Fuzzy Incorporated Noise Compensation Technique for Autoregressive Spectral Estimation

A thesis submitted to the Department of Electrical and Electronic Engineering of
Bangladesh University of Engineering and Technology
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

by
Khawza Iftekhar Uddin Ahmed

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
June 2000

The thesis entitled **"Fuzzy Incorporated Noise Compensation Technique For Autoregressive Spectral Estimation"** submitted by Khawza Iftekhar Uddin Ahmed Roll No.: 9406202F, Session: 1993-94-95 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING.

BOARD OF EXAMINERS

1. _____ 14.6.2000 — **Chairman**
(Dr. Md. Kamrul Hasan)
Assistant Professor
Department of Electrical and
Electronic Engineering, BUET
Dhaka-1000, Bangladesh

2. _____ 14.6.2000 — **Member**
(Dr. Md. Rezwan Khan)
Professor
Department of Electrical and
Electronic Engineering, BUET
Dhaka-1000, Bangladesh

3. _____ 14.6.2000 — **Member** (Ex-officio)
(Dr. Shahidul Islam Khan)
Professor and Head
Department of Electrical and
Electronic Engineering, BUET
Dhaka-1000, Bangladesh

4. _____ 14.06.2000 — **Member** (External)
(Dr. Chowdhury Mofizur Rahman)
Associate Professor and Head
Department of Computer Science
and Engineering, BUET
Dhaka-1000, Bangladesh

Declaration

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Signature of the supervisor

14.6.2000

(Dr. Md. Kamrul Hasan)
Assistant Professor
Department of Electrical and
Electronic Engineering, BUET
Dhaka-1000, Bangladesh

Signature of the candidate

Khawza Iftekhar U. Ahmed

(Khawza Iftekhar Uddin Ahmed)

Acknowledgements

I would like to thank Dr. Md. Kamrul Hasan, Assistant Professor, Dept. of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Bangladesh, for his friendly supervision, constructive suggestions and constant support during this research. I also appreciate his nagging and chiding me for enhancing my research effort. Thanks to Professor Mohammad Ali Choudhury and Dr. Kazi Mujibur Rahman for providing me their helping hand whenever required. I am also grateful to Mr. Fayyaz Khan, Head of the Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Bangladesh and Mr. Golam Toaha Rasul, Assistant Professor of the same university as they endowed me with their advocacy and encouragement for successful completion of this research.

Of course, I am indebted to my parents and my wife for their patience and love. I dedicate this thesis to my parents.

Contents

List of Tables

List of Figures

List of Principal Symbols

$x(t)$	Continuous time signal
$x(n)$	Discrete time AR signal
$R_{xx}(m)$	m-th lag Autocorrelation function of $x(n)$
P	Power density spectrum
P_{AR}	Autoregressive power density spectrum
$u(n)$	Input driving sequence (white noise)
$v(n)$	Additive white noise
$y(n)$	Observed noisy signal
σ_u^2	Variance of the sequence $u(n)$
σ_v^2	Variance of the sequence $v(n)$
a_k	k-th Autoregressive (AR) coefficient
p	True AR model order
q	Higher AR model order
K_k	k-th order reflection coefficient
$E[\cdot]$	Expectation operator
$A(z)$	True order AR polynomial
$\widehat{A}(z)$	Estimated AR polynomial by proposed method
$\widehat{B_i}(z)$	i-th number of estimated AR polynomial by previous method
$C(z)$	Higher order AR polynomial
$\alpha(k)$	Auxiliary variable for noise variance at k-th iteration
$\bar{\sigma}_v^2$	Estimated pseudo noise variance
$\mathbf{x}(n)$	$M \times 1$ output vector
$\mathbf{u}(n)$	$M \times 1$ input vector
$\mathbf{v}(n)$	$M \times 1$ additive noise vector
$\mathbf{y}(n)$	$M \times 1$ noise corrupted observed vector
$\mathbf{\Sigma_v}$	Diagonal covariance matrix of observation noise $\mathbf{v}(n)$
$\mathbf{A}^{(l)}$	l-th $M \times M$ autoregressive parameter matrices

$\mathbf{A}_p(z)$	p-th order AR matrix polynomial
$\mathbf{P}_{AR}$	Mulitchannel power spectral density matrix
$\underline{\mathbf{R}}$	$p \times p$ block Toeplitz matrix
$\mathbf{R}_{xx}(m)$	m-th lag Block entry of $\underline{\mathbf{R}}$, with $M \times M$ dimension
$\mathbf{B}_{pk}(z)$	p-th order matrix polynomial obtained at k-th iteration, can be regarded as inverse filter of $\mathbf{A}_p^{-1}(z)$
$\mathbf{\Sigma}_{\boldsymbol{\alpha}}(k)$	$M \times 1$ auxiliary variable vector for noise covariance vector at k-th iteration
$\bar{\mathbf{\Sigma}}_{\mathbf{d}}(k)$	$M \times 1$, estimated pseudo noise variance vector

Abstract

This thesis deals with the problem of autoregressive (AR) spectral estimation from a finite set of noisy observations without a priori knowledge of additive noise power. For single channel case a joint technique is proposed based on the high-order and true-order AR model fitting to the observed noisy process. The first approach utilizes the uncompensated lattice filter algorithm to estimate the parameters of the over-parameterized AR model and is one-pass. The latter uses the noise compensated low-order Yule-Walker (LOYW) equations to estimate the true-order AR model parameters and is iterative. The desired AR parameters, equivalently the roots, are extracted from the over-parameterized model roots using a root matching technique that utilizes the results obtained from the second approach. This method is highly accurate and is particularly suitable for cases where the system of unknown equations are strongly nonlinear at low SNR and uniqueness of solution from LOYW equations cannot be guaranteed. In addition, an approach based on fuzzy logic is adopted for calculating the step size adaptively with the cost function to reduce the computational time of the iterative total search technique. An extension of the above method for the estimation of multichannel autoregressive power spectrum from a finite set of noisy observations is also proposed. In this case the method is based on the Yule-Walker equations and estimates the autoregressive parameters from a finite set of measured data and then the power spectrum. An inverse filtering technique is used to estimate the observation noise variance and AR parameters simultaneously. Two different algorithms are proposed to estimate the noise variances of all channels. First algorithm is based on the gradient search technique of solving nonlinear equations and the second one is based on fuzzy incorporated iterative search.

Chapter 1

Introduction

1.1 Spectal Estimation: Background

Spectral analysis considers the problem of determining the spectral content (i.e., the distribution of power over frequency) of a time series from a finite set of measurements, by means of either nonparametric or parametric techniques. Spectral analysis finds applications in many diverse fields. In *vibration monitoring*, the spectral content of measured signals give information on the wear and other characteristics of mechanical parts under study. In *economics, meteorology, astronomy* and several other fields, the spectral analysis may reveal "hidden periodicities" in the studied data, which are to be associated with cyclic behavior or recurring processes. In *speech analysis,* spectral models of voice signals are useful in better understanding the speech production process, and in addition can be used for speech synthesis (or compression) and speech recognition. In *radar and sonar systems,* the spectral contents of the received signals provide information on the location of the sources (or targets) situated in the field of view. In *medicine* spectral analysis of various signals measured from a patient, such electrocardiogram (ECG) or electroencephalogram (EEG) signals, can provide useful material for diagnosis. In *seismology,* the spectral analysis of the signals recorded prior to and during seismic event (such as a volcano eruption or an earthquake) gives useful information on the ground movement associated with such events and may help in predicting them. Seismic spectral estimation is also used to predict subsurface geologic structure in gas and oil exploration. In *control systems,* there is a resurging interest in spectral analysis methods as a means of characterizing the dynamical behavior of a given system, and ultimately synthesizing a controller

for that system [1].

The essence of the spectral estimation problem is captured by the following informal formulation.

From a finite record of a wide sense stationary data sequence estimate how the total power is distributed over frequency (1.1)

There are two broad approaches to spectral analysis. One of these derives its basic idea directly from definition (1.1): the studied signal is applied to a bandpass filter with a narrow bandwidth, which is swept through the frequency band of interest and filter output power divided by the filter bandwidth is used as a measure of the spectral contents of the input to the filter. This is essentially what the classical (or nonparametric) methods of spectral analysis do. The second approach to spectral estimation, called the parametric approach, is to postulate a model for the data, which provides a means of parametrizing the spectrum, and to thereby reduce the spectral estimation problem to that of estimating the parameters in the assumed model. Parametric methods may offer more accurate spectral estimates than the nonparametric ones in the cases where the data indeed satisfy the model assumed by the former methods.

The nonparametric method is usually based on fast fourier transform method (Periodgram, Blackman-Tukey). This approach to spectrum analysis is computationally efficient and produces reasonable results for a large class of signal processes. In spite of these advantages, there are several inherent performance limitations of the FFT approach. The most prominent limitation is that of frequency resolution, i.e., the ability to distinguish the spectral responses of two or more signals. The frequency resolution in hertz is roughly the reciprocal of the time interval in seconds over which sampled data is available. A second limitation is due to the implicit windowing of the data that occurs when processing with the FFT. Windowing manifests itself as "leakage" in the spectral domain, i.e., energy in the main lobe of a spectral response "leaks" into the sidelobs, obscuring and distorting other spectral responses that are present. In fact, weak signal spectral responses can be masked by higher sidelobes from stronger spectral responses. Skillful selection of tapered data windows can reduce the sidelobe leakage, but always at the expense of reduced resolution [2].

These two performance limitations of the FFT approach are particularly troublesome when analyzing short data records. Short data records occur frequently in practice because many measured processes are brief in duration or slowly time-varying spectra that may be considered constant only for short record lengths. In, radar, for example, only a few data samples are available from each received radar pulse. In sonar, the motion of targets results in a time-varying spectral response due to Doppler effects.

In an attempt to alleviate the inherent limitations of the FFT approach, many alternative spectral estimation procedures have been proposed. Major techniques are Autoregressive (AR) Power Spectral Density (PSD) estimation, Moving Average (MA) PSD Estimation, Autoregressive Moving Average (ARMA) PSD Estimation, Pisarenko Harmonic Decomposition (PHD), Prony Spectral Line Estimation, Maximum Likelihood Method (MLM).

Some historical perspective is instructive for an appreciation of the basis for modern spectral estimation. The advent of spectrum analysis based on Fourier analysis can be traced to Schuster, who was the first to coin the term "periodogram", [3], [4]. Schuster made attempt to find "hidden periodicities" in the measured data. The next pioneering step was described in Norbert Wiener's classic paper on "generalized harmonic analysis" [5]. This work established the theoretical framework for the treatment of stochastic processes by using a Fourier transform approach. A major result was the introduction of the autocorrelation function with the power spectral density.

Blackman and Tukey, in a classical publication in 1958 [6] provided a practical implementation of Wiener's autocorrelation approach to power spectrum estimation when using sampled data sequences. The method first estimates the auto-correlation lags from the measured data, windows (or tapers) the autocorrelation estimates in an appropriate manner, and then Fourier transforms the windowed lag estimates to obtain the PSD estimate. This approach was the most popular spectral estimation technique until the introduction of the FFT algorithm in 1965, generally credited to Cooley and Tukey [7]. Conventional FFT spectral estimation is based on a Fourier series model of the data, that is, the process is assumed to be composed of a set of harmonically related sinusoids.

Other time series models have been used in nonengineering fields for many

years. Yule [8] and Walker [9] both used AR models to forecast trends in economic time series. Other models have arisen in the statistical and numerical analysis fields. The modern spectral estimators have their roots in these nonengineering fields of time series modeling.

The use of nontraditional spectral estimation techniques in a significant manner began in the 1960's. Parzen [10], in 1968, formally proposed AR spectral estimation. Independently in 1967, Burg [11] introduced the maximum entropy method, motivated by his work with linear prediction filtering in geosismological applications. The one-dimensional MEM (maximum entropy method) was shown formally by Van den Bos [12] to be equivalent to the AR PSD estimator.

1.2 Objective of This Research

The objective of this research is to propose a noise compensation technique for autoregressive (AR) spectral estimation from a finite set of noisy observations with unknown additive noise. Therefore the informal formulation of (1.1) can be modified in this research as stated below.

> *From a finite record of a wide sense stationary white noise corrupted data sequence estimate how the total power of the noise-free data sequence is distributed over frequency* (1.2)

Many researches have focused on this topic for noise-free observations [8]-[16]. In practical cases, however, observations contain additive noise and its effect cannot be neglected. A few methods are available for noise-corrupted observations [17]-[26]. In particular, the correlation-based methods are widely used in estimating the parameters of AR system which are then used for AR spectral estimation. However, if the observations are noisy they require a priori knowledge of the additive noise variance [17], [19]. Although it is possible to avoid the need of the knowledge of noise variance in estimating the AR parameters using the high-order Yule-Walker equations (HOYW), this method is not constrained to be nonsingular [27]. The possible singularity of the auto-correlation matrix leads to a substantial increase in the variance of the AR spectral estimate [17]. Most of the previous methods either assume that the noise variance is known [17], [19] or subtract a suboptimal amount of noise power [28] to compensate for the noise

effect. Determination of the optimal amount of noise power to remove the bias effect of noise is of utmost importance for high-resolution AR spectral estimation from noisy observations.

To combat this problem, Yahagi and Hasan [20] have proposed a method using low-order Yule-Walker equations (LOYW) to compensate for the influence of noise in determining the AR parameters and to estimate the noise variance from a given set of noisy measurements. But the case of strong nonlinearity of LOYW equations at low SNR was not addressed. Unfortunately, due to the inherent nonlinearity of LOYW equations when used for noisy observations with unknown noise variance and system parameters, the solutions found may not be unique depending upon the SNRs and the system characteristics. In these cases, the method fails to estimate the actual solution from the set of multiple solutions. Moreover, the method is computationally expensive as very small constant step size is maintained throughout the total range of search for better accuracy.

In this research a new joint method for estimating the AR parameters and hence the AR spectrum of the true signal (noise free observations) is proposed. The parameters of the over-fitted AR model to the observed noisy process are estimated by using the lattice filter which possesses high-resolution property. It is well known that stability cannot be guaranteed while using the noise compensated lattice filter algorithm[17], [20]. As such, we utilize the uncompensated lattice filter algorithm with the cost paid by adding noise roots to the transfer function of the original AR model. The desired AR parameters of the noise free obersvations are extracted from the over-parameterized AR model parameters by using a root matching technique which utilizes all the solutions obtained from iterative method described by Yahagi and Hasan [20]. As a by product, the proposed scheme has also resolved this problem of multiple solutions case of the LOYW equations which generally arises at low SNR. Furthermore, to reduce greater computing time due to large number of iterations fuzzy modeling of the cost function for step size variation is incorporated in the method of Yahagi and Hasan [20]. A comparative study with the previous method [20] is provided to demonstrate the effiectiveness of the proposed scheme. An extension of the method to the multichannel AR spectrum estimation is also investigated. A new close form equation between the auxiliary noise variances and the estimated noise

variances is also obtained. This close form relation is utilized by two different algorithm to estimate the different noise variances of all different channels. First algorithm is based on gradient search technique of solving non linear equations. Second algorithm is based on fuzzy incorporated iterative search where effect of noise of one channel is considered at a time for minimizing the cost function. Eventually the corresponding AR parameters and their spectra are obtained from the estimated noise variances.

1.3 Organization of the Thesis

In Chapter 2 a brief review of spectral estimation techniques are presented. These include spectral density definition and basics, traditional methods, modeling and the parameter identification approach for rational spectra. The effect of additive white noise on AR spectral estimation is also illustrated in this chapter.

In Chapter 3 a new approach of autoregressive apectrum astimation from observations at the presence of additive white noise is investigated. This approach is based on the high-order and true-order AR model fitting to the observed noisy process. The first approach utilizes the uncompensated lattice filter algorithm to estimate the parameters of the over-parameterized AR model and is one-pass. The latter uses the noise compensated low-order Yule-Walker (LOYW) equations to estimate the true-order AR model parameters and is iterative. The desired AR parameters, equivalently the roots, are extracted from the over-parameterized model roots using a root matching technique that utilizes the results obtained from the second approach. This method is highly accurate and is particularly suitable for cases where the system of unknown equations are strongly nonlinear at low SNR and uniqueness of solution from LOYW equations cannot be guaranteed. In addition, an approach based on fuzzy logic is adopted for calculating the step size adaptively with the cost function to reduce the computational time of the iterative total search technique.

In Chapter 4 the problem of multichannel AR spectrum estimation from noisy observation is addressed. A closed form equation of cost function is obtained which was not derived in [29]. Moreover two different iterative techniques are proposed for estimating multichannel AR spectra. One iterative technique is based on searching in a predefined manner over the error surface and the rest one

is gradient based.

The thesis concludes by presenting an overall discussion on the work and pointing out some unsolved problems for future work in Chapter 5.

Chapter 2

Review of Spectral Estimation Techniques

2.1 Spectral Density Definitions and Basics

2.1.1 Energy spectral density of deterministic signal

Traditional spectrum estimation, as currently implemented using the FFT, is characterized by many tradeoffs in an effort to produce statistically reliable spectral estimates. There are tradeoffs in windowing, time-domain averaging, and frequency-domain averaging, and frequency domain averaging of sampled data obtained from random processes in order to balance the needs to reduce sidelobes, to perform effective ensemble averaging, and to ensure adequate spectral resolution. To summarize the basics of conventional spectrum analysis, the case of a deterministic analog waveform $x(t)$, that is a continuous function of time is considered first. For generality, $x(t)$ is complex valued. If $x(t)$ is absolute integrable, i.e., the signal energy ξ is finite

$$\xi = \int_{-\infty}^{\infty} |x(t)|^2 dt < \infty \tag{2.1}$$

then the continuous Fourier transform (CFT) $X(f)$ of $x(t)$ exists and is given by

$$X(f) = \int_{-\infty}^{\infty} x(t) exp(-j2\pi ft) dt \tag{2.2}$$

Eqn. (2.1) is a sufficient, but not a necessary condition for the existence of Fourier transform [30]. The squared modulus of the Fourier transform is Energy Spectral Density, $\mathcal{E}(f)$ of $x(t)$,

$$\mathcal{E}(f) = |X(f)|^2 \tag{2.3}$$

Parseval's energy theorem, expressed as

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 \tag{2.4}$$

is a statement of the conservation of energy; the energy of the time domain signal is equal to the energy of the frequency domain transform, $\int_{-\infty}^{\infty} |X(f)|^2 df$. Thus $\mathcal{E}(f)$ is an *energy spectral density* ESD in that it represents the distribution of energy as a function of frequency. If the signal $x(t)$ is sampled at equispaced intervals of Δt sec. to produce a discrete sequence $x(n) = x(n\Delta t)$ for $-\infty < n < \infty$ then the sampled sequence can be represented as the product $x(t)$ and an infinite set of equispaced Dirac delta functions $\delta(t)$. The Fourier transform of this product may be written using distribution theory [30], as

$$\begin{aligned} X'(f) &= \int_{-\infty}^{\infty} \left[\sum_{-\infty}^{\infty} x(t)\delta(t - n\Delta t) \right] exp(-j2\pi f t) dt \\ &= \Delta t \sum_{-\infty}^{\infty} x(n) exp(-j2\pi f n \Delta t) \end{aligned} \tag{2.5}$$

Eqn. (2.5) corresponds to a rectangular integration approximation of Eqn. (2.2); the factor Δt ensures conservation of integrated area between Eqns. (2.2) and (2.5) as $\delta t \to 0$. Eqn. (2.5) will be identical in value to the transform $X(f)$ of (2.2) over the interval $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$ Hz, as long as $x(t)$ is band limited and all frequency components are in this interval. Thus the continuous energy spectral density

$$\mathcal{E}'(f) = |X'(f)|^2 \tag{2.6}$$

for data sampled from a band-limited process is identical to that of Eqn. (2.3).

If the data sequence is available from only a finite time window over a) $n = 0$ to $n = N - 1$, and b) the transform is discretized also for N values by taking samples at the frequencies $f = m\Delta f$ for $m = 0, 1, \cdots, N-1$ where $\Delta f = 1/N\Delta t$, then one can develop the familiar discrete Fourier transform (DFT) [30] from Eqn. (2.5),

$$\begin{aligned} X(m) &= \Delta t \sum_{n=0}^{N-1} x(n) exp(-j2\pi m \Delta f n \Delta t) \\ &= \Delta t \sum_{n=0}^{N-1} x(n) exp(-j2\pi mn/N) \\ &\quad \text{for } m = 0, 1, \cdots, N-1 \end{aligned} \tag{2.7}$$

where $X(m)$ is DFT of $x(n)$.

The inverse DFT is given by

$$x(n) = \Delta f \sum_{m=0}^{N-1} X(m) exp(j2\pi mn/N), \text{ for } n = 0, 1, \cdots, N-1 \tag{2.8}$$

and the energy theorem is given by

$$\sum_{n=0}^{N-1} |x(n)|^2 \Delta t = \sum_{n=0}^{N-1} |X(m)|^2 \Delta f \tag{2.9}$$

Both Eqn. (2.7) and its associated inverse transform are cyclic with period N. Thus by using Eqn. (2.7), we have forced a periodic extension to both the discretized data and the discretized transform values, even though the original continuous data may not have been periodic. The discrete ESD may then be defined as

$$\mathcal{E}(m) = |X(m)|^2, \text{ for } m = 0, 1, \ldots, N-1 \tag{2.10}$$

Both the discrete $\mathcal{E}(m)$ and the continuous $\mathcal{E}'(f)$ have been termed *periodogram* spectral estimates. However $\mathcal{E}(m)$ and $\mathcal{E}'(f)$, when evaluated at $f = m/N\Delta t$ for $m = 0, 1, \cdots, N-1$, do not yield identical values. $\mathcal{E}(m)$ is, in effect, a sampled version of a spectrum determined from the convolution of $X(f)$ with the transform of the rectangular window that contains the data samples. Thus the discrete spectrum $\mathcal{E}(m)$ based on a finite data set is a distorted version of the continuous spectrum $\mathcal{E}'(f)$ based on an infinite data set.

2.1.2 Power Spectral Density of Random Signals

In applications most of the signals encountered are indeterministic, i.e., their variation in the future cannot be known exactly. Only the probabilistic statements can be made about the variation. Signals of such category are described by random sequence which consists of an ensemble of possible realizations, each of which has some associated probability of occurrence. When the process $x(t)$ is wide sense stationary (WSS), stochastic process with zero mean rather than a deterministic, finite energy signal a different viewpoint is required. The quantity of interest is the power (time average of energy) distribution with frequency rather than energy distribution as the stochastic processes usually contain infinite energy. As such, the integration of Eqns. (2.2) and (2.7) do not exist in the case

of random signal and the autocorrelation function

$$R_{xx}(\tau) = E\left[x(t+\tau)x^*(t)\right] \tag{2.11}$$

provides the basis for spectrum analysis, rather than the random process itself. Here $E[\cdot]$ denotes the expectation operator which averages over the ensemble of realization. $R_{xx}(\tau)$ depends only on the lag between two samples averaged and $x^*(t)$ is the complex conjugate of $x(t)$. The Wiener-Khinchin theorem gives us power density spectrum PSD $P(f)$ such that

$$P(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) exp(-j2\pi f\tau) d\tau \tag{2.12}$$

To be honest, practically one does not know the statistical autocorrelation function. Therefore one has to assume that the process is ergodic in the first and the second moment which permits the substitution of time averages for ensemble averages. For an ergodic process Eqn. (2.11) becomes

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+\tau)x^*(t)dt \tag{2.13}$$

With the help of Eqn. (2.13) following alternate form of Eqn. (2.12) can be obtained [31]-[33].

$$P(f) = \lim_{T\to\infty} E\left\{ \frac{1}{2T} \left| \int_{-T}^{T} x(t) exp(-j2\pi ft) dt \right|^2 \right\} \tag{2.14}$$

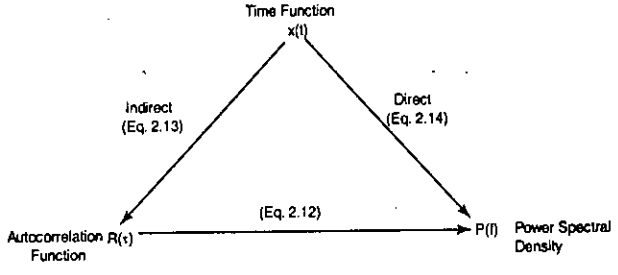

Fig. 2.1: Direct and indirect methods for obtaining PSD

Fig. 2.1 shows the direct and inirect approaches for obtaining PSD of stochastic processes.

2.2 Conventional Methods (Nonparmetric Methods) for Estimating PSD

The indirect approach of PSD estimation using an autocorrelation was introduced by Blackman and Tukey and the spectral estimator is known as *correlogram* . The

other PSD estimator, based on the direct approach using the FFT algorithm is known as periodogram.

When the data sequence is finite, only finite number of discrete autocorrelation function values, or lags are estimated. The spectral estimate proposed by Blackman and Tukey

$$\hat{P}_{BT}(f) = \Delta t \sum_{m=-M}^{M} \hat{R}_{xx}(m) exp(-2j\pi f m \Delta t) \tag{2.15}$$

uses the available autocorrelation lag estimates $\hat{R}_{xx}(m)$, where $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$ and $\hat{}$ denotes an estimate. This spectral estimate is the discrete-time version of the Wiener-Khinchin expression given by Eqn. (2.12). Autocorrelation estimate, using Eqn. (2.13) is the unbiased estimator

$$\begin{aligned} \hat{R}_{xx}(m) &= \frac{1}{N-M} \sum_{n=0}^{N-m-1} x(n+m)x^*(n) \\ &\quad \text{for } m = 0, 1, \cdots, M, \text{ where } M \leq N-1 \end{aligned} \tag{2.16}$$

The negative lag estimates can be determined from the positive lag estimates by the following relation which in accordance with the conjugate symmetric property of the autocorrelation function of a stationary process.

$$\hat{R}_{xx}(-m) = \hat{R}^*_{xx}(m) \tag{2.17}$$

Instead of Eqn. (2.16) following autocorrelation estimate is more justified according to both Jenkins-Watts [31] and Parzen [10],[34].

$$\begin{aligned} \hat{R}'_{xx}(m) &= \frac{1}{N} \sum_{n=0}^{N-m-1} x(n+m)x^*(n) \\ &\quad \text{for } m = 0, 1, \cdots, M, \text{ where } M \leq N-1 \end{aligned} \tag{2.18}$$

The reason behind the justification is, it tends to have less mean square error than Eqn. (2.16) for many finite data sets. $\hat{R}'_{xx}(m)$ is a biased estimator since $E\left[\hat{R}'_{xx}(m)\right] = \left[(N-M)/N\right] R_{xx}(m)$.

The modern version of Schuster's periodogram is the direct method of spectral analysis. A sampled data version of Eqn. (2.14) is as follows.

$$\begin{aligned} \hat{P}_{\text{PER}}(f) &= \frac{1}{N\Delta t} \left| \Delta t \sum_{n=0}^{N_1} x(n) exp(-j2\pi f n \Delta t) \right|^2 \\ &\quad \text{for } n = 0, 1, \cdots, N-1 \end{aligned} \tag{2.19}$$

It is defined for frequency interval $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$. Using the fast fourier transform Eqn. (2.19)can be evaluated at the discrete set of N equally spaced frequencies $f(m) = m\Delta f$ Hz, for $m = 0, 1, \cdots, N-1$ and $\Delta f = 1/N\Delta t$,

$$\hat{P}(m) = \hat{P}_{\text{PER}}(f(m)) = \frac{1}{N\Delta t}|X(m)|^2 \tag{2.20}$$

Here $X(m)$ is the DFT of Eqn. (2.7). $\hat{P}(m)$ is identical to the energy spectral density $\mathcal{E}(m)$ of Eqn. (2.10) except for the division by the time interval of $N\Delta t$ seconds to make $\hat{P}(m)$ a power spectral density. The total power in the process, which is assumed periodic due to the DFT property, is

$$Power = \sum_{m=0}^{N-1} \hat{P}(m)\Delta f \tag{2.21}$$

based on rectangular integration approximation of $\hat{P}_{\text{PER}}$. If the Δf factor is incorporated into $\hat{P}(m)$, then

$$\begin{aligned} \hat{\hat{P}}(m) &= \hat{P}(m)\Delta f = \frac{1}{(N\Delta t)^2}|X(m)|^2 \\ &= \left|\frac{1}{N}\sum_{n=0}^{N-1} x(n)exp(-j2\pi mn/N)\right|^2 \end{aligned} \tag{2.22}$$

Eqn. (2.22) is often known as the periodogram, but it is not scaled appropriately as PSD. Using Eqn. (2.22) it is the peak in the PSD plot, rather than the area under the plot, that is equal to the power of the assumed periodic signal. The computational economy of the FFT algorithm has made this a popular approach [2].

Many of the problems of the periodogram PSD estimation technique can be traced to the assumption made about the data outside the measurement interval. The finite data sequence may be viewed as being obtained by windowing an infinite length sample sequence with a boxcar function [2]. The use of only this data implicitly assumes the unmeasured data to be zero, which is usually not the case. This multiplication of the actual time series by a window function means the overall transform is the convolution of the desired transform with the transform of the window function. If the true power of a signal is concentrated in a narrow bandwidth, this convolution operation will spread that power into adjacent frequency regions. This phenomena, termed leakage, is a consequence of the tacit windowing inherent in the computation of periodogram.

In addition to the distorting effects of leakage on the spectral estimate, leakage has detrimental impact on power estimation and detectability of sinusoidal components [35]-[37]. Sidelobes from adjacent frequency cells add in a constructive or destructive manner to the main lobe of a response in another frequency cell of the spectrum, affecting the estimate of power in that cell. In extreme cases, the sidelobes from strong frequency components can mask the main lobe of weak frequency components in adjacent cells.

Data windowing is also the fundamental factor that determines the frequency resolution of the periodogram. The convolution of the window transform with that of the actual signal transform means that most narrow spectral response of the resultant transform is limited to that of the main-lobe width of the window transform, independent of the data. For recrtangular window, the main-lobe width between 3-dB levels (and therefore, the resolution) of the resulting $(sin\pi f)/\pi f$ transform is approximately the inverse of the observation time of $N\Delta t$ seconds. Other windows may be used, but the resolution will always be proportional to $1/N\Delta t$ Hz. Leakage effects due to data windowing can be reduced by the selection of windows with nonuniform weighting. Harris [38]has provided a good summary of the merits of various windows. Nuttal [39] gives a correction to the sidelobe behavior for some of the windows described by Harris. However, The price paid for a reduction in the sidelobes is always a broadening in the main lobe of the window transform, which in turn means a decrease in the resolution of the spectral estimate.

In a nutshell, the conventional BT and periodogram approaches to spectral estimation have the following advantages:

1. Computationally efficient if only a few lags are needed (BT) or if the FFT is used (periodogram).

2. PSD estimate is directly proportional to power for sinusoid processes

The disadvantages of these techniques are:

1. Suppression of weak signal main-lobe responses by strong signal sidelobes.

2. Frequency resolution limited by the available data record duration, independent of the characteristics of the data or its SNR.

3. Introduction of distortion in the spectrum due to sidelobe leakage.

4. Need for some sort of pseudo ensemble averaging to obtain statistically consistent spectra

5. The appearance of negeative PSD values with the BT approach when some autocorrelation sequence estimates are used.

The high variance of the periodogram and correlogram methods motivates the development of modified methods that have lower variance, at a cost of reduced resolution.

2.3 Parametric Methods for Rational Spectra

The principal difference between the spectral estimation non-parametric methods and parametric methods is that in the first case, no assumption on the studied signal is made(excepts for its stationarity). The parametric or model-based methods of spectral estimation assume that the signal satisfies a generating model with known functional form, and then proceed by estimating the parameters in the assumed model. The signal's spectral characteristics of interest are then derived from the estimated model. In those cases where the assumed model is a close approximation to the reality, it is no wonder that the parametric methods provide more accurate spectral esitmates than the nonparmetric techniques.

In the parametric methods an input driving sequence $u(n)$ and the output sequence $x(n)$ that is to model the data are related by the linear difference equation,

$$x(n) = \sum_{l=0}^{q} b_l u(n-l) - \sum_{k=1}^{p} a_k x(n-k) \tag{2.23}$$

This most general linear model is termed an ARMA (Autoregressive Moving Average Model) model. The interest in these models stems from their relationship to linear filters with rational transfer functions. Here $\{a_k\}$ is Autoregressive (AR) coefficient and $\{b_k\}$ is Moving Average (MA) coefficient.

The system function $H(z)$ between the input $u(n)$ and output $x(n)$ for the ARMA process of Eqn. (2.23) is the rational expression

$$H(z) = \frac{\mathcal{B}(z)}{\mathcal{A}(z)} \tag{2.24}$$

where

$$\mathcal{A}(z) = \text{z–transform of AR branch} = \sum_{k=0}^{p} a_k z^{-k}$$

$$\mathcal{B}(z) = \text{z–transform of MA branch} = \sum_{l=0}^{q} b_l z^{-l}$$

It is well known that the power spectrum at the output of a linear filter, $\mathcal{P}_x(z)$, is related to the power spectrum of the input stochastic process, $\mathcal{P}_u(z)$, as follows:

$$\mathcal{P}_x(z) = H(z)H^*(1/z^*)\mathcal{P}_u(z) = \frac{\mathcal{B}(z)\mathcal{B}^*(1/z^*)}{\mathcal{A}(z)\mathcal{A}^*(1/z*)} \cdot \mathcal{P}_u(z) \tag{2.25}$$

Eqn. (2.25) is normally evaluated along the unit circle, $z = exp(j2\pi f \Delta t)$ for $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$. Often the driving process $u(n)$is assumed to be a white-noise sequence of zero mean and variance σ_u^2. The PSD of the noise is then $\sigma_u^2 \Delta t$. Here Δt factor is included in the expression for power spectral density of the noise so that $P_x(exp[j2\pi f \Delta t])$, when integrated over $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$, yields the true power of an analog signal. However for convenience and simplicity from now on wards, the factor Δt is omitted and f is regarded as normalized frequency by assuming $\Delta t = 1$. The PSD of the ARMA output process is then

$$P_{\text{ARMA}}(f) = P_x(f) = \sigma_u^2 \left| \frac{B(f)}{A(f)} \right|^2 \tag{2.26}$$

where $A(f) = \mathcal{A}(exp[j2\pi f])$ and $B(f) = \mathcal{B}(exp[j2\pi f])$. Without loss of generality, one can assume $a_0 = 1$ and $b_0 = 1$ since any filter gain can be incorporated into σ_u^2.

If all the $a_k = 0$, for $k = 1, 2, \cdots, p$, then

$$x(n) = \sum_{l=0}^{q} b_l u(n-l). \tag{2.27}$$

the process given by Eqn. (2.27) is known as moving average of order q, and

$$P_{\text{MA}}(f) = \sigma_u^2 |B(f)|^2 \tag{2.28}$$

This model is sometimes termed an all-zero model [40].

If all the $b_l = 0$, for $l = 1, 2, \cdots, q$, then

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + u(n) \tag{2.29}$$

The process given by Eqn. (2.29) is strictly an autoregressive (AR) of order p. The process is termed AR in that the sequence $x(n)$ is a linear regression on itself with $u(n)$ representing the error. With this model, the present value of the process is express as a weighted sum of past values plus a noise term. The PSD is

$$P_{\mathrm{AR}}(f) = \frac{\sigma_u^2}{|A(f)|^2} \tag{2.30}$$

This model is sometimes termed an all-pole model.

The Wold decomposition theorem [40] relates the ARMA, MA and AR models. Basically the theorem asserts that any stationary ARMA or MA process of finite variance can be represented as a unique AR model of possibly infinite order; likewise, any ARMA or AR process can be represented as MA process of possibly infinite order. This theorem is important because if we choose the wrong model among the three, we may still obtain a reasonable approximation by an AR model of higher order. Thus an ARMA model can be approximated by an AR model of higher order. Since the estimation of parameters for an AR model results in linear equations, it has a computational advantage over ARMA and MA parameter estimation techniques. The largest portion of research effort on rational transfer function modeling has therefore been concerned with AR model. As this research is concentrated on AR spectral estimation from noisy observations a brief discussion on AR PSD estimation is given below.

2.3.1 Autoregressive PSD estimation

2.3.1.1 Yule-Walker equations

The derivation of Yule-Walker (YW) equations for AR spectral estimation is first given. The YW equations describe the linear relationship between the AR parameters and the auto-correlation function. The solution of these equations is provided by the computationally efficient Levinson-Durbin algorithm, the details of which reveal some fundamental properties of AR processes.

If autoregression is a reasonable model for the data, then the AR power spectral density estimate based on Eqn. (2.30) can be rewritten as

$$P_{\mathrm{AR}}(f) \;=\; |H(exp[j2\pi f])|^2 \, P_u(f)$$

$$= \frac{\sigma_u^2}{\left|1 + \sum_{k=1}^{p} a_k exp(-j2\pi f k)\right|^2} \tag{2.31}$$

It reveals that to estimate PSD one need only estimate $a_1, a_2, \cdots, a_p, \sigma_u^2$. To do this, a relationship between the AR parameters and the autocorrelation function (known or estimated) of $x(n)$ is now presented. This relationship is known as the Yule-Walker equations [41]. The derivation proceeds as follows:

$$\begin{aligned} R_{xx}(k) &= E\left[x(n+k)x^*(n)\right] = E\left[x^*(n)\left(-\sum_{l=1}^{p} a_l x(n-l) + u(n+k)\right)\right] \\ &= -\sum_{l=1}^{p} a_l R_{xx}(k-l) + E\left[u(n+k)x^*(n)\right] \end{aligned}$$

Since $H(z)$ is assumed to be stable, causal filter, we have

$$\begin{aligned} E\left(u(n+k)x^*(n)\right) &= E\left[u(n+k)\sum_{l=0}^{\infty} h^*(l)u^*(n-l)\right] \\ &= \sum_{l=0}^{\infty} h^*(l)\sigma_u^2\delta(k+1) \\ &= \sigma_u^2 h_{-k}^* \\ &= \begin{cases} 0 & \text{for } k > 0 \\ h_0^*\sigma_u^2 & \text{for } k = 0 \end{cases} \end{aligned}$$

Note that $\delta(m)$ is the discrete delta function where

$$\begin{aligned} \delta(m) &= 1 \quad \text{if } m = 0 \\ &= 0 \quad \text{if } m \neq 0 \end{aligned}$$

But $h_0 = \lim_{z\to\infty} H(z) = 1$, and therefore,

$$R_{xx}(k) = \begin{cases} -\sum_{l=1}^{p} a_l R_{xx}(k-l) & \text{for } k > 0 \\ -\sum_{l=1}^{p} a_l R_{xx}(-l) + \sigma_u^2 & \text{for } k = 0 \end{cases} \tag{2.32}$$

Eqn. (2.32) is the Yule-Walker equations. To determine the AR parameters, one need only choose p equations from Eqn. (2.32) for $k > 0$, solve for $a_1, a_2, \cdots, a_p$ and then find σ_u^2 from Eqn. (2.32) for $k = 0$. The set of equations which require the fewest lags of the autocorrelation function is the selection $k = 1, 2, \cdots, p$. They can be expressed in matrix form as

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \cdots & R_{xx}(-(p-1)) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(-(p-2)) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ \vdots \\ R_{xx}(p) \end{bmatrix} \tag{2.33}$$

Eqns. (2.33) is known as low order Yule-Walker (LOYW) equations. It is obvious that the above autocorrelation, R_{xx} is hermitian ($R_{xx}^H = R_{xx}$) and it is Toeplitz since the elements along diagonal are identical. Also, the matrix is positive definite (assuming $x(n)$ is not purely harmonic) which follows from the positive definite property of the autocorrelation function [42],[43].

It should be noted that Eqn. (2.33) can also be augmented to incorporate the σ_u^2 equation, yielding

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \cdots & R_{xx}(-p) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(-(p-1)) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p) & R_{xx}(p-1) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_u^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.34}$$

which follows from (2.32). Thus to determine the AR parameters and σ^2 one must solve Eqn. (2.34) with the $p+1$ estimated autocorrelation lags $R_{xx}(0), \cdots, R_{xx}(p)$ and use $R_{xx}(-m) = R_{xx}^*(m)$.

The Levinson-Durbin [44]-[47] algorithm provides an efficient solution for Eqn. (2.34). The algorithm requires only order p^2 operations, denoted $o(p^2)$, as opposed to $o(p^3)$ for Gaussian elimination. Although appearing at first to be just an efficient algorithm, it reveals fundamental properties of AR processes. The algorithm proceeds recursively to compute the parameter sets

$$\{a_{11}, \sigma_1^2\}, \{a_{21}, a_{22}, \sigma_2^2\}, \cdots, \{a_{p1}, a_{p2}, \cdots, a_{pp}, \sigma_p^2\}$$

An additional subscript has been added to the AR coefficients to denote the order. The final set at order p is the desired solution. In particular, the recursive algorithm is initialized by

$$a_{11} = -Rxx(1)/R_{xx}(0) \tag{2.35}$$

$$\sigma_1^2 = (1 - |a_{11}|^2)R_{xx}(0) \tag{2.36}$$

$$\tag{2.37}$$

with the recursion for $k = 2, 3, \cdots, p$ given by

$$a_{kk} = -\left[R_{xx}(k) + \sum_{l=1}^{k-1} a_{k-1,l}\right] \Big/ \sigma_{(k-1)}^2 \tag{2.38}$$

$$a_{ki} = a_{k-1,i} + a_{kk}a^*k - 1, k - i \tag{2.39}$$

$$\sigma_k^2 = (1 - |a_{kk}|^2)\sigma_{k-1}^2 \tag{2.40}$$

The parameter set $\{a_{k1}, a_{k2}, \cdots, a_{kk}, \sigma_k^2\}$ thus computed is the same as would be obtained by using Eqn. (2.34) for $p = k$. Thus the Levinson-Durbin algorithm also provides the AR parameters for all the lower order AR model that fits to the data. This is useful property when one does not know a priori the correct model order, since one can use Eqn. (2.35)-(2.40) to generate successively higher order models until the modeling error σ_k^2 is reduced to a desired value. In particular, if a process is actually an AR(p) process (an AR process of order p), then $a_{p+1,k} = a_{pk}$ for $k = 1, 2, \cdots, p$ and hence $a_{p+1,p+1} = 0$. In general for an AR(p) process, $a_{kk} = 0$ and $\sigma_k^2 = \sigma_p^2$ for $k > p$.

Table 2.1: Summary of AR process properties excluding purely harmonic processes

- Autocorrelation matrix is positive definite $\mathbf{X}^{\mathbf{H}}\mathbf{R}_{xx}\mathbf{X} > 0$ for all $\mathbf{X}$ vectors.
- Reflection coefficient sequence satisfies $|K_i| < 1$ for $i = 1, 2, \cdots, p$.
- Zeros of $A(z)$ lie within unit circle: $|z_i| < 1$ for $i = 1, 2, \cdots, p$
- Prediction error powers monotically decrease: $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_p^2 \leq 0$

Hence, the variance of the excitation noise is a constant for model order equal to or greater than the correct order. Thus the point at which σ_k^2 does not change would appear to be a good indicator of the correct model order. It can be shown that σ_k^2 first reaches its minimum at the correct model order.

The parameters $\{a_{11}, a_{22}, \cdots, a_{pp}\}$ are often called the reflection coefficients and are designed as $K_1, K_2, \cdots, K_p$. They have the property that for

$$\{R_{xx}(0), R_{xx}(1), \cdots, R_{xx}(p)\}$$

to be a valid autocorrelation sequence, i.e., the autocorrelation matrix is positive semidefinite, then it is necessary and sufficient that $|a_{kk}| = |K_k| \leq 1$ for $k = 1, 2, \cdots, p$. Furthermore a necessary and sufficient condition for the poles of $A(z)$ to be or within the unit circle of the z plane is $|K_k| \leq 1$ for $k = 1, 2, \cdots, p$. It should be noted that if $|K_k| = 1$ for some k, then the recursion (2.35)-(2.40) must terminate since $\sigma_k^2 = 0$. The process in this case is purely harmonic (consists

only of sinusoids).

2.3.1.2 Lattice filter algorithm

The problem of AR parameter estimation is closely related to the theory of linear prediction. If $x(n)$ is assumed to be an AR(p) process, the prediction of $x(n)$ on the basis of previous samples [48] can be done in the following manner.

$$\hat{x}(n) = -\sum_{k=1}^{p} \alpha(k) x(n-k) \tag{2.41}$$

then $\{\alpha_1, \alpha_2, \cdots, \alpha_p\}$ can be chosen to minimize the prediction error power Q_p where

$$Q_p = E\left[|x(n) - \hat{x}(n)|^2\right] \tag{2.42}$$

By the orthogonality principle [33]

$$E\left[(x(n) - \hat{x}(n))x^*(k)\right] = 0 \quad \text{for } k = n-1, \cdots, n-p$$

or

$$R_{xx}(k) = -\sum_{l=1}^{p} \alpha_l R_{xx}(k-l) \quad \text{for } k = 1, 2, \cdots, p \tag{2.43}$$

The minimum prediction error power is

$$Q_{p_{\min}} = E\left[(x_n - \hat{x}(n))x^*(n)\right] = R_{xx}(0) + \sum_{k=1}^{p} \alpha_k R_{xx}(-k) \tag{2.44}$$

(2.43) and (2.44) are identical to (2.32). Thus it must be true that $\alpha_k = a_{pk}$ for $k = 1, \cdots, p$ and $Q_{p_{\min}} = \sigma_p^2$ so that the best linear predictor is just $\hat{x}(n) = -\sum_{k=1}^{p} a_{pk} x(n-k)$. The error sequence, although uncorrelated with the linear estimate, is not necessarily a white process (it will be if $x(n)$ is a AR(p) process). In the limit as $p \to \infty$, the error sequence becomes white. Therefore $\{a_{k1}, a_{k2}, \cdots, a_{kk}\}$ and σ_k^2 constitute the parameters for the optimum k-th order linear predictor and the corresponding minimum prediction error power, respectively. Therefore, AR parameter estimation and linear prediction of an AR(p) process yield identical results and the theory of one is applicable to the other.

The most popular approach for AR parameter estimation with N data samples was introduced by Burg in 1967. The Burg algorithm may be viewed as

recursive lattice method based on the forward and backward errors in linear predictions with the constraint that the AR parameters satisfy the Levinson-Durbin recursion.

To derive the estimator, suppose that we are given the data $x(n), n = 0, 1, \cdots, N-1$ and let us consider the forward and backward linear prediction estimates of order k, given as

$$\hat{x}(n) = -\sum_{i=1}^{k} a_{ki} x(n-i) \tag{2.45}$$

$$\hat{x}(n-k) = -\sum_{i=1}^{k} a_{ki}^{*} x(n+i-k) \tag{2.46}$$

and the corresponding forward and backward errors e_{kn} and b_{kn} defined as

$$e_{kn} = x(n) - \hat{x}(n) \tag{2.47}$$

and

$$b_k(n) = x(n-k) - \hat{x}(n-k) \tag{2.48}$$

where a_{ki}, $0 \leq i \leq k-1, k = 1, 2, \cdots, p$, are the prediction coefficients. The sum of the square of forward and backward error is

$$\varepsilon_k = \sum_{n=k}^{N-1} \left[|e_{kn}|^2 + |b_{kn}|^2 \right] \tag{2.49}$$

This error is to be minimized by selecting the prediction coefficients, subject to the constraint that they satisfy the Levinson-Durbin recursion given by

$$a_{ki} = a_{k-1,i} + K_k a_{k-1,k-i}^{*} \qquad 1 \leq i \leq k-1, 1 \leq k \leq p \tag{2.50}$$

$$\tag{2.51}$$

where $K_k = a_{kk}$ is the k-th reflection coefficient in the lattice filter realization of the predictor. When Eqn. (2.50) is substituted into the Eqns.(2.48) and (2.49), order-recursive Eqns. (2.52) and (2.53) , respectively, for the forward and backward prediction errors are obtained.

$$\begin{aligned} e_{kn} &= x(n) + \sum_{i=1}^{k} a_{ki} x_{n-i} \\ &= x(n) + \sum_{i=1}^{k-1} (a_{k-1,i} + K_k a_{k-1,k-i}^{*}) x_{k-i} + K_k x_{n-k} \\ &= e_{k-1,n} + K_k b_{k-1,n-1} \end{aligned} \tag{2.52}$$

where

$$b_{kn} = x_{n-k} + \sum_{i=1}^{k} a_{ki}^{*} x_{n-k+i}$$

Similarly it can be shown that

$$b_{kn} = b_{k-1,n-1} + K_k^{*} e_{k-1,n} \tag{2.53}$$

Here $e_{0n} = b_{0n} = x(n)$. It is seen that the predictor coefficients for the backwards predictor are complex conjugates of those of the forward predictor, which is a consequence of the stationary autocorrelation function [42]. The relationships of Eqns.(2.52) and Eqns.(2.53) give rise to the so called lattice filter structure shown in Fig. 2.2. It is to be noted that the transfer function of the entire filter

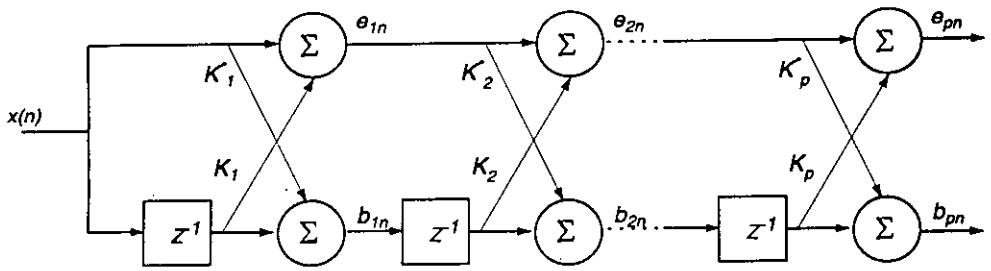

Fig. 2.2: Lattice formulation of prediction error filter

is just

$$A(z) = 1 + \sum_{i=1}^{p} a_{pi} z^{-i}$$

which is the inverse of $H(z) = 1/A(z)$. This follows from Eqn. (2.48). This filter is termed either the "inverse" filter, or" whitening" filter or "prediction error" filter.

Now if Eqn. (2.52) and Eqn. (2.53) are substituted in Eqn. (2.50) and the minimization of ε_m with respect to the complex-valued reflection coefficient K_m is performed following expression can be obtained.

$$\widehat{K}_k = \frac{-\sum_{n=k}^{N-1} e_{k-1,n} b_{k-1,n-1}^{*}}{\frac{1}{2}\sum_{n=k}^{N-1} [|e_{k-1,n}|^2 + |b_{k-1,n-1}|^2]} \tag{2.54}$$

The term in the numerator of Eqn. (2.54) is an estimate of the crosscorrelation between the forward and backward prediction errors. With the normalization factors in the denominator of Eqn. (2.54), it is apparent that $|K_k| < 1$, so that the all pole model obtained from the data is stable. As the denominator in Eqn. (2.54) is simply the least-squares estimate of the forward and backward errors, E_{k-1}^f and E_{k-1}^b, respectively, the Eqn. (2.54) can be expressed as

$$\widehat{K}_k = \frac{-\sum_{n=k}^{N-1} e_{k-1,n} b_{k-1,n-1}^*}{\frac{1}{2}\sum_{n=k}^{N-1}\left[\hat{E}_{k-1}^f + \hat{E}_{k-1}^b\right]} \tag{2.55}$$

where $\hat{E}_{k-1}^f + \hat{E}_{k-1}^b$ is an estimate of the total squared error E_k. The denominator term in Eqn. (2.55) can be computed in an order-recursive fashion according to the relation

$$\hat{E}_k = (1 - |\widehat{K}_k|^2)\hat{E}_{k-1} - |e_{k-1,k-1}|^2 - |b_{k-1,k-2}|^2 \tag{2.56}$$

where $\hat{E}_k \equiv \hat{E}_k^f + \hat{E}_k^b$ is the total least-squares error.

To summarize, the Burg algorithm computes the reflection coefficients in the equivalent lattice structure as specified by Eqns. (2.55) and (2.56). Eqn. (2.50) is used to obtain the AR model parameters. From the estimates of the AR parameters, we form the power spectrum estimate

$$P_{xx}^{BU}(f) = \frac{\hat{E}_p}{|1 + \sum_{i=1}^p \hat{a}_i e^{-j2\pi fi}|^2} \tag{2.57}$$

The major advantages of the Burg method for estimating the parameters of the AR model are:

1. it results in high frequency resolution
2. it yields a stable AR model
3. it is computationally efficient.

Conventional periodogram and BT analysis lead to spectral estimates that are characterized by many "hills and valleys", since the Fourier transform of a zero mean random process. Autocorrelation lag windowing or spectral window smoothing will substantially reduce the fluctuations but not eliminate them. An AR spectral estimator can be used to smooth these fluctuations since a pth-order AR spectral estimate is constrained to have p or less peaks (or trouhgs). For p small, a smoothed spectral estimate will result. Fig. 2.3 shows smooth PSD estimates of AR estimation over FFT based rough PSD estimation.

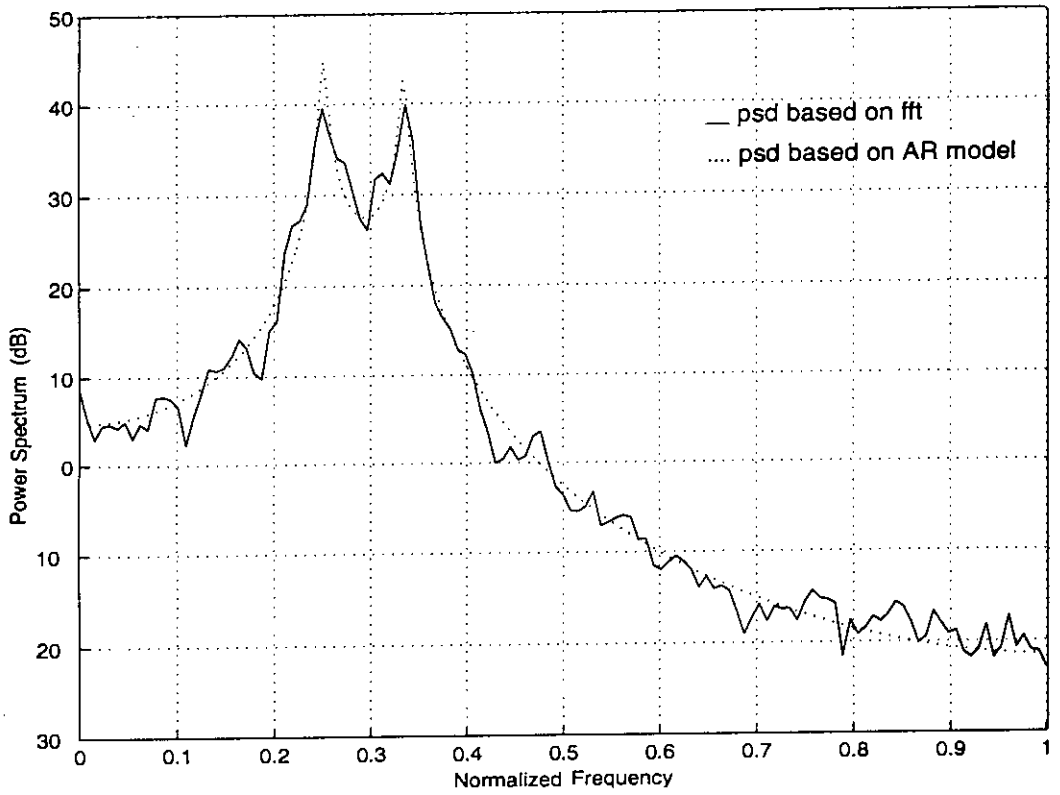

Fig. 2.3: FFT and AR model based PSD estimation of random processes

2.4 Multichannel Spectral Estimation

The spectral estimation method in previous sections it is assumed that we wish to estimate the power spectral density (PSD) of a single time series. In many practical situations the data that are available are not limited to the output of a single channel but may well be the result of observations at the output of several channels. It is common in the fields of sonar [49], radar [50], and seismic exploration [51] to record data from multiple sensors. With this additional information it is then possible to estimate cross-spectra as well as auto-spectra. The cross spectra are important in establishing linear filtering relationship between the time series [31]. *The multichannel spectral estimation problem is to estimate the auto- spectra of the individual channel and the cross-spectra between all pairs of channels.* Many of the techniques described in the previous sections are easily extended to multichannel spectral estimation by replacing scalar functions by

suitable matrix functions of a vector valued time series [52].

Let, $\mathbf{x}(n)$ be a multichannel wide sense stationary(WSS) random process. That is, the mean of the process is a constant and the auto correlation function depends only on the lag. Specifically,

$$E\left[\mathbf{x}(n)\right] = \boldsymbol{\mu}(n) = \boldsymbol{\mu}$$

which will henceforth be assumed equal to zero, and the ACF

$$\mathbf{R_{xx}}(m) = E[\mathbf{x}(n)\mathbf{x}^T(n-m)] \tag{2.58}$$

does not depend on n. The mth sample of the ACF is the $M \times M$ matrix

$$\mathbf{R_{xx}}(m) = \begin{bmatrix} r_{11}(m) & r_{12}(m) & \cdots & r_{1M}(m) \\ r_{21}(m) & r_{22}(m) & \cdots & r_{2M}(m) \\ \vdots & \ddots & \vdots & \\ r_{M1}(m) & r_{M2}(m) & \cdots & r_{MM}(m) \end{bmatrix} \tag{2.59}$$

where the (i,j) element is the cross-correlation function (CCF) between $x_i(n)$ and $x_j(n)$ at lag m or

$$r_{ij}(m) = E[x_i^*(n)x_j(n-m)] \tag{2.60}$$

The power spectral density matrix or *cross-spectral matrix* is defined as

$$\mathbf{P_{xx}}(f) = \begin{bmatrix} P_{11}(f) & P_{12}(f) & \cdots & P_{1M}(f) \\ P_{21}(f) & P_{22}(f) & \cdots & P_{2M}(f) \\ \vdots & \ddots & \vdots & \\ P_{M1}(f) & P_{M2}(f) & \cdots & P_{MM}(f) \end{bmatrix} \tag{2.61}$$

The diagonal elements $P_{ij}(f)$ are the power spectral densities (PSDs) of the individual channels or auto-PSDs, while the off-diagonal elements $P_{ij}(f)$ for $i \neq j$ are the cross-PSDs between $x_i(n)$ and $x_j(n)$, which are defined as

$$P_{ij}(f) = \sum_{m=-\infty}^{\infty} r_{ij}(m) exp(-j2\pi fm) \tag{2.62}$$

The magnitude of the cross-PSD describes whether frequency components in $x_i(n)$ are associated with large or small amplitudes at the same frequency in $x_j(n)$, and the phase of the cross-PSD indicates the phase lag or lead of $x_i(n)$ with respect to $x_j(n)$ for a given frequency component. The cross-spectral matrix may also be written as

$$\mathbf{P_{xx}}(f) = \sum_{m=-\infty}^{\infty} \mathbf{R_{xx}}(m) exp(-j2\pi fm) \tag{2.63}$$

Multichannel white noise is defined as the process whose ACF satisfies

$$\mathbf{R_{xx}}(m) = \Sigma\delta(m) \tag{2.64}$$

so that the cross spectral matrix is a constant matrix or

$$\mathbf{P_{xx}}(f) = \Sigma \tag{2.65}$$

For multichannel white noise the individual processes are each white noise processes with variance or PSD $[\Sigma]_{ii}$. The cross-correlation given by the off-diagonal elements of Σ.

The output of a multichannel filter is [52]

$$\mathbf{x}(n) = \sum_{m=-\infty}^{\infty} \mathbf{H}(m)\mathbf{u}(n-m) \tag{2.66}$$

where $\mathbf{u}(n)$ is the $M \times 1$ input sequence vector and $\mathbf{H}(m)$ is a complex $M \times M$ matrix defined as

$$\mathbf{H}(m) = \begin{bmatrix} h_{11}(m) & h_{12}(m) & \cdots & h_{1M}(m) \\ h_{21}(m) & h_{22}(m) & \cdots & h_{2M}(m) \\ \vdots & \ddots & \vdots & \\ h_{M1}(m) & h_{M2}(m) & \cdots & h_{MM}(m) \end{bmatrix} \tag{2.67}$$

and $h_{ij}(m)$ is the impulse response of the filter between the jth input and the ith output. The cross-spectral matrix is obtained by evaluating on the unit circle [17]

$$\mathbf{P}_{xx}(f) = \mathbf{H}^*(f)\mathbf{P}_{uu}(f)\mathbf{H}^T(f) \tag{2.68}$$

where $\mathbf{P}_{uu}(f)$ is the cross-spectral matrix for the input signal. Eq. 3.5 shows an important relationship in spectral estimation of a scalar time series between the input PSD and the output PSD of a LSI filter.

2.4.1 Classical spectral estimation

An estimator of the cross-PSD $P_{ij}(f)$ is provided by the cross-periodogram [31]

$$\hat{P}_{ij}(f) = \frac{1}{N}X_i^*(f)X_j(f) \qquad i,j = 1,2,\cdots,M \tag{2.69}$$

where $X_i(f)$ is the DFT of $x_i(n)$ and given as

$$X_i(f) = \sum_{n=0}^{N-1} x_i(n)exp(-j2\pi fn) \tag{2.70}$$

For the auto-PSD or for $i = j$ this reduces to the usual periodogram. In matrix notation the cross-spectrtal matrix is estimated as

$$\hat{\mathbf{P}}_{PER}(f) = \frac{1}{N}\mathbf{X}^*(f)\mathbf{X}^T(f) \tag{2.71}$$

where

$$\mathbf{X}(f) = \sum_{n=0}^{N-1} \mathbf{x}(n)exp(-j2\pi fn) \tag{2.72}$$

2.4.2 Autoregressive spectral estimation

The estimate of the cross-spectral matrix based on a multichannel AR model is given by

$$\hat{\mathbf{P}}_{AR}(f) = \widehat{\mathbf{A}}^{*-1}(f)\hat{\mathbf{\Sigma}}\widehat{\mathbf{A}}^{T-1}(f) \tag{2.73}$$

where the "hats" denote estimators. The most sraightforward means of estimating the unkown parameters is by using the Yule-Walker equations as given by Eqn. 2.74 with a suitable ACF estimator.

$$\underline{\mathbf{a}}\,\underline{\mathbf{R}} = -\underline{\mathbf{r}} \tag{2.74}$$

$$\begin{aligned} \underline{\mathbf{R}} &= \begin{bmatrix} \mathbf{R_{xx}}(0) & \cdots & \mathbf{R_{xx}}(p-1) \\ \mathbf{R_{xx}}(-1) & \cdots & \mathbf{R_{xx}}(p-2) \\ \vdots & \ddots & \vdots \\ \mathbf{R_{xx}}(-p+1) & \cdots & \mathbf{R_{xx}}(0) \end{bmatrix} \\ \underline{\mathbf{r}} &= [\mathbf{R_{xx}}(1), \mathbf{R_{xx}}(2), \cdots, \mathbf{R_{xx}}(p)], \end{aligned}$$

where $\underline{\mathbf{R}}$ is the $p \times p$ block Toeplitz matrix (i.e., a matrix with equal sub matrices, or blocks, on any block diagonal) with block entries $\mathbf{R_{xx}}(m)$ each of which has dimension $M \times M$, $\underline{\mathbf{r}}$ is the $1 \times p$ block row vector and $\mathbf{R_{xx}}(m) = E[\mathbf{x}(n)\mathbf{x}^T(n-m)]$ $(m = 0, 1, \cdots, p)$ with $\mathbf{R_{xx}}(-m) = \mathbf{R}_{\mathbf{xx}}^T(m)$.

2.5 Effect of Noise on AR Spectral Estimation

A very important problem with the AR spectral estimator is its sensitivity to the addition of observation noise to the time series [26]. If a signal $x(n)$ is contaminated by zero mean white noise $v(n)$, the observed signal $y(n)$ is described as follows:

$$y(n) = x(n) + v(n) \tag{2.75}$$

Assume that $x(n)$ is the output signal of a pth-order AR model excited by white noise $u(n)$,

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + u(n) \tag{2.76}$$

where, $u(n)$ is the zero mean white noise uncorrelated with $v(n)$ and,

$$E\{u^2(n)\} = \sigma_u^2$$
$$E\{v^2(n)\} = \sigma_v^2$$

and $E[\cdot]$ is the expectation operator.

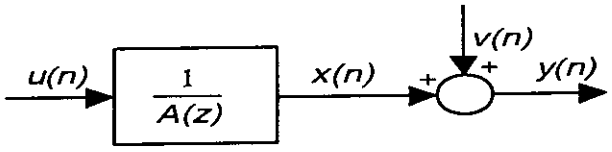

Fig. 2.4: Autoregressive process with noise

To illustrate the noise effect on AR spectral estimation an example is given in Fig. 2.5. As expected AR(p) PSD obtained from $y(n)$ is different from AR(p) PSD obtained from $x(n)$. It is seen that the spectral peaks are indistinguishable,

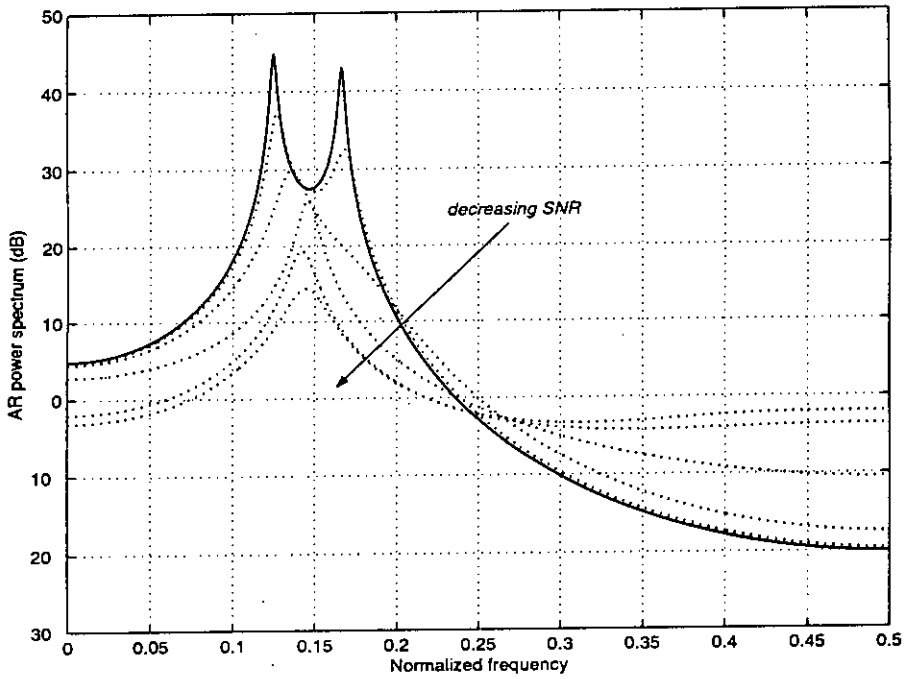

Fig. 2.5: Noise effect on AR spectral estimation; —:without noise; ...:with noise at different SNRs

broadened and unexpectedly the AR spectrum of the noisy signal lies below the AR spectrum of the noise-free signal for a wide frequency range. These are due to incorrect modeling of the observed noisy signal $y(n)$. Accurate modeling of $y(n)$ by an AR system will be addressed in the next chapter.

To reduce the degredation of the AR spectral estimate in the presence of noise, four general approaches have been proposed. They are as follows.

1. use the ARMA spectral estimate
2. filter the data to reduce the noise
3. use a large order AR model
4. compensate either the autocorrelation function estimates or the reflection coefficient estimates for the noise effects.

Approach (1) and (3) are used to get the PSD of $y(n)$. However, this research is carried out to estimate the AR spectrum of the actual data sequence $x(n)$ instead of noise corrupted $y(n)$ where it is assumed that data sequence $y(n)$ is only available. Moreover, the noise power σ_v^2 is also assumed to be unknown. Noise cancellation schemes that compensate the autocorrelation lags for the noise can be found in [17], [21]-[23]. A serious deficiency is that, one does not know how much noise power to remove. Thus, if the subtracted noise power is too large, the estimated AR spectrum will exhibit sharper peaks than the true spectrum. Moreover the system may become unstable. In the subsequent chapters we propose accurate noise compensation schemes for spectral estimation of $x(n)$ solely from $y(n)$.

2.6 Conclusion

In this chapter different methods for estimating power spectral density with their merits and demerits have been discussed. If the assumed model is accurate then the parametric methods give higher resolution even with short data length. Moreover, they do not suffer from spectral leakage. Among the three linear models AR model is widely used. Two popular methods namely the Yule-Walker mehthod and the Burg method for estimating AR parameters from noise-free observations have been focused. The Burg algorithm yields better resolution. However, in

case of noise contaminated observations, estimated AR spectrum deviates from the estimated AR spectrum of noise-free actual observations. This research is mainly concerned with the estimation of AR spectrum of the actual signal which is contaminated by additive white noise with unknown power. In the subsequent chapters we propose noise cancellation schemes for autoregressive spectral estimation of single and multichannel systems from noise corrupted observations.

Chapter 3

Autoregressive Spectral Estimation from Noisy Observations

3.1 Introduction

Spectral estimation is one of the key issues of modern signal processing, providing a powerful means of extracting useful information from a given data set. The autoregressive (AR) spectral estimation has received much attention lately in many diverse fields [1], [53]. Maximum entropy spectral estimation [11] used in seismic signal processing, and linear prediction [54] used in speech processing are identical to AR spectral estimation, though they are based on different theoretical foundations. The principal advantage of the AR spectral estimators over conventional Fourier-based spectral estimators is their enhanced resolution property [55]. From the statistical point of view, the best way to improve the resolution capability is to reduce the number of unknown parameters in the spectral estimator. This implies that we need a parsimonic model to interpret the observed data. In the decomposition theorem [56], Wold reveals that any stationary process can be expressed as a linear sum of an innovation sequence. Fuller [57] shows that the power spectrum of a stationary sequence can be fitted with an arbitrary accuracy by an autoreggressive (AR) process as long as appropriate model order is used. Using AR modeling, the observed process can be considered to have been generated by an all-pole filter driven by a white noise sequence, allowing one to relate the desired power spectrum to the noise spectral density and the frequency transfer function of the filter [58].

Many researches have focused on this topic for noise-free observations [11], [13]-[16]. In practical cases, however, observations contain additive noise and its effect cannot be neglected. A few methods are available for noise-corrupted observations [17]-[26]. In particular, the correlation-based methods are widely used in estimating the parameters of AR systems. However, if the observations are noisy they require a priori knowledge of the additive noise power [17], [19]. Although it is possible to estimate the AR parameters without knowing the noise power by using the high-order Yule-Walker equations (HOYW), it is not constrained to be nonsingular [27]. The possible singularity of the auto-correlation matrix leads to a substantial increase in the variance of the AR spectral estimate [17]. Most of the previous methods either assume that the noise power is known [17], [19] or subtract a suboptimal amount of noise power [28] to compensate for the noise effect. Determination of the optimal amount of noise power to remove the bias effect of noise is of utmost importance for high-resolution AR spectral estimation from noisy observations.

To combat this problem, Yahagi and Hasan [20] have proposed a method using the low-order Yule-Walker equations (LOYW) to compensate for the effect of noise in determining the AR parameters and to estimate the noise power from a given set of noisy observations. But the case of strong nonlinearity of the LOYW equations at low SNR was not addressed. Unfortunately, due to inherent nonlinearity of the LOYW equations when used for noisy observations with unknown noise power and system parameters, the solution may not be unique depending on SNRs and system characteristics. In these cases, the method fails to estimate the actual solution from the set of multiple solutions. Moreover, the method is computationally expensive as the step size is maintained constant and very small throughout the total range of search.

In this chapter, we propose a 2-step method to estimate the AR parameters from the signal which is contaminated by additive white noise. In step-1, the AR parameters are estimated by using the technique proposed by Yahagi and Hasan [20] incorporating fuzzy modeling of the cost function for step size variation to reduce the number of iterations. The previous technique is briefly reviewed in Sec.3.3. In step-2, we estimate the parameters of an over-fitted AR model to the observed process using the lattice filter. The desired AR parameters,

equivalently the roots, are extracted from the over-fitted model roots using a 'root matching' technique described in Sec.3.4.1 which utilizes the solution(s) found in step-1. This technique solves the problem of multiple solutions and at the same time improves the estimation accuracy of AR parameters. Moreover, instead of using constant step size we use variable step size obtained as the output of a fuzzy inference system described in Sec.3.4.2. Doing this reduces the number of iterations as compared to the previous one. Finally, in Sec.3.5, several examples of computer simulation are presented to demonstrate the effectiveness of the proposed method.

3.2 Problem Formulation

If a signal $x(n)$ is contaminated by a white noise process $v(n)$ with distribution $\mathcal{N}(0, \sigma_v^2)$, the observed signal $y(n)$ is obtained as

$$y(n) = x(n) + v(n) \tag{3.1}$$

Assume that $x(n)$ is the output signal of a pth-order AR model excited by a sequence of white noise $u(n)$ with distribution $\mathcal{N}(0, \sigma_u^2)$ and is given by

$$x(n) = -\sum_{i=1}^{p} a_i x(n-i) + u(n) \tag{3.2}$$

The observation noise $v(n)$ is assumed to be independent of the input noise $u(n)$, i.e., $E[u(n)v(n-t)] = 0$ for all t, where $E[\cdot]$ denotes the expectation operator. The order p of the AR model is assumed to be known.

The z-transform of Eqn. (3.2) can be obtained as

$$X(z) = \frac{1}{A(z)} U(z) \tag{3.3}$$

where, $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}$, and $X(z)$ and $U(z)$ denote the z-transforms of $x(n)$ and $u(n)$, respectively. The transfer function $H(z)$ of the AR system in Eqn. (3.3) is given by

$$H(z) = \frac{1}{A(z)} \tag{3.4}$$

The AR power spectrum [53] is defined by

$$P_{AR}(z) = \frac{\sigma_u^2}{|A(z)A(z^{-1})|} \tag{3.5}$$

where, $A(z)$ is evaluated on the unit circle in the z-plane. Since we are interested in estimating the system characteristics, we can set $\sigma_u^2 = 1$ as it acts only as a scaling factor. Eqn. (3.5), therefore, reduces to

$$P_{AR}(z) = \frac{1}{|A(z)A(z^{-1})|} \tag{3.6}$$

For the noise-less case, $\{a_k\}$ can be obtained from the following Yule-Walker equations [53]:

$$R_{xx}(m) = -\sum_{k=1}^{p} a_k R_{xx}(m-k) \qquad m \geq 1 \tag{3.7}$$

where, $R_{xx}(m)$ is the auto-correlation function defined as

$$R_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-1-|m|} x(n)x(n+|m|) \tag{3.8}$$

and N is the number of data points. Clearly, any p equations are sufficient to determine the AR parameters. Generally, $m = 1, 2, \cdots, p$ is chosen which results in a set of symmetric Toeplitz equations. When noise is present, however, we cannot estimate $R_{xx}(m)$ since only $y(n)$ is available. But exploiting the following relation we can calculate $R_{xx}(m)$ for all values of m except $m = 0$ as σ_v^2 is unknown.

$$R_{xx}(m) = \begin{cases} R_{yy}(0) - \sigma_v^2 & m = 0 \\ R_{yy}(m) & m \neq 0 \end{cases} \tag{3.9}$$

Although the high-order Yule-Walker equations where $R_{xx}(0)$ is absent, can be used to estimate the AR parameters [17], this approach is not constrained to be nonsingular [27]. The possible singularity of the auto-correlation matrix leads to a substantial increase in the variance of the AR spectral estimate [17].

From Eqn. (3.6) it is clear that the problem of spectrum estimation by AR method (i.e., finding $P_{AR}(\omega)$) solely from a set of noisy observations $y(n)$ depends on finding the values of σ_v^2, the additive noise power, and $\{a_i\}$, the AR parameters, where $i = 1, 2, \cdots, p$.

3.3 Review of The Previous Method

Considering the case of singularity of auto-correlation matrix and the accuracy of estimates, compensation was done utilizing Eqn. (4.4) rather than using the

high-order Yule-Walker equations.

$$\begin{bmatrix} R_{yy}(0)-\sigma_v^2 & R_{yy}(1) & \cdots & R_{yy}(p-1) \\ R_{yy}(1) & R_{yy}(0)-\sigma_v^2 & \cdots & R_{yy}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{yy}(p-1) & R_{yy}(p-2) & \cdots & R_{yy}(0)-\sigma_v^2 \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{yy}(1) \\ R_{yy}(2) \\ \vdots \\ R_{yy}(p) \end{bmatrix} \quad (3.10)$$

In [20], an iterative method was developed for solving Eqn. (4.4), termed as the noise compensated low-order Yule-Walker equations. The limitations of the method proposed in [20] are briefly discussed below.

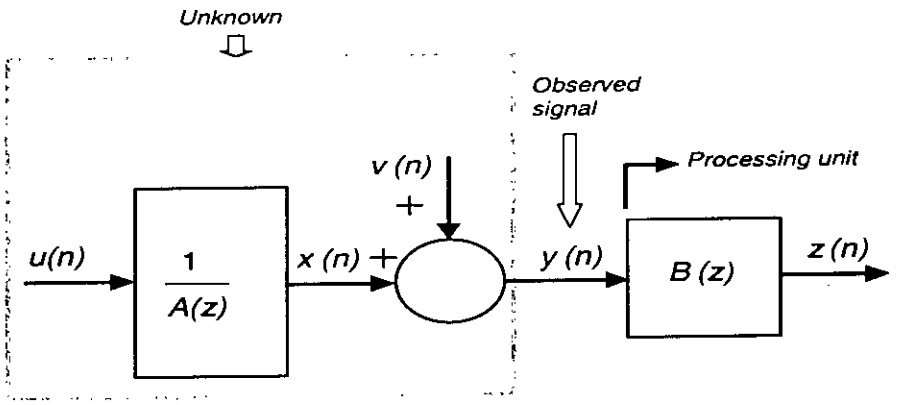

Fig. 3.1: FIR filtering of AR(p) plus noise process.

Consider the AR plus noise process with a processing filter $B(z)$ shown in Fig. 3.1. In order to estimate σ_v^2 and AR parameters from the observed signal $y(n)$, this signal is processed using an auxiliary filter $B(z)$. The polynomial $B(z)$ is defined as

$$B(z) = 1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_p z^{-p} \quad (3.11)$$

Note that the order of $B(z)$ is set equal to the known order p of the AR system and $b_0 = 1$ is assumed without loss of generality. Our objective here is to estimate the noise power σ_v^2 and $\{a_i\}$ $(i = 1, 2, \cdots, p)$ by equating $B(z)$ with $A(z)$. Therefore, $B(z)$ can be regarded as the inverse filter of $H(z)$ in Eqn. (3.4), i.e.,

$$H(z)B(z) = 1 \quad (3.12)$$

when $B(z) = A(z)$ holds. The filtered output $z(n)$ (see Fig. 3.1) of the observed signal $y(n)$ is given by

$$z(n) = y(n) + \sum_{k=1}^{p} b_k y(n-k) \quad (3.13)$$

Substituting Eqn. (3.1) in Eqn. (4.6) gives

$$z(n) = x(n) + \sum_{k=1}^{p} b_k x(n-k) + v(n) + \sum_{k=1}^{p} b_k v(n-k) \tag{3.14}$$

The noise power is estimated in the following way.

The one-step time delayed autocorrelation of $z(n)$ is given by

$$E\left[z(n)z(n-1)\right] = E\left[u'(n)u'(n-1)\right] + \left(b_1 + b_1 b_2 + \cdots + b_{(p-1)} b_p\right) \sigma_v^2 \tag{3.15}$$

where

$$u'(n) = x(n) + \sum_{k=1}^{p} b_k x(n-k) \tag{3.16}$$

Using Eqn. (3.15), the noise power σ_v^2 can be expressed as

$$\sigma_v^2 = \frac{E[z(n)z(n-1)] - E[u'(n)u'(n-1)]}{b_1 + b_1 b_2 + \cdots + b_{(p-1)} b_p} \tag{3.17}$$

As $u'(n)$ cannot be measured, it is impossible to use Eqn. (4.29) directly. Neglecting the term $E[u'(n)u'(n-1)]$ in Eqn. (4.29) gives

$$\bar{\sigma}_v^2(k) = \frac{E\left[z(n)z(n-1)\right]}{b_1 + b_1 b_2 + \cdots + b_{(p-1)} b_p}, \quad k = 1, 2, \cdots \tag{3.18}$$

where k is the iteration number and $\bar{\sigma}_v^2(k)$ is the pseudo noise power obtained from each iteration. The iterative total search technique described in [20] is outlined below.

Since the noise power σ_v^2 is unknown, substituting an auxiliary variable $\alpha(k)$ for σ_v^2 in Eqn. (4.4) we obtain

$$\begin{bmatrix} R_{yy}(0) - \alpha(k) & R_{yy}(1) & \cdots & R_{yy}(p-1) \\ R_{yy}(1) & R_{yy}(0) - \alpha(k) & \cdots & R_{yy}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{yy}(p-1) & R_{yy}(p-2) & \cdots & R_{yy}(0) - \alpha(k) \end{bmatrix}$$
$$\times \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = - \begin{bmatrix} R_{yy}(1) \\ R_{yy}(2) \\ \vdots \\ R_{yy}(p) \end{bmatrix}, \quad \alpha(k) \geq 0 \tag{3.19}$$

Eqn. (4.39) is calculated repeatedly for different arbitrary values of $\alpha(k)$ and every time $\bar{\sigma}_v^2(k)$ is determined from Eqn. (4.8). As $\alpha(k)$ approaches to σ_v^2, $\{b_i\}$ and $u'(n)$ approach to $\{a_i\}$ and $u(n)$, respectively. At the point of equality, the relationship $E[u'(n)u'(n-1)] = 0$ holds. Under this condition both Eqns. (4.29)

and (4.8) give the same value of the noise power. Therefore, use of Eqn. (4.8) instead of Eqn. (4.29) does not affect the desired result. A cost function is then defined as

$$e(k) = |\bar{\sigma}_v^2(k) - \alpha(k)| \tag{3.20}$$

which gives a measure of the error between the auxiliary variable $\alpha(k)$ for noise power and the pseudo noise power $\bar{\sigma}_v^2(k)$. An estimated value of the noise power $\widehat{\sigma}_v^2$ can be obtained from the point where $e(k)$ becomes zero. Here iteration starts from $\alpha(k) = \alpha_{min} = 0$ and finishes at $\alpha(k) = \alpha_{max}$. The maximum limit of $\alpha(k)$, i.e., α_{max} can be calculated by forming a $(p+1) \times (p+1)$ Toeplitz matrix from the vector $[R_{yy}(0), R_{yy}(1), \cdots, R_{yy}(p)]$ and taking the $(p+1)$-th element from the vector found from the singular value decomposition of the Toeplitz matrix. Determination of α_{max} in this way ensures that $\alpha_{max} > \sigma_v^2$ [59]. In each iteration $\alpha(k)$ is incremented by s, where $s = (\alpha_{max} - \alpha_{min})/h$. The value of h depends on the accuracy desired. Depending on the value of s, $\alpha(k)$ can be made to pass through σ_v^2. Obviously, the precision of the iterative algorithm depends on the value of s. The smaller the value of s, the better the resolution one can achieve.

Now we are going to address two problems in the above mentioned iterative technique.

1. As σ_v^2 as well as $\{a_i\}$ are unknown and they are in product form, Eqn. (4.4) possesses inherent nonlinearity. Here $\alpha(k)$ is increased linearly with a step size defined by s and $\bar{\sigma}_v^2(k)$ is calculated in each step and compared with $\alpha(k)$. A solution is reached where the distance as expressed by Eqn. (4.9) is ideally zero. The question arises as the iteration proceeds from $\alpha(0) = \alpha_{min} = 0$ to $\alpha(k) = \alpha_{max}$ whether there exist more than one iterations where the distances are zero. If multiple solutions exist, how can we extract or differentiate the actual solution? Due to the presence of inherent nonlinearity, solutions found throughout the range $0 - \alpha_{max}$ may not be unique. We show in our simulation part that particularly at low SNR the nonlinearity becomes very strong and there exist multiple solutions. Therefore, there might exist more than one solution, unlike the unique solution case described in [20].

2. A large number of iterations is required due to the constant small step size

maintained throughout the total range of search to obtain the desired accuracy. This is computationally expensive and hence a method for reducing the number of iterations without affecting the accuracy is desired.

3.4 The Proposed Method

A new 2-step approach is formulated to overcome the problems mentioned in the preceding section. First, we estimate the true-order AR model parameters and the additive noise power using the iterative method described by Yahagi and Hasan [20] incorporating fuzzy modeling of the cost function for step size variation to reduce the number of iterations. By the term 'true-order AR model' we refer to the AR model with transfer function $H(z) = 1/A(z)$ from input $u(n)$ to output $x(n)$ (see Fig. 3.1). It is assumed that multiple solutions are obtained particularly at low SNRs. Next, we estimate the parameters of the over-fitted AR model to the observed noisy process. By the term 'over-fitted AR model' we refer to the AR model with transfer function $H^o(z) = 1/C(z)$ from input $w(n)$ to output $y(n)$ (refer to Eqn. (3.25)). The parameters of this model are estimated using the lattice filter which possesses high-resolution property. It is known that stability cannot be guaranteed while using the noise compensated lattice filter algorithm [17], [20]. As such, we utilize the uncompensated lattice filter algorithm with the cost paid by adding noise roots to the transfer function of the original AR model. The desired AR parameters are extracted from the over-fitted AR model parameters by using all the solutions obtained from the previous approach. As a by product, we also determine which set, i.e., $\hat{\sigma}_v^2$ and $B(z)$ is closest to the unknown actual values, i.e., σ_v^2 and $A(z)$, respectively, from multiple sets of solution.

3.4.1 Extracting the desired AR parameters

Any AR system corrupted by noise can be modeled by another high-order AR system. As the AR model order increases, y(n) can be more and more accurately characterized by an AR process [60]. It can be shown as follows.

The z-transform of Eqn. (3.1) is given by

$$Y(z) = X(z) + V(z) \tag{3.21}$$

From Eqns. (3.3) and (3.21), we can write

$$Y(z) = \frac{1}{A(z)}U(z) + V(z) \tag{3.22}$$

Pagano [26] and Kay [60] have shown that an AR process corrupted by additive white Gaussian noise can be exactly represented by an ARMA process. As described in [26] and [60], the power spectral density (PSD) of the observed process $y(n)$ is

$$\begin{aligned} P_y(z) &= \frac{\sigma_u^2}{A(z)A(z^{-1})} + \sigma_v^2 \\ &= \frac{\sigma_u^2 + \sigma_v^2 A(z)A(z^{-1})}{A(z)A(z^{-1})} \end{aligned} \tag{3.23}$$

If we write $\sigma_w^2 F(z)F(z^{-1}) = \sigma_u^2 + \sigma_v^2 A(z)A(z^{-1})$, then $y(n)$ may be modeled as

$$\sum_{i=0}^{p} a_i y(n-i) = \sum_{j=0}^{p} f_j w(n-j), \;\; a_0 = b_0 = 1 \tag{3.24}$$

where $F(z) = 1 + \sum_{j=1}^{p} f_j z^{-j}$ and $w(n)$ is a white noise sequence with distribution $\mathcal{N}(0, \sigma_w^2)$. The z-transform of Eqn. (3.24) can be obtained as

$$\begin{aligned} Y(z) &= \frac{F(z)}{A(z)}W(z) \\ &= \frac{1}{A(z)/F(z)}W(z) \\ &\cong \frac{1}{A(z)D(z)}W(z) \\ &\cong \frac{1}{C(z)}W(z) \end{aligned} \tag{3.25}$$

Here, $W(z)$ is the z-transform of the new white noise sequence $w(n)$, $D(z) \cong F(z)^{-1}$ (as only finite number of terms are considered) and $C(z) = A(z)D(z)$. We may write $C(z)$ as

$$C(z) = 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_q z^{-q} \tag{3.26}$$

Since $C(z)$ is the product of $A(z)$ and $D(z)$, it is clear that q, the order of $C(z)$ must be greater than p, the order of $A(z)$. From Eqns. (3.25) and (3.26), we may assert that $C(z)$, consisting of the parameters $\{c_i\}$, where $i = 1, 2, \cdots, q$ can be obtained directly from the noisy data $y(n)$. Moreover, when q is reasonably greater than p, $C(z)$ includes p roots of $A(z)$ and $(q-p)$ roots of $D(z)$ where

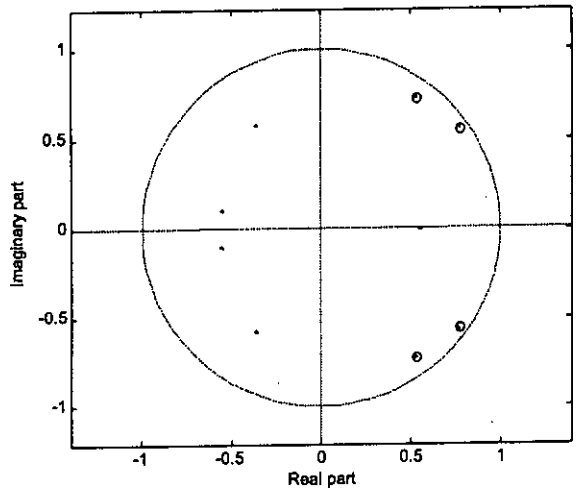

Fig. 3.2: Roots of $A(z)$ and $C(z)$; 'o':roots of $A(z)$, $p = 4$; 'x':roots of $C(z)$, $q = 9$.

the roots of $D(z)$ are solely due to the additive noise $v(n)$. The parameters of $C(z)$ will be estimated by using the lattice filter algorithm [15]. To illustrate this point, let us consider an AR system, $x(n) - 2.595x(n-1) + 3.339x(n-2) - 2.2x(n-3) + 0.731x(n-4) = u(n)$. To obtain $y(n) = x(n) + v(n)$, we assume that $x(n)$ is corrupted by additive white noise $v(n)$ at an SNR of 20 dB. Here, SNR is defined by

$$SNR = 10 \log \frac{\sigma_x^2}{\sigma_v^2} \qquad \text{(dB)} \tag{3.27}$$

$p = 4$, and $A(z) = 1 - 2.595z^{-1} + 3.339z^{-2} - 2.2z^{-3} + 0.731z^{-4}$. Now, using the lattice filter algorithm for high-order AR model with $q = 9$, we can find $C(z) = 1.0000 - 1.3584z^{-1} + 0.6510z^{-2} + 0.4249z^{-3} - 0.1026z^{-4} - 0.1503z^{-5} + 0.0341z^{-6} + 0.2098z^{-7} - 0.0137z^{-8} - 0.0586z^{-9}$ from $y(n)$. In Fig. 3.2, we see that $p = 4$ roots of $A(z)$ match 4 roots of $C(z)$, out of $q = 9$ roots of $C(z)$. The other roots are due to noise.

We now illustrate that using $C(z)$ how the ambiguity raised due to the presence of strong nonlinearity particularly at low SNR can be eliminated. In the case of multiple sets of solution found in step-1 described in the previous section, one solution must be fairly close to the actual noise power σ_v^2 and AR parameters $\{a_i\}$, where $i = 1, 2, \cdots, p$. As $C(z)$ includes the roots of $A(z)$ as well as some extra roots due to noise, we need to develop a method of elimination. To extract the roots of $A(z)$ from $C(z)$, we use the solutions obtained by using the

noise compensated LOYW equations. We denote these solutions as $B_i(z)$, where $i = 1, 2, \cdots, M$, and M denotes the number of solutions. From these multiple solutions, if exist, one solution should be considered at a time for matching with all of the roots of $C(z)$. Obviously, only one solution which has roots closer to the roots of $A(z)$ will match more closely than those of the other solutions. In this way, an estimate of $A(z)$ can be obtained from $C(z)$ using the multiple solutions $B_i(z)$. We propose the following algorithm for separating the desired roots from $C(z)$.

The Algorithm:

1. Find the coefficients of $C(z)$ of the over-fitted AR model using the lattice filter algorithm from the observed sequence $y(n)$. Here, the order of lattice filter q should be reasonably larger than the AR system order p. An estimate of q can be obtained using any standard technique [61].

2. Find the roots of $C(z)$. Let R be the total number of real roots and λ be the total number of complex roots.

3. Determine $B_i(z)$ using the LOYW equations, where $i = 1, 2, \cdots, M$ and M denotes the number of solutions. Let r_i be the total number of real roots and τ_i be total number of complex roots of the ith solution.

4. Calculate the distances, $d^i_{jl} = \left|\beta^i_j - \phi_l\right|$, $j = 1, 2, \cdots, r_i$, $l = 1, 2, \cdots, R$, where

   - β^i_j is the jth real root of $B_i(z)$, $j = 1, 2, \cdots, r_i$.
   - ϕ_l is the lth real root of $C(z)$, $l = 1, 2, \cdots, R$.

5. Find the minimum distances

$$D^i_j = \min_l(d^i_{jl}), \quad j = 1, 2, \cdots, r_i.$$

   Let the real roots of $C(z)$ be R^i_j corresponding to the minimum distances D^i_j, $j = 1, 2, \cdots, r_i$.

6. Calculate the distances, $h^i_{km} = |\zeta^i_k - \psi_m|$, $k = 1, 2, \cdots, \tau_i/2$, $m = 1, 2, \cdots, \lambda/2$, where

- ζ_k^i is the kth complex root of $B_i(z)$, $k = 1, 2, \cdots, \tau_i/2$. Here only the complex roots with positive imaginary parts are considered.
- ψ_m is the mth complex root of $C(z)$, where $m = 1, 2, \cdots, \lambda/2$. Here only the complex roots with positive imaginary parts are considered.

7. Find the minimum distances

$$H_k^i = \min_m(h_{km}^i), \quad k = 1, 2, \cdots, \tau_i/2.$$

Let the complex roots with positive imaginary parts of $C(z)$ be λ_k^i corresponding to the minimum distances H_k^i, $k = 1, 2, \cdots, \tau_i/2$.

8. Calculate the total distance for the ith solution, $e_i = \sum_{j=1}^{r_i} D_j^i + 2\sum_{k=1}^{\tau_i/2} H_k^i$, $i = 1, 2, \cdots, M$. Find the value of $i = i_{min}$ at which e_i is minimum.

9. Form the polynomial $\widehat{A}(z)$ by using the real roots $R_j^{i_{min}}$, $j = 1, 2, \cdots, r_{i_{min}}$, complex roots $\lambda_k^{i_{min}}$, $k = 1, 2, \cdots, \tau_{i_{min}}/2$ along with their respective complex conjugate roots using the following equation:

$$\widehat{A}(z) = [\prod_{j=1}^{r_{min}} (1 - R_j^{i_{min}} z^{-1})] \times [\prod_{k=1}^{\tau_{i_{min}}/2} (1 - \lambda_k^{i_{min}} z^{-1}) \times (1 - \mathrm{conj}(\lambda_k^{i_{min}})\, z^{-1}]$$

where conj$(\cdot)$ denotes the conjugate operation. This algorithm ensures extraction of the roots from $C(z)$, that shows minimum distance with one complete set of roots of $B_i(z)$, $i = 1, 2, \cdots, M$. The extracted roots correspond to the roots of $\widehat{A}(z)$. An estimated value of the noise variance (i.e., $\widehat{\sigma}_v^2$) can be obtained from the i_{min}th solution.

3.4.2 Determining the step size: Fuzzy approach

In the iterative technique explained in Sec.3.3, the step size s is constant throughout the entire range of $\alpha(k)$ from 0 to α_{max}. Depending on the value of s the number of iterations may become large. If high accuracy of the estimated values, i.e., $\widehat{\sigma}_v^2$ and $\widehat{A}(z)$ is desired, the step size s should me made smaller. On the other hand using a larger step size may cause fast convergence with poor estimation or even without any convergence at all. Therefore, the choice of the step size reflects a trade-off between the accuracy of estimation and the speed of convergence.

The value of the cost function defined in Eqn. (4.9) should decrease when a solution is approached and increase whenever a solution is crossed. With this in

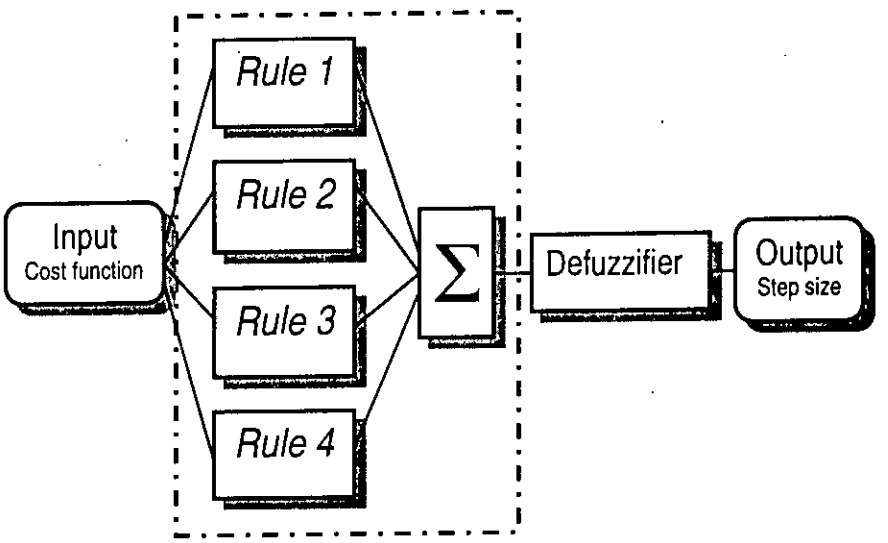

Fig. 3.3: Block diagram for a fuzzy inference system: System consists of 1 input, 1 output, and 4 rules.

mind, we can determine the step size as a function of the cost function instead of keeping it to be constant throughout the iteration range. That is if the cost function becomes smaller then the step size will be reasonably smaller to achieve the solution (to pass $\alpha(k)$ through the point of solution) and if the cost function becomes larger then the step size will be larger to skip the less concerned region of iteration. We should be careful in selecting the maximum step size of an iteration so that it does not miss any solution (i.e., a minimum). The minimum step size depends on our desired accuracy of estimation. In this way, we can reduce the number of iterations significantly without sacrificing the accuracy of estimation since we keep the step size smaller in the region of iteration where we approach a solution.

For determining the step size in accordance with the value of the cost function a Fuzzy Inference System (FIS) is designed which maps nonlinearly the input, cost function, to the appropriate output, the step size. We consider the following points in designing the FIS.

1. A single-input single-output Mamdani fuzzy inference system [62], [63] with 4 rules is chosen.

2. The input to the FIS is the cost function (error index) calculated using Eqn. (4.9). We select the universe of discourse of the cost function to be the value obtained from Eqn. (4.9) by executing the first iteration setting $\alpha(k) = 0$ in Eqn. (4.39).

3. We define four membership functions (MFs) for the input. They are 'very small', 'small', 'significant' and 'large'. The type of membership function and their ranges are depicted in Fig. 4.6. Normalized cost function can be obtained by dividing the cost function by the universe of discourse calculated in step 2.

4. The output from the FIS is the effective step size s. We have calculated the universe of discourse of step size to be $(1/K) \times \alpha_{max}$. We have found $K = 10$ as an well performed approximation. But in some cases where the cost function falls rapidly we may have to increase the value of K to avoid any chance of missing a solution. Where the cost function changes slowly we can choose a smaller value of K which makes the total number of iterations less.

5. We define four membership functions (MFs) for the output also. They are 'very small', 'small', 'significant' and 'large'. The type of membership function and their ranges are shown in Fig. 4.7. Step size is shown in normalized scale. Normalized step size can be obtained by dividing the step size by the universe of discourse calculated in step 4. Actually, we are more concerned of the 'very small' membership function of the output since the accuracy of the iterative technique depends on how the 'very small' MF is defined. We choose the 'very small' MF as a zero-mean Gaussian type membership function

$$f(x; \sigma, c) = e^{-\frac{(x-c)^2}{2\sigma^2}}, \quad c = 0 \tag{3.28}$$

An increase in the value of σ will increase the value of the step size in the region where the cost function is 'very small'. It decreases the total number of iterations but the cost must be paid in estimation accuracy.

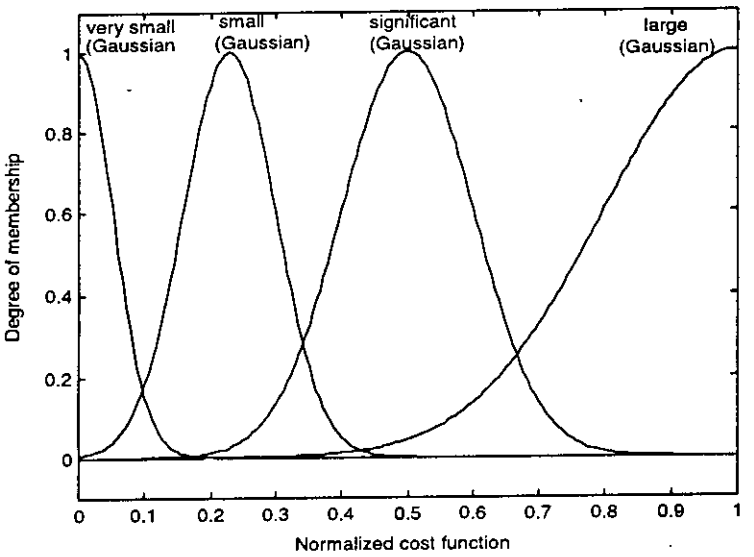

Fig. 3.4: Membership function of the input variable, the cost function, to the FIS.

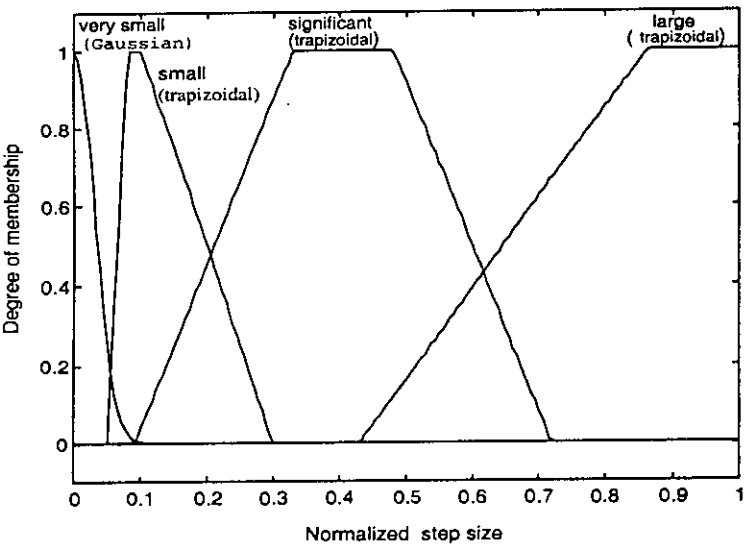

Fig. 3.5: Membership function of the output variable, the step size, to the FIS.

6. The following rules may be implemented in the FIS:

   - Rule 1: If (cost function is very small) then (step size is very small)
   - Rule 2: If (cost function is small) then (step size is small)
   - Rule 3: If (cost function is significant) then (step size is significant)
   - Rule 4: If (cost function is large) then (step size is large)

In this fuzzy model no logical operator is used as there is only one input. The implication method we choose is min (minimum), the aggregation method is max (maximum), and we obtain the crisp output via centroid defuzzification strategy. Fig. 3.6 shows the input-output mapping of the proposed FIS. The following

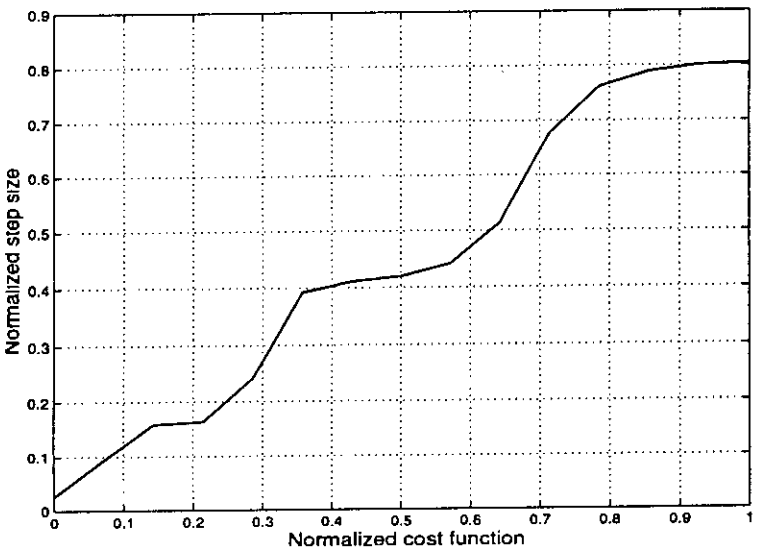

Fig. 3.6: Input-output mapping of the FIS.

steps can be followed to implement the above mentioned FIS.

1. Set the maximum range of the cost function and the maximum step size (universe of discourses) by executing, respectively, step 2 and step 4 of the aforementioned criteria of the FIS. Set the desired accuracy of estimation by adjusting the parameter σ of the Gaussian membership function 'very small' of step size according to step 5.

2. Set the initial value of $\alpha(0) = 0$ and the final value of $\alpha(k) = \alpha_{max}$

3. Calculate the cost function using Eqn. (4.9).

4. Supply 'cost function' calculated in step 3 to the input of the FIS.

5. Determine s, the 'step size', as the output of the FIS.

6. Calculate $\alpha(k) = \alpha(k-1) + s, \qquad k = 1, 2, \cdots$

7. if $\alpha(k) \leq \alpha_{max}$, repeat steps 3 to 6, otherwise terminate.

3.5 Simulation Results

In this section, we examine and illustrate the proposed method using several numerical examples. Consider the following AR system.
Example 1:

$$x(n) = 2.595x(n-1) - 3.339x(n-2) + 2.2x(n-3) \\ -0.731x(n-4) + u(n) \tag{3.29}$$

The observed sequence is $y(n) = x(n) + v(n)$, where $v(n)$ is the additive white noise. We assume that the input noise power σ_u^2 is normalized to unity without loss of generality and the number of data samples $N = 4000$. For this example the order of the AR system is $p = 4$. The estimated AR parameters obtained by the previous method at various SNRs defined by Eqn. (3.27) are presented in Table 3.1. The results shown are the mean value of 20 simulation runs. The quantity in bracket represents standard deviation of respective samples obtained from 20 simulation runs. As can be seen from Table 3.1, though at SNR=30dB the solution is unique, we get two solutions at SNR=20dB and SNR=15dB using the previous method. The term 'Solution 1' in Table 1 refers to the values of $\alpha(k)$ and $B(z)$ corresponding to the first intersection point between $\alpha(k)$ and $\bar{\sigma}_v^2(k)$ (e.g., see Fig. 7). Similarly, 'Solution 2' refers to the values of $\alpha(k)$ and $B(z)$ corresponding to the second intersection point between $\alpha(k)$ and $\bar{\sigma}_v^2(k)$ and so on. The solution points obtained by the previous method at SNR=20dB are graphically shown in Fig. 3.7 for a typical simulation run. From Fig. 3.7 we see that there exist two intersection points. The value of the cost function at the first intersection point is 3.4251×10^{-4} and at the second intersection point is 3.9805×10^{-4}. But only the second intersection corresponds to the actual solution. Therefore, we see that even the value of the cost function is not the least at the actual solution, that is global minimum cannot be guaranteed at the

Table 3.1: Results on noise power and AR parameter estimation by the previous method

SNR (dB)		15	20	30
Solution 1	$\widehat{b_1}$	-1.7564 (0.0547)	-1.7371 (0.0344)	-2.5777 (0.1177)
	$\widehat{b_2}$	1.3973 (0.1026)	1.3556 (0.0662)	3.2508 (0.2799)
	$\widehat{b_3}$	-0.3339 (0.0855)	-0.2967 (0.0555)	-2.1400 (0.274)
	$\widehat{b_4}$	0.0262 (0.0310)	0.0146 (0.0169)	0.6670 (0.1073)
	$\widehat{\sigma_v^2}$	2.784 (0.1928)	0.6202 (0.063)	0.0942 (0.0161)
Solution 2	$\widehat{b_1}$	-2.5200 (0.2148)	-2.5341 (0.0927)	Not found
	$\widehat{b_2}$	3.1660 (0.4908)	3.1986 (0.2147)	Not found
	$\widehat{b_3}$	-2.0356 (0.4683)	-2.0665 (0.2063)	Not found
	$\widehat{b_4}$	0.6694 (0.1733)	0.6815 (0.0778)	Not found
	$\widehat{\sigma_v^2}$	3.1084 (0.1934)	0.9702 (0.0658)	Not found

Table 3.2: Results on noise power and AR parameter estimation by the proposed method

SNR (dB)	15	20	30
$\widehat{a_1}$	-2.6388 (0.0152)	-2.6304 (0.0144)	-2.5946 (0.0253)
$\widehat{a_2}$	3.4054 (0.0363)	3.4029 (0.0356)	3.3381 (0.0538)
$\widehat{a_3}$	-2.2409 (0.0356)	-2.2487 (0.0386)	-2.1983 (0.05)
$\widehat{a_4}$	0.7324 (0.0146)	0.7322 (0.0183)	0.7301 (0.0189)
$\widehat{\sigma_v^2}$	3.0445 (0.2278)	1.0051 (0.0867)	0.0992 (0.0082)
σ_v^2(true value)	3.0767	0.9985	0.1008

[Note: numerical value in the bracket denotes the standard deviation of the respective samples obtained from the 20 simulation runs]

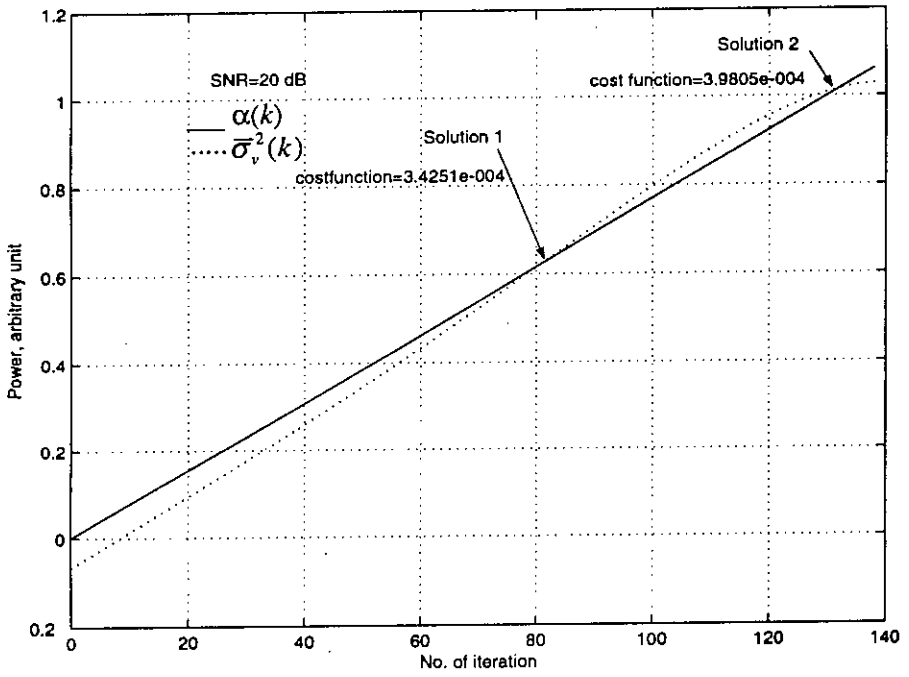

Fig. 3.7: Two cross over of $\alpha(k)$ and $\bar{\sigma}_v^2(k)$ imply two solutions at $SNR = 20$ dB, obtained from a typical simulation run. Here $\alpha(k)$ increases linearly.

real solution. Therefore, this cost function alone is blind to the actual solution when multiple solutions persist.

The results of AR parameter estimation obtained by using the proposed joint technique for the same example and simulation conditions are given in Table 3.2. The outcome of the proposed technique is unique as shown. From Table 3.1 and Table 3.2 it is also evident that the estimated AR parameters by the proposed method are very close to the Solution 1 at 30 dB SNR, Solution 2 at 20 dB as well as 15 dB SNR, obtained by the previous method. But in each case $\hat{A}(z)$ obtained by using the proposed joint technique is more accurate than the solution directly obtained from $B(z)$ of the previous method.

The effect of fuzzy modeling the cost function on step size can be observed vividly from Fig. 3.8. In the proposed method the step size follows the behavior of the cost function as shown in Fig. 3.8. It is small when the cost function is small and it is large when the cost function is large as desired. This results nonlinear behavior in the auxiliary variable $\alpha(k)$ as depicted in Fig. 3.9, unlike its linear behavior shown in Fig. 3.7. Thus it ensures iteration requirement be less and also gives better accuracy of estimation as the step size becomes small or very small where the cost function is smaller in value.

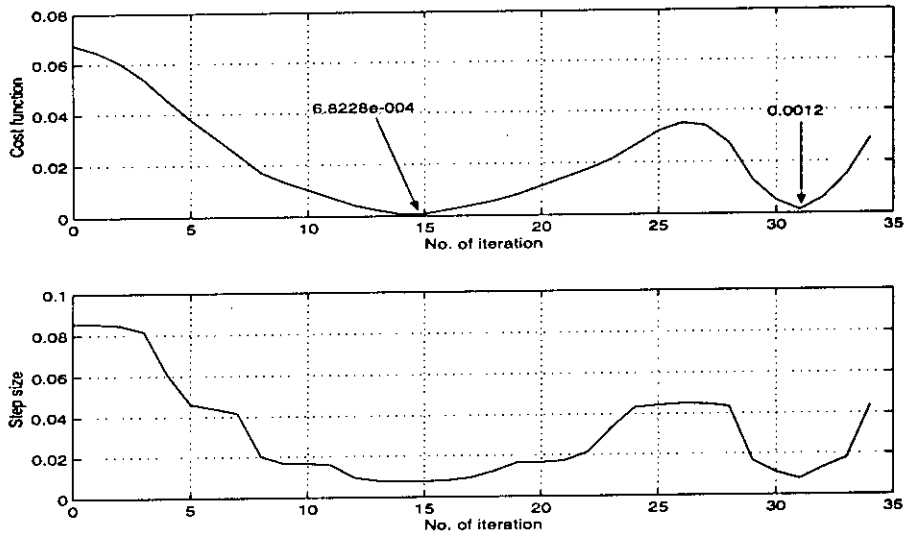

Fig. 3.8: Variation of the cost function and the step size with iteration using the proposed method at $SNR = 20$ dB, for a typical simulation run

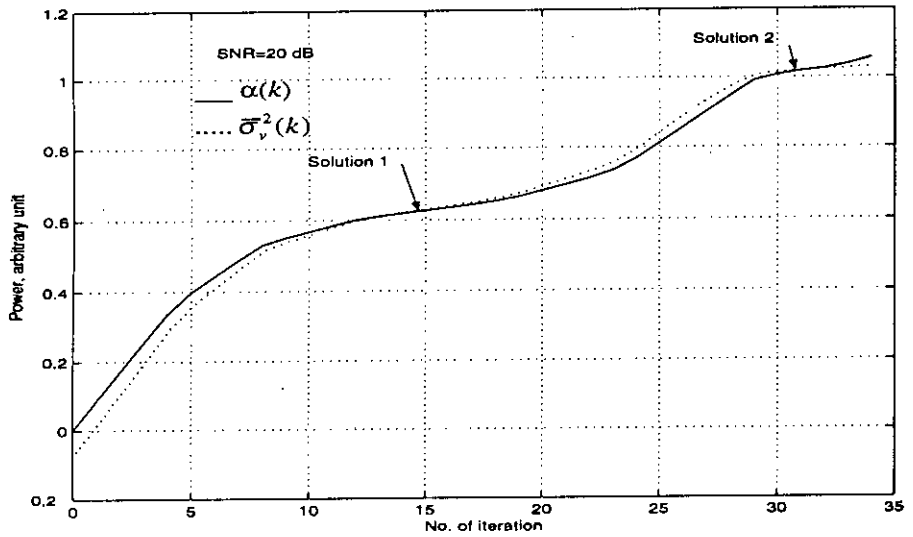

Fig. 3.9: Two cross over of $\alpha(k)$ and $\bar{\sigma}_v^2(k)$ imply two solutions at $SNR = 20$ dB, obtained from a typical simulation run. Here $\alpha(k)$ increases nonlinearly.

Table 3.3: Comparison of the accuracy of estimation and the number of iterations

SNR (dB)	15	20	30
$e^2_{previous}$	0.0664	0.0437	0.0158
$e^2_{proposed}$	0.0080	0.0077	4.6700e-006
Minimum step size	0.0081	0.0077	0.0039
No. of iterations: previous method	391 (23.4819)	133 (7.6784)	38 (2.7198)
No. of iterations: proposed method	32 (9.6206)	38 (3.1473)	22 (0.9403)
Time (sec.) required in iterations: previous method	34.6060 (7.4266)	9.2580 (0.5257)	3.929 (0.2826)
Time (sec.) required in iterations: proposed method	2.4500 (0.7722)	3.7340 (0.3187)	1.4695 (0.0690)

Note: numerical value in the bracket denotes the standard deviation of the respective samples obtained from the 20 simulation runs

A comparison of the previous and the proposed joint methods on accuracy of parameter estimation and computational complexity is presented in Table 3.3. We find that at all SNRs the error $e^2_{proposed}$ is less than the error $e^2_{previous}$. The index $e^2_{previous}$ denotes the sum of the squared error among the true and the estimated AR parameters obtained by the previous method. And $e^2_{proposed}$ is the sum of the squared error among the true and the estimated AR parameters obtained by the proposed method. When multiple solutions are obtined we cannot identify the true solution using the previous method. However, for fair comparison $e^2_{previous}$ is calculated considering the solution closest to the actual one. In the proposed method the AR parameters are calculated from the matched roots of $C(z)$. It is also evident from Table 3.3 that in our method, the required number of iterations as well as the time required for these iterations is less than that of the previous constant step size total search technique. Here the mean value of no. of iterations obtained from 20 simulation runs is rounded to the next integer value. For the system described by Eqn. (3.29) we notice that the variation of the cost function is rather flat, particularly at high SNR, e.g., 30 dB, and consequently the step size variation becomes less. Hence, as the step size remains more or less constant the reduction of the number of iterations is less at higher SNR for this system using the proposed method. But we see that at SNR=20 dB and SNR=15 dB

the number of iterations has reduced significantly by our method.

In Table 3.4 and Table 3.5, we present summarized results for other examples keeping the simulation conditions same. We see that for each example the proposed method performs better both in terms of accuracy and computational requirement than the previous one.

Finally, the power spectrum plot is shown in Fig. 3.10 for the system in Example 1. We see that the estimated spectrum by our method is much closer to the actual spectrum than that of the previous method. The variance of the estimation is also found satisfactory which can be observed from Fig. 3.11. For other examples also we have obtained similar results.

3.6 Conclusion

A hybrid method for autoregressive (AR) spectral estimation from noisy observations has been proposed. The major focus of the paper has been to develop a highly accurate estimation algorithm for AR parameters from the noise corrupted signal to eliminate the smoothing effect of noise on AR spectrum. To achieve this goal, unlike conventional approaches, noise uncompensated lattice filter (LF) algorithm, to avoid any chance of instability, is employed to estimate the parameters of a high-order AR model fitted to the observed process. It is shown that this model preserves the true AR system roots precisely with some extra roots due to noise. A root matching technique has been adopted to extract the true system roots from the mixture of noise plus system roots. This technique has utilized the solution(s), single or multiple, from the low-order Yule-Walker (LOYW) equations as auxiliary solution. It has been shown that the iterative method based on the noise compensated LOYW equations [20] is blind to the desired solution of AR parameters and noise power particularly at low SNR due to strong inherent nonlinearity. The proposed scheme has also resolved this problem of multiple solutions case of the LOYW equations.

We have also proposed a nonlinear iterative technique for finding the auxiliary solution using the LOYW equations. Nonlinear iteration is performed by fuzzy modeling the cost function. The incremental step size in each successive iteration is obtained adaptively with the cost function as the output of a fuzzy inference system. This has reduced the number of iterations significantly as compared to

Table 3.4: Results on noise power and AR parameter estimation by the previous and proposed methods for AR(2) and AR(6) systems

AR system	Previous method					Proposed method			
	SNR (dB)		5	15	30	SNR (dB)	5	15	30
Example 2	Solution 1	$\hat{b}_1$	-0.7356 (0.0091)	-1.7983 (0.0104)	-1.7981 (0.0045)	$\hat{a}_1$	-1.7692 (0.0042)	-1.7989 (0.0056)	-1.7985 (0.0036)
		$\hat{b}_2$	0.0009 (0.0017)	0.9616 (0.0103)	0.9681 (0.0043)	$\hat{a}_2$	0.9364 (0.0042)	0.9662 (0.0054)	0.9686 (0.0040)
$a_1 = -1.8$		$\hat{\sigma}_v^2$	7.2060 (2.0124)	3.2936 (0.5045)	0.1006 (0.0113)	$\hat{\sigma}_v^2$	31.7009 (4.1391)	3.1782 (0.4114)	0.0979 (0.0119)
	Solution 2	$\hat{b}_1$	-1.7587 (0.0546)	Not found	Not found	σ_v^2(true value)	31.8209	3.2006	0.0978
$a_2 = 0.97$		$\hat{b}_2$	0.9324 (0.0491)	Not found	Not found				
		$\hat{\sigma}_v^2$	31.3907 (3.5834)	Not found	Not found				
Example 3	Solution 1	$\hat{b}_1$	-0.2920 (0.0293)	-0.8587 (0.0389)	-0.8435 (0.0260)	$\hat{a}_1$	-0.8490 (0.1088)	-0.8492 (0.0357)	-0.8584 (0.0274)
		$\hat{b}_2$	0.4149 (0.0395)	1.6512 (0.0593)	1.6473 (0.0385)	$\hat{a}_2$	1.6615 (0.2539)	1.6626 (0.0322)	1.6717 (0.0269)
		$\hat{b}_3$	0.3269 (0.0244)	-0.7602 (0.0787)	-0.7248 (0.0544)	$\hat{a}_3$	-0.7483 (0.3097)	-0.7541 (0.0432)	-0.7538 (0.0316)
$a_1 = -0.8484$		$\hat{b}_4$	0.1967 (0.0138)	1.4191 (0.0693)	1.3827 (0.0479)	$\hat{a}_4$	1.3778 (0.2601)	1.3913 (0.0509)	1.4148 (0.0337)
		$\hat{b}_5$	0.0365 (0.0200)	-0.5694 (0.0393)	-0.5400 (0.0366)	$\hat{a}_5$	-0.5621 (0.1242)	-0.5458 (0.0342)	-0.5639 (0.0216)
$a_2 = 1.6590$		$\hat{b}_6$	-0.0287 (0.0243)	0.6256 (0.0290)	0.6089 (0.0204)	$\hat{a}_6$	0.6328 (0.0710)	0.6044 (0.0274)	0.6260 (0.0179)
		$\hat{\sigma}_v^2$	1.0834 (0.6082)	0.4487 (0.0431)	0.0133 (0.0052)	$\hat{\sigma}_v^2$	4.4221 (0.3575)	0.4494 (0.0320)	0.0158 (0.0050)
$a_3 = -0.7372$	Solution 2	$\hat{b}_1$	-0.3298 (0.0339)	Not found	Not found	σ_v^2(true value)	4.4959	0.4506	0.0142
		$\hat{b}_2$	0.5050 (0.0398)	Not found	Not found				
$a_4 = 1.4044$		$\hat{b}_3$	0.3255 (0.0368)	Not found	Not found				
		$\hat{b}_4$	0.2257 (0.0233)	Not found	Not found				
$a_5 = -0.5556$		$\hat{b}_5$	0.0558 (0.0269)	Not found	Not found				
		$\hat{b}_6$	0.0167 (0.0237)	Not found	Not found				
$a_6 = 0.6236$		$\hat{\sigma}_v^2$	2.2367 (0.4900)	Not found	Not found				
	Solution 3	$\hat{b}_1$	-0.8610 (0.1495)	Not found	Not found				
		$\hat{b}_2$	1.6525 (0.3553)	Not found	Not found				
		$\hat{b}_3$	-0.7587 (0.4029)	Not found	Not found				
		$\hat{b}_4$	1.3985 (0.4456)	Not found	Not found				
		$\hat{b}_5$	-0.5657 (0.2445)	Not found	Not found				
		$\hat{b}_6$	0.6059 (0.2017)	Not found	Not found				
		$\hat{\sigma}_v^2$	4.4618 (0.2654)	Not found	Not found				

Note: numerical value in the bracket denotes the standard deviation of the respective samples obtained from the 20 simulation runs

Table 3.5: Comparison of the accuracy of estimation and the number of iterations for AR(2) and AR(6) systems

AR System	SNR (dB)	5	15	30
Example 2	$e^2_{previous}$	0.0031	7.3450e-005	7.2200e-006
	$e^2_{proposed}$	0.0021	1.5650e-005	4.2100e-006
$a_1 = -1.8$	Minimum step size	0.0571	0.0190	0.0015
$a_2 = 0.97$	No. of iterations: Previous method	559 (62.774)	185 (24.8995)	198 (7.2909)
	No. of iterations: Proposed method	28 (2.3725)	40 (1.5174)	43 (1.6255)
	Time (sec.) required in iterations: previous method	44.07730 (4.8969)	12.7010 (1.7113)	13.54 (0.4998)
	Time (sec.) required in iterations: proposed method	1.8920 (0.1695)	2.7670 (0.1019)	3.0935 (0.1614)
Example 3	$e^2_{previous}$	0.0011	0.0011	0.0012
	$e^2_{proposed}$	9.6427e-004	9.3550e-004	7.1966e-004
$a_1 = -0.8484$ $a_2 = 1.6590$ $a_3 = -0.7372$ $a_4 = 1.4044$ $a_5 = -0.5556$ $a_6 = 0.6236$	Minimum step size	0.0389	0.0023	0.0017
	No. of iterations: Previous method	119 (5.7866)	232 (16.3272)	60 (2.5189)
	No. of iterations: Proposed method	83 (12.2022)	144 (2.2804)	43 (6.8748)
	Time (sec.) required in iterations: previous method	8.2345 (0.4058)	16.16 (1.5502)	4.1833 (0.1785)
	Time (sec.) required in iterations: proposed method	5.8610 (0.8976)	10.232 (0.1599)	3.0280 (0.4990)

Note: numerical value in the bracket denotes the standard deviation of the respective samples obtained from the 20 simulation runs

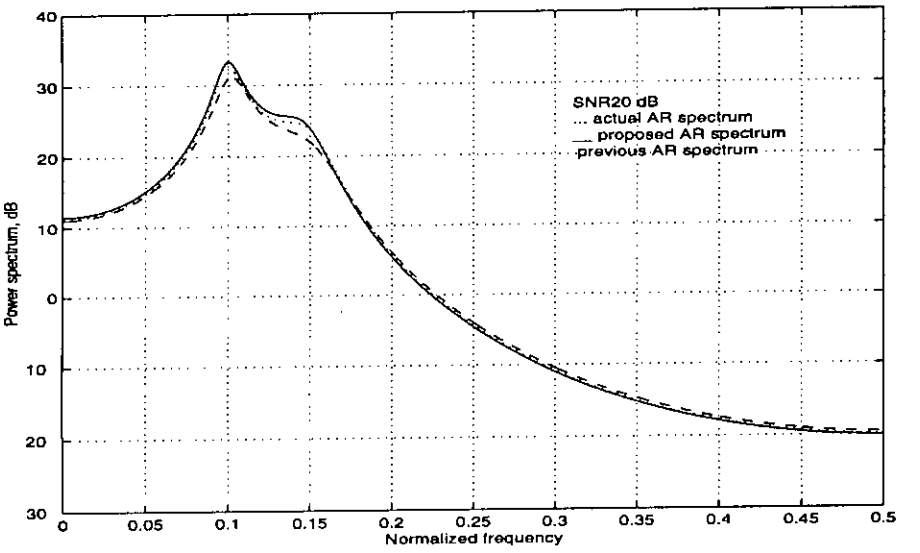

Fig. 3.10: Comparison of the spectral estimation results obtained from the previous and the proposed methods.

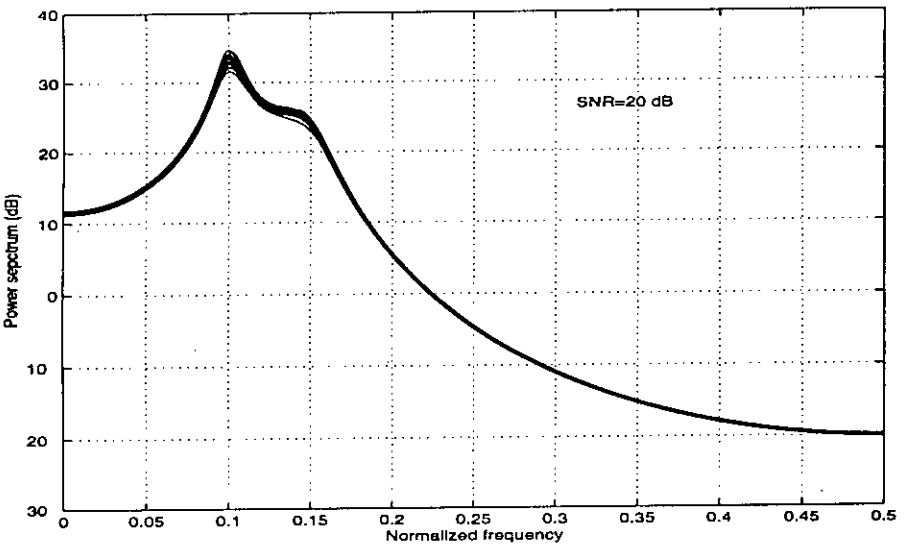

Fig. 3.11: Estimated power spectrum by the proposed method for twenty simulation runs.

the previous one where a constant step size has been maintained throughout the total range of iteration. Clearly, the proposed method is not one-pass and utilizes a joint approach for spectral estimation. Nevertheless, if one is ready to pay the price, the gain in terms of performance, especially in difficult situations such as low signal-to-noise ratio or closely "spaced" signals may be significant.

Chapter 4

Multichannel Autoregressive Spectral Estimation from Noisy Observations

4.1 Introduction

The spectral estimation of a system using a parametric method involves fitting any one of the following; 1) autoregressive (AR) model and 2) autoregressive moving average (ARMA) model. The unknown parameters of the assumed model must be identified using a finite set of data obtained from the physical system. When noise-free input-output data are available, many of the algorithms proposed so far can be used to estimate the system parameters and consequently the power spectrum. However in most practical cases the output signals cannot be measured accurately due to the presence of noise. Therefore, the problem of estimation of multichannel AR spectrum using noise-corrupted output is one of the central issues of modern signal processing, providing a powerful means of extracting useful information from a given data set.

Several methods have been proposed for the spectral estimation of single-input single-output AR system using noise-corrupted output signals. If the additive noise is white, then the bias effect of the noise may be reduced using noise power cancellation schemes [17], [64].

Multichannel generalization, which is particularly useful in the sonar and geo-seismic fields, of the Burg method and its variants has been examined by Marple and Nuttall [65] for the noise-free case and by Gupta and Fairman [66] for the case with noise. The latter approach, however, requires input data as well as the

noise-contaminated output data. Recently, a method based on the maximum likelihood estimation was proposed for the identification of multichannel AR models [67]. However, in this method it is assumed that the observations are noise-free. In other methods based on the scalar representation of the multichannel processes [68], it is also considered that the observations are noise-free, an assumption that is seldom valid in practice.

In this paper, we propose two methods for the spectral estimation of multichannel AR processes contaminated by additive white observation noise. The noise-compensated multichannel Yule-Walker method is used to estimate the parameters. Accurate noise compensation is shown to be feasible by means of new methods for estimating the multichannel observation noise variances. The methods, based on Newton-Raphson technique and Fuzzy iterative searh, are proposed for estimating the observation noise variances. The multichannel spectral estimation scheme that we introduce here is based on the method to estimate the parameters proposed by Yahagi and Hasan [20] for single-channel AR processes.

4.2 Problem Formulation

Consider the M-channel pth-order AR process

$$\mathbf{x}(n) = -\sum_{l=1}^{p} \mathbf{A}^{(l)}\mathbf{x}(n-l) + \mathbf{u}(n) \tag{4.1}$$

$$\mathbf{A}^{(l)} = \begin{bmatrix} a_{11}^{(l)} & a_{12}^{(l)} & \cdots & a_{1M}^{(l)} \\ a_{21}^{(l)} & a_{22}^{(l)} & \cdots & a_{2M}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}^{(l)} & a_{M2}^{(l)} & \cdots & a_{MM}^{(l)} \end{bmatrix}$$

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_M(n) \end{bmatrix}, \quad \mathbf{u}(n) = \begin{bmatrix} u_1(n) \\ u_2(n) \\ \vdots \\ u_M(n) \end{bmatrix},$$

where $\mathbf{A}^{(l)}$ $(l = 1, 2, \cdots, p)$ are the $M \times M$ autoregressive parameter matrices, $\mathbf{x}(n)$ represents the $M \times 1$ output vector and $\mathbf{u}(n)$ is the $M \times 1$ vector representing the input signal. It is assumed to be zero-mean stationary multichannel white noise. The signal components $x_i(n)$ and $u_i(n)$ $(i = 1, 2, \cdots, M)$ denote, respectively, the output and input of the ith channel. The order p of the autoregression is assumed

to be known and the AR parameter matrices $\mathbf{A}^{(l)}$ are constrained such that the roots of

$$\det\ \mathbf{A}_p(z) = 0 \tag{4.2}$$

lie inside the unit circle [69], where

$$\mathbf{A}_p(z) = \mathbf{I} + \mathbf{A}^{(1)}z^{-1} + \mathbf{A}^{(2)}z^{-2} + \cdots + \mathbf{A}^{(p)}z^{-p} \tag{4.3}$$

$\mathbf{I}$ is the $M \times M$ identity matrix, and z^{-1} is the backward shift operator, i.e., $z^{-1}x(n) = x(n-1)$. The matrix transfer function from input to output of an M-channel pth-order AR process can be represented as

$$\mathbf{H}(z) = \mathbf{A}_p^{-1}(z) \tag{4.4}$$

Now the multichannel AR power spectrum [65] of $\mathbf{x}(n)$ is given by

$$\mathbf{P}_{AR}(z^{-1}) = (\mathbf{H}^*(\frac{1}{z^*}))^{-1}(\mathbf{H}^T(z))^{-1} \tag{4.5}$$

evaluated on unit circle i.e., $z = e^{j2\pi f}$. Clearly, $\mathbf{P}_{AR}(f)$ is Hermitian, i.e., $\mathbf{P}_{AR}^{ij}(f) = (\mathbf{P}_{AR}^{ji}(f))^*$.

In many practical situations, observation noise corrupts the data samples. We assume that $\mathbf{y}(n)$ is the noise-corrupted output of an M-channel AR process

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) \tag{4.6}$$

$$\mathbf{y}(n) = \begin{bmatrix} y_1(n) \\ y_2(n) \\ \vdots \\ y_M(n) \end{bmatrix}, \quad \mathbf{v}(n) = \begin{bmatrix} v_1(n) \\ v_2(n) \\ \vdots \\ v_M(n) \end{bmatrix},$$

where the $M \times 1$ vector $\mathbf{v}(n)$ denotes additive zero-mean stationary white observation noise of unknown covariance matrix $\mathbf{\Sigma_V} = \text{diag}\left(\sigma_{v_1}^2, \sigma_{v_2}^2, \cdots, \sigma_{v_M}^2\right)$. The signal components $y_i(n)$ and $v_i(n)$ $(i = 1, 2, \cdots, M)$ denote, respectively, the noise-corrupted output and the observation noise on the ith channel. It is assumed that the observation noise $\mathbf{v}(n)$ is uncorrelated with the input $\mathbf{u}(n)$, i.e., $E[\mathbf{v}(n)\mathbf{u}^T(t)] = \mathbf{0}$ for all n and t, where $E[\cdot]$ is the expectation operator and T denotes the transpose operation.

Let

$$\mathbf{Y} = [\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(N-1)]^T \tag{4.7}$$

denote the vector containing N observed samples from the M-channel stationary AR process. In row block vector form the unknown AR coefficient matrices are defined as

$$\underline{\mathbf{a}} = [\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(p-1)}, \mathbf{A}^{(p)}] \tag{4.8}$$

Therefore, from Eq. (4.5) it is clear that the problem of spectrum estimation by AR method solely from a set of noisy observations $\mathbf{Y}$ depends on finding the values of $\mathbf{A}^{(l)}$, the AR parameters, where $l = 1, 2, \cdots, p$.

4.3 Parameter Estimation Method

If $\mathbf{v}(n) = \mathbf{0}$, then the Yule-Walker equations for an M-channel AR process [1, Chap. 15] are

$$\underline{\mathbf{a}}\,\underline{\mathbf{R}} = -\underline{\mathbf{r}} \tag{4.9}$$

$$\begin{aligned} \underline{\mathbf{R}} &= \begin{bmatrix} \mathbf{R_{xx}}(0) & \cdots & \mathbf{R_{xx}}(p-1) \\ \mathbf{R_{xx}}(-1) & \cdots & \mathbf{R_{xx}}(p-2) \\ \vdots & \ddots & \vdots \\ \mathbf{R_{xx}}(-p+1) & \cdots & \mathbf{R_{xx}}(0) \end{bmatrix} \\ \underline{\mathbf{r}} &= [\mathbf{R_{xx}}(1), \mathbf{R_{xx}}(2), \cdots, \mathbf{R_{xx}}(p)], \end{aligned}$$

where $\underline{\mathbf{R}}$ is the $p \times p$ block Toeplitz matrix (i.e., a matrix with equal submatrices, or blocks, on any block diagonal) with block entries $\mathbf{R_{xx}}(m)$ each of which has dimension $M \times M$, $\underline{\mathbf{r}}$ is the $1 \times p$ block row vector and $\mathbf{R_{xx}}(m) = E[\mathbf{x}(n)\mathbf{x}^T(n-m)]$ $(m = 0, 1, \cdots, p)$ with $\mathbf{R_{xx}}(-m) = \mathbf{R}_{\mathbf{xx}}^T(m)$. A recursive technique can be used to solve Eq. (4.9) which greatly reduces the amount of computation required [70], [71].

When noise is present, we cannot estimate $\mathbf{R_{xx}}(m)$ since only $\mathbf{y}(n)$ is known. However, using the relation

$$\mathbf{R_{yy}}(m) = \begin{cases} \mathbf{R_{xx}}(0) + \mathbf{R_{vv}}(0), & m = 0 \\ \mathbf{R_{xx}}(m), & m \neq 0 \end{cases} \tag{4.10}$$

the block matrix $\underline{\mathbf{R}}$ in Eq. (4.9) can be calculated by replacing $\mathbf{R_{xx}}(0) = \mathbf{R_{yy}}(0) - \mathbf{\Sigma_v}$ and $\mathbf{R_{xx}}(m) = \mathbf{R_{yy}}(m)$ for $m \neq 0$. where $\mathbf{R_{yy}}(m) = E[\mathbf{y}(n)\mathbf{y}^T(n-m)]$ $(m = 0, 1, \cdots, p)$ and $\mathbf{\Sigma_v} = \mathbf{R_{vv}}(0)$ is the $M \times M$ diagonal matrix if the noise is uncorrelated from channel to channel, i.e., $E[v_i(n)v_j(t)] = 0$ for $i \neq j$ and for all n and t. The diagonal element $\sigma_{v_i}^2$ $(i = 1, 2, \cdots, M)$ denote the variance of

the observation noise on the ith channel. We are now faced with another problem since the noise variance on each channel must be known *a priori* in order to calculate $\underline{\mathbf{R}}$ from Eq. (4.9). In most practical cases, noise variances are not known in advance. Therefore, in order to estimate the AR parameter matrices a method for estimating the noise variances is required. This can be done using the following algorithm. Consider an M-channel AR process corrupted by noise

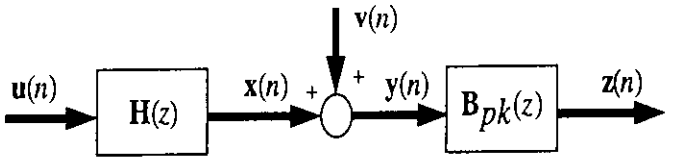

Fig. 4.1: Inverse filtering of AR(p) plus noise system.

with a filter $\mathbf{B}_{pk}(z)$ as shown in Fig. 4.1. The matrix polynomial $\mathbf{B}_{pk}(z)$ is defined as

$$\mathbf{B}_{pk}(z) = \mathbf{I} + \mathbf{B}^{(1k)}z^{-1} + \mathbf{B}^{(2k)}z^{-2} + \cdots + \mathbf{B}^{(pk)}z^{-p} \tag{4.11}$$

where k is the iteration index. Note that the order of $\mathbf{B}_{pk}(z)$ is set equal to the known order p of the M-channel AR process and $\mathbf{B}^{(0k)} = \mathbf{I}$ is assumed without loss of generality. Our objective here is to estimate the noise covariance matrix $\mathbf{\Sigma_V}$ and $\underline{\mathbf{a}}$ by matching $\mathbf{B}_{pk}(z)$ with $\mathbf{A}_p(z)$. Therefore, $\mathbf{B}_{pk}(z)$ can be regarded as the inverse filter of $\mathbf{H}(z)$ in Eq. (4.4), i.e.,

$$\mathbf{H}(z)\mathbf{B}_{pk}(z) = \mathbf{I} \tag{4.12}$$

The filtered output $\mathbf{z}(n)$ (see Fig. 4.1) is given by

$$\mathbf{z}^{(k)}(n) = \mathbf{y}(n) + \sum_{l=1}^{p} \mathbf{B}^{(lk)}\mathbf{y}(n-l). \tag{4.13}$$

Substituting Eq. (4.6) into Eq. (4.13) gives

$$\mathbf{z}^{(k)}(n) = \mathbf{x}(n) + \sum_{l=1}^{p} \mathbf{B}^{(lk)}\mathbf{x}(n-l) + \mathbf{v}(n) + \sum_{l=1}^{p} \mathbf{B}^{(lk)}\mathbf{v}(n-l). \tag{4.14}$$

Then using the time-delayed autocorrelation lags of $\mathbf{z}(n)$, we get

$$E\left[\mathbf{z}^{(k)}(n)\mathbf{z}^{(k)T}(n-1)\right] = E\left[\mathbf{u}^{(k)}(n)\mathbf{u}^{(k)T}(n-1)\right] + \left(\mathbf{B}^{(1k)}\mathbf{\Sigma_V} + \mathbf{B}^{(2k)}\mathbf{\Sigma_V}\mathbf{B}^{(1k)T} + \cdots + \mathbf{B}^{(pk)}\mathbf{\Sigma_V}\mathbf{B}^{((p-1)k)T}\right) \tag{4.15}$$

where

$$\mathbf{u}^{(k)}(n) = \mathbf{x}(n) + \sum_{l=1}^{p} \mathbf{B}^{(lk)}\mathbf{x}(n-l) \tag{4.16}$$

Substituting $E\left[\mathbf{u}^{(k)}(n)\mathbf{u}^{(k)T}(n-1)\right] = \mathbf{R}_{\mathbf{u}}^{(k)}(1)$ and $E\left[\mathbf{z}^{(k)}(n)\mathbf{z}^{(k)T}(n-1)\right] = \mathbf{R}_{\mathbf{z}}^{(k)}(1)$ into Eq. (4.15), we get

$$\mathbf{B}^{(1k)}\mathbf{\Sigma_V} + \mathbf{B}^{(2k)}\mathbf{\Sigma_V}\mathbf{B}^{(1k)T} \cdots + \mathbf{B}^{(pk)}\mathbf{\Sigma_V}\mathbf{B}^{((p-1)k)T} = \mathbf{R}_{\mathbf{z}}^{(k)}(1) - \mathbf{R}_{\mathbf{u}}^{(k)}(1) \tag{4.17}$$

If the additive noise on each channel is unequal, in general, we can find the noise covariances (diagonal elements of matrix $\mathbf{\Sigma_V}$) as follows. We define

$$\mathbf{B}^{(lk)} = \begin{bmatrix} b_{11}^{(lk)} & b_{12}^{(lk)} & \cdots & b_{1M}^{(lk)} \\ b_{21}^{(lk)} & b_{22}^{(lk)} & \cdots & b_{2M}^{(lk)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{M1}^{(lk)} & b_{M2}^{(lk)} & \cdots & b_{MM}^{(lk)} \end{bmatrix} \tag{4.18}$$

then lth $(l = 1, 2, \cdots, p)$ term of left side of Eq. (4.17)

$$\begin{aligned} \mathbf{T}^{(lk)} &= \mathbf{B}^{(lk)}\mathbf{\Sigma_V}\mathbf{B}^{((l-1)k)T} \\ &= \begin{bmatrix} b_{11}^{(lk)} & b_{12}^{(lk)} & \cdots & b_{1M}^{(lk)} \\ b_{21}^{(lk)} & b_{22}^{(lk)} & \cdots & b_{2M}^{(lk)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{M1}^{(lk)} & b_{M2}^{(lk)} & \cdots & b_{MM}^{(lk)} \end{bmatrix} \times \begin{bmatrix} \sigma_{v_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{v_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{v_M}^2 \end{bmatrix} \\ &\quad \times \begin{bmatrix} b_{11}^{((l-1)k)} & b_{12}^{((l-1)k)} & \cdots & b_{1M}^{((l-1)k)} \\ b_{21}^{((l-1)k)} & b_{22}^{((l-1)k)} & \cdots & b_{2M}^{((l-1)k)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{21}^{((l-1)k)} & b_{22}^{((l-1)k)} & \cdots & b_{2M}^{((l-1)k)} \end{bmatrix} \\ &= \begin{bmatrix} T_{11}^{(lk)} & T_{12}^{(lk)} & \cdots & T_{1M}^{(lk)} \\ T_{21}^{(lk)} & T_{22}^{(lk)} & \cdots & T_{2M}^{(lk)} \\ \vdots & \vdots & \ddots & \vdots \\ T_{M1}^{(lk)} & T_{M2}^{(lk)} & \cdots & T_{MM}^{(lk)} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} T_{11}^{(lk)} &= \sigma_{v_1}^2 b_{11}^{(lk)} b_{11}^{((l-1)k)} + \sigma_{v_2}^2 b_{12}^{(lk)} b_{12}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{1M}^{(lk)} b_{1M}^{((l-1)k)} \\ T_{12}^{(lk)} &= \sigma_{v_1}^2 b_{11}^{(lk)} b_{21}^{((l-1)k)} + \sigma_{v_2}^2 b_{12}^{(lk)} b_{22}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{1M}^{(lk)} b_{2M}^{((l-1)k)} \\ T_{1M}^{(lk)} &= \sigma_{v_1}^2 b_{11}^{(lk)} b_{M1}^{((l-1)k)} + \sigma_{v_2}^2 b_{12}^{(lk)} b_{M2}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{1M}^{(lk)} b_{MM}^{((l-1)k)} \\ T_{21}^{(lk)} &= \sigma_{v_1}^2 b_{21}^{(lk)} b_{11}^{((l-1)k)} + \sigma_{v_2}^2 b_{22}^{(lk)} b_{12}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{2M}^{(lk)} b_{1M}^{((l-1)k)} \end{aligned}$$

$$
\begin{aligned}
T_{22}^{(lk)} &= \sigma_{v_1}^2 b_{21}^{(lk)} b_{21}^{((l-1)k)} + \sigma_{v_2}^2 b_{22}^{(lk)} b_{22}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{2M}^{(lk)} b_{2M}^{((l-1))k} \\
T_{2M}^{(lk)} &= \sigma_{v_1}^2 b_{21}^{(lk)} b_{M1}^{((l-1)k)} + \sigma_{v_2}^2 b_{22}^{(lk)} b_{M2}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{2M}^{(lk)} b_{MM}^{((l-1)k)} \\
T_{M1}^{(lk)} &= \sigma_{v_1}^2 b_{M1}^{(lk)} b_{11}^{((l-1)k)} + \sigma_{v_2}^2 b_{M2}^{(lk)} b_{12}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{MM}^{(lk)} b_{1M}^{((l-1)k)} \\
T_{M2}^{(lk)} &= \sigma_{v_1}^2 b_{M1}^{(lk)} b_{21}^{((l-1)k)} + \sigma_{v_2}^2 b_{M2}^{(lk)} b_{22}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{MM}^{(lk)} b_{2M}^{((l-1)k)} \\
T_{MM}^{(lk)} &= \sigma_{v_1}^2 b_{M1}^{(lk)} b_{M1}^{((l-1)k)} + \sigma_{v_2}^2 b_{M2}^{(lk)} b_{M2}^{((l-1)k)} + \cdots + \sigma_{v_M}^2 b_{MM}^{(lk)} b_{MM}^{((l-1)k)}
\end{aligned}
$$

Adding the matrices of the left side of Eq. (4.17) and equating the diagonal elements of the resultant matrix to the corresponding diagonal elements of the right side difference matrix, we get

$$\sigma_{v_1}^2 b_{11}^{(1k)} + T_{11}^{(2k)} + \cdots + T_{11}^{(pk)} = Z_{11}^{(k)}(1) - U_{11}^{(k)}(1) \tag{4.19}$$

$$\sigma_{v_2}^2 b_{22}^{(1k)} + T_{22}^{(2k)} + \cdots + T_{22}^{(pk)} = Z_{22}(1) - U_{22}^{(k)}(1) \tag{4.20}$$

$$\vdots$$

$$\sigma_{v_M}^2 b_{MM}^{(1k)} + T_{MM}^{(2k)} + \cdots + T_{MM}^{(pk)} = Z_{MM}(1) - U_{MM}^{(k)}(1) \tag{4.21}$$

Putting the values of $T_{ij}^{(lk)} (l = 1, 2, \cdots, p \text{ and } i = j = 1, 2, \cdots, M)$ in above Eqs. and rearranging, we get

$$\begin{aligned}
&\sigma_{v_1}^2 \left(b_{11}^{(1k)} + b_{11}^{(2k)} b_{11}^{(1k)} + \cdots + b_{11}^{(pk)} b_{11}^{((p-1)k)} \right) + \sigma_{v_2}^2 \left(b_{12}^{(2k)} b_{12}^{(1k)} + \cdots + b_{12}^{(pk)} b_{12}^{((p-1)k)} \right) \\
&+ \cdots + \sigma_{v_M}^2 \left(b_{1M}^{(2k)} b_{1M}^{(1k)} + \cdots + b_{1M}^{(pk)} b_{1M}^{((p-1)k)} \right) = Z_{11}^{(k)}(1) - U_{11}^{(k)}(1)
\end{aligned} \tag{4.22}$$

$$\begin{aligned}
&\sigma_{v_1}^2 \left(b_{21}^{(2k)} b_{21}^{(1k)} + \cdots + b_{21}^{(pk)} b_{21}^{((p-1)k)} \right) + \sigma_{v_2}^2 \left(b_{22}^{(1k)} + b_{12}^{(2k)} b_{12}^{(1k)} + \cdots + b_{12}^{(pk)} b_{12}^{((p-1)k)} \right) \\
&+ \cdots + \sigma_{v_M}^2 \left(b_{1M}^{(2k)} b_{1M}^{(1k)} + \cdots + b_{1M}^{(pk)} b_{1M}^{((p-1)k)} \right) = Z_{22}^{(k)}(1) - U_{22}^{(k)}(1)
\end{aligned} \tag{4.23}$$

$$\begin{aligned}
&\sigma_{v_1}^2 \left(b_{M1}^{(2k)} b_{M1}^{(1k)} + \cdots + b_{M1}^{(pk)} b_{M1}^{((p-1)k)} \right) + \sigma_{v_2}^2 \left(b_{M2}^{(2k)} b_{M2}^{(1k)} + \cdots + b_{M2}^{(pk)} b_{M2}^{((p-1)k)} \right) \\
&+ \cdots + \sigma_{v_M}^2 \left(b_{MM}^{(1k)} + b_{MM}^{(2k)} b_{MM}^{(1k)} + \cdots + b_{MM}^{(pk)} b_{MM}^{((p-1)k)} \right) = Z_{MM}^{(k)}(1) - U_{MM}^{(k)}(1)
\end{aligned} \tag{4.24}$$

Above simultaneous linear equations can be expressed in matrix form as

$$\mathbf{C}\underline{\Sigma_{\mathbf{d}}} = \mathbf{D} \tag{4.25}$$

where noise co-variance vector (diagonal of $\Sigma_{\mathbf{v}}$) $\Sigma_{\mathbf{d}}$ is

$$\Sigma_{\mathbf{d}} = [\sigma_{v_1}^2, \sigma_{v_2}^2, \cdots, \sigma_{v_M}^2]^T,$$

$$\mathbf{D} = \begin{bmatrix} Z_{11}^{(k)}(1) - U_{11}^{(k)}(1) \\ Z_{22}^{(k)}(1) - U_{22}^{(k)}(1) \\ \vdots \\ Z_{MM}^{(k)}(1) - U_{MM}^{(k)}(1) \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1M} \\ C_{21} & C_{22} & \cdots & C_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ C_{M1} & C_{M2} & \cdots & C_{MM} \end{bmatrix}$$

and

$$\begin{aligned}
C_{11} &= \left(b_{11}^{(1k)} + b_{11}^{(2k)} b_{11}^{(1k)} + \cdots + b_{11}^{(pk)} b_{11}^{((p-1)k)}\right) \\
C_{12} &= \left(b_{12}^{(2k)} b_{12}^{(1k)} + \cdots + b_{12}^{(pk)} b_{12}^{((p-1)k)}\right) \\
C_{1M} &= \left(b_{1M}^{(2k)} b_{1M}^{(1k)} + \cdots + b_{1M}^{(pk)} b_{1M}^{((p-1)k)}\right) \\
C_{21} &= \left(b_{21}^{(2k)} b_{21}^{(1k)} + \cdots + b_{21}^{(pk)} b_{21}^{((p-1)k)}\right) \\
C_{22} &= \left(b_{22}^{(1k)} + b_{22}^{(2k)} b_{22}^{(1k)} + \cdots + b_{22}^{(pk)} b_{22}^{((p-1)k)}\right) \\
C_{2M} &= \left(b_{2M}^{(2k)} b_{2M}^{(1k)} + \cdots + b_{2M}^{(pk)} b_{2M}^{((p-1)k)}\right) \\
C_{M1} &= \left(b_{M1}^{(2k)} b_{M1}^{(1k)} + \cdots + b_{M1}^{(pk)} b_{M1}^{((p-1)k)}\right) \\
C_{M2} &= \left(b_{M2}^{(2k)} b_{M2}^{(1k)} + \cdots + b_{M2}^{(pk)} b_{M2}^{((p-1)k)}\right) \\
C_{MM} &= \left(b_{MM}^{(1k)} + b_{MM}^{(2k)} b_{MM}^{(1k)} + \cdots + b_{MM}^{(pk)} b_{MM}^{((p-1)k)}\right)
\end{aligned}$$

$$\begin{aligned}
\mathbf{C} = & \begin{bmatrix} b_{11}^{(1k)} & 0 & \cdots & 0 \\ 0 & b_{22}^{(1k)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{MM}^{(1k)} \end{bmatrix} \\
& + \begin{bmatrix} b_{11}^{(2k)} b_{11}^{(1k)} & b_{12}^{(2k)} b_{12}^{(1k)} & \cdots & b_{1M}^{(2k)} b_{1M}^{(1k)} \\ b_{21}^{(2k)} b_{21}^{(1k)} & b_{22}^{(2k)} b_{22}^{(1k)} & \cdots & b_{2M}^{(2k)} b_{2M}^{(1k)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{M1}^{(2k)} b_{M1}^{(1k)} & b_{M2}^{(2k)} b_{M2}^{(1k)} & \cdots & b_{MM}^{(2k)} b_{MM}^{(1k)} \end{bmatrix} + \cdots \\
& + \begin{bmatrix} b_{11}^{(pk)} b_{11}^{((p-1)k)} & b_{12}^{(pk)} b_{12}^{((p-1)k)} & \cdots & b_{1M}^{(pk)} b_{1M}^{((p-1)k)} \\ b_{21}^{(pk)} b_{21}^{((p-1)k)} & b_{22}^{(pk)} b_{22}^{((p-1)k)} & \cdots & b_{2M}^{(pk)} b_{2M}^{((p-1)k)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{M1}^{(pk)} b_{M1}^{((p-1)k)} & b_{M2}^{(pk)} b_{M2}^{((p-1)k)} & \cdots & b_{MM}^{(pk)} b_{MM}^{((p-1)k)} \end{bmatrix}
\end{aligned}$$

Using Eq. (4.18) $\mathbf{C}$ can be expressed as

$$\mathbf{C} = \mathbf{B}^{(1k)} .* \mathbf{I} + \mathbf{B}^{(2k)} .* \mathbf{B}^{(1k)} + \cdots + \mathbf{B}^{(pk)} .* \mathbf{B}^{((p-1)k))} \tag{4.26}$$

where ".*" denotes element to element multiplication of matrices. Therefore, the noise covariance vector is given by

$$\boldsymbol{\Sigma}_{\mathbf{d}} = \mathbf{C}^{-1} \mathrm{diag}\left(\mathbf{R}_{\mathbf{z}}^{(k)}(1) - \mathbf{R}_{\mathbf{u}}^{(k)}(1)\right) \tag{4.27}$$

Since $\mathbf{u}^{(k)}(n)$ cannot be measured, it is impossible to use Eq. (4.27) directly. Consequently, we define

$$\bar{\mathbf{\Sigma}}_{\mathbf{d}}(k) = \mathbf{C}^{-1}\mathrm{diag}\left(\mathbf{R}_{\mathbf{z}}^{(k)}(1)\right) \tag{4.28}$$

where

$$\bar{\mathbf{\Sigma}}_{\mathbf{d}}(k) = [\bar{\sigma}_{v_1}^2, \bar{\sigma}_{v_2}^2, \cdots, \bar{\sigma}_{v_M}^2]^T \tag{4.29}$$

Note that $\mathbf{R}_{\mathbf{u}}^{(k)}(1) = \mathbf{0}$ holds when $\mathbf{B}_{pk}(z) = \mathbf{A}_p(z)$ is satisfied. This is the only case for which Eqs. (4.27) and (4.28) become identical, they differ in all other instances. We exploit this point in estimating the noise covariance vector and hence the AR coefficient vector.

Since our aim is to match $\mathbf{B}_{pk}(z)$ with $\mathbf{A}_p(z)$, the coefficient matrices of $\mathbf{B}_{pk}(z)$ can be calculated in the same way as those of $\mathbf{A}_p(z)$. We rewrite Eq. (4.9) as

$$\underline{\mathbf{b}}\ (\underline{\mathbf{R}}' - \underline{\boldsymbol{\alpha}}_k) = -\underline{\mathbf{r}}' \tag{4.30}$$

$$\begin{aligned}
\underline{\mathbf{R}}' &= \begin{bmatrix} \mathbf{R_{yy}}(0) & \cdots & \mathbf{R_{yy}}(p-1) \\ \mathbf{R_{yy}}(-1) & \cdots & \mathbf{R_{yy}}(p-2) \\ \vdots & \ddots & \vdots \\ \mathbf{R_{yy}}(-p+1) & \cdots & \mathbf{R_{yy}}(0) \end{bmatrix} \\
\underline{\mathbf{r}}' &= \left[\mathbf{R_{yy}}(1), \mathbf{R_{yy}}(2), \cdots, \mathbf{R_{yy}}(p)\right] \\
\underline{\mathbf{b}} &= \left[\mathbf{B}^{(1k)}, \mathbf{B}^{(2k)}, \ldots, \mathbf{B}^{((p-1)k)}, \mathbf{B}^{(pk)}\right],
\end{aligned}$$

where $\underline{\mathbf{R}}'$ and $\underline{\mathbf{r}}'$ are the $p \times p$ block matrix and the $1 \times p$ block row vector, respectively, calculated from the observed data sequence $\mathbf{y}(n)$, $\underline{\mathbf{b}}$ is the $1 \times p$ block row vector to be determined and $\underline{\boldsymbol{\alpha}}_k$ is the $p \times p$ block diagonal matrix with identical block entries $\boldsymbol{\alpha}(k)$. The $M \times M$ diagonal matrix $\boldsymbol{\alpha}(k) = \mathrm{diag}\,(\alpha_1(k), \alpha_2(k), \cdots, \alpha_M(k))$. Eq. (4.30) is solved iteratively for different arbitrary matrices $\boldsymbol{\alpha}(k)$. Using these values of $\mathbf{B}^{(1k)}, \mathbf{B}^{(2k)}, \cdots, \mathbf{B}^{(pk)}$ and the corresponding filtered output vector$\mathbf{z}^{(k)}(n)$, we can determine each time $\bar{\mathbf{\Sigma}}_{\mathbf{d}}(k)$ from Eq. (4.28). Thus $\alpha_1(k), \alpha_2(k), \cdots, \alpha_M(k)$ are the auxiliary variables of elements of matrix $\mathbf{\Sigma}_{\mathbf{d}}$. As $\alpha_1(k), \alpha_2(k), \cdots$ and $\alpha_M(k)$ approach $\sigma_{v_1}^2, \sigma_{v_2}^2, \cdots$ and $\sigma_{v_M}^2$, respectively and simultaneously, $\underline{\mathbf{b}}$ approaches $\underline{\mathbf{a}}$. Therefore, at equality, the relationship $\mathbf{R}_{\mathbf{u}}^{(k)}(1) = \mathbf{0}$ holds due to the properties of white noise. Under this condition, Eqs. (4.17) and (4.28) give the same value for the noise variances. Therefore, using Eq. (4.28) instead of Eq. (4.17) does not affect the desired

result. Since the true noise covariance matrix $\mathbf{\Sigma_V}$ is unknown, the key idea is to match $\alpha_1(k), \alpha_2(k), \cdots$ and $\alpha_M(k)$ with $\bar{\sigma}^2_{v_1}, \bar{\sigma}^2_{v_2}, \cdots$ and $\bar{\sigma}^2_{v_M}$, respectively and simultaneously instead. When these quantities become identical simultaneously, at least theoretically, an estimate $\hat{\mathbf{\Sigma}}_\mathbf{d}$ of the observation noise covariance matrix can be determined from $\alpha_1(k), \alpha_2(k), \cdots$ and $\alpha_M(k)$. In addition, the block row vector $\underline{\mathbf{b}}$ obtained by substituting $\hat{\mathbf{\Sigma}}_\mathbf{V} = \text{diag}\left(\hat{\sigma}^2_{v_1}, \hat{\sigma}^2_{v_2}, \cdots, \hat{\sigma}^2_{v_M}\right)$ for $\boldsymbol{\alpha}(k)$ in Eq. (4.30) gives an estimate of the desired block row vector $\underline{\mathbf{a}}$. Since $\alpha_1(k), \alpha_2(k), \cdots$ and $\alpha_M(k)$ denotes the auxiliary variables for the elements of matrix $\bar{\mathbf{\Sigma}}_\mathbf{d}$, we can write

$$\begin{aligned}
\bar{\sigma}^2_{v_1} &= f_1(\alpha_1, \alpha_2, \cdots, \alpha_M) \\
\bar{\sigma}^2_{v_2} &= f_2(\alpha_1, \alpha_2, \cdots, \alpha_M) \\
&\vdots \\
\bar{\sigma}^2_{v_M} &= f_M(\alpha_1, \alpha_2, \cdots, \alpha_M)
\end{aligned}$$

4.4 Estimation of Noise Variance

In this section, we propose two methods to estimate the noise variances of different channels. The AR parameters are obtained simultaneously as a by product which are then used to compute the power spectrum.

4.4.1 Approach I: Gradient search technique

In order to match $\alpha_1(k), \alpha_2(k), \cdots$ and $\alpha_M(k)$ with $\bar{\sigma}^2_{v_1}, \bar{\sigma}^2_{v_2}, \cdots$ and $\bar{\sigma}^2_{v_M}$, respectively and simultaneously, we can write at matching condition

$$\alpha_1 - \bar{\sigma}^2_{v_1} = 0 \tag{4.31}$$

i.e,

$$\begin{aligned}
\alpha_1 - f_1(\alpha_1, \alpha_2, \cdots, \alpha_M) &= 0 \\
\Rightarrow F_1(\alpha_1, \alpha_2, \cdots, \alpha_M) &= 0
\end{aligned} \tag{4.32}$$

$$\alpha_2 - \bar{\sigma}^2_{v_2} = 0 \tag{4.33}$$

i.e,

$$\begin{aligned}
\alpha_2 - f_2(\alpha_1, \alpha_2, \cdots, \alpha_M) &= 0 \\
\Rightarrow F_2(\alpha_1, \alpha_2, \cdots, \alpha_M) &= 0
\end{aligned} \tag{4.34}$$

$$\vdots$$

$$\alpha_M - \bar{\sigma}^2_{v_M} = 0 \tag{4.35}$$

i.e,

$$\begin{aligned} \alpha_M - f_M(\alpha_1, \alpha_2, \cdots, \alpha_M) &= 0 \\ \Rightarrow F_M(\alpha_1, \alpha_2, \cdots, \alpha_M) &= 0 \end{aligned} \tag{4.36}$$

We do not know the analytical expression of $F_1, F_2, \cdots$ and F_M. But we can use numerical method to solve these M simultaneous nonlinear equations. In this approach, we propose a numerical method based on Newton-Raphson technique to solve the above set of nonlinear simultaneous equations. And consequently we can estimate the noise covariances. In the review of Newton-Raphson method, let us consider the initial guess value of the roots of these non linear Equations are $\alpha_i^0 (i = 1, 2 \cdots, M)$. If the exact roots are $\alpha_i^0 + \triangle\alpha_i$ then we get,

$$F_i(\alpha_1^0 + \triangle\alpha_1, \alpha_2^0 + \triangle\alpha_2, \cdots, \alpha_M^0 + \triangle\alpha_M) = 0 \tag{4.37}$$

Expanding the functions as a Taylor series about the point $(\alpha_1^0, \alpha_2^0, \cdots, \alpha_M^0)$ we get,

$$F_i(\alpha_1^0, \cdots, \alpha_M^0) + \sum_{j=1}^{M} \frac{\partial F_i}{\partial \alpha_j} \triangle\alpha_j + \sum_{j=1}^{M} \frac{\partial F_i}{\partial \alpha_j} (\triangle\alpha_j)^2 + \cdots = 0 \tag{4.38}$$

Truncating the series gives,

$$\sum_{j=1}^{M} \frac{\partial F_i}{\partial \alpha_j} \triangle\alpha_j \cong -F_i(\alpha_1^0, \alpha_2^0, \cdots, \alpha_M^0) \tag{4.39}$$

From Eq. (4.39), we can write for $i = 1, 2, \cdots, M$

$$\begin{bmatrix} \frac{\partial F_1}{\partial \alpha_1} & \frac{\partial F_1}{\partial \alpha_2} & \cdots & \frac{\partial F_1}{\partial \alpha_M} \\ \frac{\partial F_2}{\partial \alpha_1} & \frac{\partial F_2}{\partial \alpha_2} & \cdots & \frac{\partial F_2}{\partial \alpha_M} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_M}{\partial \alpha_1} & \frac{\partial F_M}{\partial \alpha_2} & \cdots & \frac{\partial F_M}{\partial \alpha_M} \end{bmatrix} \begin{bmatrix} \triangle\alpha_1 \\ \triangle\alpha_2 \\ \vdots \\ \triangle\alpha_M \end{bmatrix} = - \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_M \end{bmatrix} \tag{4.40}$$

Since we do not know the analytical expression of the functions, we will approximate the partial derivatives of the functions by recalculating the functions with a small perturbation as follows:

$$\frac{\partial F_i}{\partial \alpha_j} = \frac{F_i(\alpha_1, \cdots, \alpha_j + \delta, \cdots, \alpha_M) - F_i(\alpha_1, \cdots, \alpha_i, \cdots \alpha_M)}{\delta} \tag{4.41}$$

where δ is small increment of variables' values. To achieve our goal, we propose the following algorithm.

The Algorithm:

---

1. Initialize $\alpha_i(k) = 0, \;\; i = 1, 2, \cdots, M$.
2. Calculate $\bar{\sigma}^2_{v_i}, \;\; i = 1, 2, \cdots, M$ from Eq. (4.28).
3. Calculate $F_i = (\alpha_i(k) - \bar{\sigma}^2_{v_i})$ at $\alpha_i(k) = 0, \;\; i = 1, 2, \cdots, M$.
4. Similarly calculate F_i at $\alpha_i(k) + \delta, \;\; i = 1, 2, \cdots, M$.
5. Calculate partial derivatives of functions with respect to $\alpha_1, \alpha_2, \cdots, \alpha_M$ from Eq. (4.41) with the calculated values of function at steps 3 and 4.
6. Calculate $\triangle\alpha_1, \triangle\alpha_2, \cdots, \triangle\alpha_M$ from Eq. (4.40).
7. Calculate $\alpha_i(k+1) = \alpha_i(k) + \triangle\alpha_i, \;\; i = 1, 2, \cdots, M$.
8. Repeat steps 2 to 7 until $F_1, F_2, \cdots, F_M$ converge simultaneously to zero.

Though this method converges vary fast its final result depends on the initial value of the noise power. Moreover the method may fail to converge to the global minimum. In the next approach we overcome the problem of dependency of the estimated noise power on the initialization of the search technique.

4.4.2 Approach II: Fuzzy iterative search

Arbitrary matrix $\boldsymbol{\alpha}(k)$ is changed iteratively and as $\boldsymbol{\alpha}(k)$ approaches the actual matrix of noise variance, i.e., $\boldsymbol{\Sigma_v}$, $\underline{\mathrm{b}}$ approaches $\underline{\mathrm{a}}$ and $\mathbf{R}^{(k)}_{\mathbf{u}}(1)$ approaches to zero due to the properties of white noise. Consequently $\boldsymbol{\Sigma_\alpha}(k) - \bar{\boldsymbol{\Sigma}}_{\mathbf{d}}(k)$ also approaches zero. The performance index for fuzzy iterative search technique is therefore defined as

$$J(k) = \left[\boldsymbol{\Sigma_\alpha}(k) - \bar{\boldsymbol{\Sigma}}_{\mathbf{d}}(k)\right]^T \left[\boldsymbol{\Sigma_\alpha}(k) - \bar{\boldsymbol{\Sigma}}_{\mathbf{d}}(k)\right] \tag{4.42}$$

where $\boldsymbol{\Sigma_\alpha}(k) = [\alpha_1(k), \alpha_2(k), \cdots, \alpha_M(k)\;]^T$. Now we are going to present fuzzy incorporated iterative search technique which will give the desired solution, i.e.,

$\Sigma_{\boldsymbol{\alpha}}(k) = \Sigma_{\mathbf{d}}$ and $\underline{\mathbf{b}} = \underline{\mathbf{a}}$ at the global minimum of the above defined performance index. Here it is assumed that the desired solution lies where the global minimum occurs. In this approach only one noise power on each channel of the noise power vector $\Sigma_{\boldsymbol{\alpha}} = [\alpha_1, \alpha_2, \cdots, \alpha_M]^T$ is changed at a time. Noise power $\alpha_i(k)$ of the i-th channel where $i = 1, 2, \cdots, M$ is changed iteratively from 0 to $\alpha_{i_{max}}$ with a variable incremental step size $s_i(k)$ such that $\alpha_i(k+1) = \alpha_i(k) + s_i(k)$. Here $\alpha_{i_{max}}$ is the maximum limit of $\alpha_i(k)$ which is assumed to be known. While $\alpha_i(k)$ is changed, all noise powers of the other channels kept unchanged. A Fuzzy Inference System (FIS) determines $s_i(k)$ in accordance with value of $J(k)$ which is discussed in the following section. Through out the range of iteration $\alpha_i' = \arg\{\min J(\alpha_i(k))\}$ is determined which corresponds to the minimum value of $J(k)$. When $\alpha_i(k+1)$ exceeds $\alpha_{i_{max}}$ the iteration of the i-th channel stops and iteration of the noise power $\alpha_{i+1}(k)$of the next $i+1$-th channel starts keeping $\alpha_i = \alpha_i'$ and all the noise power of the other channel kept unchanged. In this way when the iterations of channel-1 to channel-M are completed one cycle is completed and another new cycle starts. The concerned channel whose noise power is changed iteratively is a particular phase of the cycle. To illustrate the approach additional superscripts can be assigned on the variable $\Sigma_{\boldsymbol{\alpha}}, \alpha_i, s_i$. For example $\Sigma_{\boldsymbol{\alpha}}^{qi}$ denotes the iteratively changed noise power vector $\Sigma_{\boldsymbol{\alpha}}$ of i-th phase of the q-th cycle. In the i-th phase of q-th cycle we get $\alpha_i^{q'}$ by using $\Sigma_{\boldsymbol{\alpha}}^{qi}(k) = [\alpha_1^{q'}, \alpha_2^{q'}, \cdots, \alpha_i^q(k), \cdots, \alpha_M^{(q-1)'}]^T$. $\alpha_i^q(k)$ is varied from 0 to known $\alpha_{i_{max}}$ with a variable incremental step size $s_i^q(k)$ according to the following equation

$$\alpha_i^q(k+1) = \alpha_i^q(k) + s_i^q(k), \qquad i = 1, 2, \cdots, M \tag{4.43}$$

The superscript on α_i and s_i denotes the number of cycle. The value of $\alpha_i^{q'} = \arg\{\min J(\alpha_i^q(k))\}$ is determined which corresponds to the minimum value of $J(k)$ for the i-th phase of the q-th cycle . When $\alpha_i^q(k)$ is changed iteratively, the noise of the previous channels kept unchanged to their respective values where $J(k)$ was minimum in the q-th cycle in each respective phase. The noise power of the channels following the i-th channels remain also unchanged to their respective values where $J(k)$ was minimum in the $(q-1)$-th cycle in each respective phase.

After completion of the q-th cycle $\Sigma_{\boldsymbol{\alpha}}^q = [\alpha_1^{q'}, \alpha_2^{q'}, \cdots, \alpha_M^{q'}]^T$ is obtained. $(q+1)$-th cycle starts with $\Sigma_{\boldsymbol{\alpha}}^{(q+1)1}(k) = [\alpha_1^{(q+1)}(k), \alpha_2^{q'}, \cdots, \alpha_M^{q'}]^T$. Again $\alpha_1^{q+1}(k)$ is varied from 0 to $\alpha_{1_{max}}$. After completion of the $(q+1)$-th cycle

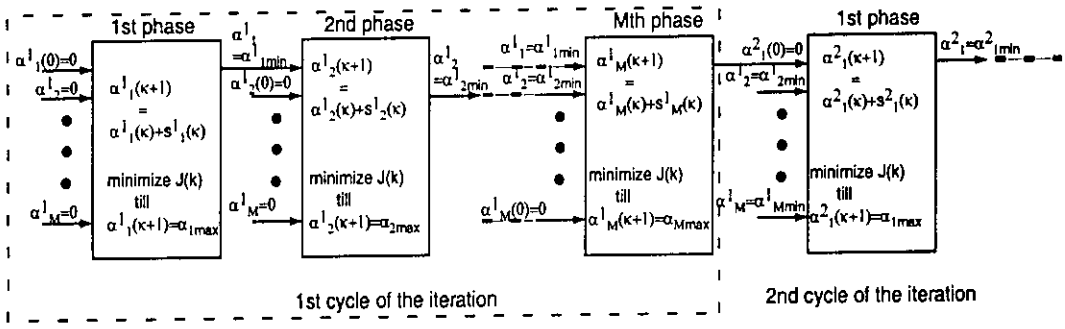

Fig. 4.2: Flow diagram of fuzzy iterative search

$\Sigma_{\boldsymbol{\alpha}}^{(q+1)1} = [\alpha_1^{(q+1)'}(k), \alpha_2^{(q+1)'}, \cdots, \alpha_M^{(q+1)'}]^T$ is obtained. The iterative method converges to the value of global minimum after the l-th cycle when $\Sigma_{\boldsymbol{\alpha}}^{l} = \Sigma_{\boldsymbol{\alpha}}^{(l-1)}$. First cycle of the Fuzzy iterative approach is shown in Fig. 4.2. The algorithm of iteration technique is summarized as follows.

The Algorithm:

---

1. Initialization:

$$\Sigma_{\boldsymbol{\alpha}}^{11}(0) = \left[\alpha_1^1(0), \alpha_2^1, \cdots, \alpha_M^1\right]^T$$

where $\alpha_1^1(0) = \alpha_2^1 = \cdots = \alpha_M^1 = 0$

2. Find $\Sigma_{\boldsymbol{\alpha}}^{qi}(k) = [\alpha_1^{q'}, \alpha_2^{q'}, \cdots, \alpha_i^q(k), \cdots, \alpha_M^{(q-1)'}]^T$ and hence find

$$\alpha_i^{q'} = \arg\{\min J(\alpha_i^q(k))\}$$

where $i = 1, 2, \cdots, M$. When $i = M$, i.e. after completion of the q-th cycle we get $\Sigma_{\boldsymbol{\alpha}}^{q} = [\alpha_1^{q'}, \alpha_2^{q'}, \cdots, \alpha_M^{q'}]^T$

3. Repeat step 2 until $q = 1, 2, \cdots, l$ such that $\Sigma_{\boldsymbol{\alpha}}^{l} = \Sigma_{\boldsymbol{\alpha}}^{(l-1)}$

We can visualize the way of convergence of Approach-2 from the surface plot shown in Fig. 4.3 and Fig. 4.4. At the starting, noise power of ch-2 is kept

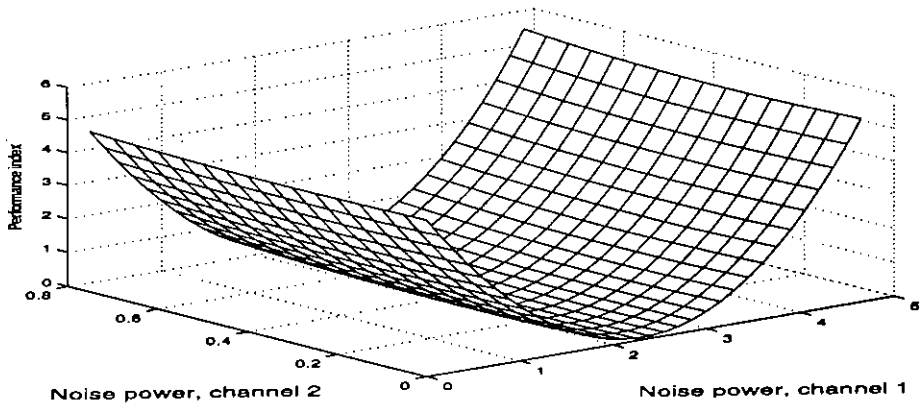

Fig. 4.3: Surface plot of performance index vs. noise power at $SNR_1 = 10$ dB, $SNR_2 = 20$ dB for first-order, two-channel, AR process where $a_{11}^{(1)} = -0.85, a_{12}^{(1)} = 0.75, a_{21}^{(1)} = -0.65, a_{22}^{(1)} = -0.55$

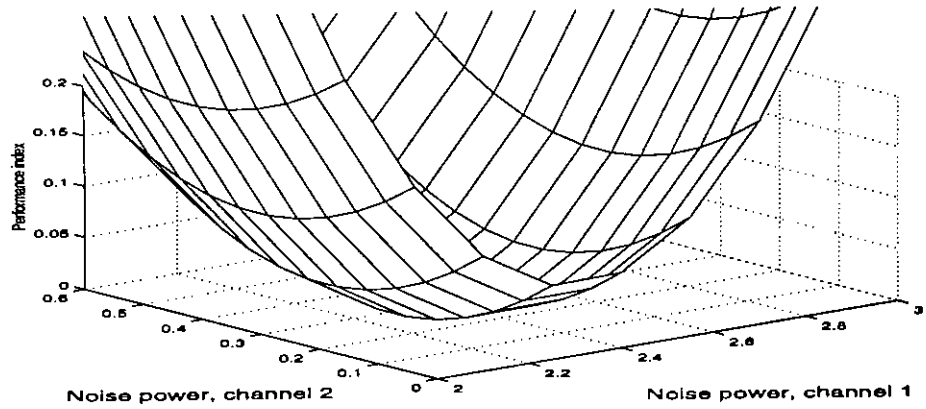

Fig. 4.4: Closer view of surface plot of Fig. 4.3 at the neighborhood of the global minimum

zero, we search along the axis of ch-1 to get the minimum value of performance index. Again,the obtained minimum value of ch-1 is kept constant and search starts along ch-2. In this way after few search we reach the global minimum of the error surface. The searching technique can be thought as if one ball is thrown from the top of the surface and it starts rolling according to the curvature of the surface and ultimately it stops at the lowest point of the bottom.

4.4.2.1 Determining the step size: Fuzzy approach

In Eqn. (4.43) $s_i^q(k)$ denotes step size whose value depends on the performance index $J(k)$. We are searching the minimum value of $J(k)$ in the range 0 to known $\alpha_{i_{max}}$, by increasing $\alpha_i^q(k)$ with a step size $s_i^q(k)$. The step size can be determined as a function of the performance index instead of keeping it to be constant throughout the iteration range. That is if the performance index becomes smaller then the step size can be adjusted reasonably smaller to achieve the minimum value of $J(k)$ and the corresponding $\alpha_i^{q'}$. If the performance index becomes larger then the step size will be larger to skip the less concerned region of iteration. We should be careful in selecting the maximum step size of the iteration so that the iteration does not miss the minimum. The minimum step size depends on our desired accuracy of the estimation. In this way, we can reduce the number of iterations as well as computation time significantly without sacrificing the accuracy of estimation since we keep the step size smaller in the region of iteration where we approach the minimum.

For determining the step size in accordance with the value of the performance index a Fuzzy Inference System (FIS) is designed which maps non-linearly the input, performance index $J(k)$, to the appropriate output, the step size $s_i^q(k)$ for the i-th phase of the q-th cycle. We consider the following criteria in designing the FIS:

1. This is a Mamdani fuzzy model.

2. The input to the FIS is the performance index calculated using Eqn. (4.42). We select the universe of discourse of the input to be the value obtained from Eqn. (4.42) when $\Sigma_{\boldsymbol{\alpha}} = [\alpha_1, \alpha_2, \cdots, \alpha_M]^T$, where $\alpha_1 = \alpha_2 = \cdots = \alpha_M = 0$. Universe of discourse for the input is same for each channel.

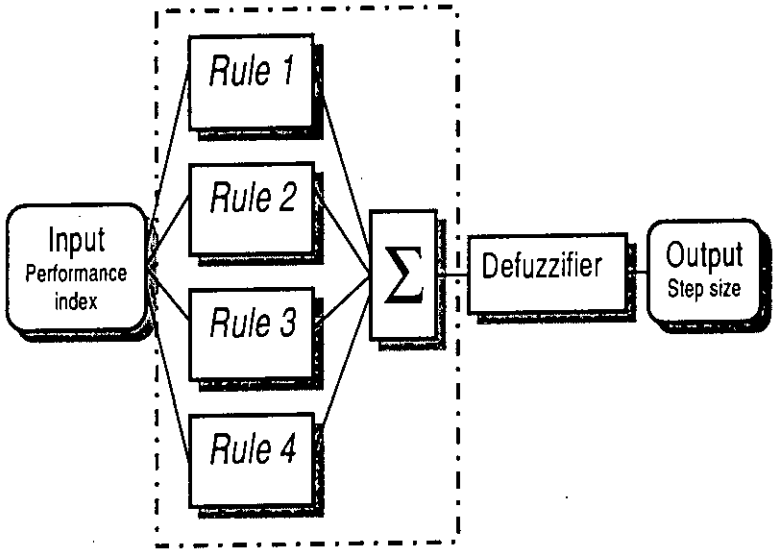

Fig. 4.5: Designing a fuzzy inference system for determining the step size: System consists of 1 input, 1 output, and 4 rules.

3. We define four membership functions (MFs) for the input. They are 'verysmall', 'small', 'significant' and 'large'. The type of membership function and their ranges are depicted in Fig. 4.6. Normalized error can be obtained dividing the error by universe of discourse calculated in step 2.

4. The output from the FIS is the effective step size $s_i^q(k)$. We have calculated the universe of discourse of the step size to be $(1/K) \times \alpha_{i_{max}}$ for $i = 1, 2, \cdots, M$. It is apparent that the universe of discourse of output is different for different channel. We have found $K = 10$ as an well performed approximation. But in some cases where the error function falls rapidly we may have to increase the value of K to avoid any chance of missing the minimum. We can choose a smaller value of K where the cost function changes slowly. It makes the total number of iteration less.

5. We define four membership functions(MFs) for the output also. They are 'verysmall', 'small', 'significant' and 'large'. The type of membership function and their ranges are shown in Fig. 4.7. Step size is shown in normalized scale. Normalized step size is the ratio of stepsize to the universe of discourse. The accuracy of the iterative technique depends on how the 'verysmall' MF is defined. We choose the 'verysmall' MF as a zero-mean gaussian type membership function

$$f(x; \sigma, c) = e^{-\frac{(x-c)^2}{2\sigma^2}}, \quad c = 0 \tag{4.44}$$

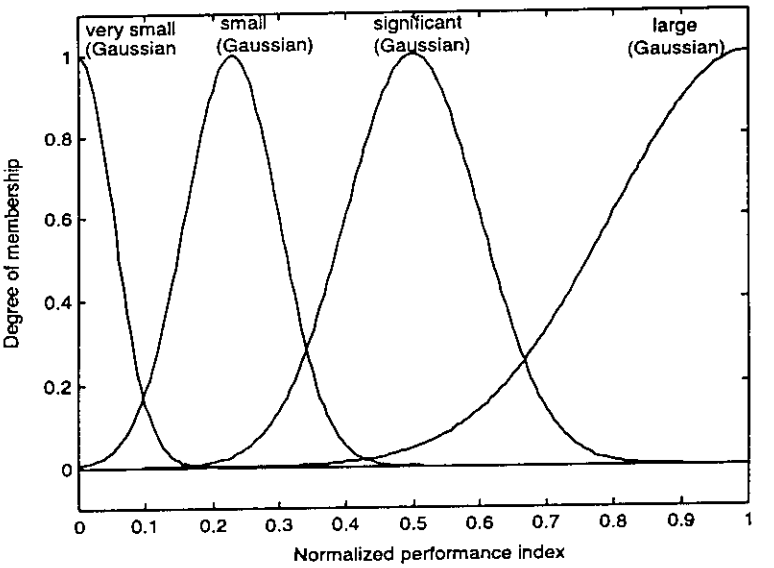

Fig. 4.6: Membership function of the input variable, the cost function, to the FIS.

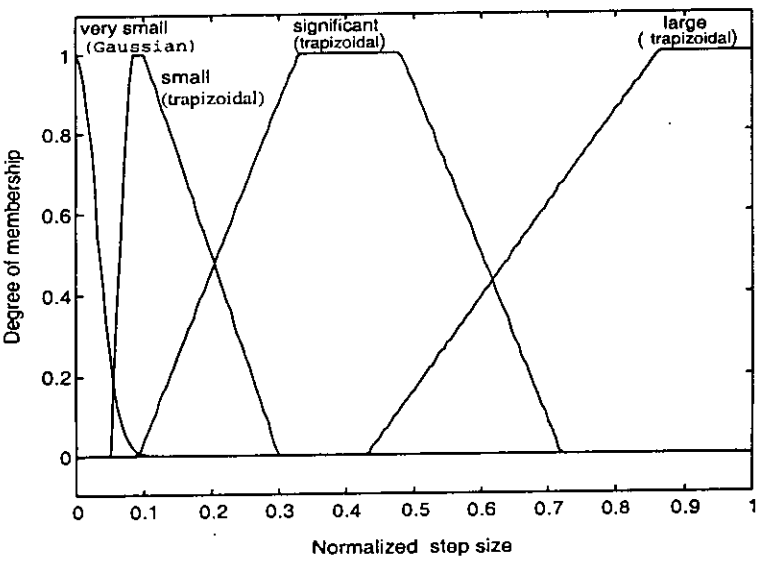

Fig. 4.7: Membership function of the output variable, the step size, to the FIS.

An increase in the value of σ will increase the value of the step size in the region where the error is 'verysmall'. It decreases the total number of iterations but the cost must be paid in estimation accuracy.

6. Following rules are implemented in the FIS
   - If (error is verysmall) then (step size is verysmall)
   - If (error is small) then (step size is small)
   - If (error is significant) then (step size is significant)
   - If (error is large) then (step size is large)

In this Fuzzy model no logical operator is used as there is only one input. The implication method we choose is min (minimum), the aggregation method is max (maximum), and we defuzzify the output by centroid calculation. The following algorithm can be followed to implement the above mentioned FIS.

1. Set the maximum error range and the maximum step size (universe of discourses) by executing respectively step 2 and step 4 of the aforementioned criteria of the FIS. Set the desired accuracy of the estimation by adjusting the parameter σ of the gaussian membership function 'verysmall' of step size according to step 5 for each channel.

2. Set initial value of $\alpha_i^q(0) = 0$ and final value to $\alpha_{i_{max}}$ for i-th phase of the q-th cycle. Where $i = 1, 2, \ldots, M$ and $q = 1, 2, \cdots, l$.

3. Calculate the error using Eqn. (4.42).

4. Supply 'error' calculated in step 3 as input to the FIS.

5. Determine $s_i^q(k)$, the 'step size', as the output of the FIS.

6. Calculate $\alpha_i^q(k+1) = \alpha_i^q(k) + s_i^q(k)$.

7. If $\alpha_i^q(k+1) \leq \alpha_{i_{max}}$, repeat steps 3 to 6, otherwise terminate.

This approach is particularly suitable for a single minimum error surface or when the multiple minima align at one particular noise power of one of the channels. Another point is that the maximum limit of the noise power of each channel should be known a priori. Most of these preconditions can be satisfied at high SNR level but at low SNR these could be hardly realized.

4.5 Simulation Results

To illustrate the performance of the proposed estimation methods, a second-order ($p = 2$) two-channel ($M = 2$) auto regressive process was used to generate data samples. Data was generated according to the recursion

$$\begin{aligned} \mathbf{x}(n) &= -\mathbf{A}^{(1)}\mathbf{x}(n-1) - \mathbf{A}^{(2)}\mathbf{x}(n-2) + \mathbf{u}(n) \\ \mathbf{y}(n) &= \mathbf{x}(n) + \mathbf{v}(n) \end{aligned}$$

in which the auto regressive coefficient matrices are

$$\mathbf{A}^{(1)} = \begin{bmatrix} 0.30 & 0.038 \\ -0.18 & 0.78 \end{bmatrix}, \quad \mathbf{A}^{(2)} = \begin{bmatrix} 0.97 & -0.108 \\ 0.9 & 0.50 \end{bmatrix}$$

and $\mathbf{u}(n)$ is two-channel stationary white noise, uncorrelated between channels and of unit variance on each channel. It is also assumed that $\mathbf{v}(n)$ is two-channel stationary white noise, uncorrelated between channels and with $\mathbf{u}(n)$. The variance on each channel is adjusted to obtain different signal-to-noise ratios (SNRs) defined as $SNR_i = 10\log_{10}\left(\sigma_{x_i}^2/\sigma_{v_i}^2\right)$ (dB), where $\sigma_{x_i}^2$ denotes the ith channel signal power and SNR_i represents the signal-to-noise ratio on the ith channel.

4.5.1 Simulation results of gradient based technique

In the experiment $N = 4000$ data samples for each channel was used. The results obtained using the proposed method for different combination of SNRs and initial conditions are presented in Table (4.1) and Table (4.2). The results shown are the mean value of 20 simulation runs. The quantity in bracket represents standard deviation of 20 simulation runs. From Table (4.1) and Table (4.2) it is evident that the mean values of estimated noise power and AR parameters are very close to their true counter parts. The standard deviation is also reasonably good.

4.5.2 Simulation results of fuzzy iterative search

The simulation conditions were kept unchanged. The results obtained using this method for different combination of SNRs are presented in Table (4.3) and Table (4.4). The entries represent the mean value of 20 simulation runs. The quantity in bracket represents standard deviation of 20 simulation runs. As can be seen from Table (4.3) and Table (4.4) the mean values of estimated noise power and

Table 4.1: Estimated noise variances for 2-channel AR(2) process by gradient based technique ($N = 4000$, $\alpha_1(0) = \alpha_2(0) = 0$)

Channel 1			Channel 2		
SNR (dB)	$\sigma^2_{v_1}$ (true value)	$\widehat{\sigma^2_{v_1}}$ (estimate)	SNR (dB)	$\sigma^2_{v_2}$ (true value)	$\widehat{\sigma^2_{v_2}}$ (estimate)
5	1.2156	1.2627 (0.1418)	10	1.0640	1.0560 (0.1008)
10	.3847	.4076 (0.0640)	5	3.4077	3.3620 (0.2213)

Table 4.2: Estimated AR coefficients for 2-channel AR(2) process by gradient based technique ($N = 4000$, $\alpha_1(0) = \alpha_2(0) = 0$)

True AR coefficients	Estimated AR coefficients	
	$SNR1$=5dB,$SNR2$=10dB	$SNR1$=10dB,$SNR2$=5dB
$a^{(1)}_{11} = 0.3$	0.3095 (0.0266)	0.3070 (0.0256)
$a^{(1)}_{12} = 0.038$	0.0307 (0.0144)	0.0356 (0.0151)
$a^{(1)}_{21} = -0.18$	-0.1791 (0.0308)	-0.1875 (0.0452)
$a^{(1)}_{22} = 0.78$	0.7827 (0.0187)	0.7844 (0.0294)
$a^{(2)}_{11} = 0.97$	0.9923 (0.0669)	0.9868 (0.0363)
$a^{(2)}_{12} = -0.108$	-0.1213 (0.0292)	-0.1174 (0.0200)
$a^{(2)}_{21} = 0.90$	0.9188 (0.0649)	0.9130 (0.0348)
$a^{(2)}_{22} = 0.50$	0.4931 (0.0338)	0.4934 (0.0191)

[Note: numerical value in the bracket denotes the standard deviation of the respective samples obtained from the 20 simulation runs]

Table 4.3: Estimated noise variances for 2-channel AR(2) process by fuzzy iterative search technique ($N = 4000$, $\alpha_1(0) = \alpha_2(0) = 0$)

Channel 1				Channel 2			
SNR (dB)	$\sigma^2_{v_1}$ (true value)	$\widehat{\sigma}^2_{v_1}$ (estimate)	minimum stepsize	SNR (dB)	$\sigma^2_{v_2}$ (true value)	$\widehat{\sigma}^2_{v_2}$ (estimate)	minimum stepsize
5	1.2366	1.2174 (0.1260)	0.0076	10	1.0827	1.0795 (0.0844)	0.0076
10	.3790	.3919 (0.0725)	0.0078	5	3.3645	3.3966 (0.1842)	0.0228

Table 4.4: Estimated AR coefficients for 2-channel AR(2) process by fuzzy iterative search technique ($N = 4000$, $\alpha^1_1(0) = \alpha^1_2(0) = 0$)

True AR coefficients	Estimated AR coefficients	
	$SNR1$=5dB,$SNR2$=10dB	$SNR1$=10dB,$SNR2$=5dB
$a^{(1)}_{11} = 0.3$	0.2945 (0.0345)	0.3014 (0.0202)
$a^{(1)}_{12} = 0.038$	0.0398 (0.0169)	0.0350 (0.0144)
$a^{(1)}_{21} =$ -0.18	-0.1913 (0.0267)	-0.1928 (0.0291)
$a^{(1)}_{22} = 0.78$	0.7846 (0.0208)	0.7868 (0.0199)
$a^{(2)}_{11} = 0.9$	0.9672 (0.0613)	0.9733 (0.0398)
$a^{(2)}_{12} =$ -0.108	-0.1085 (0.0339)	-0.1104 (0.0199)
$a^{(2)}_{21} = 0.90$	0.8864 (0.0380)	0.9115 (0.0524)
$a^{(2)}_{22} = 0.50$	0.5034 (0.0163)	0.5005 (0.0255)

[Note: numerical value in the bracket denotes the standard deviation of the respective samples obtained from the 20 simulation runs]

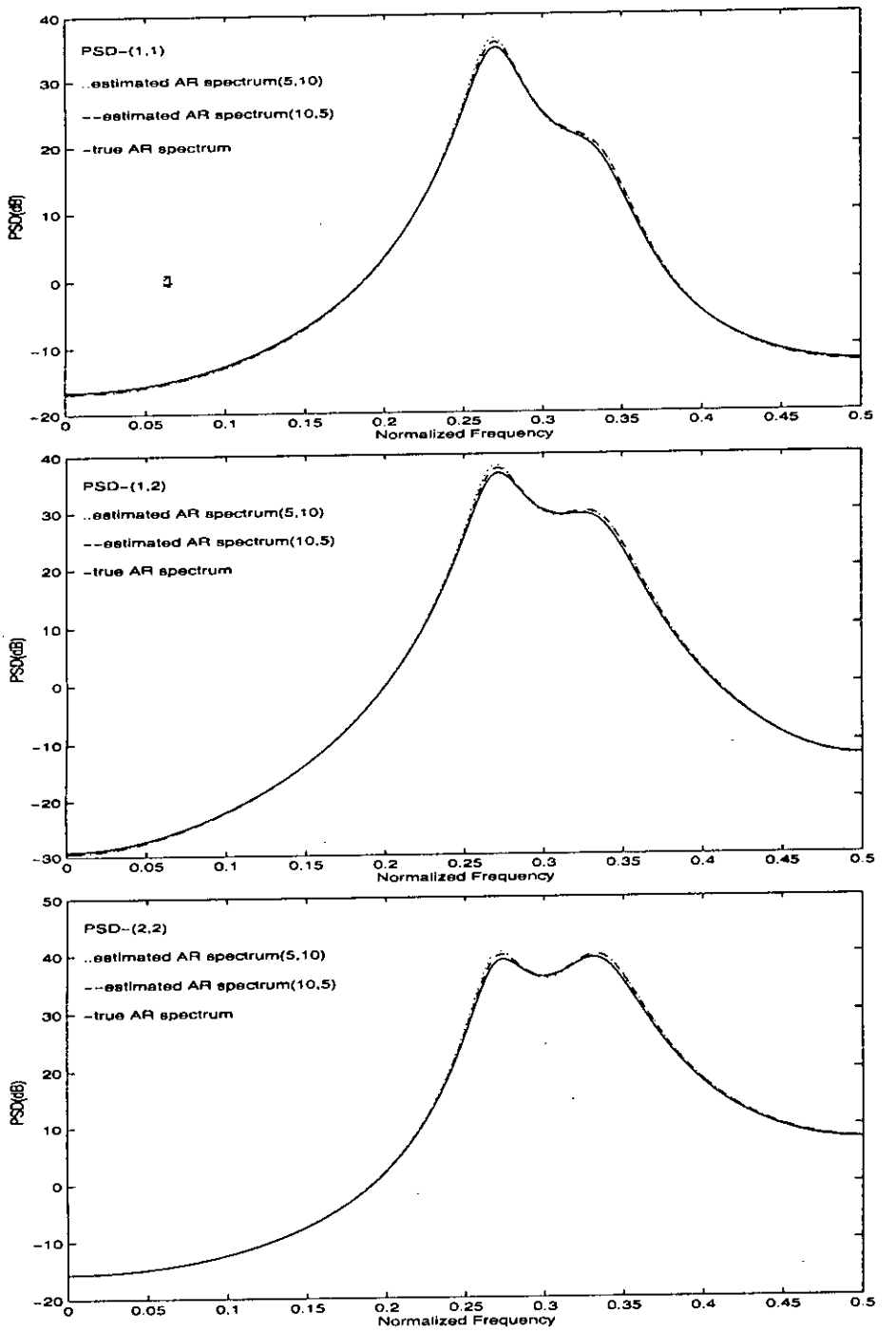

Fig. 4.8: True and estimated AR spectrum by the proposed gradient search method for $SNR_1 = 5$ dB, $SNR_2 = 10$ dB

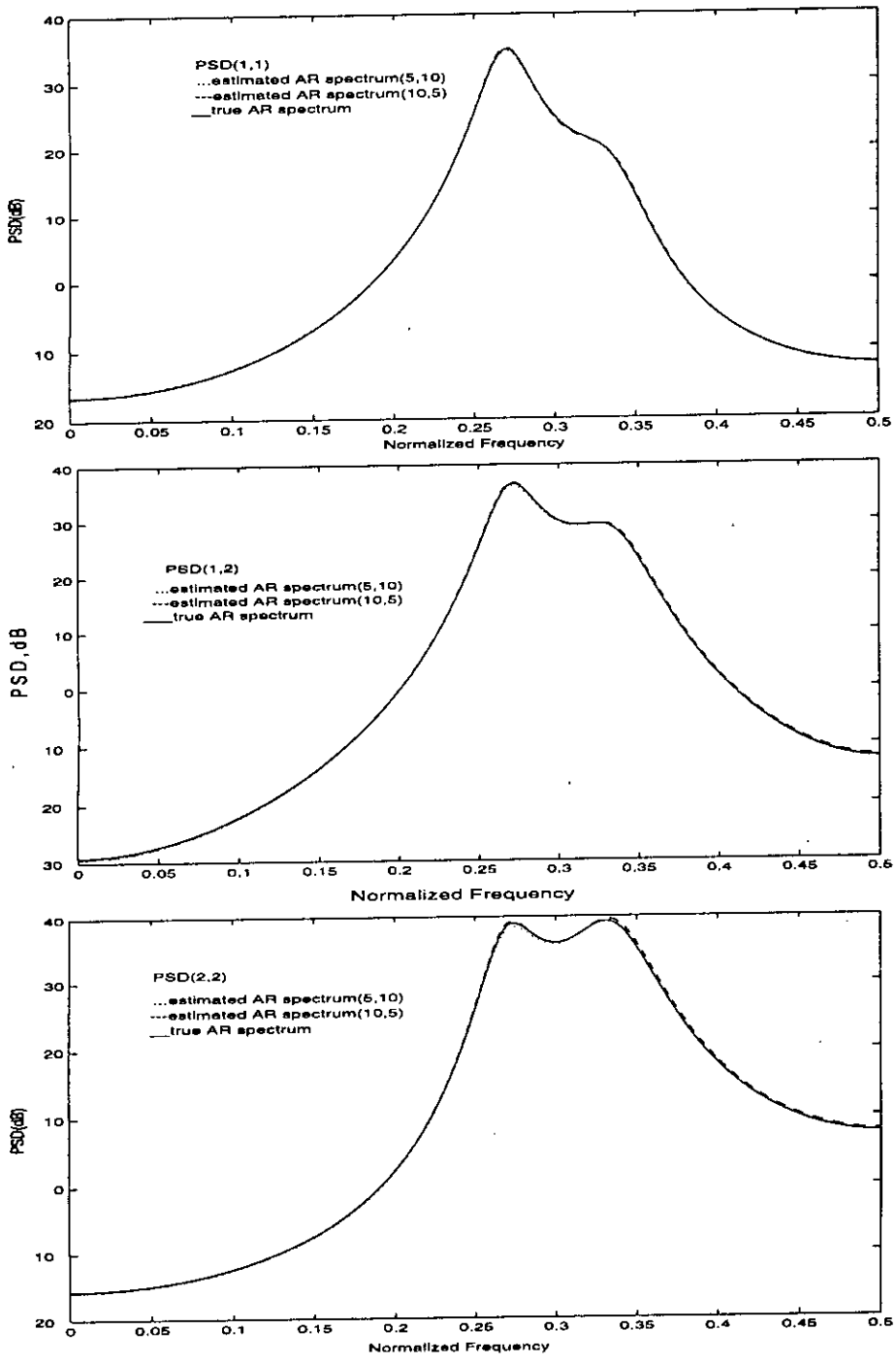

Fig. 4.9: True and estimated AR spectrum by the proposed fuzzy iterative search method for $SNR_1 = 5$ dB, $SNR_2 = 10$ dB

AR parameters are very close to their true counter parts. The standard deviation is also within the acceptable limit.

The power spectrum plots using the estimated AR parameters are shown in Fig. 4.8 and Fig. 4.9 for gradient search and fuzzy iterative search techniques, respectively. We see that estimated multichannel AR spectrum by the proposed methods is very close to true spectrum.

4.6 Conclusion

In this chapter the problem of power spectral estimation of multichannel AR processes corrupted by white observation noise of unknown power has been addressed. The noise-compensated Yule-Walker method has been used to solve the problem. Accurate noise compensation has been shown to be feasible by introducing new methods for estimating the multichannel observation noise variance on each channel. A new close form relation between the auxiliary variable for true and estimated noise variance vectors paves the way of estimating the noise power. Two different techniques have been proposed for estimating the observation noise variance. The proposed gradient based method converges very fast but depending on the initial value of the noise power it may converge to a false solution and even the estimated noise power may become negative. Fuzzy iterative search overcomes the problem of initialization. This approach is mostly suitable for an error surface with single minimum. The results shown for different SNRs demonstrate that the proposed schemes are very effective for multichannel AR spectral estimation from noisy observations.

Chapter 5

Conclusion

A new fuzzy incorporated noise compensation technique for autoregressive (AR) spectral estimation from noisy observations has been proposed. The major focus of the research was to develop a highly accurate estimation algorithm for AR parameters from the noise corrupted signal to eliminate the smoothing effect of noise on AR spectrum. To achieve this goal, unlike conventional approaches, uncompensated lattice filter (LF) algorithm, to avoid any possible instability, is employed to estimate the parameters of a high-order AR model fitted to the observed process. It is shown that this model preserve the true AR system roots precisely with some extra roots due to noise. As such a root matching technique has been adopted to extract the true system roots from the mixture of noise plus system roots. This technique has utilized the solution(s), single or multiple, from the low-order Yule-Walker (LOYW) equations as auxiliary solution. It has been shown that the iterative method based on the noise compensated LOYW equations [20] is blind to the desired solution of AR parameters particularly at low SNR due to strong inherent nonlinearity. As a by product, the proposed scheme has also resolved this problem of multiple solutions case of the LOYW equations.

We have also proposed a nonlinear iterative technique for finding the auxiliary solution using the LOYW equations. Nonlinear iteration is performed by fuzzy modeling the error index for calculating the step size adaptively. This has reduced the number of iterations significantly as compared to the previous one where a constant step size has been maintained throughout the total range of iteration. Clearly, the proposed method is not one-pass and utilizes a joint approach for spectral estimation. Nevertheless, if one is ready to pay the price, the gain in

terms of performance, especially in difficult situations such as low signal-to-noise ratio, or closely "spaced" signals, may be significant.

The power spectrum estimation problem of multichannel AR processes with additive white observation noise of unknown power is also addressed. The noise-compensated Yule-Walker method has been used to solve the problem. Accurate noise compensation has been shown to be feasible by introducing new methods for estimating the multichannel observation noise variance. Two different techniques have been proposed for estimating the observation noise variance. Both these methods have some advantages and disadvantages. Gradient based method converges very fast but depending on the initial value of the noise power it may converge to false solution and even the estimated noise power may become negative. Fuzzy iterative search overcomes the problem of initialization. This approach is mostly suitable for an error surface with single minimum.

However, unlike single channel AR processes, the case of multiple solutions particularly at low SNR have not been addressed. Further investigation is required to solve this problem.

The order of the process has been assumed to be known. However, the estimation of order of the system from noisy observation should be given further consideration.

Bibliography

[1] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis,*Upper Saddle River, NJ: Prentice-Hall (1997).

[2] S. M. Kay and S. L. Jr. Marple, "Spectrum Analysis-A modern Perspective",*Proceeding of the IEEE,* vol.69, no.11, pp.1380-1419, November 1980.

[3] A. Schuster,"On the investigation of hidden periodicities with application to supposed 26 day period of meteorological phenomena," *Terrestrial Magnetism,* vol. 3,pp. 13-41, Mar. 1898.

[4] A. Schuster, "The periodogram of magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871-1895," *Trans. Cambridge Philosophical Soc.,* vol. 18, pp. 107-135,1899.

[5] N. Wiener, "Generalized harmonic analysis," *Acta Mathematics,* vol. 55, pp. 117-258, 1930.

[6] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra From the Point of View of Commuinications Engineering.* New York: Dover, 1959.

[7] J. W. Cooley and J. W, Tukey, "An algorithm for machine calculation of complex Forier series," *Math. Comput.,* vol. 19, pp. 297-301, Apr. 1965.

[8] G. U. Yule, "On a method of investing periodicities in disturbes series, with specical reference to Wolfer's sunspot numbers," *Philosophical Trans. Royal Soc. London, Series A,* vol. 226, pp. 267-298, July 1927.

[9] G. Walker,"On periodicites in series of related terms,", *Proc. Roy. Soc. London, Series A,* vol. 131,pp. 518-532, 1931.

[10] E. Parzen, "Statistical spectral analysis (single channel case in 1968," Dep. Statistics, Standford Univ., Standford, CA, Tech. Rep. 11, June 10, 1968.

[11] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meeting Society of Exploration Geophysicists* (Oklahoma City, OK), Oct. 31, 1967.

[12] A. Vendenbos, "Alternative interpretation of maximum entropy spectral analysis," *IEEE Trans. Inform. Theory,* vol. IT-17, pp.493-494, July 1971.

[13] Akaike, H., "A new look at the statistical model identification", *IEEE Trans. Automat. Contr.*, vol.AC-19, no.6, pp. 716-723, Dec. 1974.

[14] S. Li, S. and Dickinson, B. W., "An efficient method to compute consistent estimates of the AR parameters of an ARMA model", *IEEE Trans. Automat. Contr.*, vol.AC-31, no.3, pp.275-278, Mar. 1986.

[15] S. Li, S. and Dickinson, B. W., "Application of the lattice filter to robust estimation of AR and ARMA models", *IEEE Trans. Acoust., Speech, Signal Processing*, vol.36, no.4, pp.502-512, Apr. 1988.

[16] Konstantinides, K., "Threshold bounds in SVD and a new iterative algorithm for order selection in AR models", *IEEE Trans. Signal Processing*, vol.39, no.5, pp. 1218-1221, May 1991.

[17] S. M. Kay, "Noise compensation for autoregressive spectral estimates", *IEEE Trans. Acoust., Speech, Signal Processing*, vol.28, no.3, pp.292-303, Jun. 1980.

[18] Kanai, H., Abe, M. and Kido, K., "Accurate autoregressive spectrum estimation at low signal-to-noise ratio using a phase matching technique", *IEEE Trans. Acoust., Speech, Signal Processing*, vol.35, no.9, pp.1264-1272, Sep. 1987.

[19] Sun, W. and Yahagi, T., "A new overfitting lattice filter for ARMA parameter estimation with additive noise", *IEICE Trans. Fundamentals*, vol.E75-A, no.2, pp.247-254, Feb. 1992.

[20] Yahagi, T., Hasan, M. K., "Estimation of noise variance from noisy measurements of AR and ARMA systems: Application to blind identification of linear time-invariant systems", *IEICE Trans. Fundamentals*, vol.E77, no.5, pp.847-855, May 1994.

[21] S. L. Marple, "High resolution autoregressive spectrum analysis using noise power cancellation,"in *Rec. 1978 IEEE Int. Conf. Speech and Signal Processing.*, pp. 345-348.

[22] J. M. Melsa and J.D. Tomick, "Linear predictive coding with additive noise for application to speech digitization,"in *Proc. 14th Allerton Conf. Circuits and Systems Theory,pp. 500-508,* Sept. 27, 1976.

[23] E.H. Satorius and J. T. Alexander, "High resolution spectral analysis sinusoids in corrlated nooise," in *Rec. 1978 ICASSP*(Tulsa, OK), pp. 349-351, Apr. 10-12, 1978.

[24] J. Treichler, "γ-LMS and its use in a noise-compensating adptive spectral analysis technique." in*Rec. 1979 ICASSP,* pp. 933-936.

[25] B. widrow et al., "Adaptive noise cancelling: Principles and application," *Proc. IEEE,* vol. 63 pp. 1692-1716, Dec. 1975.

[26] M. Pagano, "Estimation of models of autoregressive signal plus white noise,"*Ann. Statistics*, vol. 2, pp.99-108, 1974.

[27] Luis, V.-D., "New insights into the higher-order Yule-Walker equations", *IEEE Trans. Accoust., Speech, Signal Processing,* vol. 38, no.9, pp.1649-1651, Sep. 1990.

[28] Hu, H.-T., "Linear prediction analysis of speech signals in the presence of white Gaussian noise with unknown variance", *IEE Proc.-Vis. Image Signal Process.",* vol. 145, no. 4, pp.303-308, Aug. 1998.

[29] Hasan, M. K. and Yahagi T., "An iterative method for the identification of multichannel autoregressive processes with addititive observation noise", *IEICE Trans. Fundamentals*, vol. E79-A, no. 5,pp. 674-680, May 1996.

[30] E. Oran Brigham, *The Fast Fourier transform.* Englewood Cliffs, NJ: Prentice-Hall. 1974.

[31] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications.* San Francisco, CA: Holden-Day, 1968

[32] L. H. Coopmans, *The Spectral Analysis of Time Series.* New York: Academic Press, 1974.

[33] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 1965.

[34] E. Parzen, "Mathematical Considerations in the estimation of spectra," *Technometrics,* vol.3, pp. 167-190, May 1961.

[35] J. F. Prewitt,"Amplitude bias in the Fourier transforms of noisy signals," it IEEE Trans. Antennas Propagat., vol. AP-26, pp. 730-731, Sept. 1978.

[36] N. R. Strader, II, "Effects of subharmonic frequencies on DFT coefficient," *Proc. IEEE,* vol. 68, pp. 285-286, Feb. 1980.

[37] K. Toman,"The spectral shift of truncated sinusoids," *J. Geo-physical Res.,* vol. 70,pp. 1749-1750, Apr. 1,1965.

[38] F. J. Herris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE,* vol. 66,pp. 51-83, Jan. 1978.

[39] A. H. Nuttal, "Some windows with very good sidelobe behaviour," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. ASSP-29, pp. 84-89, Feb. 1981.

[40] E. C. Whitman,"The spectral analysis of discrete time series in terms of linear regressive models," Naval Ordnance Labs Rep. Noltr-70-109, White Oak, MD, June 23,1974.

[41] G.E.P. Box and G. M. Jenkins,*Time Series Analysis: Forcasting and Control.* San Francisco, CA:Holden-Day, 1970

[42] J. P. Burg ,"Maximum entropy spectral analysis,"Ph. D. dissertation, Dep. Geophysics, Stanford Univ., Standford, CA, May 1975

[43] R. M. Gray, "Toeplitz and circulant matrices: A review," Information Systems Laboratory, Center for Systems Research, Stanford University, Tech. Rep. No. 6502-1, June 1971.

[44] G.Cybenko,"Round-off error propagation in Durbin's , Levinson's, and Trench's Algorithms," *Rec. 1979 IEEE Int. Conf. Acoustics, Speech, and Signal Processing,* pp. 498-501

[45] J. Durbin,"The fitting of time series models," *Rev. Inst. Int. de Stat.,* vol. 28, pp. 233-244,1960

[46] N. Levinson,"The Wiener (root mean square) error criterion in filter design and prediction,"*J. Math. Phys.,*vol. 25, pp. 261-278,1947

[47] R.A.Wiggins and E.A.Robinson,"Recursive solution to the multi-channel filtering problem," *J. Geophysical Res.,* vol. 70, pp. 1885-1891,Apr. 1965.

[48] S. S. Haykin, Ed., *Nonlinear Methods of Spectral Analysis.* New York: Springer-Verlag, 1979

[49] Knight, W. C., R. G. Pridham, and S. M. Kay,"Digital Signal Processing for Sonar," Proc. IEEE, Vol. 69,pp. 1451-1506, Nov.1981.

[50] Mongingo, R. A., and T. W. Miller, Introduction to Adaptive Arrayas, Wiley, New York,1980.

[51] Justice, J. H.,"Array Processing in Exploration Seismology," in Array Signal Processimg, S Haykin et al., eds., Prentice-Hall, Englewood Cliffs, N.J.,1985.

[52] S. M.Kay, *Modern Spectral Analysis,*Upper Saddle River, NJ: Prentice-Hall (1997).

[53] Kay, S. M., *Modern Spectral Estimation, Theory and Application*, Englewood Cliffs, New Jersy, Prentice-Hall, 1988.

[54] Makhoul, J., "Linear prediction:A tutorial review", *Proc. IEEE*, vol.63, pp.561-580, Apr. 1975.

[55] Lacoss, R. T., "Data adaptive spectral analysis methods", *Geo-phys.*, vol.36, pp.661-675, Aug. 1971.

[56] Koopmans, L. H., *The Spectral Analysis of Time Series*, New York: Academic, 1974.

[57] Fuller, W. A., *Introduction to Statistical Time Series*, Wiley, 1976.

[58] Zhang, Q. T., "A statistical resolution theory of the AR method of mpectral analysis", *IEEE Trans. Signal Processing*, vol.46, no.10, pp.2757-2766, Oct. 1998.

[59] Cadzow, J. A., Baseghi, B., and Hsu, T., "Singular-value decomposition approach to time series modelling", *IEE Proc.*, vol. 130, Pt. F, no. 3, pp.202-210, Aug. 1983.

[60] S. M. Kay, "The Effects of noise on the autoregressive spectral esitmator",*IEEE Trans. Accoust., Speech, Signal Processing,* vol. 27, no. 5, pp.478-485, Oct. 1979.

[61] P. M. Djuric and S. M. Kay, "Order selection of autoregressive models", *IEEE Trans. Signal Processing,* vol.40, no.11, pp.2829-2833, Nov.1992.

[62] Mamdani, E. H. and Assilian, S., "An experiment in linguistic synthesis with a fuzzy logic controller", *Int. Journal Man-Machine studies*, vol.7, no.1, pp.1-13, 1975.

[63] Jang, J.-S. R., Sun., C.-T and Mizutani, E., *Neuro-Fuzzy and Soft Computing: A Comprehensive Approach to Learning and Machine Intelligence*, New Jersey, Prentice-Hall, 1997.

[64] S. L. Jr. Marple, Digital Spectral Analysis with Applications, Prentice-Hall, New Jersey, 1987.

[65] S. L. Jr. Marple and A. H. Nuttall, "Experimental comparison of three multichannel linear prediction spectral estimators," Proc. IEE, vol.130, no.3, pp.218-229, April 1983.

[66] R. D. Gupta and F. W. Fairman, "Parameter estimation for multivariable systems," IEEE Trans. Automat. Contr., vol.AC-19, no.5, pp.546-548, Oct. 1974.

[67] D. T. Pham and D. Q. Tong, "Maximum likelihood estimation for a multivariate autoregressive model," IEEE Trans. Signal Processing, vol. 42, no. 11, pp. 3061-3072, Nov. 1994.

[68] H. Sakai, "Circular lattice filtering using Pagano's method," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-30, no. 2, pp. 279-287, April 1982.

[69] P. Whittle, "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," Biometrika, vol.50, no.1, pp.129-134, 1963.

[70] R. Wiggins and E. A. Robinson, "Recursive solution to the multichannel filtering problem," J. Geophys. Res., vol.70, no.8, pp.1885-1891, April 1965.

[71] E. A. Robinson, Multichannel Time Series Analysis with Digital Computer Programs, Goose Pond Press, Texas, 1983.