to the fifth singular value is 31.5966 and there after becomes approximately one for the successive values. Hence, the choice of the dominant singular values is quite obvi-

ous.

4.2.2.1 Effect of Noise

As the SNR of the signal decreases the spread of the singular values begins to deteriorate. In other words, the signal begins to smear over the entire space and the signal space becomes less and less distinguishable from the noise space. Under these circumstances the SVD method becomes ineffective as a SNR enhancement technique.

The signal in equation (4.4) is considered again. The standard deviation of the noise is increased from 10^{-4} to 10^{-3} while maintaining all other signal parameters same as in the case study 4. The distribution of zeroes is shown in figure 4.19.

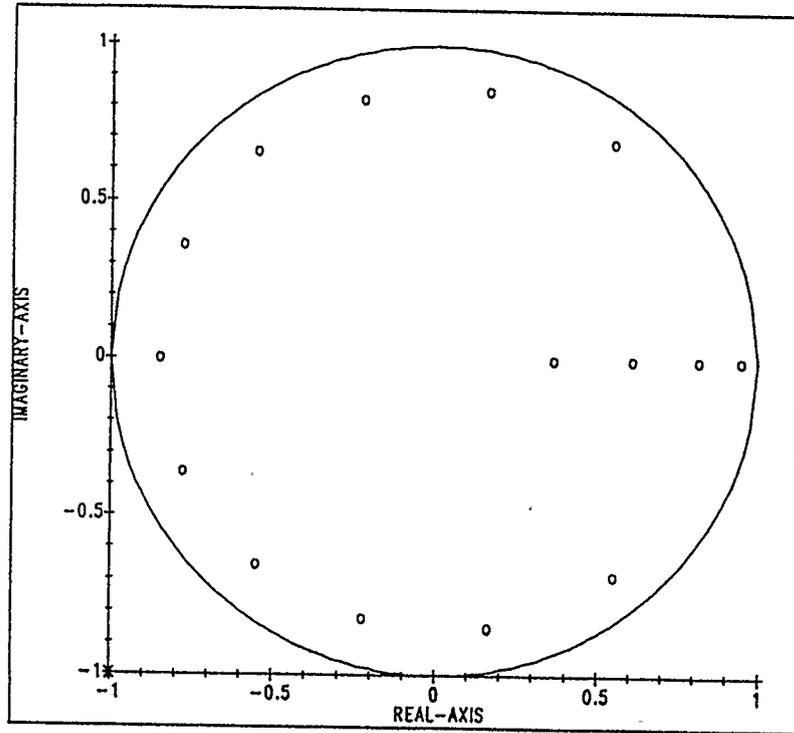


Figure 4.19 Distribution of the zeroes of the AR polynomial for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

The ratio of fourth to fifth singular value is now 3.154 and becomes one there after. Therefore, the four signal components are still distinguishable. The four exponents are accurately estimated as 0.5019, 1.9698, 4.8646 and 9.9285. None of the other techniques considered in this thesis were able to accurately estimate the four exponents in this noise environment.

When the standard deviation of the noise is increased to 10^{-2} only three dominant singular values become distinguishable. The ratio of fourth to fifth singular

value becomes just about unity. Selecting three dominant singular values gives the estimate of the exponents as 0.6038, 2.7153 and 8.6727. The location of the zeroes of the polynomial indicating only three signal components is shown in figure 4.20. No improvement in the estimates was obtained by increasing the predictor order to $\frac{1}{3}N$, or increasing the sampling rate by a factor of two or selecting more than three dominant singular values.

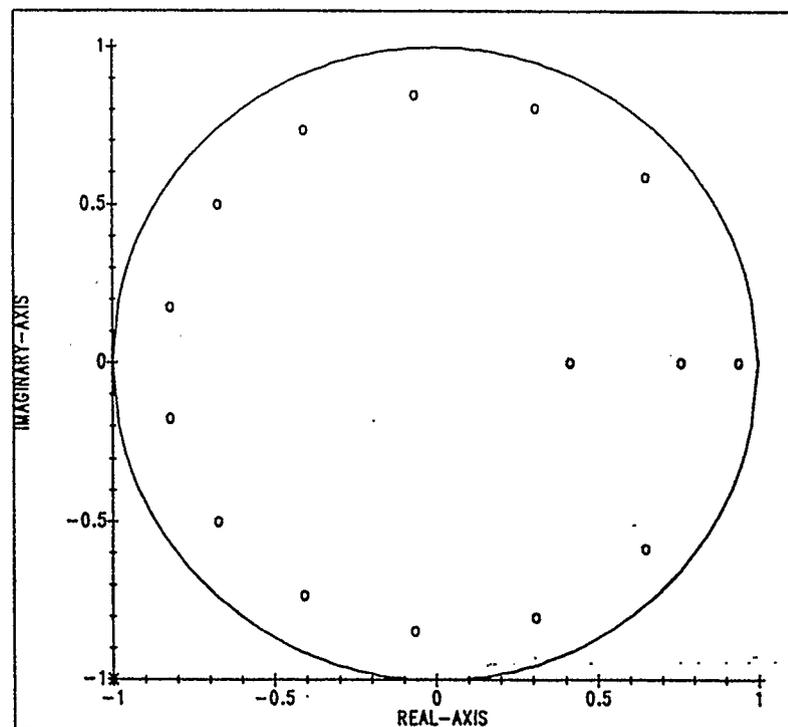


Figure 4.20 Distribution of the zeroes of the AR polynomial for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-2}$.

It is further noticed that the level of noise that can be tolerated is a function of the number of signal components and the closeness of the exponents. To emphasize this point, the signal in equation (4.3) is studied with $\sigma_n = 10^{-2}$. As shown in figure 4.21 the three signal components are still resolvable.

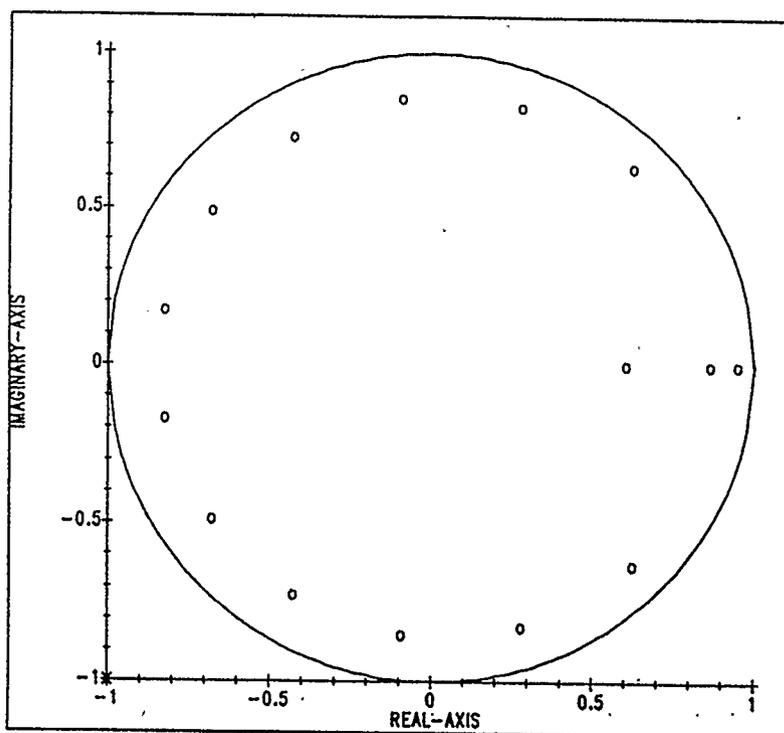


Figure 4.21 Distribution of the zeroes of the AR polynomial for $\underline{\mathbf{A}} = \underline{\boldsymbol{\lambda}} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-2}$.

From the location of the zeroes in the interval zero and one the exponents are estimated as 1.0058, 2.8835 and 9.9358. With noise level increased to 10^{-1} only two components were resolved with both exponents being grossly inaccurate. No

other combination of the predictor order, the sampling rate and the number of dominant singular values was able to accurately estimate the number of signal components.

4.2.2.2 Comments on the Total Least Squares Method

The following important properties of the Total Least Squares method clearly stand out.

1. The roots of $A(z)$ always lie inside the unit circle. Therefore, the predictor determined by this method is always stable.
2. The parameters are estimated directly from the original data with a minimum user input. Hence, the technique is very robust and does not involve nonlinear transformation and interpolation as is the case with the Gardner's method.
3. Unlike the iterative techniques this method of analysis requires neither a prior knowledge of the number of signal components nor the initial parameter estimates.
4. Examination of the singular values spread of the data matrix gives an indication of the number of signal components. It has been found experimentally

that in order to be able to resolve the signal components the ratio $\frac{\sigma_M}{\sigma_{M+1}} \geq 2$,

where σ_M denotes the M^{th} . singular value.

Furthermore, the use of the the SVD in inverting a matrix always ensures a stable inversion method.

4.2.3 The Weighted Total Least Squares Method (An Algebraic Method)

In the case of Total Least Squares the decay rates of the signal components are estimated by evaluating the prediction error filter, $A(z)$, between the interval zero and one. It is, however, possible to obtain the estimate of the signal decay-rates in a manner similar to the transform method by mapping the interval zero to one on to the unit circle. This is accomplished by minimizing the weighted-norm of the predictor coefficients defined by equation (2.32). It is the weighting matrix P in this equation that defines the mapping. Though the derivation of this matrix is not attempted in this assertion, a closed form solution for obtaining the predictor coefficients, if the matrix P is known and is symmetric, is presented in section 2.5.

The solution is validated by showing that if the matrix P is chosen to be an identity matrix, then the solution presented by equation (2.44) converges to the one given by equation (2.30) for the conventional norm-minimization. For this purpose the test signal of equation (4.4) is again considered with a noise standard deviation of 10^{-3} . The data is fitted with 15th. and 20th. order one step forward-predictor and the predictor coefficients determined in the two cases are tabulated in table 4.1.

Table 4.1 Comparison of predictor coefficients derived using the Total Least Squares and the Weighted Total Least Squares ($P = I$) methods.

Predictor Coeff.	Norm-Minimization predictor order		Weighted Norm-Minimization predictor order	
	15	20	15	20
a_0	1.0000	1.0000	1.0000	1.0000
a_1	-0.2419	-0.1249	-0.2419	-0.1249
a_2	-0.2165	-0.1175	-0.2165	-0.1175
a_3	-0.2043	-0.1200	-0.2043	-0.1200
a_4	-0.1637	-0.1125	-0.1637	-0.1125
a_5	-0.1503	-0.1132	-0.1503	-0.1132
a_6	-0.1000	-0.1111	-0.1000	-0.1132
a_7	-0.0521	-0.0952	-0.0521	-0.0952
a_8	-0.0020	-0.0871	-0.0020	-0.0871
a_9	0.0373	-0.0637	0.0373	-0.0637
a_{10}	0.0893	-0.0556	0.0893	-0.0556
a_{11}	0.0996	-0.0277	0.0996	-0.0277
a_{12}	0.0624	-0.0017	0.0624	-0.0017
a_{13}	-0.0089	0.0249	-0.0089	0.0249
a_{14}	-0.1041	0.0450	-0.1041	0.0450
a_{15}	0.0339	0.0685	0.0339	0.0675
a_{16}	-	0.0675	-	0.0675
a_{17}	-	0.0349	-	0.0349
a_{18}	-	-0.0192	-	-0.0192
a_{19}	-	-0.0798	-	-0.0798
a_{20}	-	0.0291	-	0.0291

Thus, the derivation of solution equation (2.44) for the Weighted

Total Least Squares has been indirectly validated.

4.2.4 Estimating the Amplitudes Using a Least Squares (LS) Method

Once the exponents have been estimated by any of the three aforementioned techniques, the amplitudes associated with these exponents can easily be estimated by a least squares method. One such method, namely the Prony's extended LS method has been implemented to estimate the amplitudes. Again, the SVD is used to obtain a LS fit of the data to obtain the amplitude vector, \underline{A} .

The two signals considered here are given by equations (4.3) and (4.4) with the maximum tolerable noise levels of $\sigma_n = 10^{-2}$ and $\sigma_n = 10^{-3}$ respectively. The decay rates estimates used were obtained in section 4.2.2.1. The estimates of the amplitudes obtained are tabulated in tables 4.2 and 4.3 respectively.

Table 4.2 Least Squares estimates of the amplitudes for
 $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-2}$.

Exponents λ_i	Amplitude estimates (A_i) for different number of data points, N				
	N = 100	N = 30	N = 20	N = 10	N = 3
1.0058	0.9714	0.9760	0.9755	1.0096	-0.3214
2.8835	2.9156	2.9065	2.9074	2.8474	4.7283
9.9358	10.1122	10.1176	10.1172	10.1460	9.5870

Table 4.3 Least Squares estimates of the amplitudes for
 $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

Exponents λ_i	Amplitude estimates (A_i) for different number of data points, N				
	N = 100	N = 30	N = 20	N = 10	N = 4
0.5019	0.5007	0.5011	0.5013	0.5080	0.4745
1.9697	1.9154	1.9140	1.9131	1.8919	1.9551
4.8646	4.8611	4.8632	4.8646	4.8906	4.8563
9.9285	10.2226	10.2215	10.2208	10.2090	10.2134

It is seen that in both cases with $N = 30$ the estimates of the amplitudes obtained are reasonably accurate even at these high noise levels.

In the next two sections the results from two different methods namely, the nonlinear least squares and the maximum likelihood methods are compared. For the reasons outlined in Chapter 2 these methods are classified as iterative methods and require the knowledge of the number of signal components, M , and the initial estimates of the signal parameters, A_i 's and λ_i 's. Because of the inherent nature of iterative methods they have the tendency of converging to a local solution or

diverging if the initial estimates are in large errors.

4.2.5 The Nonlinear Least Squares (NLS) Method

Given the time series $x(n)$; $1 \leq n \leq N$ the $2M$ -dimensioned objective function is formed as

$$f(A_1, \dots, A_M, \lambda_1, \dots, \lambda_M) = \sum_{j=1}^N [x_j - \sum_{i=1}^M A_i e^{-\lambda_i t_j}]^2$$

and its minimum is sought.

The technique, justifiably, derives its name "Nonlinear Least Squares" from the facts that

1. the function defined above is nonlinear in the decay rates, λ_i 's, and
2. minimizing the function in a $2M$ vector space is equivalent to performing a least squares fit of the data with respect to the signal parameters.

From the knowledge of the number of signal components (M) and the initial estimates of the $2M$ signal parameters, the function minimum is searched using Powell's method, outlined in Chapter 2, in $2M$ dimensions. The convergence, and for that matter convergence to the global minimum depends on the accuracy of the initial parameter estimates.

Two separate case studies are undertaken to analyze the effect of the noise and the effect of the errors in the initial estimates on the convergence of the algo-

rithm.

Case study 1: The test signal given in equation (4.3) and shown below is considered again.

$$x(\tau) = 1.0 e^{-1.0 \tau} + 3.0 e^{-3.0 \tau} + 10.0 e^{-10.0 \tau} + n(\tau) \quad (4.8)$$

Fifty data samples are obtained by uniformly sampling the signal between the interval zero and 10.0 seconds. Gaussian noise with a zero mean and a standard deviation (σ_n) of 10^{-4} is added to the samples. The algorithm is started with 5%, 10% and 20% errors in the initial estimates. Table 4.4 shows the estimated signal component parameters.

Table 4.4 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-4}$.

Parameters	5% Error		10% Error		20% Error	
	Seed	Estimates	Seed	Estimates	Seed	Estimates
A_1	0.95	1.0022	0.9	1.0022	0.8	1.0022
λ_1	1.05	1.0006	1.1	1.0006	1.2	1.0006
A_2	2.85	3.0007	2.7	3.0007	2.4	3.0007
λ_2	3.15	3.0044	3.3	3.0044	3.6	3.0044
A_3	9.5	9.9970	9.0	9.9970	8.0	9.9970
λ_3	10.5	10.0002	11.0	10.0002	12.0	10.0002

To study the effect of noise on the estimates, the standard deviation of the noise is increased to 10^{-3} . The seed values and the parameter estimates obtained

are tabulated in table 4.5.

Table 4.5 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

Parameters	5% Error		10% Error		20% Error	
	Seed	Estimates	Seed	Estimates	Seed	Estimates
A_1	0.95	1.0222	0.9	1.0222	0.8	1.0223
λ_1	1.05	1.0057	1.1	1.0057	1.2	1.0057
A_2	2.85	3.0091	2.7	3.0091	2.4	3.0092
λ_2	3.15	3.0445	3.3	3.0445	3.6	3.0447
A_3	9.5	9.9681	9.0	9.9680	8.0	9.9679
λ_3	10.5	10.0032	11.0	10.0032	12.0	10.0033

Case study 2: The test signal in this case consists of four exponentials with closely-related exponents and is given as

$$x(\tau) = 0.5 e^{-0.5 \tau} + 2.0 e^{-2.0 \tau} + 5.0 e^{-5.0 \tau} + 10.0 e^{-10.0 \tau} + n(\tau). \quad (4.9)$$

One hundred sample values are obtained by uniformly sampling the signal between the interval 0.0 and 10.0 seconds. Additive Gaussian noise has a zero mean and a standard deviation of 10^{-4} . The simulation results are shown in table 4.6.

Table 4.6 Parameter estimates obtained using NLS method for
 $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$.

Parameters	5% Error		10% Error		20% Error	
	Seed	Estimates	Seed	Estimates	Seed	Estimates
A_1	0.475	0.4991	0.45	0.4991	0.40	0.5343
λ_1	0.525	0.4997	0.55	0.4997	0.60	0.5138
A_2	1.90	1.9850	1.80	3.9850	1.60	2.6782
λ_2	2.10	1.9930	2.20	1.9930	2.40	2.2659
A_3	4.75	4.9727	4.50	4.9728	4.00	7.1803
λ_3	5.25	4.9754	5.50	4.9754	6.00	6.3878
A_4	9.50	10.0431	9.00	10.0431	8.00	7.1074
λ_4	10.50	9.9852	11.00	9.9852	12.00	10.9009

For the noise standard deviation levels of 10^{-3} and 10^{-2} the parameter estimates obtained are shown in tables 4.7 and 4.8 respectively.

Table 4.7 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

Parameters	5% Error		10% Error	
	Seed	Estimates	Seed	Estimates
A_1	0.475	0.4996	0.45	0.5315
λ_1	0.525	0.4997	0.55	0.5126
A_2	1.90	2.0111	1.80	2.6288
λ_2	2.10	2.0071	2.20	2.2517
A_3	4.75	4.8653	4.50	6.2846
λ_3	5.25	4.9580	5.50	6.1049
A_4	9.50	10.1239	9.00	8.0561
λ_4	10.00	9.9563	11.00	10.5408

Table 4.8 Parameter estimates obtained using NLS method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-2}$.

Parameters	5% Error		10% Error	
	Seed	Estimates	Seed	Estimates
A_1	0.475	0.5104	0.45	0.5209
λ_1	0.525	0.5030	0.55	0.5075
A_2	1.90	2.6879	1.80	2.8881
λ_2	2.10	2.2522	2.20	2.3253
A_3	4.75	4.5430	4.50	4.9107
λ_3	5.25	5.6100	5.50	6.0035
A_4	9.50	9.7590	9.00	9.1805
λ_4	10.50	9.96880	11.00	10.0899

The parameter estimates in these cases when the seed values were 20% in error had high errors and are, therefore, not included in tables 4.7 and 4.8. As seen from the results the performance of this method deteriorates when

1. the signal components are greater than four,
2. the exponents are closely-related,
3. the noise level is high and
4. when the seed values are grossly inaccurate.

Furthermore, it was found that for signals with more than four components

this technique of analysis becomes computationally inefficient.

4.2.6 The Maximum Likelihood (ML) Method

Given the time series $x(n)$; $1 \leq n \leq N$ the maximum likelihood method estimates the parameter vector \mathbf{g} , defined in Chapter 2, by its most plausible values. In other words the logarithm of the conditional joint probability density function, more commonly known as the log-likelihood function, is maximized with respect to the parameter vector.

If for the reasons outlined in Chapter 2, the data is assumed to follow Poisson distribution then the log-likelihood function is characterized by equation (2.5). Differentiating equation (2.5) with respect to A_i and λ_i , for $1 \leq i \leq M$, gives a set of $2M$ nonlinear equations in $2M$ unknowns, M being the number of signal components. This set of equations is solved iteratively using Newton-Raphson method.

Two different case studies are performed with the test signals same as in the case of the nonlinear least squares method.

Case study 1: The test signal considered is given in equation (4.8) and consists of three signal components. One hundred data samples are obtained by uniformly sampling the data between the interval 0.0 and 5.0 seconds. The additive noise is again Gaussian with a zero mean and a standard deviation of 10^{-4} . The seed values used for the initial parameters estimates are 5%, 10% and 20% in

error. The parameter estimates obtained are shown in table 4.9.

Table 4.9 Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-4}$.

Parameters	-5% Error		+10% Error		+20% Error	
	Seed	Estimates	Seed	Estimates	Seed	Estimates
A_1	0.95	0.9984	1.10	0.9984	1.20	0.9984
λ_1	0.95	0.9995	1.10	0.9995	1.20	0.9995
A_2	2.85	2.9991	3.30	2.9991	3.60	2.9991
λ_2	2.85	2.9968	3.30	2.9968	3.60	2.9968
A_3	9.50	10.0025	11.00	10.0025	12.00	10.0025
λ_3	9.50	9.9988	11.00	9.9988	12.00	9.9988

The standard deviation of the noise is next increased from 10^{-4} to 10^{-3} while maintaining all other parameters to be the same. The estimates obtained in this case are listed in table 4.10.a.

Table 4.10.a Parameter estimates obtained using ML method for
 $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

Parameters	-5% Error		+10% Error		+20% Error	
	Seed	Estimates	Seed	Estimates	Seed	Estimates
A_1	0.95	0.9838	1.10	0.9838	1.20	0.9838
λ_1	0.95	0.9952	1.10	0.9952	1.20	0.9952
A_2	2.85	2.9918	3.30	2.9918	3.60	2.9918
λ_2	2.85	2.9690	3.30	2.9690	3.60	2.9690
A_3	9.50	10.0245	11.00	10.0245	12.00	10.0246
λ_3	9.50	9.9883	11.00	9.9883	12.00	9.9883

Parameters estimates obtained using different seed values while maintaining the same noise level are shown in table 4.10.b.

Table 4.10.b Parameter estimates obtained using ML method for
 $\underline{A} = \underline{\lambda} = [1.0, 3.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

Parameters	+5% Error		-10% Error		-20% Error	
	Seed	Estimates	Seed	Estimates	Seed	Estimates
A_1	1.05	0.9838	0.90	1.5125	0.80	0.9838
λ_1	1.05	0.9952	0.90	-3.6846	0.80	0.9953
A_2	3.15	2.9919	2.70	-2.5118	2.40	2.9918
λ_2	3.15	2.9690	2.70	-3.7192	2.40	2.9690
A_3	10.50	10.0245	9.00	3.2326	8.00	10.0245
λ_3	10.50	9.9883	9.00	-2.0620	8.00	9.9883

Case study 2: The signal considered here consists of four components and is given by equation (4.9). One hundred data samples are obtained by uniformly sampling the signal between the interval 0.0 and 10.0 seconds. The additive Gaussian noise has a zero mean and 10^{-4} standard deviation. With seed values 20% in error the algorithm did not converge even after 50 iterations. Consequently, Table 4.11 shows parameter estimates only when seed values are in 5% and 10% error.

Table 4.11 Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-4}$.

Parameters	+5% Error		+10% Error	
	Seed	Estimates	Seed	Estimates
A_1	0.525	0.4987	0.55	0.9596
λ_1	0.525	0.4996	0.55	0.6421
A_2	2.10	1.9774	2.20	3.8303
λ_2	2.10	1.9898	2.20	6.1190
A_3	5.25	4.8653	5.50	-0.0142
λ_3	5.25	4.9619	5.50	6.1190
A_4	10.50	10.0658	11.00	11.9052
λ_4	10.50	9.9780	11.00	6.1190

With the noise level increased to 10^{-3} convergence was obtained only for the case with seed values in 5% error as shown in Table 4.12.

Table 4.12 Parameter estimates obtained using ML method for $\underline{A} = \underline{\lambda} = [0.5, 2.0, 5.0, 10.0]^T$ and $\sigma_n = 10^{-3}$.

Parameters	+5% Error		-5% Error	
	Seed	Estimates	Seed	Estimates
A_1	0.525	0.4867	0.475	0.4867
λ_1	0.525	0.4960	0.475	0.4960
A_2	2.10	1.7812	1.90	1.7813
λ_2	2.10	1.8976	1.90	1.8977
A_3	5.25	4.6639	4.75	4.6643
λ_3	5.25	4.6521	4.75	4.6524
A_4	10.50	10.5673	9.50	10.5668
λ_4	10.50	9.8096	9.50	9.8098

For a noise level with standard deviation greater than 10^{-3} either no convergence is achieved or the estimates obtained are quite inaccurate. Furthermore, the process of evaluating the error vector $\Delta \underline{x}$ in equation (2.10) may result in the matrix \mathbf{A} being singular for some given initial conditions. In this respect the maximum likelihood method is found to be numerically less stable than the nonlinear least squares method.

It is further observed that both the nonlinear least squares and the maximum likelihood methods failed to improve the parameter estimates when the seed values used were those obtained using the Total Least Squares and the Prony's LS tech-

niques.

4.3 DECOMPOSITION OF WOOD-NMR DATA

The distribution and interaction of water in the wood are of great importance in its commercial utilization. To this end, proton nuclear magnetic resonance (H-NMR) imaging of water in wood on the basis of spin-spin relaxation time, T_2 , is the only method available for the direct measurement of the distribution of water in the wood.

Proton NMR yields an accurate measurement of the moisture contents and allows an identification of anatomically different regions of the wood. Consequently, the method finds varied commercial applications in the wood industry, such as Kiln drying of wood, identification of defects and different species of wood for production-line analysis and measurement of the lignin content of pulp.

4.3.1 NMR Data Acquisition

The wood NMR data considered in this section is supplied by a group of researchers, at the University of British Columbia, working on the measurement of moisture contents and the distribution of water in the anatomically different regions of the wood using H-NMR technique [12]. The raw data contains 128 points of pre-baseline and 512 points of post-baseline data. A total of 864 NMR echos are processed and each echo is sampled 4 times in the vicinity of the peak. The first 224 echos are selected by sampling every echo while the next 640 are obtained by

retaining every 4th echo. This results in the raw data files containing a total of 4096 data points.

The nonlinear sampling of the signal ensures that sufficient data points are obtained in the time-interval where the signal is decaying rapidly.

4.3.2 Data Processing and Decimation

The baseline or the constant d.c. offset is removed from the data set by subtracting from each point the average of 512 post-baseline points. The pre-baseline points are discarded and the magnitude of each echo is calculated as the average of the four samples corresponding to the echo. The time of this echo is referenced to the last T-90 degree pulse. All echos whose magnitudes are less than 1.0% of the first echo are discarded.

After the data has been processed as outlined above, it is further decimated so that it is computationally manageable while still being representative of the original signal. The reduced data set is obtained by taking the first 200 point, plus additional 100 points by decimating the remaining data points by two and decimating whatever remains by four. In this way, a total of 416 points are obtained for the

final analysis.

4.3.3 Parameter Estimation

If the fast T_2 , spin-spin relaxation time, component is identified with water in and on the cell wall fibrils, the medium component is attributed to the water in the ray and latewood tracheid lumens, and the long component to the water in the earlywood tracheid lumens. Then the Wood-NMR decay curve is modeled [12] as

$$Y = A e^{-Bt} + C e^{-Dt} + E e^{-Ft} \quad (4.10)$$

where B , D and F are the inverses of the three T_2 values for the three components and A , C and E are their respective populations. The estimates of these parameters have been used by the aforementioned researchers in estimating the moisture contents at different stages of Kiln drying of wood, defect identification and species determination.

Two case studies are carried out in this section to compare the performance of the Transform and the Total Least Squares methods. The parameter estimation techniques using the nonlinear least squares and the maximum likelihood approaches are omitted in this section because they are both iterative techniques and require initial parameters estimates which are not available under the current circumstances.

Case study 1: The data file 'dope1lms' is considered for this case study. The Fourier transform of the distribution function obtained according to equation (3.11) is shown in figure 4.22. It is apparent, in this case, that even the choice of the "Good portion" of the deconvolved data is not always so obvious. The first local minimum occurs at the 11th. point but the global or absolute minimum occurs at the 21st. point.

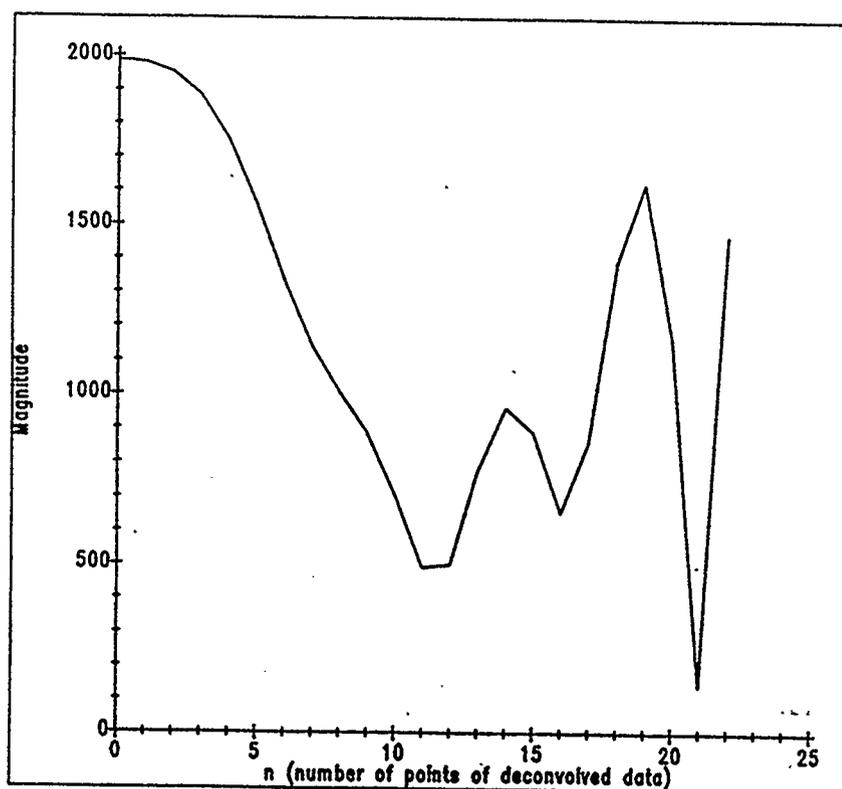


Figure 4.22 Deconvolved data obtained using file 'dope1lms'.

Selecting "Good portion" of the data as 21 ($2 \times 10 + 1$) points, predictor order 17 and three dominant singular values the two exponents are estimated as 10.0996

and 41.5360.

The distribution function estimate obtained using 41 ($2 \times 20 + 1$) points, 30th order predictor and 6 dominant singular values is shown in figure 4.23.

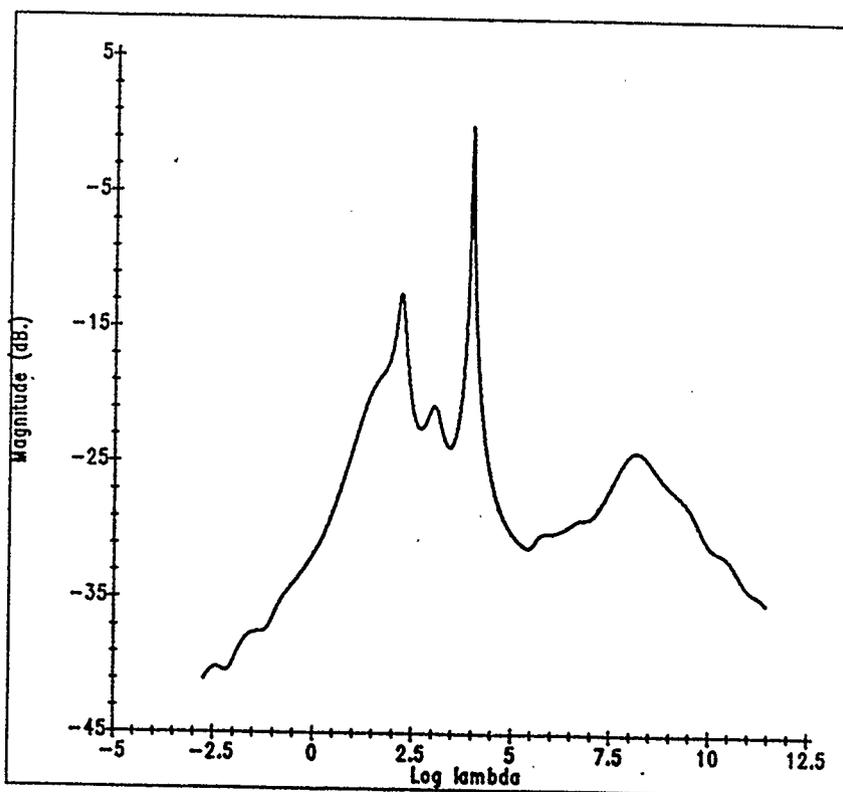


Figure 4.23 Estimate of the parameter distribution function for 'dope11ms'.

The exponents are estimated as 8.9129, 20.4017 and 52.9192. In all of the cases considered so far, uniform samples of the transformed data are obtained by fitting the data with 30 cubic splines. However, when the data is fitted with 35 cubic splines, the "Good portion" of the deconvolved data is selected as 23 points, predictor order is 17 and 7 singular values are selected, the exponents are estimated

as 7.2179, 15.6426 and 49.3254. Thus, again, showing a strong dependence of the estimates on the number of cubic splines and hence the selection of nodes.

Uniform samples of the decimated data, required for the Total Least Squares method, are acquired by fitting the data with 30 cubic splines and resampling at 100 Hz. Basically, the decay rate of the fastest decaying component dictates the sampling frequency while the decay rate of the slowest decaying component dictates the sampling time. These relationships are such that at a given sampling frequency one must obtain a sufficient number of samples of the fastest decaying component and the sampling time, for best results, must be close to ten times the longest time constant i.e. the reciprocal of smallest decay rate. Out of 113 points so obtained only the first 100 are considered and the predictor order is chosen to be 15. The zeroes of the AR polynomial are plotted in the complex plane in figure 4.24.

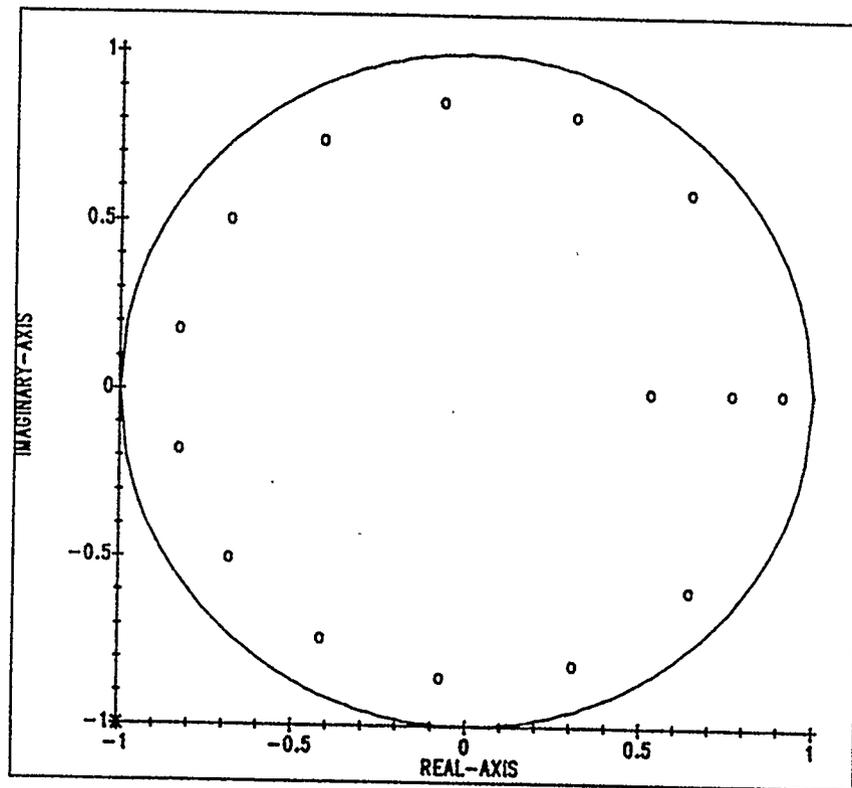


Figure 4.24 Distribution of the zeroes of the AR polynomial for 'dope11ms'.

Three zeroes corresponding to the three signal components lie on the positive real-axis between the interval zero and one. From the location of these zeroes the exponents are estimated as 9.4184, 26.7685 and 63.3322. Furthermore, fitting the data with 35 cubic splines produced the estimates as 9.4116, 26.8034 and 63.7590. This exhibits the robustness of the Total Least Squares method to the selection of the nodes and hence the arithmetic errors.

Case study 2: The data file 'cpmg10' is considered for this case study. The estimated distribution function shown in figure 4.25 is obtained by fitting the

transformed data with 30 cubic splines.

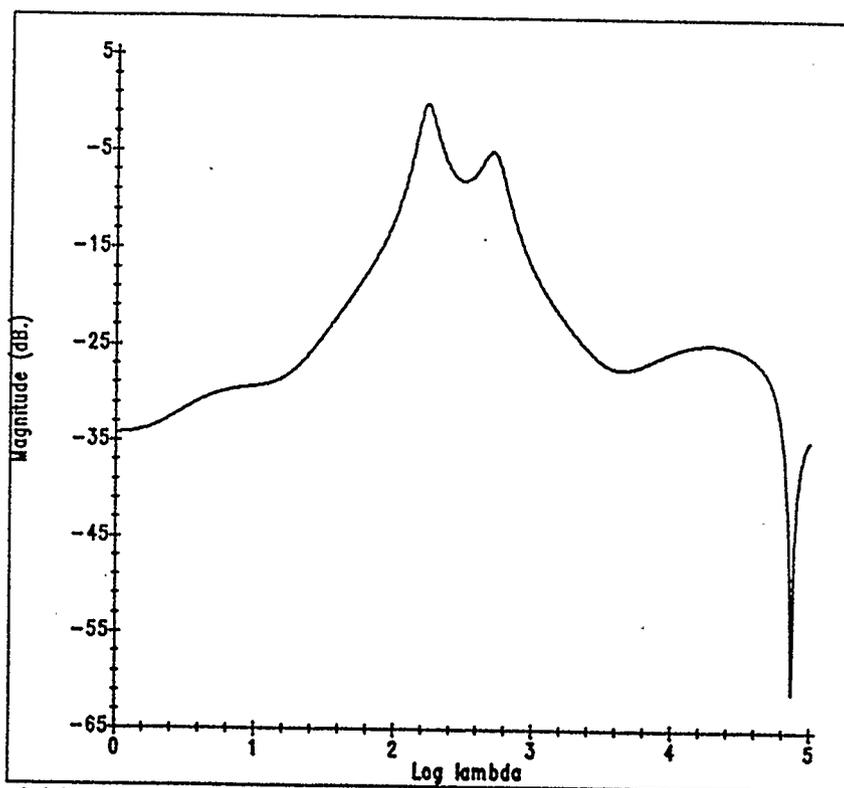


Figure 4.25 Estimate of the parameter distribution function for 'cpmg10'.

Only two signal components are resolved with exponent estimates as 9.3407 and 14.9262.

Using the Total Least Squares method, with uniform samples obtained as outlined in case study 1 with 30 cubic splines, the results are shown in figure 4.26. From the location of the zeroes three signal components are estimated with exponents 9.3828, 17.6188 and 155.5412. Sampling the same data at the same rate but with 40 splines produces the estimates as 9.3809, 17.6489 and 155.9698.

Again, showing minimal sensitivity of this method to the spline fitting.

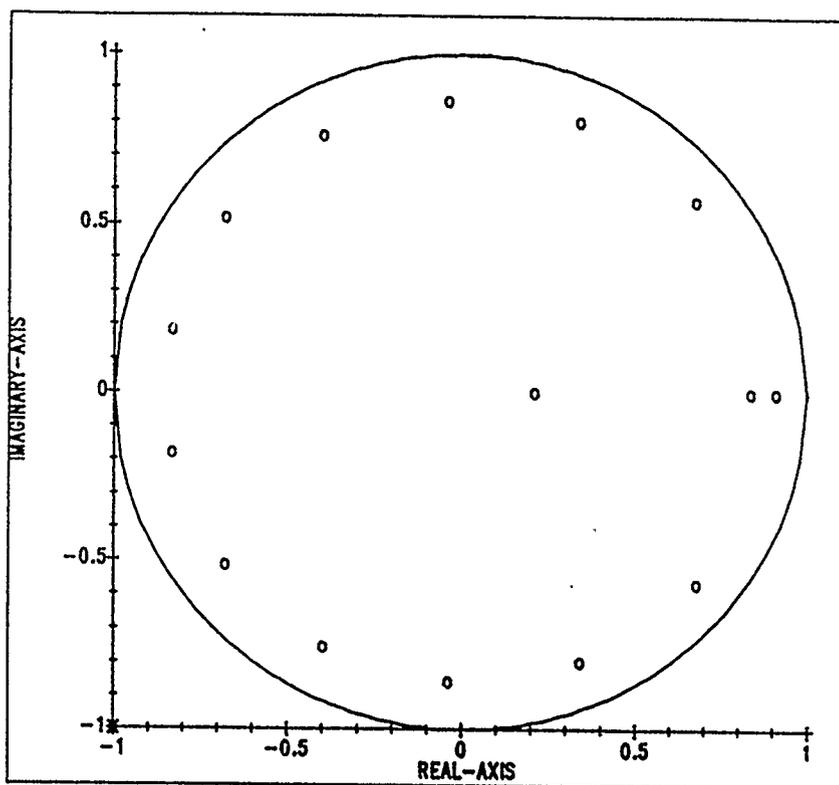


Figure 4.26 Distribution of the zeroes of the AR polynomial for 'cpmg10'.

4.3.4 An Alternate Approach to Selecting the Dominant Singular Values

Theoretically, the number of nonzero singular values of a data matrix is equal to the column rank and to the number of signal components in the case of multi-exponential signal. Therefore, in the absence of noise selecting the number of dominant singular values using a simple ratio test is quite trivial. It may be noted that l_2 norm criterion, discussed in this Chapter, is also very helpful in the presence of

moderate levels of additive noise.

However, when the amplitudes of the exponentials are excessively large, as in the present case, it becomes exceedingly difficult to get a clear fix on the dominant singular values. In such cases, the property that singular values corresponding to signal space are more robust to noise than the singular values corresponding noise space [40], can be successfully exploited. This implies that the addition of noise to the data should have the least effect on the dominant singular values. Thus, this test when used along with the other two tests namely, the ratio and the l_2 norm test can provide much better clues to the number of dominant singular values. Table 4.13 shows the effect of zero mean Gaussian noise on the singular values. From this table it becomes clear that there are only 3 dominant singular values.

Table 4.13 Effect of zero mean Gaussian noise on the singular values.

N	Singular values of noise added data		
	S.D. = 0.0	S.D. = 25.0	S.D. = 40.0
1	307245.000	307316.000	307358.000
2	28175.600	28175.500	28177.500
3	3068.380	2958.950	2908.960
4	575.087	640.370	754.750
5	294.926	447.370	684.742
6	140.471	428.114	599.142
7	78.551	372.874	573.193
8	69.119	346.106	548.073
9	44.121	340.967	533.153
10	27.460	329.646	523.046
11	15.473	313.368	499.024
12	7.518	297.471	463.097
13	2.260	280.584	438.460
14	0.984	268.883	423.689
15	0.585	257.371	403.449
16	0.499	216.995	349.426

4.3.5 Estimating the Amplitudes of the Exponentials

Using the decay rate estimates obtained using the Total Least Squares method, the amplitudes estimates are obtained using Prony's LS method. Only the first 50 uniformly sampled data points are used for the least squares fit. The estimates of the amplitudes obtained along with the corresponding exponents used are listed in Tables 4.14 and 4.15

Table 4.14 Amplitude estimates obtained using LS fit (dopellms).

Exponent estimates used. λ_i	Amplitude estimates obtained. A_i
9.4184	10344.37
26.7685	15905.30
63.3322	21246.30

Table 4.15 Amplitude estimates obtained using LS fit (cpmg10).

Exponent estimates used. λ_i	Amplitude estimates obtained. A_i
9.3809	35662.00
17.6489	29512.90
155.9698	42462.70

Synthetic data is generated using these estimates. The parameters of the fit are shown in table 4.16 and the data fit is shown in figures 4.27 and 4.28.

Table 4.16 Parameters of synthetic data fit.

Parameters	File = dopellms	File = cpmg10
Data energy	83002204094.04	444428890810.75
Error energy	897997.06	48298940.71
% Misfit error	0.001082	0.010868

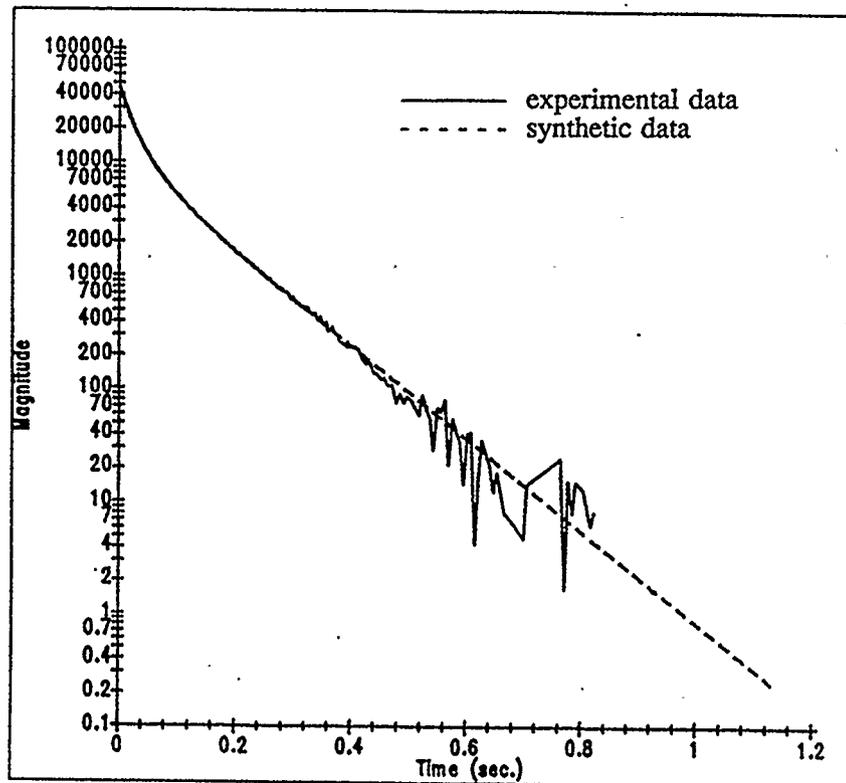
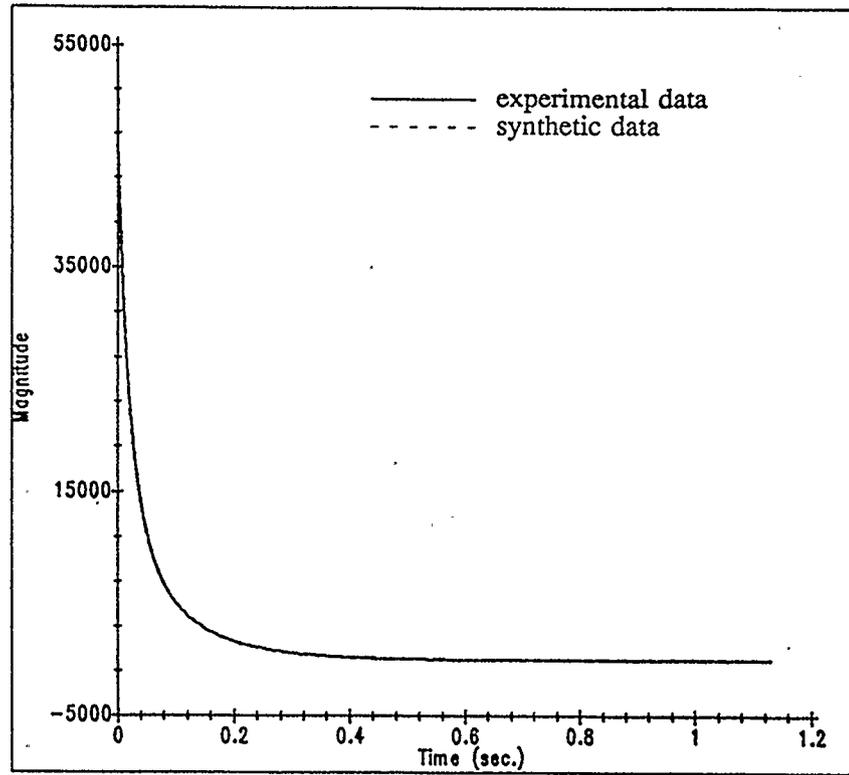


Figure 4.27 Synthetic data fit of 'dopellms'.

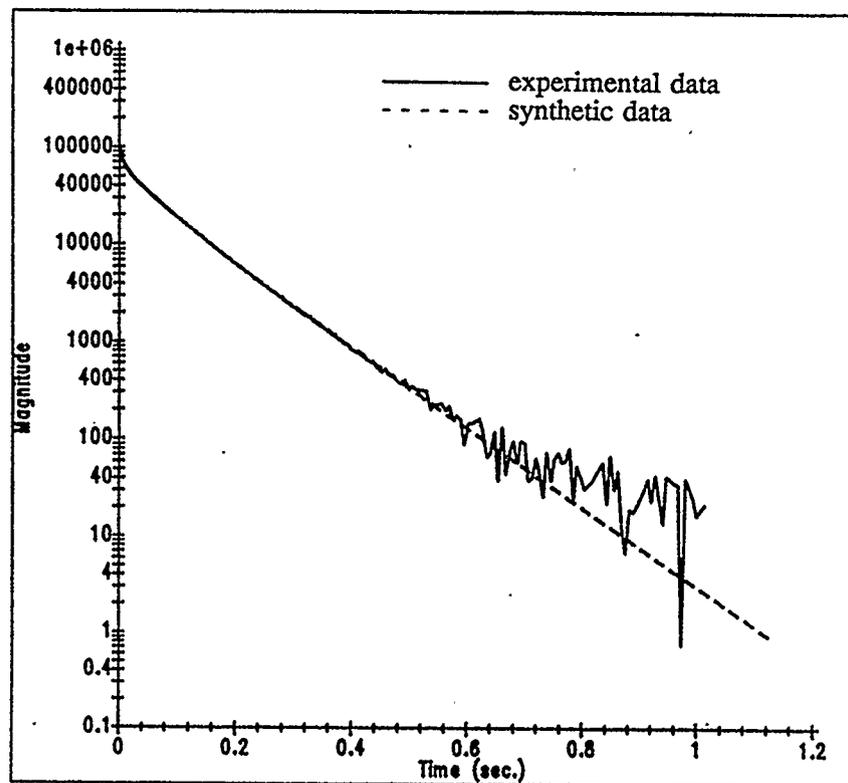
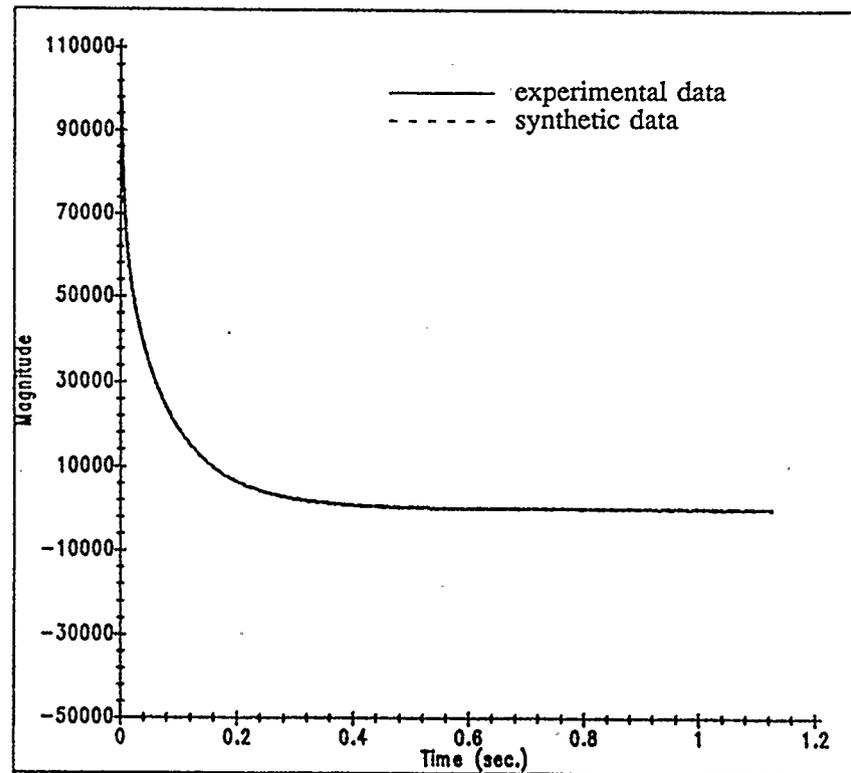


Figure 4.28 Synthetic data fit of 'cpmg10'.

The results show an excellent agreement of the synthetic data with the experimental one suggesting accurate estimate of the signal parameters.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

A novel, efficient and accurate technique of estimating the parameters of a multi-exponential signal is presented. The method requires neither a priori knowledge of the number of signal components nor the initial parameter estimates and belongs to the class of time-domain parametric methods. The significance of this method lies in its ability to enhance the SNR of the noise corrupted data by discarding the noise space and retaining the signal space.

The discrete multi-exponential data is modeled as an AR process and the AR coefficients derived using the singular value decomposition (SVD) of the data matrix have a minimum possible norm resulting in minimum possible prediction error variance. This technique is referred to in the scientific literature as a Total Least Squares method and produces a stable predictor.

A polynomial in AR coefficients determined by this method is formed next. The roots of this polynomial between zero and one indicate the number of signal components and the decay rates are related to the locations of these zeroes. Once the number of components and the decay rates have been estimated, the amplitude estimates are obtained using a linear least squares method.

The performance of this technique is directly compared to the frequency domain Gardner's transform method. Both these methods are tested on the simulated and experimentally obtained Wood-NMR data. In the case of Gardner's method, the uniform samples of the non-linearly transformed data are obtained by the cubic spline interpolation and the "Good portion" of the deconvolved data is modeled as an ARMA process using the Transient Error Method. The AR parameters are obtained using a SVD based least squares method and the transient error terms constitute the MA portion.

The results conclusively indicate a superior performance of the Total Least Squares (TLS) method over the Gardner's transform method, both at low and high noise levels. It is found that the Gardner's method is extremely sensitive to the data values at the tail-end where the SNR of the data is very low. Consequently, at low signal to noise ratio (SNR) the parameter distribution functions obtained selecting different nodes were markedly different. The parameter estimates obtained using the TLS method, on the other hand, exhibited a strong robustness to the choice of nodes and hence also to the arithmetic errors. It is further seen that the simple selection of the "Good portion" of the deconvolved data is not always so straight forward and frequently needs some trial runs.

In the case of simulated data, the choice of the optimal values for the weighting factor α , and the sampling interval, as reported by Salami [24], did not improve the performance of the transform method. In all the cases studied, the new method produced better estimates when the number of signal components were

greater than two and SNR was low.

For the experimental (Wood-NMR) data the transform method failed to accurately estimate the number of signal components and consequently, the estimates of the parameters obtained were in gross error. On the other hand, the Total Least Squares method successfully detected all three signal components and a close agreement of the synthetic data with the experimental data indicates an accurate estimate of the signal parameters.

A predictor order of $N/2$ in the case of the transform method and $N/6$ in the case of the TLS method produced the best estimates. Both methods used the SVD of the data matrix to obtain the AR filter coefficients.

For the sake of completeness the nonlinear least squares and the maximum likelihood methods have also been implemented. Given a priori knowledge of the number of signal components and the initial parameters estimates, these methods iteratively improve upon the initial estimates. The results show that when the number of signal components is less than or equal to three and the additive noise is moderate, these techniques are quite effective.

The TLS method has been extended to the Weighted Total Least Squares method with the objective of mapping the finite interval between zero and one on to the unit circle, thus reducing the complex task of polynomial root finding to performing the DFT. A closed form solution to determine the AR filter coefficients is obtained, if the weighting matrix is known and is symmetric. The determination of this matrix will complement the work in this thesis.

REFERENCES

1. Isenberg, I., Dyson, R.D. and Hanson, R., "Studies on the analysis of fluorescence decay data by the methods of moments", *Biophys. Journal*, Vol. 13, pp. 1090-1115, 1973.
2. Pizer, S.M., Ashare, A.B., Callahan, A.B. and Brownell, G.L., "Fourier transform analysis of tracer data", Chapter 3, *Concepts and models of biomathematics*, Edited by F. Heinmets, Marcel Dekker Inc., N.Y., 1969.
3. Lin, T.K. and Dutt, J.E., "A new method for the analysis of sums of exponential decay curves", *Math. Biosciences*, Vol. 20, pp. 381-391, 1974.
4. Provencher, S.W., "Numerical solution of linear integral equations of the first kind. Calculation of molecular weight distribution from sedimentation equilibrium data", *The journal of Chem. Physics*, Vol. 46, No. 8, pp. 3229-3236, 1967.
5. Markel, J.D. and Gray, A.H., "Linear prediction of speech", New York, Springer-Verlag, 1976.
6. Moffat, D.L. and Mains, R.K., "Problem and Solutions associated with Prony's method for processing transient data", *IEEE Trans. Antennas Propagat.*, Vol. AP-26, pp 174-182, Jan. 1978.
7. Robbins, G.M. and Huang, T.S., "Inverse filtering for linear shift-invariant imaging systems", *Proc. IEEE*, Vol. 60, No.7, pp. 862-872, July 1972.

8. Twomey, S., "The application of Numerical filtering to the solution of integral equations encountered in indirect sensing measurement", J. Franklin Inst., Vol. 274, No. 2, pp. 95-109, Feb. 1965.
9. Bracewell, R.N. and Roberts, J.A., "Aerial smoothing in radio astronomy", Aust. J. Phys., Vol. 7, pp. 615-640, 1954.
10. Neissner, H., "Multiexponential Fitting Methods", Numerical Analysis, ISNM 37, Birkhanser Verlag, Basel UND Stuttgart, pp. 63-75, 1977.
11. Sandor, T., Bleier, A.R., Ruenzel, P.W., Adams, D.F. and Jolesz F.A., "Application of the maximum likelihood principle to separate exponential terms in T_2 relaxation of nuclear magnetic resonance", Magnetic Resonance Imaging, Vol. 6, pp. 27-40, 1988.
12. Menon, R.S., Mackay, A.L., Hailey, J.R.T., Bloom, M., Burgess, A.E. and Swanson, J.S., "An NMR determination of the Physiological water Distribution in Wood during Drying", Journal of Applied Polymer Science, Vol. 33, pp. 1141-1155, 1987.
13. Kirchner, P.P. et al, "The Analysis of Exponential and Non-Exponential Transients in Deep-Level Transient Spectroscopy", J. Applied Physics, 52, (11), Nov. 1981.
14. Monimoto, K.T. et al, "Multi-Exponential Analysis of Deep-Level Transient Spectroscopy", Applied Physics A, 39, 1986.

15. Shapiro, F.R. et al, "The use of LP modeling for the Analysis of Transients from the Semiconductor Defects", J. Appl. Physics, 55 (10), 1984.
16. Simon, W., "Mathematical Techniques for Physiology and Medicine", Academic Press, New York, 1972.
17. Lemaitre, A. and Melange, J.P., "An efficient method for multiexponential fitting with a computer", Computers and Biomedical Res., No. 4, pp. 555-560, 1971.
18. Mancini, P. and Pilo, A., "A computer program for multiexponential fitting by peeling method", Computers and Biomedical Res., Vol. 3, pp. 1-14, 1970.
19. Thomasson, W.M. and Clark Jr., J.W., "Analysis of exponential decay curves: A three-step scheme for computing exponents", Math. Biosci., Vol. 22, pp. 179-195, 1974.
20. Gardner, D.G., Gardner, J.C., Lausch, G. and Meinke, W.W., "Method for the analysis of multicomponent exponential decays", J. Chem. Physics, Vol. 31, pp. 978-986, 1959.
21. Schlsinger, J., "Fit to the exponential data with exponential functions using the fast Fourier transform", Nucl. Inst. Methods, Vol. 106, pp. 503-508, 1973.
22. Cohn-Sfetcu, S., Smith, M.R. and Nichols S.T., "A digital technique for analyzing a class of multicomponent signals", Proc. IEEE, Vol. 63, No. 10, pp. 1460-1467, Oct. 1975

23. Arunachalam, V., "Multicomponent signal analysis", Ph.D. Dissertation, Dept. of Elect. Engg., The University of Calgary, Alta., 1980.
24. Salami, M.J.E., "Application of ARMA models in multicomponent signal analysis", Ph.D. Dissertation, Dept. of Elect. Engg., The University of Calgary, Alta., 1985.
25. Nichols, S.T., Smith, M.R. and Salami, M.J.E., "Application of ARMA modeling to multicomponent signals", IFAC Identification and System Parameter Estimation", York U.K., pp. 1473-1478, 1985.
26. Pagano, M., "Estimation of models of autoregressive signal plus white noise", Ann. Statistics, Vol. 2, pp. 99-108, 1974.
27. Salami, M.J.E., Nichols, S.T. and Smith, M.R., "A SVD-based transient error method for analyzing noisy multicomponent exponential signals", IEEE - International Conference on ASSP, Dallas U.S.A., Vol. 1, pp. 677-680, Apr.6 1987.
28. Kumaresan, R. and Tufts, D., "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Look Like Maximum Likelihood", Proc. IEEE, Vol. 70, No. 9, June 1982.
29. Kumaresan, R. and Tufts, D., "Estimating the Parameters of Exponentially damped Sinusoids and Pole-Zero Modeling in Noise", IEEE Trans. ASSP-30, No. 6, Dec. 1982.

30. Sohie, R.L. and Mirchandani, A., "Accurate estimation of closely spaced, real, decaying exponentials in noise", IEEE - International Conference on ASSP, Dallas U.S.A., Vol. 1, pp. 669-672, Apr.6 1987.
31. Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., "NUMERICAL RECIPES in C: The Art of Scientific Computing", New York, Cambridge University Press, 1988.
32. Cohn-Sfetcu, S., Smith, M.R. and Nichols, S.T., "On the presentation of signal by basis kernels with product argument", Proc. IEEE, Vol. 63, pp 326-327, 1975.
33. Callahan, A.B. and Pizer S.M., "The applicability of Fourier transform analysis to biological compartmental models", Natural Automata and Useful Simulation (H.H. Pattee, E.A. Edelsack, L. Fein and A.B. Callahan Eds.), Spartan, Washington, D.C. 1966.
34. Provencher, S.W., "A Fourier method for the analysis of exponential decay curves", Biophys. J., Vol. 16, pp 27-41, 1976.
35. Smith, P.J., "FITLOS: A FORTRAN PROGRAM FOR FITTING LOW-ORDER POLYNOMIAL SPLINES BY THE METHOD OF LEAST SQUARES", NASA technical note NASA TN D-6401, Aug. 1971.
36. Haykin, S., "Adaptive filter theory", Prentice-Hall, Englewood Cliff, New Jersey - U.S.A., 1986.

37. Childers, D.G., "Modern Spectral Analysis", IEEE Press, New York, 1978.
38. Marple Jr., S.L., "A new autoregressive spectrum analysis algorithm", IEEE Trans. ASSP, Vol. 28, pp. 441-453, Aug. 1980.
39. Barrodale, I. and Erickson, R.E., "Algorithms for least-squares linear prediction and maximum entropy spectral analysis - part 1: Theory", Geophysics, Vol. 45, pp. 420-432, Mar. 1980.
40. Tufts, D.W. and Kumaresan, R.E., "Singular value decomposition and improved frequency estimation using linear prediction", IEEE Trans. ASSP, Vol. 30, pp 671-675, Aug. 1982.
41. Wilkinson, J.H., "The Algebraic Eigenvalue Problem", Oxford, England: Clarendon, 1965.
42. Lawson, C.L. and Hanson, R.J., "Solving Least Squares problems", Englewood Cliffs., N.J., Prentice-Hall, 1974.
43. Cadzow, J.A., Baseghi, B., and Hsu, T., "Singular-value decomposition approach to time series modeling", IEEE Proc., Vol. 130, pp 202-210, April 1983.
44. Kmenta, J., Elements of econometrics, Macmillan, New York, 1971.