A New Recursive High-Resolution Parametric Method for Power Spectral Density Estimation

by

Jijoong Kim, B.E. (Hons.)

Thesis submitted for the degree of

Master of Engineering Science

The University of Adelaide

Faculty of Engineering
Department of Electrical and Electronic Engineering

April, 1995

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in and University or other tertiary institution, and to the best of the author's knowledge and belief contains no material previously published or written by another person, except where due reference has been made in the text.

Should the thesis be accepted for the award of Degree, the author hereby consents to this copy, when deposited in the University Library, being made available for loan and photo-copying.

Signature _____

Date April 19, 1995

CONTENTS

ABSTRACT

The estimation of power spectral density (PSD) has progressed through several stages since the turn of the century. Research in this area has led to the development of a wide range of techniques and increased the number of applications.

This thesis reviews the conventional PSD estimation techniques developed in the last three decades. They are largely partitioned into two main streams: traditional non-parametric techniques and AR model based techniques, which are being applied to wide range of data by virtue of its potential to achieve increased spectral resolution. Resolution is an important consideration in PSD estimation. A technique is classified as a high-resolution if it can resolve two or more closely spaced signals at low signal to noise ratio (SNR). This thesis also includes the attributes of these conventional techniques, and some comparisons.

This brief review is followed by the introduction of a new recursive high resolution parametric method which stems from the well known *normal equations*. In the first part, we address the formulation of this method and derivation of the recursive equation, and study the stability of the system described by the recursive equation.

To show the advantages of the recursive method, an extensive performance test is conducted using short data which comprise one or two sinusoids corrupted with additive gaussian noise, and the results are evaluated and compared with those obtained from the Yule-Walker method, which was chosen because it falls under the same frame work as the proposed method.

Finally, the implementation of this method using FFT is discussed; we use pruning to reduce the computational complexity. The computational complexity achieved using this method is thoroughly evaluated and compared with that of existing techniques.

ACKNOWLEDGEMENT

I wish to express my sincere gratitude and deepest appreciation to my supervisor, Dr. Abdesselam Bouzerdoum, for his generous support and patient guidance throughout this research. The incorporation of his insightful suggestions and valuable comments have made this research all possible.

I am deeply indebted to several of my colleagues, Mr. Jihan Zhu, Mr. Carmine Pontecorvo, Mr. Richard Beare, for their cooperation and friendship.

Professor Robert E. Bogner introduced me to the field of signal processing, when he taught me my first signal processing class, and he accelerated my interest towards this field while supervising for my 4^{th} year project. He has always been inspiring and friendly. I am very grateful for his contribution to my education.

I wish to thank my parents, Mr. Pilki Kim and Mrs. Soonok Kim for their everlasting love, support, and prayer. I also give my appreciation to my brother and sister, and my dearest love Miss. Christine Jang for their love, faith, and patience. This work is dedicated to them. I will forever be indebted to them for their love and sacrifice throughout.

Last, but most, I wish to thank the Lord for making all this possible.

LiST OF FIGURES

LIST OF TABLES

NOTATION AND NOMENCLATURE

Most of the notation used throughout this thesis is standard to the signal processing area. For clarification, the list of the important notations in this thesis is provided in the following.

Matrix and Vector Notation

X	:	Matrix
$\boldsymbol{x}$	:	Vector
X^H	:	Hermitian transpose of X
X^T	:	Transpose of X
x^*	:	Complex conjugate of x
$\hat{x}$	:	Estimate of x
X^{-1}	:	Inverse of X
$x \circledast y$	:	Circular convolution of x and y
$FFT\{x\}$	:	Fast Fourier transformation of x
$E[x]$	:	Expected value of x
∇	:	Gradient Operator
$diag(X)$	:	Diagonal matrix whose elements are the main diagonal of X
$eig(X)$	:	Eigenvalues of X
X_{ij}	:	the ijth element of X
$\|x\|$	:	Modulus of x
$var[x]$	:	Variance of x

Parameters and Symbols

N	:	Number of data points in a data record
p	:	AR model order
h	:	Step size parameter in the recursive algorithm
Γ	:	Autocorrelation matrix
R	:	Estimate of autocorrelation matrix
U	:	Normalization factor of the Welch periodogram
P_{xz}	:	Power spectral density of a sequence $x(n)$
a	:	AR model parameter
f	:	Relative frequency in cycles per sample (cps)
$\gamma_{xx}(m)$	:	Autocorrelation of a sequence $x(n)$
$r_{xx}(m)$	:	Estimate of autocorrelation of a sequence $x(n)$
σ_x^2	:	Variance of a sequence $x(n)$
f_m	:	Error of forward linear prediction of order m
g_m	:	Error of backward linear prediction of order m
E_m	:	Total least squared error of linear prediction of order m
$\boldsymbol{r}$	:	Autocorrelation vector – $[\gamma_{xx}(1) \;\; \gamma_{xx}(2) \;\ldots\; \gamma_{xx}(p)]^T$
$\boldsymbol{y}$	:	Estimate of autocorrelation vetor – $[r_{xx}(1) \;\; r_{xx}(2) \;\ldots\; r_{xx}(p)]^T$

CHAPTER 1 *Introduction*

This chapter introduces Power Spectral Density (PSD) estimation, and presents the organisation of the thesis. The general description and the issues of the power spectral density are given in Section 1.1 on page 1 and the organisation of thesis is given in Section 1.2 on page 3.

1.1 Power Spectral Density

This thesis deals with the estimation of the spectral characteristics of signals that are characterised as stationary random processes. Many of the phenomena that occur in nature are best characterised statistically in terms of time averages. For example, meteorological phenomena such as the fluctuations in air temperature and pressure are best characterised statistically as random processes [76]. Thermal noise voltage generated in resistors and electronic devices are additional examples of physical signals that are well modelled by random processes [76].

With such signals, we must adopt a statistical view point which deals with the average characteristics of random signals. In particular, the autocorrelation function of a random process is an appropriate statistical average for characterizing random signals in the time

domain; spectral analysis provides a means for characterising random signals in the frequency domain.

In its simplest form, spectral analysis involves estimating the amplitudes of the harmonics of a periodic signal from a finite set of data samples, i.e., making an estimate of the energy in each Fourier component. Since practical signals are usually non periodic, it is preferable to adopt the term *energy density* rather than energy as a measure of frequency content of the signal. Thus consideration is made into the energy in a band of the frequencies rather than at a particular frequency.

If the signal is of infinite energy, a more appropriate measure is the *power spectral density*, (PSD), or *power density spectrum,* because for an infinite energy signal the Fourier transformation does not exist. Thus in order to deal with the widest class of signals, which would include deterministic signals as well as random signals, spectral analysis is very often defined as a technique for estimating the power spectral density of the signal to be analysed [31].

The mathematical framework for the theoretical analysis of random signals is provided by the theory of probability and stochastic processes.

Spectral analysis [78] is a large and much researched subject with numerous applications. Power spectral density analysis techniques have been widely applied in geophysical data processing [5], [13], [14], [16], [58], [59], [71], [78], [89], [95], [96], radar [36], [40], [49], [53], speech-communications [4], [32], [35], [42], [61], [63], [65], medical systems [27], [102], direction finding [25], [90], oceanography [39], radio astronomy [103], [104], sonar, and imaging as well as vibration analysis, non destructive testing and seismic signal investigations [31], [33]. The main purpose of spectrum analysis is to aid signal interpretation. With the knowledge of where the significant components of a signal occur, it is often possible to improve the coding of the signal, i.e. represents the signal with less bits of information [31].

Research in power density spectrum estimation has led to a variety of *parametric and non parametric techniques*, extensions to multi-dimensional, multi-channel, and spatio-temporal processing, and development of computationally fast algorithms [29]. Concur-

rently, various techniques have been developed for estimation of signal parameters, such as frequencies of sinusoids, poles and zeros of a signal. In some cases, parameter estimation is an intermediate step in spectrum analysis.

1.2 Organization of the Thesis

This thesis is organised as follows:

Chapter 2. In this chapter, we start off by briefly discussing the limitations associated with the classical PSD estimation methods and the increased applications and demands for short data PSD analysis, which prompted the advent of various model based PSD estimation methods. We then provide a comprehensive review of the conventional PSD estimation methods, tracking back to the evolution of the PSD estimation methods from the classical concepts based on the FFT and the autocorrelation techniques to the model based concepts (i.e. parametric PSD estimation), which are the spring board for the contemporary PSD estimators. The discussion is limited to a small subset of a large set of PSD estimation methods that appeared in the literature.

Chapter 3. In this chapter, we present a new method named the *recursive method* for the AR model based PSD estimation. This method enhances the spectral peak of a sinusoid by reducing the 3-dB bandwidth of the peak and increasing its height. We then address the mathematical formulation of this method by adopting the concepts of adaptive filtering for linear prediction [64], [66]. Using the well-known normal equations as a back bone and incorporating the steepest descent method, we derive a compact expression for the proposed recursive equations.

Chapter 4. In this chapter, we present the simulation results, and analyse the performance of the proposed method and compare it with the Yule-Walker method for PSD estimation [100], [106]. The Yule-Walker method was chosen for comparison because it falls under the same framework as the proposed method, i.e., both methods use the normal equations.

The objective is to compare these two methods on the basis of their bandwidth resolution, frequency bias and their robustness in the presence of additive noise, and then to identify their relative merits through discussion and illustrative simulations. The test data records used for performance analysis are short in duration and comprise one or two sinusoids in additive Gaussian noise.

The test performed are the following:

- Effect of signal to noise ratio (SNR): The SNR of the data is varied and the consequences of these variations on the quality of the spectrum are observed and discussed.
- Relative frequency: The relative frequency of the sinusoid is reduced until the two algorithms are unable to detect the corresponding peak in the spectrum and the behaviour in the low frequency region is examined.
- Step multiplication factors: A simple test is carried out to evaluate the appropriate step multiplication factors for the best performance of the recursive method.
- Initial phase of the sinusoid: We tested the methods by varying the initial phase of the sinusoid over the range $0 - 360°$ (3-dB bandwidth and frequency bias).
- 2 Sinusoids test: The resolving power of the recursive method and the Yule-Walker method were examined using two sinusoids very close to each other.

Chapter 5. This chapter addresses the implementation and the computational complexity of the recursive method, by incorporating the widely used FFT algorithm [17], [76] into the recursive equation. We study the analogy between matrix-vector multiplication and circular convolution of two sequences and then devise a pruned FFT algorithm [67], [73], [81]-[85] to compute the iterative equation in the frequency domain. We analyse the computational complexity of this method and make comparisons with other conventional methods.

Chapter 6. We restate briefly the significance of this work and our overall contribution. We then present some concluding remarks and summarize the thesis. Related topics for future research are suggested at the end of this chapter.

CHAPTER 2 *Review of Conventional Spectrum Estimation Methods*

This chapter contains a brief review of the conventional power spectral estimation techniques. They are largely divided into two categories: the classical non-parametric methods and the modern parametric (model-based) methods. The non-parametric methods, including periodograms and correlogram, are described in Section 2.2. Section 2.3 covers the three well known parametric methods: the Yule-Walker, Burg, and Unconstrained least-squares method. The conventional model order selection criteria are discussed in Section 2.4. Finally some test results of these conventional PSD estimators are included in Section 2.5.

2.1 Introduction

The estimation of PSD of a sampled stationary process is usually based on the FFT. Approaches based on the FFT are relatively efficient in computation time, and yield reasonable results for a large class of signals. The FFT approach, however, has a number of performance limitations; the most prominent one being the limit of frequency resolution due to the finite-length of the data record available. Spectral resolution, or frequency resolution, can be thought of as a measure of how closely spaced in frequency two sinusoids can become before they cannot be distinguished, and hence are detected as a single broad lobe.

A second limitation is due to the implicit windowing of the data that occurs when applying the FFT. The windowing manifests itself as *leakage* in the frequency domain obscuring and distorting other spectral components present, or even masking the weak spectral components. By adopting appropriate windows, the problem of side lobe leakage can be alleviated at the expense of degradation of spectral resolution. Harris [34] provides a good summary of the merits of various windows.

These two limitations are particularly noticeable when only short data records are available for analysis. In practice, such a situation arises frequently as most data records are finite in duration, non-stationary, or have slowly time-varying spectra that may only be considered constant for short durations. 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 [52].

In an attempt to alleviate the inherent limitations of the FFT approach, many alternative spectral estimation methods have been proposed over the last decade. The PSD estimation techniques fall into two broad classes: the classical non-parametric methods, and the modern model-based or parametric methods originating with the work of Yule [106]. The relationship among the conventional PSD estimation methods is shown in Fig. 2.1.

In this chapter, we describe the methods developed and applied by Walker [100], Bartlett [6], Blackman and Tukey [9], Burg [13] and others.

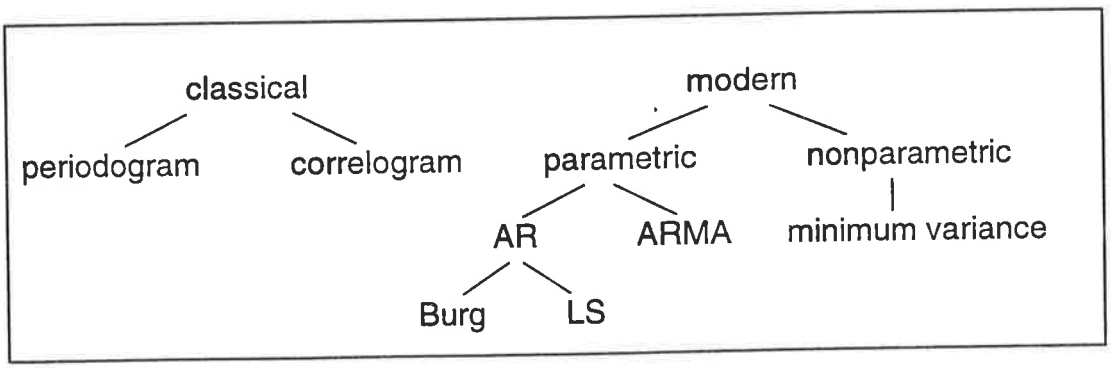

FIGURE 2.1 Overview of spectral analysis techniques

2.2 Non-Parametric Methods for PSD Estimation

The classical methods developed by Bartlett [6], Blackman and Tukey [9], and Welch [101] are considered here. These methods are referred to as non-parametric because they make no assumptions on how the data was generated. Their frequency resolution depends upon the length of the data record; the best resolution achieved is equal to the spectral width of a rectangular window of length N (approximately equal to $\frac{1}{N}$ at 3-dB points). These methods undergo a decrease in frequency resolution in order to reduce the variance of the spectral density estimator. The following subsections present the highlights only; references that provide greater detail of the classical spectral estimation techniques include Gardner [26], Jenkins and Watts[43] and Koopmans [56].

In the following two subsections, we consider the two most prominent non-parametric PSD estimates: the periodogram [6], [8], [19], [70] and the correlogram [9]. The periodogram is known as the direct method because the PSD estimate is obtained from direct computation of the Fourier transformation (FT) of the signal. The second method, the correlogram, is called the indirect method because it requires two steps. First, an autocorrelation is computed from the signal samples, and then the PSD estimate is obtained by computing the FT of the autocorrelation sequence.

2.2.1 Periodogram PSD Estimators

The simplest form of the formal definition of the PSD, based on ergodicity, has the discrete time form

$$P_{xx}(f) = \lim_{N \to \infty} E\left[\frac{1}{(2N+1)} \left| \sum_{n=-N}^{N} x(n)\, e^{-j2\pi fn} \right|^2 \right] \tag{2.1}$$

All that is available, in a practical situation, is the finite sequence of N signal samples $\{x(0), ..., x(N-1)\}$.

Ignoring the expectation operation and assuming a finite data set of N samples, we can write the estimate of the power spectral density as

$$\hat{P}_{xx}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)\, e^{-j2\pi fn}\right|^2 = \frac{1}{N}|X(f)|^2 \qquad (2.2)$$

This technique is the simplest form of what is known as a periodogram, the structure of which is illustrated in Fig. 2.2. The periodogram is appropriate for analysis of ergodic random signals; i.e., signals whose time averages are equivalent to their statistical averages. The popularity of this method stems from the existence of efficient FFT algorithms for computing the DFT.

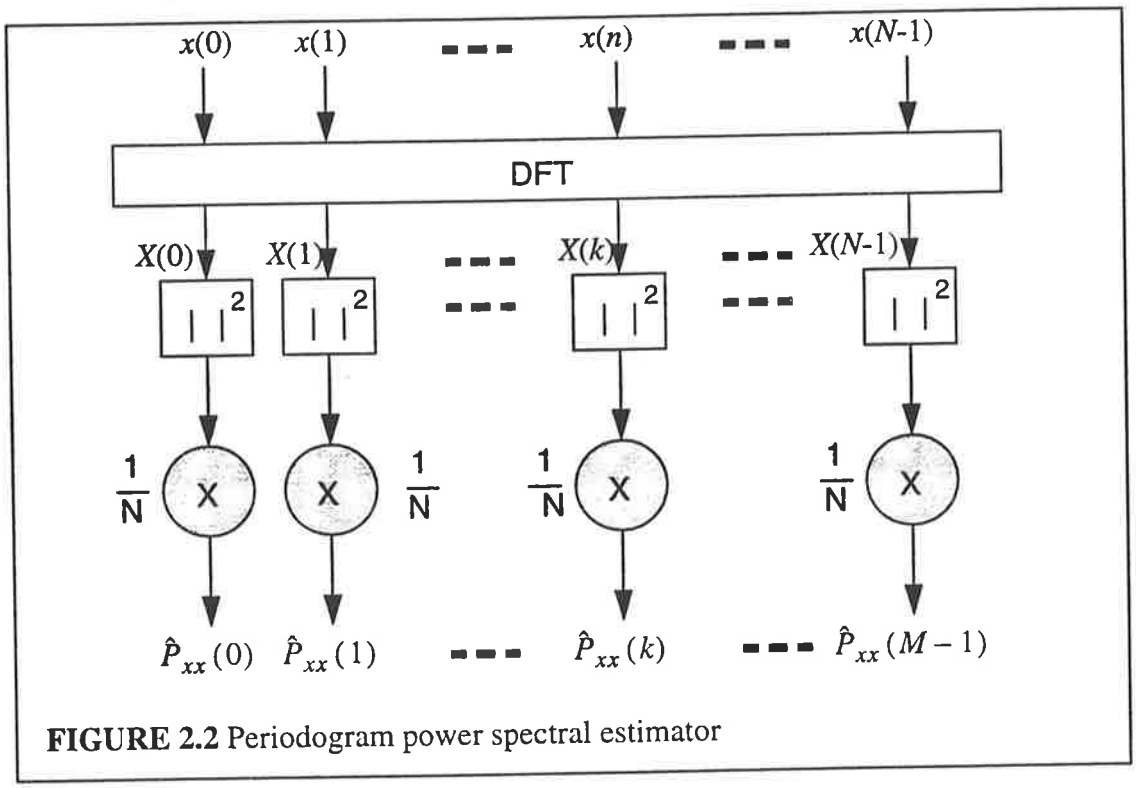

FIGURE 2.2 Periodogram power spectral estimator

The periodogram is not a consistent estimate of the true PSD, and it is only asymptotically unbiased. Furthermore, it can be shown that the mean of the estimated spectrum, using Eqn. (2.2), is the convolution of the true PSD with $W_B(f)$, the FT of the Bartlett (Rectangular) window.

$$E\left[\hat{P_{xx}}(f)\right] = \int_{-1/2}^{1/2} P_{xx}(\alpha)\, W_B(f-\alpha)\, d\alpha \quad \text{where } \mu_x = E[x(n)] = 0$$

Therefore, the mean of the estimated spectrum is a smoothed version of the true PSD. Furthermore, the PSD estimate in Eqn. (2.2) suffers from the spectral leakage problem inherent in all windowing methods.

One must resort to pseudo ensemble averaging in order to smooth the periodogram PSD estimate. Daniel's periodogram [19], Bartlett periodogram [6] and Welch periodogram [101] averaging methods are considered here.

2.2.1.1 Daniel's Periodogram

This method [19], [70] smoothes the rapid fluctuation of the sample spectrum by averaging over adjacent spectral frequencies. The modified periodogram estimate at frequency f_i can be obtained by averaging the M points on either side of f_i.

$$\hat{P}_{xx}^{D}(f_i) = \frac{1}{2M+1} \sum_{k=i-M}^{i+M} \tilde{P}_{xx}(f_k) \;,\; f_k = \frac{k}{N} \text{ and } M \le i \le N-M-1 \tag{2.3}$$

This can be expressed as the convolution of the sample spectrum with a low pass filter

$$\hat{P}_{xx}^{D}(f) = \hat{P}_{xx}(f) * H(f)$$

This concept can be thought of as passing the sample spectrum through a low pass filter with frequency response $H(f)$.

2.2.1.2 Bartlett Periodogram

To reduce the variance in the periodogram, a pseudo-ensemble is created by dividing the N-point sequence into K non-overlapping segments each of length M.

$$x_i(n) = x(n+iM) \qquad \text{where } i = 0, 1, \ldots, K-1 \quad n = 0, 1, \ldots, M-1 \tag{2.4}$$

For each segment, we compute the periodogram

$$\tilde{P}_{xx}^{(i)}(f) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x_i(n)\, e^{-j2\pi fn}\right|^2, \quad i = 0, 1, \ldots, K-1 \tag{2.5}$$

At each frequency of interest, the periodogram for the K segments are averaged to produce the Bartlett averaged periodogram

$$\hat{P}_{xx}^{B}(f) = \frac{1}{K}\sum_{i=0}^{K-1} P_{xx}^{(i)}(f) \tag{2.6}$$

Provided that $E[x(n)] = 0$ and $\sum_{n=-\infty}^{\infty} |\gamma_{xx}(n)| < \infty$, the mean of this PSD estimate is

$$E\left[\hat{P}_{xx}^{B}(f)\right] = \frac{1}{K}\sum_{i=0}^{K-1} E\left[\hat{P}_{xx}^{(i)}(f)\right] \tag{2.7}$$

$$E\left[\hat{P}_{xx}^{B}(f)\right] = E\left[\hat{P}_{xx}^{(i)}(f)\right] \tag{2.8}$$

$$E\left[\hat{P}_{xx}^{B}(f)\right] = \sum_{m=1-M}^{M-1}\left(1-\frac{|m|}{M}\right)\gamma_{xx}(m)\, e^{-j2\pi fm}, \quad \gamma_{xx} = E\left[x^*(t)\, x(t+m)\right] \tag{2.9}$$

$$E\left[\hat{P}_{xx}^{B}(f)\right] = \frac{1}{M}\int_{-\frac{1}{2}}^{\frac{1}{2}} P_{xx}(\alpha)\left(\frac{\sin(\pi(f-\alpha)M)}{\sin(\pi(f-\alpha))}\right)d\alpha \tag{2.10}$$

$$E\left[\hat{P}_{xx}^{B}(f)\right] = P_{xx}(f) * W_B(f) \tag{2.11}$$

Clearly the true spectrum is convolved with the FT of the Bartlett window. The Bartlett averaged periodogram can be shown to have variance inversely proportional to the number of segments, i.e., $var\left[\hat{P}_{xx}^{B}(f)\right] \propto \frac{P_{xx}^2(f)}{K}$.

The resolution has now been degraded due to segmentation of the data into samples each of length, M, and the effective spectral window will have a broader main lobe bandwidth. For fixed $N = K \cdot M$, there is the usual trade-off between high spectral resolution (M is as large as possible) and small estimation variance (K is as large as possible).

2.2.1.3 Welch Periodogram

Welch [101] made two basic modifications to the Bartlett window. First, the data is divided into L segments of M samples each, with a shift of D samples between adjacent segments; thus,

$$x_i(n) = x(n+iD), \qquad \begin{matrix} n = 0, 1, \ldots, M-1 \\ i = 0, 1, \ldots, L-1 \end{matrix} \qquad (D \le M) \tag{2.12}$$

where iD is the starting point of the ith segment.

The second modification made by Welch to the Bartlett Method is windowing the data prior to the computation of the segment periodogram.

$$\tilde{P}_{xx}^{(i)}(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x_i(n)\, w(n)\, e^{-j2\pi fn}\right|^2, \qquad i = 0, 1, \ldots, L-1 \tag{2.13}$$

where U is the normalization factor

$$U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n) \tag{2.14}$$

The Welch power density spectrum estimate is the average of the modified periodograms

$$\hat{P}_{xx}^{W}(f) = \frac{1}{L}\sum_{i=0}^{L-1} \tilde{P}_{xx}^{(i)}(f) \tag{2.15}$$

Provided that $E[x(n)] = 0$ and $\sum_{n=-\infty}^{\infty} |\gamma_{xx}(n)| < \infty$, the mean of this PSD estimate is

$$E\left[\hat{P}_{xx}^{W}(f)\right] = \frac{1}{L}\sum_{i=0}^{L-1} E\left[\tilde{P}_{xx}^{(i)}(f)\right] \tag{2.16}$$

$$= E\left[\tilde{P}_{xx}^{(i)}(f)\right] \tag{2.17}$$

$$= P_{xx}(f) * W(f) \tag{2.18}$$

where, by definition,

$$W(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} w(n)\, e^{-j2\pi fn}\right|^2 \tag{2.19}$$

Welch specifically proposed the use of the Hanning window [70], [101], and 50% overlap between the segments, which leads to a very efficient implementation with the FFT algorithm.

Like the Bartlett periodogram, the variance of the Welch periodogram is roughly inversely proportional to the number of segments L.

$$var\left[\hat{P}_{xx}^{W}(f)\right] \propto \frac{P_{xx}^{2}(f)}{L} \tag{2.20}$$

Due to overlap, more segments are obtained, for a fixed data record, in the Welch method than in the Bartlett method ($L \geq K$). As a result, the Welch periodogram tends to have less variance than the Bartlett periodogram.

2.2.2 Correlogram for PSD

An alternative approach to the PSD is the so called correlogram. Before we discuss this, we must first introduce the concept of correlation. The autocorrelation $\gamma_{xx}(m)$ of a sequence $x(n)$ is the expectation of the product of $x(n)$ with a time delayed replica of itself, $x(n+m)$.

$$\gamma_{xx}(m) = E[x(n)\,x(n+m)] \tag{2.21}$$

If $x(n)$ is an ergodic process, the expectation operator in Eqn. (2.21) can be replaced by the time average, which can be approximated by

$$r_{xx}(m) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\,x(n+m)\,. \tag{2.22}$$

A DFT of the set of estimated autocorrelation coefficients then provides an estimate of the PSD, $\hat{P}_{xx}(f)$, as shown below.

$$\hat{P}_{xx}(f) = \sum_{n=0}^{M-1} r_{xx}(n)\,e^{-j2\pi fn} \qquad M < N \tag{2.23}$$

This is the correlogram spectral estimator. Fig. 2.3 illustrates the structure of the correlogram.

Often the N samples are padded with extra zeros to increase the data size. This does not improve the resolution of the estimated PSD but it generates more spectral points.

A more detailed discussion on the correlogram method can be found in Blackman and Tukey [9], who were the first to extensively study and popularize the discrete time correlogram method of PSD estimation. They suggested the maximum lag of $M \approx \frac{N}{10}$ to avoid the greater statistical variance associated with higher lags of the autocorrelation estimate.

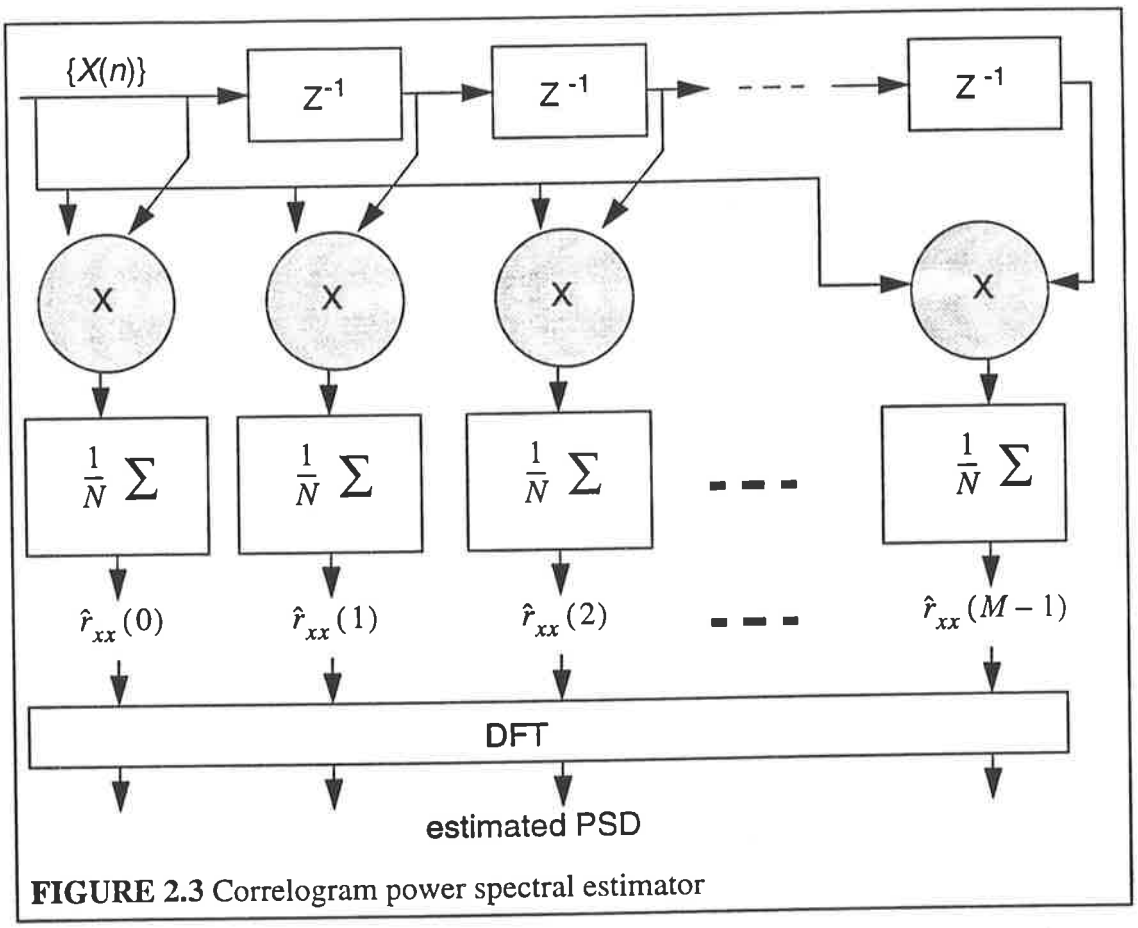

FIGURE 2.3 Correlogram power spectral estimator

2.2.3 Summary

Classical techniques for spectral analysis are based on either the periodogram or the correlogram, or hybrids of the two. They show *robustness* in the signal environment. An additional advantage is that they are well understood because of their age, and can be made computationally efficient through the use of the FFT.

Classical techniques, however, have the disadvantage that they provide a poor compromise between spectral resolution, and spectral quality which is the measure of how good the PSD is in a statistical sense. The ability to resolve two signals is fundamentally limited by the number of samples, *N*. Processing more samples improves the resolution, however, these classical techniques are not appropriate for non-stationary signals because the classical techniques implicitly assume that the signal repeats itself in a periodic manner outside the analysis window (stationarity is assumed). Also they are subject

to spectral leakage, which distorts the desired spectrum, and masks weak signals. For the Blackman-Tukey correlogram, negative PSD values may be obtained in the spectrum with some window weightings and autocorrelation estimates (e.g. unbiased) [52].

2.3 Parametric Methods for PSD Estimation

The classical approach uses the FFT operation on either windowed data or windowed lag estimates (autocorrelation estimates). The basic limitation of these non-parametric methods is the inherent assumption that the unobserved data of lag values outside the window are zeros. This assumption severely limits the frequency resolution and the resultant quality of the power spectrum.

In this section, we describe methods that use *a priori* information, or assumption, on how the data was generated. The model for the generation of the data is constructed with a number of parameters which can be estimated from the observed data. With the model and the estimated parameters we can obtain the spectral estimate by substituting the estimated model parameters into the theoretical PSD implied by the model. The parametric modelling is depicted in Fig. 2.4.

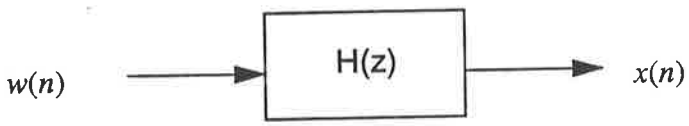

FIGURE 2.4 Parametric model of random process

where $w(n)$ is the stationary discrete-time stochastic process, and $x(n)$ is the observed signal. We may then write [75]

$$P_{xx}(f) = |H(f)|^2 \cdot P_{ww}(f) \tag{2.24}$$

where $P_{xx}(f)$ and $P_{ww}(f)$ are the power density spectra of $x(n)$ and $w(n)$ respectively. The most straight forward approach to parametric spectral estimation is to assume that

the signal from the filter $H(z)$ is auto-regressive (AR), which means that the present sample value is a weighted sum of the p previous sample values.

This method is, by far, the most widely used because firstly, the AR model is suitable for representing spectra with narrow peaks, and secondly, the AR model results in very simple linear equations to be solved for the AR parameters [76].

2.3.1 Yule-Walker Method

In the Yule-Walker method [100], [106], we estimate the autocorrelation from the data and use the estimate, that is the Yule-Walker or *normal equations*, to solve for the AR model parameters. In this method, it is desirable to use the biased form of the autocorrelation estimate

$$r_{xx}(m) = \frac{1}{N}\sum_{n=0}^{N-|m|-1} x^*(n)\,x(n+m)\,, \qquad 0 \le m \le N-1 \tag{2.25}$$

to reduce the risk of ill-conditioning. The result is a stable AR model, but at the expense of degrading the AR spectral resolution, and shifting of spectral peaks from their true locations. The Levinson-Durbin Algorithm [18], [21], [60], [105], provides the solution of the normal equations with a computational complexity of $O(p^2)$.

The corresponding power spectrum estimate is

$$P_{xx}^{YW}(f) = \frac{\hat{\sigma}_{wp}^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_p(k)\, e^{-j2\pi fk}\right|^2} \tag{2.26}$$

where $\hat{a}_p(k)$ are estimates of the AR parameters obtained from the Levinson-Durbin recursion and $\hat{\sigma}_{wp}^2$ is the estimated minimum mean-square value for the p^{th} order predictor. This method suffers degradations of spectral resolution, frequency estimation bias,

and spectral line splitting. The reasons for spectral line splitting have been documented by Kay and Marple [51].

2.3.2 Burg Method

This method was devised by Burg [13], [14], [15] for estimating the AR parameters. It may be viewed as a constrained least-squares estimation procedure using the sum of the forward and backward errors in linear prediction, with the constraint that the AR parameters satisfy the Levinson-Durbin algorithm.

Assuming an all-pole stationary stochastic process, the forward and backward linear prediction errors of order m are given by

$$f_m(n) = \sum_{k=0}^{m} a_m(k)\, x(n-k) \tag{2.27}$$

$$g_m(n) = \sum_{k=0}^{m} a^*_m(k)\, x(n-m-k)\ ,\ 0 \le k \le m-1\ ,\ m = 1, 2, \ldots, p \tag{2.28}$$

where $a_m(k)$, is the prediction coefficient. The least-square error is

$$E_m = \sum_{n=1}^{N-1} \left[|f_m|^2 + |g_m|^2 \right] \tag{2.29}$$

This error is minimised by selecting the appropriate prediction coefficients with the constraint that they satisfy the Levinson-Durbin recursion given by

$$a_m(k) = a_{m-1}(k) + a_m(m)\, a^*_{m-1}(m-k) \tag{2.30}$$

where $a_m(m)$ is the mth reflection coefficient in the lattice filter realization of the predictor.

Now if we substitute Eqn. (2.30) into Eqn. (2.27) and Eqn. (2.28), we get the following recursive relationships

$$f_m(n) = f_{m-1}(n) + a_m(m)\, g_{m-1}(n-1) \tag{2.31}$$

$$g_m(n) = g_{m-1}(n-1) + a^*_m(m) f_{m-1}(n) \quad \text{for } 1 \le n \le N-m \tag{2.32}$$

Substituting Eqn. (2.31) and Eqn. (2.32) into Eqn. (2.29) and minimizing with respect to $a_m(m)$, we obtain

$$a_m(m) = \frac{-2\sum_{n=m}^{N-1} f_{m-1}(n)\, g^*_m(n-1)}{\sum_{n=m}^{N-1}\left[\left|f_{m-1}(n)\right|^2 + \left|g_{m-1}(n-1)\right|^2\right]}, \quad m = 1, 2, \ldots, p \tag{2.33}$$

Andersen [3] discovered the recursive relationship for the denominator term that reduces the computational effort needed in computing $a_m(m)$. Since the Levinson recursion is maintained in the Burg algorithm, then

$$\hat{E}_m = \hat{E}_{m-1}\left[1 - \left|a_m(n)\right|^2\right] \tag{2.34}$$

From the estimation of the AR parameters, we form the power density spectrum estimate

$$P_{xx}(f) = \frac{\hat{E}_p}{\left|1 + \sum_{k=1}^{p} \hat{a}_p(k)\, e^{-j2\pi fk}\right|^2} \tag{2.35}$$

The Burg method results in high spectral resolution, yields a stable AR model, and is computationally efficient. However, it exhibits undesirable spectral line splitting for high SNRs and high model orders (see Fougere [23], [24]), spurious peaks for high order

models [79], [97], degradation of resolution at low SNR [50], [68], and sensitivity to the initial phase of short-length sinusoids [52], [91]. An analytical discussion of the reasons for line splitting in the Burg method has been presented by Herring [38]. Swingler [87] also provided simulations that indicated that Hamming taper function reduced the frequency bias. Several modifications involving the introduction of a weighting sequence on the squared forward and backward errors have been proposed to improve those inherent limitations.

2.3.3 Unconstrained Least-Squares Method

In contrast to the Burg method which is basically a least-squares lattice algorithm with an added constraint that the predictor coefficients must satisfy the Levinson recursion, Marple [69] proposed an unconstrained least-squares algorithm to determine the AR parameters.

The forming of the forward and backward linear prediction estimate, and their corresponding errors is identical to that of Burg method. However, the Levinson-Durbin recursion is not imposed on the AR parameters.

The unconstrained minimization of the squared error with respect to the prediction coefficients yields a set of linear equations.

$$\sum_{k=1}^{p} a_p(k)\, r_{xx}(l,k) = -r_{xx}(l,0)\,, \qquad l = 1, 2, \ldots, p \tag{2.36}$$

The resulting residual least-squares error is

$$\hat{E}_p = r_{xx}(0,0) + \sum_{k=1}^{p} \hat{a}_p(k)\, r_{xx}(0,k) \tag{2.37}$$

Hence the unconstrained least-squares power density spectrum estimate is

$$P_{xx}(f) = \frac{\hat{E}_p}{\left|1 + \hat{a}_p(k)\, e^{-j2\pi fk}\right|^2} \tag{2.38}$$

The correlation matrix implied by Eqn. (2.36) is not toeplitz, and hence the Levinson-Durbin algorithm can not be applied. Marple [69] devised an algorithm which has a lattice structure and employs Levinson-Durbin type order recursions and additional time recursions. Its performance characteristics are superior to those of the Burg algorithm in the sense that this method does not exhibit the same sensitivity to such problems as frequency bias, spurious peaks, and spectral line splitting [69], [87].

Its computational efficiency is almost comparable to that of the Levinson-Durbin algorithm and the Burg algorithm, requiring about 20% more computations than the Burg algorithm. With this method there is no guarantee of obtaining a stable model, but in spectrum estimation this is not considered to be a significant problem.

2.3.4 Model Order Selection Criteria

The best choice of model order p is generally not known *a priori,* and choosing the correct model order is a major issue in all AR modelling problems. The choice of the model order, p, is a compromise between the degree of spectral detail required, the accuracy, and the computational requirement. An adequate choice of model order is such that there is a sufficient number of poles to represent all expected resonances (two poles per sinusoid) with some additional poles to give spectral shaping, and to approximate the troughs in the spectrum. Too low a model order results in a highly smoothed spectral estimate. If the model order selected is too high, however, spurious low level peaks in the spectrum are likely to occur.

Much work has been done by researchers on this problem and many experimental results have been presented in the literature (see, Gersch and Sharpe [28], Ulrych and Bishop [97], Tong [92]-[94], Jones [45], [46], Berryman [7], Kashyap [48], Huzzii [41], Kane [47] and Kozin [57]).

Akaike [1], [2] has proposed two criteria, for selecting the model order, called the *Final Prediction Error* (FPE) criterion and the *Akaike Information Criterion* (AIC) which are given below.

$$FPE(p) = \hat{\sigma}_{wp}^2\left(\frac{N+p+1}{N-p-1}\right) \tag{2.39}$$

$$AIC(p) = N\ln\left(\hat{\sigma}_{wp}^2\right) + 2p \tag{2.40}$$

where $\hat{\sigma}_{wp}^2$ represents the estimated variance of the linear prediction error.

The model order p is chosen to minimise the performance indices FPE(p) and AIC(p) respectively. The FPE(p) criterion works fairly well with a pure AR process, but tends to underestimate the model order p against the actual process [7], [45], [46], [97]. The AIC(p) criterion adopts the maximum likelihood approach to obtain the model order which minimises the AIC(p). This criterion has the disadvantage that it is statistically inconsistent; i.e., it overestimates the model order as N becomes very large [48].

These two methods proposed by Akaike are both based on *Asymptotic Information Theory*, and they are asymptotically equivalent.

An alternative information criterion, proposed by Rissanen [77], is based on selecting the model order which *minimises the description length* (MDL), where MDL is defined as

$$MDL(p) = N\ln\hat{\sigma}_{wp}^2 + p\ln N \tag{2.41}$$

This criterion is found to be statistically consistent [82].

The fourth criterion has been proposed by Parzen [74]. This is called the *criterion autoregressive transfer* (CAT) function and is defined as

$$CAT(p) = \left(\frac{1}{N}\sum_{k=1}^{p}\frac{1}{\sigma_{wk}^2}\right) - \frac{1}{\hat{\sigma}_{wp}^2} \tag{2.42}$$

where

$$\bar{\sigma}_{wk} = \frac{N}{N-k}\hat{\sigma}_{wp} \tag{2.43}$$

Ulrych and Clayton [98] have found that for short data segments neither the FPE, AIC nor CAT work well. Also for harmonic processes in noise, the FPE and AIC tend to underestimate p if the SNR is high [37], [59].

According to other experimental results, for data records with small lengths the optimum AR model order selected, in many cases, lies in the range $N/3$ to $N/2$ [76], [92]. Another reference is Broersen [12], which describes a Finite Sample Theory for AR model order selection. This theory takes into account the distinction existing between the different estimation methods hence being more accurate for finite samples. In that paper, for most existing model order-selection criteria, equivalents are given, based on the finite sample theory. Their performance is better for small samples, and they converge to the existing criteria for increasing sample size.

As a concluding remark, it is suggested that when the prior information in regard to the physical process implied by the data, is not provided or unknown, one should try different model orders and different criteria, and ultimately interpret the different results, or use the information resulting from non-parametric methods, which in practice would disqualify many parametric results.

2.3.5 Summary

In this subsection, we briefly list the attributes of the three AR PSD estimation method.

Yule-Walker Method:

- Model order must be selected.
- Better resolution than FFT or BT, but not as good as other AR methods.
- Spectral line splitting occurs.

- Implied windowing caused by the biased autocorrelation estimate distorts the spectrum.
- No side lobes.
- Applicable to seismic, speech, radar clutter data.
- Stable linear prediction filter is guaranteed if biased lag estimates are used.

Burg Method:

- Must determine model order.
- High resolution for low noise levels.
- Good spectral fidelity for short data records.
- Spectral line splitting can occur.
- Bias in the frequency estimates of the peaks.
- No implied windowing.
- No side lobes.
- Stable linear prediction filter guaranteed.
- Adaptive filtering applicable.
- Constrained recursive least-squares approach.

LS method:

- Sharper response for narrow band processes than other AR estimates.
- No spectral line splitting observed.
- Reduced bias in frequency estimates.
- No side lobes.
- Must determine model order.
- Stable linear prediction filter not guaranteed, but stable in most cases.
- Based on exact recursive least squares solution with no constraint.

2.4 Performance Comparisons

We have highlighted the key properties of the conventional AR PSD estimation techniques. In this section, we illustrate some typical PSD estimates of these techniques to better explain what the actual spectra obtained by these techniques look like.

Fig. 2.5 illustrates the PSD estimates of the eight conventional techniques discussed in this chapter. The data record consists of a single sinusoid and additive gaussian noise. The frequency axis ranges from 0.0 cps to 0.5 cps, and represents the fraction of the sampling frequency; we will call it *relative frequency* in cycles per sample *(cps)*. The sinusoid has a relative frequency of 0.3 cps and SNR of 20dB. Fig. 2.5 is intended to illustrate properties of each technique, especially for short data records, rather than to serve as a basis for comparing relative performance among the techniques.

A PSD estimate using the periodogram is shown in Fig. 2.5 (a). When generating this periodogram, 236 zeros are padded to the data record to produce a better picture because the 256-point FFT provides finer interpolation than the 20-point FFT does. Doing this also resolves potential ambiguities, and reduces the quantization error in the accuracy of estimating the frequency of a spectral peak. The nominal resolution of the 20-point data record is 0.05 cps. In fact all the techniques mentioned in this chapter except for the Bartlett averaged periodogram in Fig. 2.5 (c), and the Welch averaged periodogram in Fig. 2.5 (d) are based on a 20-point sample sequence. The Welch and Bartlett periodograms involve segmentation. A segmentation of a 20-point sample sequence into a number of smaller sequences results in dramatically degraded resolution. What we want to demonstrate here is the typical shape of the spectra with spectral resolutions based upon a 20-point data record. The improvement on variance of the estimate due to the segmentation is not something we can put on a display. Hence, for these two methods, a 100-point sample sequence is used instead, and segmented into a number of 20-point sample sequences. Each 20-point sample sequence is then padded with 236 zeros for evaluating the PSD estimate at frequencies spaced $\frac{1}{256}$ cps, just like the other classical PSD estimates. For the Welch method, a 50% overlapping between the successive data segments, i.e., 10 samples, is tried, and the Hanning window is used. For the Daniel periodogram,

each sample spectrum is averaged over 4 adjacent spectral frequencies, i.e., two on each side.

A 20-point data record may be considered to be obtained by windowing an infinite length sample sequence with a rectangular window. The use of this data implicitly assumes that the unmeasured data is zero. It is evident in Fig 2.5 (a), (b), (c), (e), that the periodograms and the correlogram, except the Welch averaged periodogram, suffer from the distorting effect of leakage. The sidelobes exhibit a characteristic *sinc* function, which is the transform of a rectangular window. The heights of the sidelobes are relatively large, therefore if there was any weak frequency component with a height less than say 10dB, then it will certainly be masked by one of the sidelobes. The Welch averaged periodogram does not exhibit any prominent sidelobes, however the bandwidth of the mainlobe is almost doubled.

The spectral resolution with the classical PSD estimates are significantly poorer than that of the modern AR PSD estimates. The conventional AR PSD estimates are shown in Fig. 16 (f)-(g). Although all are AR PSD estimates, differing only in the manner that the AR coefficients are estimated, the resulting AR PSD estimates are different. These PSD estimates do not exhibit sidelobes, and show an improved resolution. The AR PSD estimate based on the unconstrained least-squares method in Fig. 2.5 (h) provides the sharpest response at f_0=0.3 cps. From the graphs and the previous discussion, the unconstrained least-squares method seems to be the best amongst the methods discussed in this example.

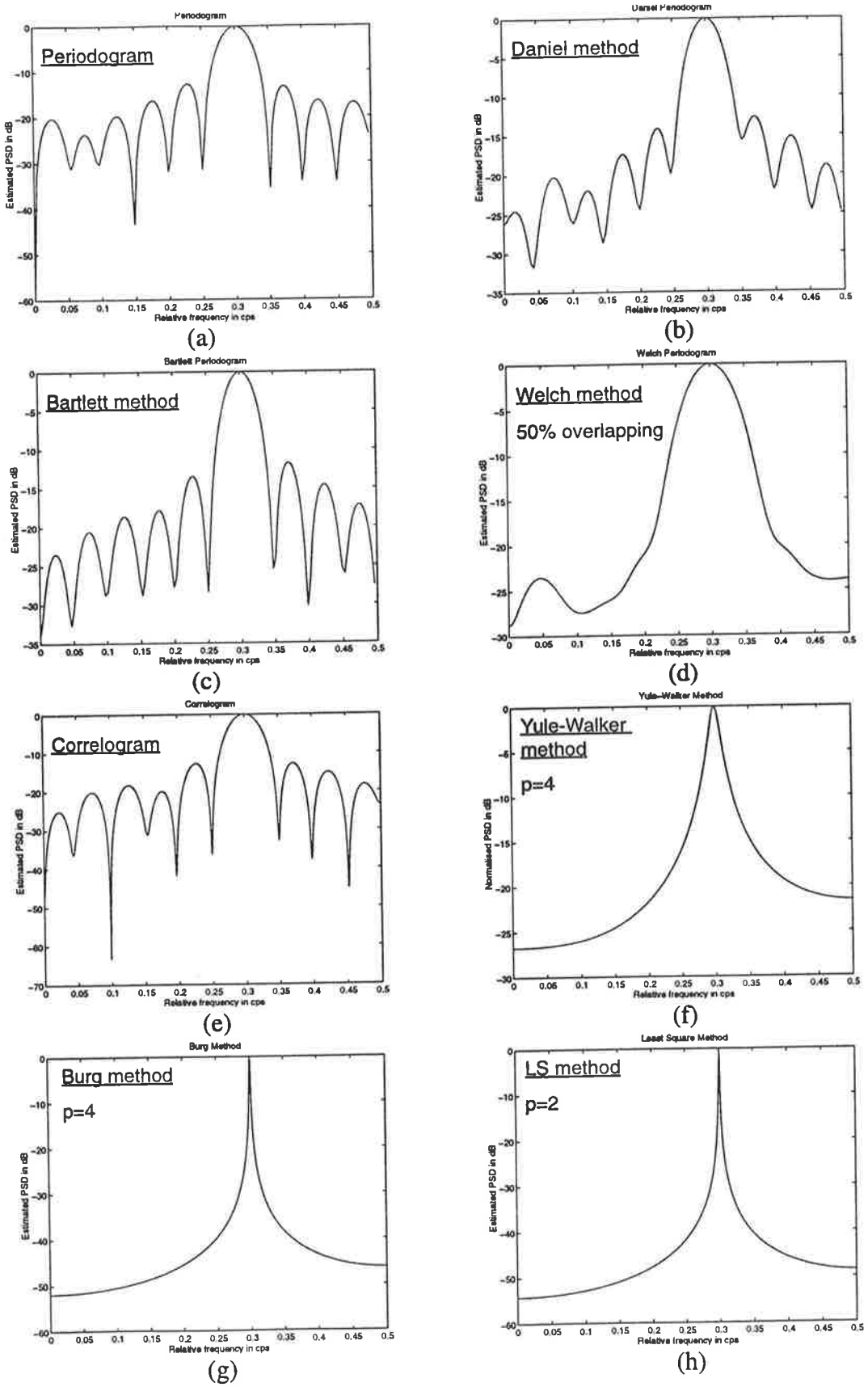

FIGURE 2.5 Conventional AR PSD estimates (f_0=0.3 cps, SNR=20dB, N=20).

CHAPTER 3 *New Parametric Method*

3.1 Introduction

With the background developed in the previous chapters, we are now in a good position to introduce a new recursive parametric method. The brief introduction of the devised method is included with some results for demonstrative purposes in the next section. The formulation of the system together with the derivation of the recursive equation are presented in Section 3.3. In section 3.4 we focus on the stability of the system and discuss preconditioning. We conclude this chapter with a treatment of the step size parameter for the recursive equation

3.2 A Recursive Parametric Method

In chapter 2, we reviewed the conventional AR PSD estimation methods and discussed briefly their advantages and disadvantages. In this section, we introduce a new AR PSD estimation method. This method is described by a simple recursive equation:

$$\boldsymbol{u}(n+1) = M \cdot \boldsymbol{u}(n) + y_1 \tag{3.1}$$

where $\boldsymbol{u}(n)$ and y_1 are vectors of length p, and M is a $p \times p$ symmetric matrix; more details about this equation will be presented in section 3.3.

This method is somewhat similar to the conventional Yule-Walker method. However instead of using the Levinson-Durbin algorithm to find the solution of the normal equations, the new method uses an iterative approach to find the AR model parameters. In general, these parameters are not a solution to the normal equations except when they converge to a steady state value (see Fig. 3.1).

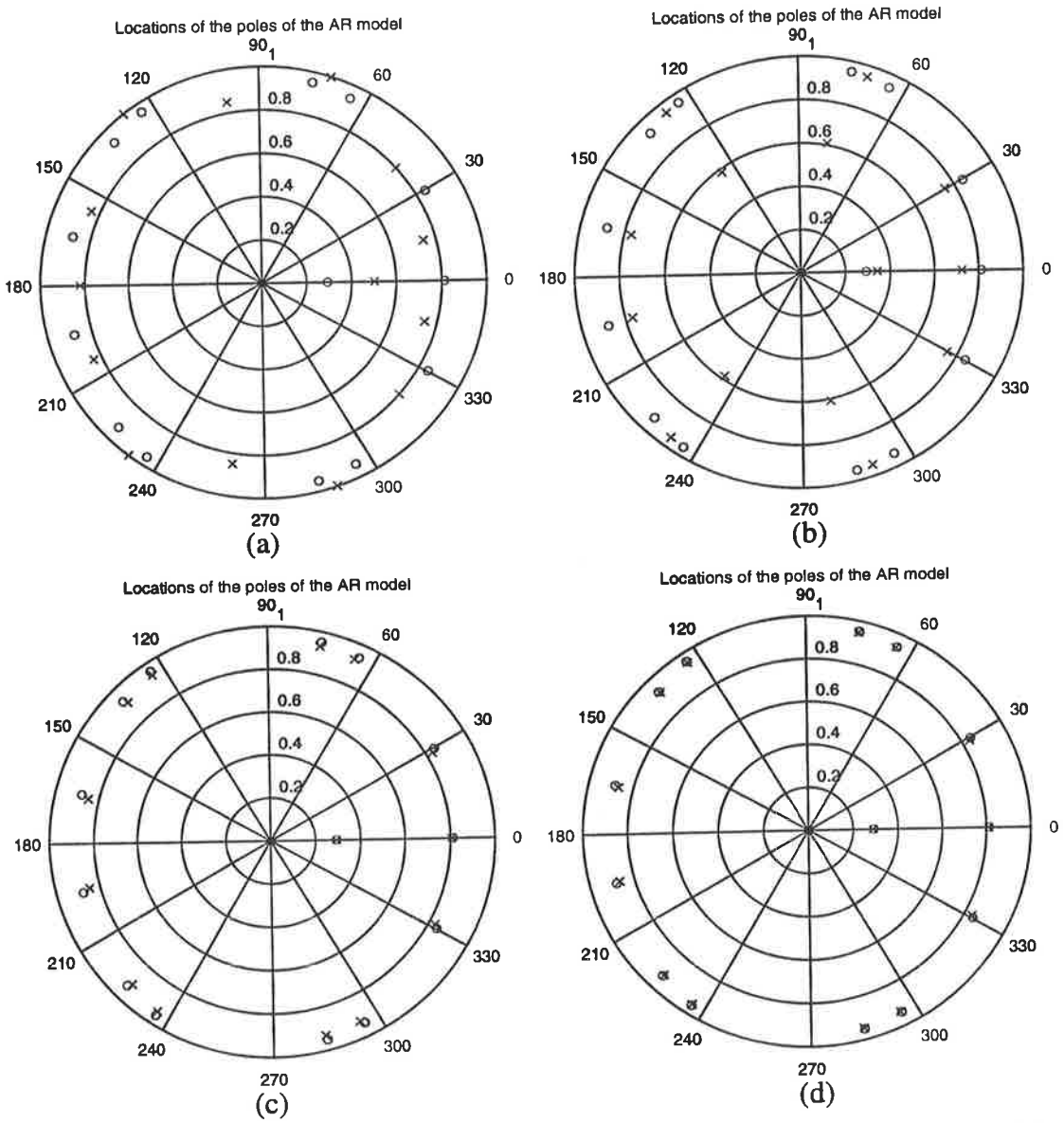

FIGURE 3.1 Comparison of poles of Yule-Walker (o) method and the recursive (x) method, N=20 points, f_0=0.2 cps & f_1=0.35 cps, p=14 & (a) 1 iteration, (b) 3 iterations, (c) 9 iterations, (d) 17 iterations.

In Fig. 3.1 (a), two poles found by the recursive method are very near the unit circle at $\theta = 71.3°$ and $\theta = 127°$. They correspond to the two spectral peaks located at

f = 0.198 cps and f = 0.353 cps. As the number of iterations increases, the poles associated with the recursive method move towards the poles associated with the Yule-Walker method. When the number of iterations reach 17, the poles of the recursive method almost coincide with those of the Yule-Walker method, as shown in Fig. 3.1 (d).

This new method yields better bandwidth resolution than the Yule-Walker method does when the model parameters are found using one iteration only. The PSD estimates of a short data record (N=20) comprising two sinusoids of equal amplitude, using the Yule-Walker method and the recursive method, are shown in Fig. 3.2. In Fig. 3.2 (a), the estimated PSD using the Yule-Walker method, represented by the dashed line, exhibits spectral line splitting due to the fact that the number of the estimated AR parameters is a large percentage of the length of data record. However, the estimated PSD using the new method, represented by the solid line, displays extremely sharp peaks located very near the true frequency positions, f_0=0.2 cps and f_1=0.35 cps, after one iteration only. This recursive method does not exhibit line splitting if the number of iterations is small. Clearly upon inspection of Fig. 3.2 (b), (c) and (d), it is apparent that the estimated AR parameters converge to the exact solution of the Yule-Walker equation as the number of iterations increases. When the estimated AR parameters reach the steady state, the two spectra look identical.

In Section 2.4, we concluded that, for short data records, the unconstrained least-squares (LS) method is the best. Since the LS method is the best, it would be appropriate to make a quick comparison between PSD estimates of the LS method and the recursive method to see if the recursive method lives up to the expectation that it will at least be comparable to conventional AR PSD estimates including the LS method.

The Marple algorithm is used here to get the exact LS solution. This algorithm selects the AR model order for itself. Each value of the AR model order is tried in an increasing order, starting from p=1, until the residual prediction error energy at p = M, (E_M) is a

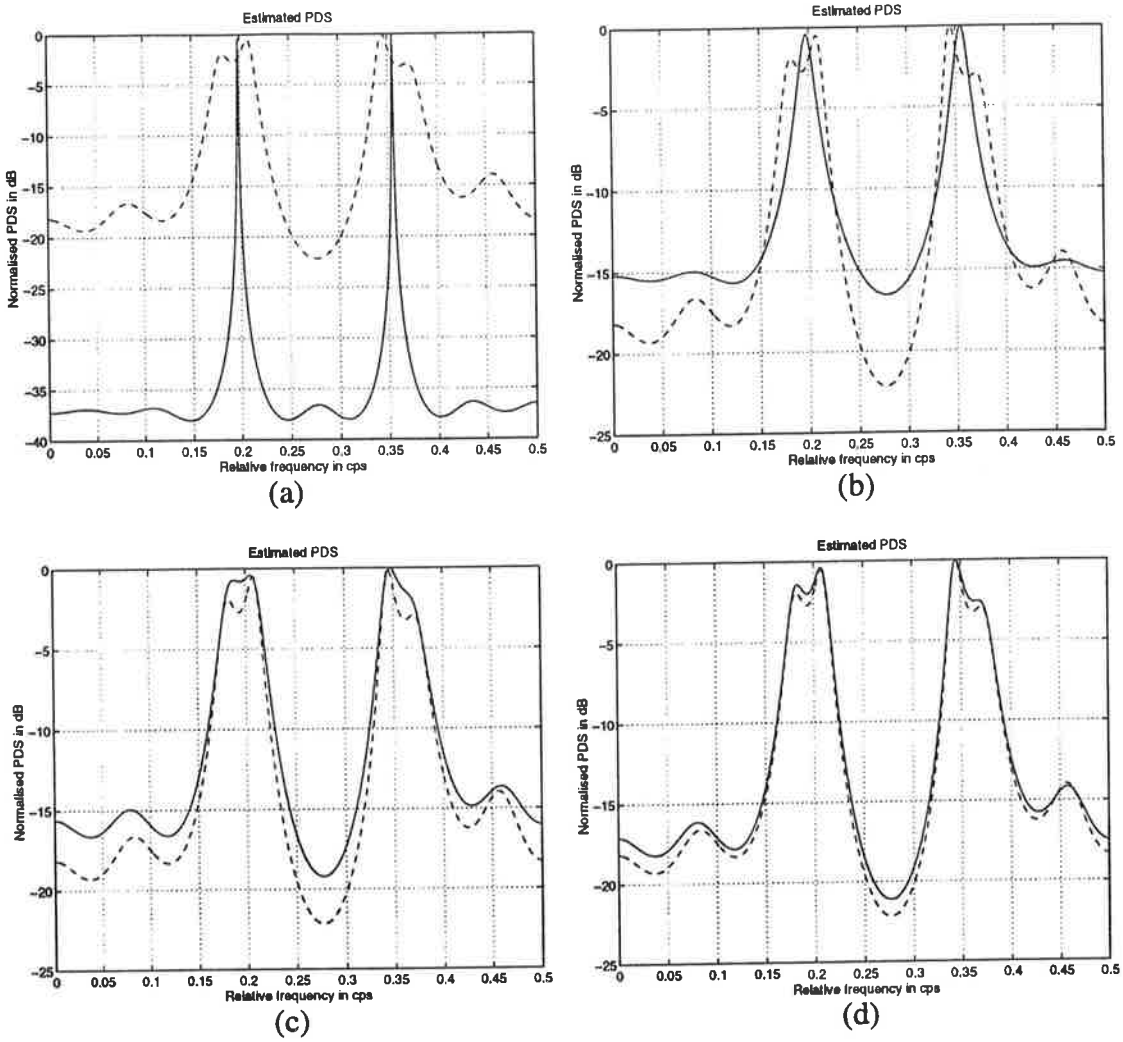

FIGURE 3.2 AR PSD estimates of the Yule-Walker method (dashed) and the recursive method (solid). N=20, f_0=0.2 cps, f_1=0.35 cps, p=14, and number of iterations =1 iteration (a), 3 iterations (b), 9 iterations (c), 17 iterations (d).

small fraction of the total signal energy (E_0), namely $\frac{E_M}{E_0} < \text{TOL1 (tolerance)}$, or it has changed by only a small fraction from the previous model order, that is

$$\frac{E_{M-1} - E_M}{E_M} < \text{TOL2 (tolerance)}$$

For this test, 3 sets of 20-point data records, having SNR of 5dB, 10dB, and 20dB, are used with two frequency values, f_0=0.3cps and f_0=0.1cps. The tolerances, TOL1 and

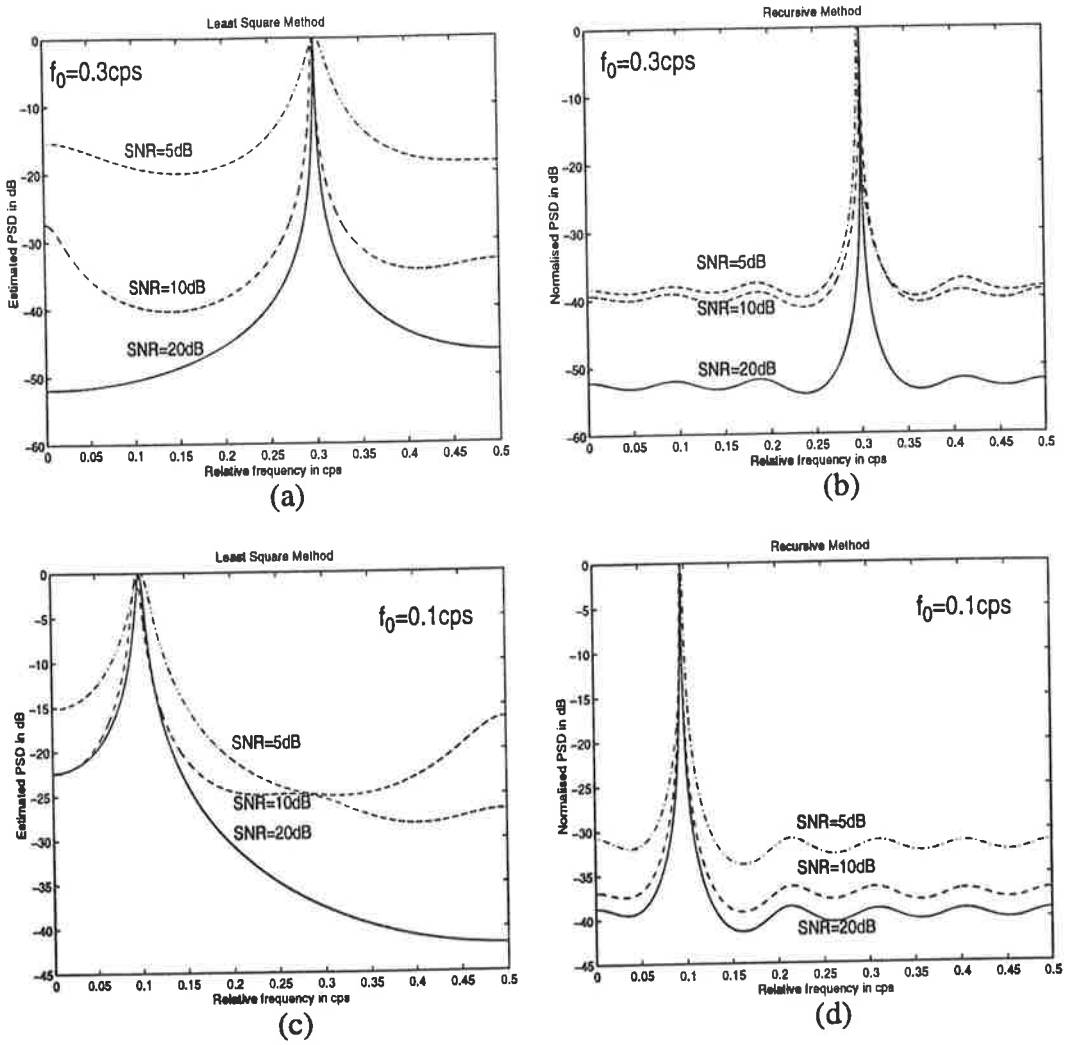

FIGURE 3.3 The LS PSD estimates, (a) and (c), and the recursive PSD estimates, (b) and (d), using N=20.

TOL2, are set to 0.1. The AR model orders used here are shown in Tables 3.1 and 3.2, together with the 3-dB bandwidth of the peaks. The LS PSD estimates with f_0=0.3cps and f_0=0.1cps are depicted in Fig. 3.3 (a) and (c), and the recursive PSD estimates with f_0=0.3cps and f_0=0.1cps are depicted in Fig. 3.3 (b) and (d). The solid, dashed, and dash-dotted lines refer to SNR values of 20dB, 10dB, and 5dB respectively.

Let us consider the case when f_0=0.3cps. Fig. 3.3 (a) and (b) show that, in the recursive method, the peaks are much sharper, and they do not get degraded as much as with the LS method as the SNR decreases. When the SNRs are10dB and 20dB, the heights of the peaks obtained from the two methods are comparable. However when the SNR is as low

as 5dB, the LS PSD estimate is much degraded; the 3-dB bandwidth is 0.195cps, and the height of the peak is about 20dB. On the other hand, the recursive PSD estimate retains its superiority; the 3-dB bandwidth is 0.0008cps (approximately 24 times narrower), and the height of the peak is about 40dB.

Usually the quality of the AR PSD estimates is poorer at low frequencies. So a low frequency value, 0.1 cps, is selected for the PSD estimate comparison at a low frequency; the results are depicted in Fig. 3.3 (c) and (d). Again the recursive PSD estimate holds its superiority over the LS PSD estimate. The recursive PSD estimates exhibit peaks of much larger heights (almost twice in dB scale) and much smaller 3-dB bandwidth. When SNR=5dB, the 3-dB bandwidth of the recursive PSD estimate is approximately 42 times smaller than that of the LS PSD estimate.

The detailed performance of the recursive method will be covered in Chapter 4. Since the LS method is computationally complex, and the Yule-Walker method is more related to the recursive method, the Yule-Walker method is used for a rigorous performance comparison in Chapter 4.

TABLE 3.1 Best 3-dB bandwidths of the LS and recursive estimates at f_0=0.3 cps.

Method	SNR=5dB	SNR=10dB	SNR=20dB
LS	0.195cps p=5	0.0023cps p=5	0.0011cps p=2
Recursive	0.008cps p=10	0.0007cps p=10	0.0002cps p=10

TABLE 3.2 Best 3-dB bandwidths of the LS and recursive estimates at f_0=0.1 cps.

Method	SNR=5dB	SNR=10dB	SNR=20dB
LS	0.098cps p=4	0.009cps p=5	0.0079cps p=5
Recursive	0.0023cps p=10	0.0018cps p=10	0.0008cps p=10

3.3 Derivation of the Recursive Method

In AR-based PSD estimation methods, the model parameters are obtained by solving a linear prediction problem. The coefficients of a one-step forward linear predictor are obtained by solving a system of linear equations known as the *normal equations* [52], [69], [76]. The solution of the normal equations can be found by direct matrix inversion or using the Levinson-Durbin recursion [18], [21], [60], [105]. A less common approach to solving the normal equations is an iterative approach. In this approach an initial guess of the solution is made, and then the solution is updated iteratively until the steady state is reached; the steady state solution forms the solution of the normal equations. In this section the iterative approach is exploited to derive the recursive equation (3.1). The goal here is not to obtain a solution of the normal equations but to obtain a set of coefficients which are used for PSD estimation. The normal equations are derived next. Then Eqn. (3.1) is derived in Subsection 3.3.2.

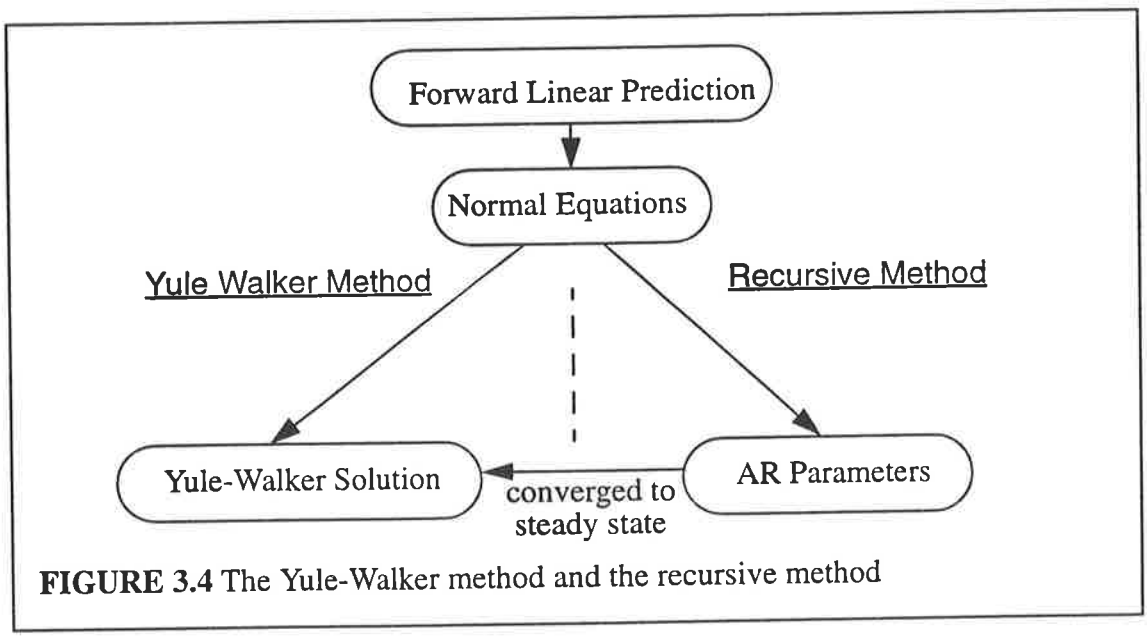

FIGURE 3.4 The Yule-Walker method and the recursive method

3.3.1 Linear Prediction

Since the recursive method originates from the linear prediction, we begin with the problem of predicting a future value of a zero-mean stationary complex process from observation of past values of the process. In particular, we consider the one-step forward linear predictor which forms the prediction of the value $x(n)$ by a weighted linear combination of the past values $x(n\text{-}1)$, $x(n\text{-}2)$,..., $x(n\text{-}p)$. Hence the linearly predicted value of $x(n)$ is

$$\hat{x}(n) = -\sum_{k=1}^{p} a_p(k)\, x(n-k) \tag{3.2}$$

where $\{-a_p(k)\}$ represent the weights in the linear combination, and are called prediction coefficients of the one step forward linear predictor of order p. The difference between the value $x(n)$ and the predicted value $\hat{x}(n)$ is called the *forward prediction error*, denoted as $f_p(n)$:

$$f_p(n) = x(n) - \hat{x}(n) = \sum_{k=0}^{p} a_p(k)\, x(n-k) \tag{3.3}$$

where $a_0 = 1$.

We view linear prediction [64], [66] as being equivalent to linear filtering where the predictor is embedded in the linear filter as shown in Fig. 3.5. In Fig. 3.5, the system consists of two basic parts: a transversal filter with adjustable tap weights, called prediction coefficients, whose values are denoted as $a_p(1)$,..., $a_p(p)$, and a mechanism for adjusting these tap weights in an adaptive manner.

We may express the estimation error $f_p(n)$ in Eqn. (3.3) using the vector notation as

$$f_p(n) = x(n) + \boldsymbol{a}^T \cdot \boldsymbol{x} \tag{3.4}$$

where the $\boldsymbol{a}$ and $\boldsymbol{x}$ are vectors of length p, and the superscript T signifies *transposition*.

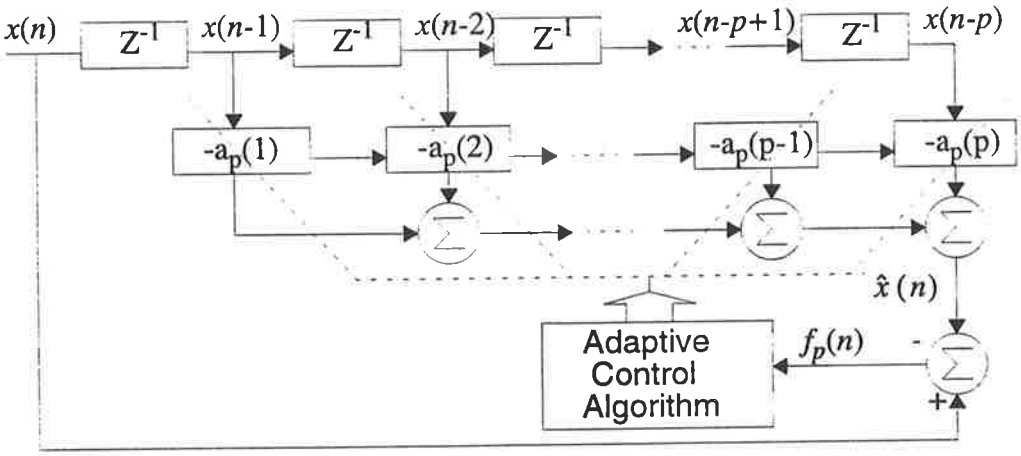

FIGURE 3.5 Adaptive filter and linear prediction

$$\boldsymbol{a}^T = [a_p(1), a_p(2), \ldots, a(p)] \tag{3.5}$$

$$\boldsymbol{x}^T = [x(n-1), x(n-2), \ldots, x(n-p)] \tag{3.6}$$

The mean square value of the forward linear prediction error $f_p(n)$ may be achieved by evaluating the expectation of the squared forward linear prediction error:

$$\xi_p^f = E\left[\left|f_p(n)\right|^2\right] \tag{3.7}$$

$$= E\left[\left|x(n) + \boldsymbol{a}^T \cdot \boldsymbol{x}\right|^2\right]$$

$$= E\left[\left(x(n) + \boldsymbol{a}^T \cdot \boldsymbol{x}\right) \cdot \left(x^*(n) + \boldsymbol{x}^H \cdot \boldsymbol{a}^*\right)\right]$$

$$= E\left[x(n) \cdot x^*(n)\right] + \boldsymbol{a}^* \cdot E\left[x(n) \cdot \boldsymbol{x}^H\right] + \boldsymbol{a}^T \cdot E\left[x^*(n) \cdot \boldsymbol{x}\right] + \boldsymbol{a}^T \cdot E\left[\boldsymbol{x} \cdot \boldsymbol{x}^H\right] \cdot \boldsymbol{a}^*$$

$$\xi_p^f = \gamma_{xx}(0) - \boldsymbol{a}^* \cdot \boldsymbol{r} - \boldsymbol{a}^T \cdot \boldsymbol{r}^H + \boldsymbol{a}^T \Gamma \boldsymbol{a}^* \tag{3.8}$$

where

$$\boldsymbol{r} = -E[x(n) \cdot \boldsymbol{x}^H] = -\begin{bmatrix} \gamma_{xx}(1) \\ \gamma_{xx}(2) \\ \vdots \\ \gamma_{xx}(p) \end{bmatrix}$$

$$\text{and } \Gamma = E[\boldsymbol{x} \cdot \boldsymbol{x}^H] = \begin{bmatrix} \gamma_{xx}(0) & \gamma_{xx}(1) & \dots & \gamma_{xx}(p-1) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \dots & \gamma_{xx}(p-2) \\ \vdots & \dots & \dots & \vdots \\ \gamma_{xx}(p-1) & \dots & \gamma_{xx}(1) & \gamma_{xx}(0) \end{bmatrix} \tag{3.9}$$

and the superscript H signifies *Hermitian transposition*. Here we use estimates $\boldsymbol{y}$ and R for $\boldsymbol{r}$ and Γ, respectively:

$$\boldsymbol{y} = -\begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \text{ and } R = \begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(2) & \dots & R(p-2) \\ \vdots & \dots & \dots & \vdots \\ R(p-1) & \dots & R(1) & R(0) \end{bmatrix} \tag{3.10}$$

where $R(m)$ is the autocorrelation estimate (see Eqn. (2.25)).

For the case of real valued data, the autocorrelation function r_{xx}(k) is also real for all k hence the autocorrelation matrix, R, is symmetric and toeplitz [37].

This simplifies Eqn. (3.8):

$$\hat{\xi}_p^f = \sigma_x^2 - (2a^T \cdot y) + a^T R a \tag{3.11}$$

where σ_x^2 is a variance of the desired response, and R is a $p \times p$ positive semidefinite symmetric matrix. Thus we may visualize the dependence of the mean-square error, ζ_p^f or $J(a)$, on a as a bowl-shaped surface with a unique minimum if R is positive definite. The global minimum of $J(a)$ clearly occurs at $a = R^{-1} \cdot y$ which may be calculated directly by inverting R. However, the inverse process becomes computationally intensive for large p and numerically unstable for ill-posed problems [11].

3.3.2 Recursive Method for PSD Estimation

An alternative to finding the minimum of the prediction error, ξ_p^f in Eqn. (3.11) is to use the gradient approach. According to the *method of steepest descent* (Murray [72]), the update value of the tap weight vector at time $n+1$ is computed using the simple recursive relation [37]

$$a(n+1) = a(n) + \frac{1}{2} \cdot h(-\nabla(n)) \tag{3.12}$$

where h is a positive value referred to as *step size parameter*, and ∇ is the gradient with respect to a of Eqn. (3.11). Rearranging Eqn. (3.12) gives

$$-\nabla(n) = \frac{2[a(n+1) - a(n)]}{h} \tag{3.13}$$

Let $a = B \cdot u$ where B is a preconditioner $p \times p$ symmetric matrix, then Eqn. (3.11) becomes

$$J(u) = \sigma_x^2 - 2u^T \cdot B^T \cdot y + u^T \cdot B^T \cdot R \cdot B \cdot u \tag{3.14}$$

and Eqn. (3.13) becomes

$$-\nabla(n) = 2B \cdot \frac{[\boldsymbol{u}(n+1) - \boldsymbol{u}(n)]}{h} \tag{3.15}$$

Differentiating $J(\boldsymbol{u})$ in Eqn. (3.14) with respect to $\boldsymbol{u}$ yields the value for the gradient vector.

$$\nabla = \frac{\partial}{\partial \boldsymbol{u}} J(\boldsymbol{u}) = -2B^T \cdot y + 2B^T \cdot (R \cdot B) \cdot \boldsymbol{u} = -2B^T \cdot [y - (R \cdot B) \cdot \boldsymbol{u}] \tag{3.16}$$

Since B is symmetric, Eqn. (3.16) can be written as

$$-\nabla = 2B \cdot [y - (R \cdot B) \cdot \boldsymbol{u}] \tag{3.17}$$

Comparing Eqn. (3.15) and Eqn. (3.16), we get the following difference equation.

$$\frac{\boldsymbol{u}(n+1) - \boldsymbol{u}(n)}{h} = y - (R \cdot B) \cdot \boldsymbol{u} \tag{3.18}$$

Let us decompose the matrix product $R \cdot B$ into diagonal and off-diagonal matrices denoted A and C respectively.

$$A = diag(R \cdot B) \tag{3.19}$$

$$C = R \cdot B - A. \tag{3.20}$$

Substituting for $R \cdot B$ by A and C into Eqn. (3.18), and estimating $\boldsymbol{u}$ by letting $A\boldsymbol{u} = A\boldsymbol{u}(n+1)$ and $C\boldsymbol{u} = C\boldsymbol{u}(n)$ (reasons for doing this come later in Section 3.4), we get

$$\frac{\boldsymbol{u}(n+1) - \boldsymbol{u}(n)}{h} = y - A \cdot \boldsymbol{u}(n+1) - C \cdot \boldsymbol{u}(n) \tag{3.21}$$

$$\boldsymbol{u}(n+1) - \boldsymbol{u}(n) = hy - hA \cdot \boldsymbol{u}(n+1) - hC \cdot \boldsymbol{u}(n)$$

$$(I + hA) \cdot \boldsymbol{u}(n+1) = (I - hC) \cdot \boldsymbol{u}(n) + hy$$

$$\boldsymbol{u}(n+1) = (I + hA)^{-1} \cdot [(I - hC) \cdot \boldsymbol{u}(n) + hy]$$

and since A is a diagonal matrix in the form of $A = kI$ (see Subsection 3.4),

$$\boldsymbol{u}(n+1) = \frac{(I - hC)}{1 + hk}\boldsymbol{u}(n) + \frac{h}{1 + hk}y \tag{3.22}$$

By rewriting Eqn. (3.22) in simplified form, we get the recursive equation (3.1).

$$\boldsymbol{u}(n+1) = M \cdot \boldsymbol{u}(n) + \boldsymbol{y}_1 \tag{3.23}$$

where

$$M = \frac{I - hC}{1 + hk} \tag{3.24}$$

and

$$\boldsymbol{y}_1 = \frac{hy}{1 + hk} \tag{3.25}$$

The purpose of decomposing the product $R \cdot B$ into matrices A and C and decomposing the state vector $\boldsymbol{u}$ into $\boldsymbol{u}(n+1)$ and $\boldsymbol{u}(n)$ is to improve the stability of the system by means of increasing the tolerable range of the step size parameter. Analytic verification of this is shown in Section 3.4.

3.4 Preconditioner Matrix

In this section, we address the issue of an appropriate choice of the preconditioner B [11]. To describe the condition of an autocorrelation matrix R quantitatively, we define

the condition number of R as the ratio of the maximum singular value to the minimum singular value of R [37], [86].

A matrix which has a condition number near unity is said to be well-conditioned whereas a large condition number indicates that the matrix is ill-conditioned which in turn implies that the system described by the matrix is susceptible to errors in the system matrix and round off errors caused by finite-word length in digital simulations.

The condition number of a matrix can be improved by multiplying the matrix by another matrix B called the preconditioner. One common choice for B [11] is

$$\beta_{ii} = \frac{k}{r_{ii}}, \; k > 0 \tag{3.26}$$

where β_{ii} and r_{ii} (or $R(0)$) are the diagonal elements of B and R respectively. This is a reasonable approximation to a scalar multiple of the inverse of R. However this diagonal preconditioning does not improve the condition number of the autocorrelation matrix. Because the diagonal elements in the autocorrelation matrix R are identical, the preconditioner will simply be $\frac{k}{R(0)} \cdot I$, which will not have any effect on the singular values of $R \cdot B$.

This preconditioner, nevertheless, controls the upper limits of the eigenvalues of $R \cdot B$. The eigenvalues of $R \cdot B$ determine the maximum permissible step size. Therefore the preconditioner, or more precisely, the value of k is the key factor determining the range of the step size which does not violate the stability or convergence of the recursive algorithm. To illustrate this, let us consider the recursive equation (3.23)

$$\boldsymbol{u}(n+1) = M \cdot \boldsymbol{u}(n) + \boldsymbol{y}_1 \qquad \text{where } M = \frac{I - h \cdot C}{1 + hk}$$

For all modes of the recursive algorithm to converge, all the eigenvalues of M must lie inside the unit circle [76].

$$|eig(M)| < 1$$

$$\left|eig\left(\frac{I - h \cdot C}{1 + hk}\right)\right| < 1$$

$$\frac{1}{1 + hk} \cdot |eig(I - h \cdot C)| < 1$$

$$|eig(I - h \cdot C)| < 1 + hk$$

$$|1 - h \cdot eig(C)| < 1 + hk \tag{3.27}$$

If $1 - h \cdot eig(C) < 0$,

$$-1 + h \cdot eig(C) < 1 + hk$$

$$h(eig(C) - k) < 2$$

$$h < \frac{2}{eig(C) - k} \tag{3.28}$$

If $1 - h \cdot eig(C) \geq 0$,

$$1 - h \cdot eig(C) \leq 1 + hk$$

$$0 \leq h(k + eig(C))$$

$$0 \leq h \cdot eig(R \cdot B)$$

and since eigenvalues of $R \cdot B$ are real and positive,

$$0 \leq h \tag{3.29}$$

Since $C = R \cdot B - A$, and $A = k \cdot I$,

$$eig(C) = eig(RB) - k \tag{3.30}$$

$$= \frac{k}{R(0)} \cdot eig(R) - k \qquad (\text{since } RB = \frac{k}{R(0)} \cdot R) \tag{3.31}$$

Therefore the necessary condition for the convergence or stability of the recursive algorithm is that the step size h satisfies the following condition.

$$0 < h < \frac{2}{eig_{max}(RB) - 2k} \tag{3.32}$$

or equivalently $$0 < h < \frac{2R(0)}{k(eig_{max}(R) - 2R(0))} \tag{3.33}$$

From Eqn. (3.33), the upper limit of the step size parameter, denoted by h_{max}, is inversely proportional to k. In other words, the permissible range of h now can be manipulated by the external user by deciding on the value of k, and this provides the system some degree of flexibility.

Now let us consider the maximum allowable step size for the recursive equation letting $\boldsymbol{u} = \boldsymbol{u}(n)$.

$$\boldsymbol{u}(n+1) - \boldsymbol{u}(n) = h\boldsymbol{y} - hR \cdot B \cdot \boldsymbol{u}(n) \tag{3.34}$$

$$\boldsymbol{u}(n+1) = (I - hR \cdot B)\boldsymbol{u}(n) + h\boldsymbol{y}$$

$$= M' \cdot \boldsymbol{u}(n) + \boldsymbol{y}' \tag{3.35}$$

where

$$M' = I - hR \cdot B$$

$$\boldsymbol{y}' = h\boldsymbol{y}$$

Now the convergence of the system described by this recursive equation is guaranteed if all eigenvalues of the system matrix M' lie inside the unit circle.

$$|eig(M')| < 1$$

$$|eig(I - hR \cdot B)| < 1$$

$$|1 - h \cdot eig(R \cdot B)| < 1 \tag{3.36}$$

If $1 - h \cdot eig(R \cdot B)$ of Eqn. (3.36) is less than zero, then

$$-1 + h \cdot eig(R \cdot B) < 1$$

$$h \cdot eig(R \cdot B) < 2$$

$$h < \frac{2}{eig(R \cdot B)} \tag{3.37}$$

If $1 - h \cdot eig(R \cdot B)$ of Eqn. (3.36) is greater than or equal to zero, then

$$1 - h \cdot eig(R \cdot B) \leq 1$$

$$h \cdot eig(R \cdot B) \geq 0$$

$$h \geq 0 \tag{3.38}$$

Putting together Eqn. (3.37) and Eqn. (3.38), we get

$$0 < h < \frac{2}{eig(R \cdot B)} \tag{3.39}$$

or equivalently $$0 < h < \frac{2R(0)}{k \cdot eig(R)} \tag{3.40}$$

By comparing Eqns. (3.32) and (3.39), we can clearly see that the former is more stable than the latter; i.e., the range of h for which (3.32) is stable is larger than that for which (3.39) is stable.

3.5 Step Size Parameter

In this section we introduce an expression for the scalar constant k, and discuss the consequences of using this particular expression. Later on we introduce a new term called the step multiplication factor.

First, the scalar quantity k in Eqn. (3.32) is chosen to be $\frac{1}{p}$. By doing this, the variable term k and the other variable term p, the AR model order, are now related reducing the number of user- variable terms. This also ensures that the smallest h range is $0 < h < 2$ so that choosing the h value less than 2 guarantees the convergence of the system in all cases. This can be explained by using *Gershgorin's theorem* [37].

Let $Q = RB$, and let q_{ij} be the elements of Q, where $i, j = 1, 2, \ldots, p$ and $\lambda_1, \lambda_2, \ldots, \lambda_p$ be its eigenvalues. Denote by D_i the i_{th} *Gershgorin disk* with centre q_{ii} and radius r_i.

$$r_i = \sum_{j=1}^{p} |q_{ij}|, \; i = 1, 2, \ldots, p \quad \& \; j \neq i \tag{3.41}$$

Denote by Ω the union of these p disks, then all the eigenvalues of Q must lie inside Ω. Since Q has real and positive eigenvalues, the eigenvalues will lie on the strip of real axis as depicted in Fig. 3.6. The diagonal elements of Q are all equal, i.e., q_{ii} =

$\frac{k}{R(0)} \times R(0) = k$. Since k is selected to be $\frac{1}{p}$, the diagonal elements q_{ii} are equal to $\frac{1}{p}$ for all i values ranging from 1 to p.

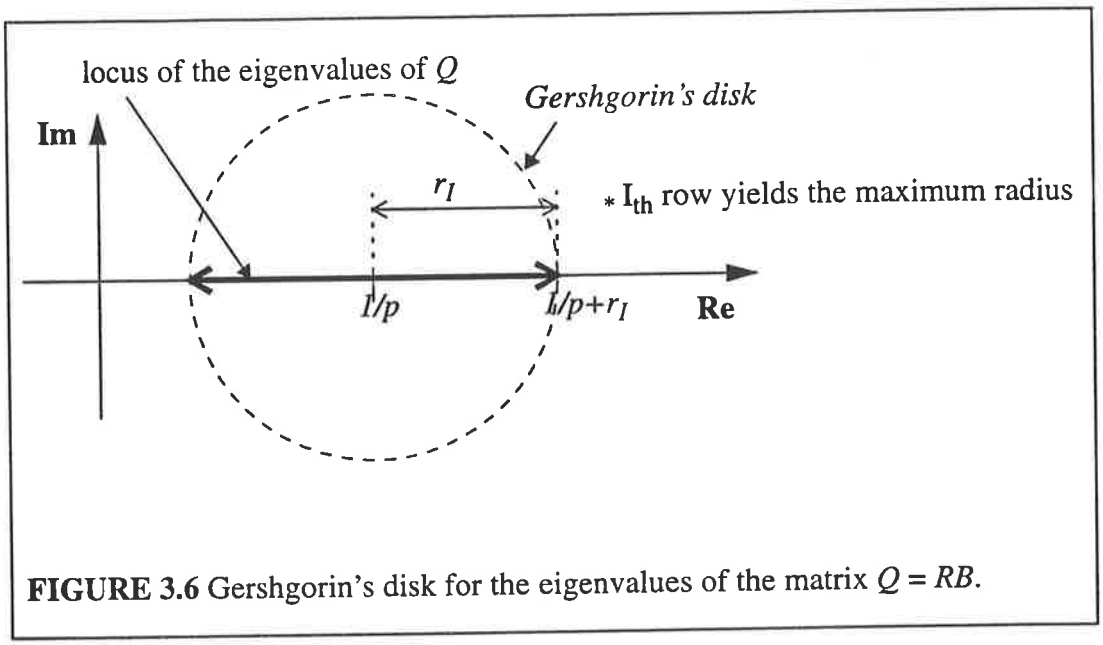

FIGURE 3.6 Gershgorin's disk for the eigenvalues of the matrix $Q = RB$.

In fact, the matrix Q is equal to $\frac{k}{R(0)} \cdot R$ since B is equal to $\frac{k}{R(0)} \cdot I$. Then, q_{ij} will be equal to $\frac{k}{R(0)} \cdot r_{ij}$ where r_{ij} refers to ij^{th} element of the autocorrelation matrix R, and $R(0)$ is the diagonal elements r_{ii} or equivalently $r_{xx}(0)$.

Let $r_I = \max_i (r_i)$.

$$r_I = \sum_{j=1}^{p} |q_{Ij}|$$

$$= \sum_{j=1}^{p} \frac{k}{R(0)} |r_{Ij}|$$

$$= \frac{1}{p}\sum_{j=1}^{p}\frac{|r_{Ij}|}{R(0)}, \; j \neq I \tag{3.42}$$

We know that the autocorrelation sequence is symmetrical about its maximum value in the middle which has zero lag. (i.e $r_{xx}(0) \geq r_{xx}(j)$, $j \neq 0$)

This indicates that $R(0)$ will always be greater than the off-diagonal elements r_{Ij} in the I^{th} row.

So $\frac{|r_{Ij}|}{R(0)} < 1$

and this lead to

$\sum_{j=1}^{p}\frac{|r_{Ij}|}{R(0)} < \sum_{j=1}^{p} 1$, $j \neq I$, therefore the summation applies to p-1 elements.

$$\sum_{j=1}^{p}\frac{|r_{Ij}|}{R(0)} < p-1 \text{ , since } \sum_{j=1}^{p} 1 = p-1 \tag{3.43}$$

therefore

$$r_I = \frac{1}{p}\sum_{j=1}^{p}\frac{|r_{Ij}|}{R(0)} < \frac{1}{p}\cdot(p-1) \tag{3.44}$$

From Fig. 3.6

$$\lambda_{max} < \frac{1}{p} + r_I \tag{3.45}$$

Substituting Eqn. (3.44) into Eqn. (3.45) we get

$$\lambda_{max} < \frac{1}{p} + \frac{p-1}{p}$$

$$\lambda_{max} < 1 \tag{3.46}$$

$$\therefore eig_{max}(RB) < 1 \tag{3.47}$$

Now referring back to Eqn. (3.32), the above result indicates that the lowest possible value for h_{max} ($h_{max} = \frac{2}{eig_{max}(RB) - 2k}$) occurs when $eig_{max}(RB)$ is nearest to 1 and the corresponding value of h_{max} is $\frac{2}{1-2k} \approx 2$.

In order to use the recursive algorithm for PSD estimation, a suitable value for the step size h must be decided by the user. Recalling the condition for convergence of the system, i.e., $0 < h < \frac{2}{eig_{max}(RB) - 2k}$, if we set $k = \frac{1}{p}$,

$$h_{max} = \frac{2}{1 - \frac{2}{p}} > 2 \,,\; p > 2 \tag{3.48}$$

We simply decide the step size parameter by multiplying h_{max} by a real number ranging between 0 and 1 so that h is inside the convergence range. We name this multiplier the *Step Multiplication Factor* (abbreviated as *smf*).

$$h = smf \times h_{max} \tag{3.49}$$

The value of the term *smf* must be decided by the user together with the model order p.

CHAPTER 4 *Performance Analysis*

4.1 Introduction

This chapter analyses the performance of the devised recursive method by measuring the spectral bandwidth at the 3-dB points and the frequency bias of the peak of the PSD estimate. We adopt the term *relative frequency* in cycles per sample (cps) as a measure of frequency f, which is defined as

$$f = \frac{F}{F_s} \tag{4.1}$$

where F is the frequency of the analogue signal and F_s is the sampling frequency, both in hertz.

The signal to noise ratio (SNR) is defined as

$$SNR = 10\log_{10}\left(\frac{A^2}{2\sigma^2}\right) \tag{4.2}$$

where σ^2 is variance of the additive noise, and A is the amplitude of the sinusoid. The objective is to compare the performances of two PSD estimation methods, namely the recursive method and the Yule-Walker method, on the basis of spectral bandwidth of the main lobe at the 3-dB points, frequency bias, and robustness in the presence of additive noise.

We define the 3-dB bandwidth as the spectral width of a peak at -3-dB points, and the frequency bias as the absolute value of the bias in the estimated frequency, i.e., $|f_0 - f_{\text{estimated}}|$.

The performance analysis using a single sinusoid is presented in Section 4.2. The analysis includes various tests: effect of varying SNR, low frequency test, step multiplication factor selection test, variation of 3-dB bandwidth with frequency, and effect of introducing a non-zero initial phase in the data. In Section 4.3, we consider the PSD estimates of two sinusoids in additive noise. Tests will concentrate on the effects of the model order and SNR on the quality of the PSD estimates, and on peak-resolving ability of the recursive method in comparison with the Yule-Walker method. The results of different tests are summarised in Section 4.4.

4.2 Single Sinusoid Test

In this section, we present some experimental results on the performance of the recursive method and the Yule-Walker method, using a single sinusoid corrupted with additive gaussian noise. As mentioned earlier, the recursive method with one iteration yields very narrow spectral bandwidth. Therefore, all experiments in this section are performed with one iteration only, except for the low frequency test in Subsection 4.2.2. In the next subsection, we look into the effect of varying SNR on the spectral bandwidth and the frequency bias. Then we analyse the performance of the two methods at low frequencies in Subsection 4.2.2. In Subsection 4.2.3, we investigate the selection of step multiplication factor (smf) and the sensitivity of the PSD estimate to smf. In Subsection 4.2.4, we study the variation of the spectral bandwidth as a function of frequency. Finally, in Subsection 4.2.5, we investigate the effect of initial phase of a sinusoid on the PSD estimate.

4.2.1 Effect of SNR

Here we observe the effect of additive noise on the quality of the PSD estimate of the Yule-Walker method and the recursive method.

The data consists of a single sinusoid with additive Gaussian noise. The sinusoid has a normalised or relative frequency of 0.3 cps and an amplitude of 5 volts peak. Before

starting the detailed analysis, let us make a quick comparison between the two methods by plotting the estimated spectra at two different SNR levels, one high and one low. The purpose of this comparison is to gain an insight into the effect of noise on spectral bandwidth and frequency bias.

Figure 4.1 presents the estimated spectra of a sinusoid of length N=20 for SNR values of 5dB and 25dB, respectively. The plotted spectra were estimated assuming AR model order of p=4, p=8, and p=16, respectively. The spectra obtained using the Yule-Walker method are represented by dashed lines, and the spectra obtained from the recursive method are represented by solid lines.

In Fig. 4.1 (a) and (b), the model order is held constant at p=4 and the SNR is increased from 5dB to 25dB. We can clearly see that when p=4 the recursive method yields much narrower spectral bandwidths at the expense of an increased frequency bias. In the Yule-Walker method, the increase of SNR seems to improve the spectral bandwidth and the frequency bias seems to remain very small, and independent of SNR variations. On the other hand, in the recursive method, the spectral resolution seems to improve slightly, but the improvement is not significant enough to convince us that it is due to the increase in SNR and not due to variations in the spectrum.

In Fig. 4.1 (c) and (d), the model order is increased to p=8. With the Yule-Walker method, the spectra do not show any perceptible differences from the earlier ones except for the slight spurious fluctuations. Again an increase in SNR improves the PSD estimate. However, with the recursive method, both the spectral resolution and the frequency bias are significantly improved. Interestingly, with SNR=25 dB the spectrum is somewhat poorer than that with SNR=5 dB. This could indicate that p=8 is not the right model order for SNR=25dB, or the smf is far from the optimum value.

Further increase of the model order to p=16 causes a split of the peak and introduces noticeable spurious peaks in the spectra of the Yule-Walker method as depicted in Fig. 4.1 (e) and (f). With SNR=25 dB, the spurious peaks in the spectrum are somewhat diminished but the improvement is not dramatic enough to completely get rid of line splitting.

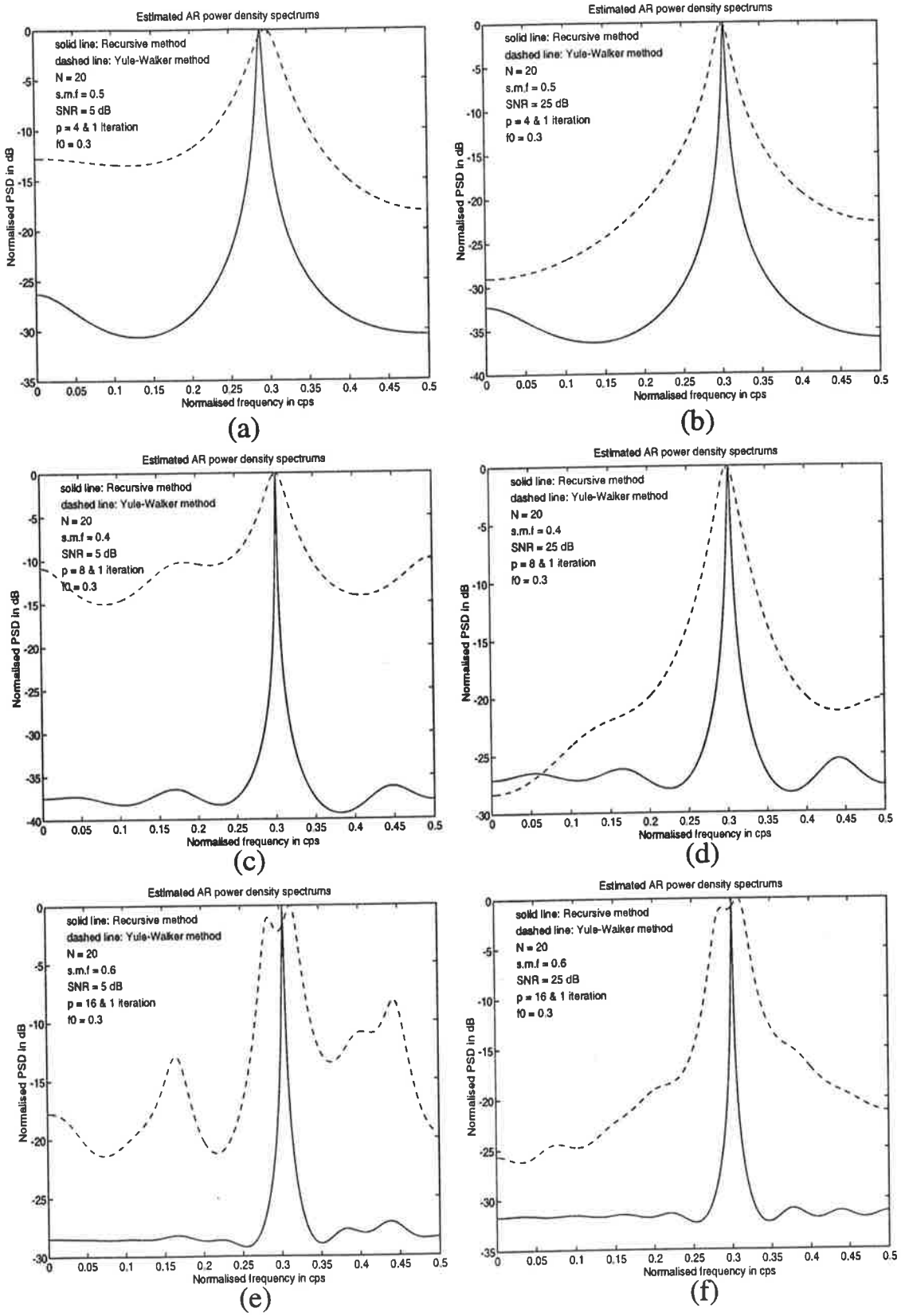

FIGURE 4.1 Spectrum estimates of a 20-point random process consisting of a single sinusoid of frequency 0.3 cps with additive white noise; The solid line is the estimate of the recursive method and the dashed line is the estimate of the Yule-Walker method.

The spectra yielded by the recursive method are again very good, having a very sharp peak with very small frequency bias and much suppressed spurious peaks. The SNR seems not to influence the PSD of the recursive method.

It can be concluded that the recursive method yields better spectra than the Yule-Walker method does, and the increase of SNR seems to improve the spectra of the Yule-Walker method but not necessarily those of the recursive method.

To verify the above conclusions, more rigorous tests are conducted to analyse the behaviour of the spectra as the SNR varies. Fig. 4.2 illustrates the results (spectral bandwidth and frequency bias) obtained using a data record of length N=100 which is assumed to be an AR process with model order p=4, 8, and16. The SNR ranges from 1dB to 30dB in regular intervals of 1 dB.

To obtain more statistically reliable results for each value of SNR, the method was tested 6 times and the results (i.e. bandwidth resolution and frequency bias) were averaged. Each SNR value is tolerable up to ±5 dB. For example, once the noise level is set to give a SNR of 20 dB and if the actual SNR measured from the data lies outside the range $19.5 \leq SNR < 20.5$ due to the finite length of data, then the data is discarded and another set is generated. This process continues until the measured SNR is in the tolerable range ($19.5 \leq SNR < 20.5$).

In Fig. 4.2 (a) and (b), it can be seen that the Yule-Walker method exhibits poor spectral bandwidth and a slight frequency bias when the SNR is small, but as the SNR increases a significant improvement occurs. At large SNRs (SNR ≥ 20dB), the frequency bias becomes so small that it is barely perceptible, and the spectral bandwidth reduces to about 0.003 cps; this is true for all three model orders. It can also be observed that the greater the model order is the better is the quality of the spectrum. For p=16, the spectral bandwidth and the frequency bias are the smallest and are more or less constant throughout the SNR range. The results obtained with p=4 are the worst.

In Fig. 4.2 (c) and (d), it is observed that the recursive method after only one iteration yields better spectrum than the Yule-Walker method does. After trying different smf values (see Section 3.5 for a definition of smf), the step sizes yielding the best results were

selected, and the corresponding results were plotted. When p=2, the results were far too large compared to those of higher model orders, and did not fit in the graphs shown in Fig. 4.2. It is well known that p=2 is the smallest possible value for the model order to represent a peak in frequency domain corresponding to a sinusoid in the time domain. It is found that the recursive method requires model orders at least greater than two to get an acceptable PSD estimate, except in the steady state. With p=2, the best possible performance achievable by the recursive method is when the parameters of the recursive method converge to the exact solution of the normal equations; the results then are identical to those obtained with the Yule-Walker method.

Fig. 4.2 (c) and (d) shows that the recursive method, with p=4, 8, 16, yields very small spectral bandwidth throughout the entire SNR range, even at SNR values as low as 1dB or 2dB. For SNR=1dB, the spectral bandwidth is smaller than 3.75×10^{-3} cps, and for high SNR values, it is smaller than 6.3×10^{-4} cps, which is about $\frac{1}{5}$ of the bandwidth yielded by the Yule-Walker method at high SNRs. The frequency bias is also very small and comparable to that of the Yule-Walker method for small SNR values. However, an increase in SNR does not contribute any significant improvement in the frequency bias. For higher SNR values, the frequency bias of the recursive method is not as good as that of the Yule-Walker method, but it is kept at a very small value, below 0.004 cps, throughout the entire SNR range.

Similar experiments were conducted for N=50 and N=20. The results are depicted in Figs. 4.3 and 4.4, respectively. These plots show that as the data length is reduced both the spectral bandwidth and the frequency bias become larger. With the recursive method, the increases in the spectral bandwidth and frequency bias are much less significant than with the Yule-Walker method. Furthermore, the recursive method does not exhibit line splitting with high model orders, which is inherent in the Yule-Walker method. This is clearly evident in Fig. 4.4 (a) and (b) where the Yule-Walker method exhibits radical increases in the spectral bandwidth and frequency bias if p=16, which is too high for a 20-point signal. In Figs. 4.4 (c) and (d) the recursive method is shown to have small spectral bandwidth and frequency bias at all SNRs. A closer examination of Fig. 4.3 (c) and Fig 4.4 (c), reveals that, for N=50, the spectral bandwidth is largest when p=16, and

for N=20, the spectral bandwidth is largest when p=8. This phenomenon appears also with the frequency bias. Selecting higher model order does not always lead to an improved result. From the observation of all the plots in Figs. 4.2, 4.3 and 4.4, we can conclude that the recursive PSD estimate yields narrower peak, but at the expense of increased bias of the estimated frequency. The performance of the recursive method is more consistent than that of the Yule-Walker method.

The step multiplication factors used in the Figs. 4.2, 4.3 and 4.4 are tabulated below. Since the frequency bias and the bandwidth resolution are approximately constant throughout the SNR range, the optimum smf can be considered to be independent of SNR. This table indicates that the optimum values for the smf lie in the range smf=0.4 to smf=0.6.

Table 4.1 Step Multiplication Factors (smf)

N	p = 2	p = 4	p = 8	p = 16
100	-	0.5	0.4	0.5
50	-	0.5	0.4	0.5
20	-	0.6	0.4	0.6

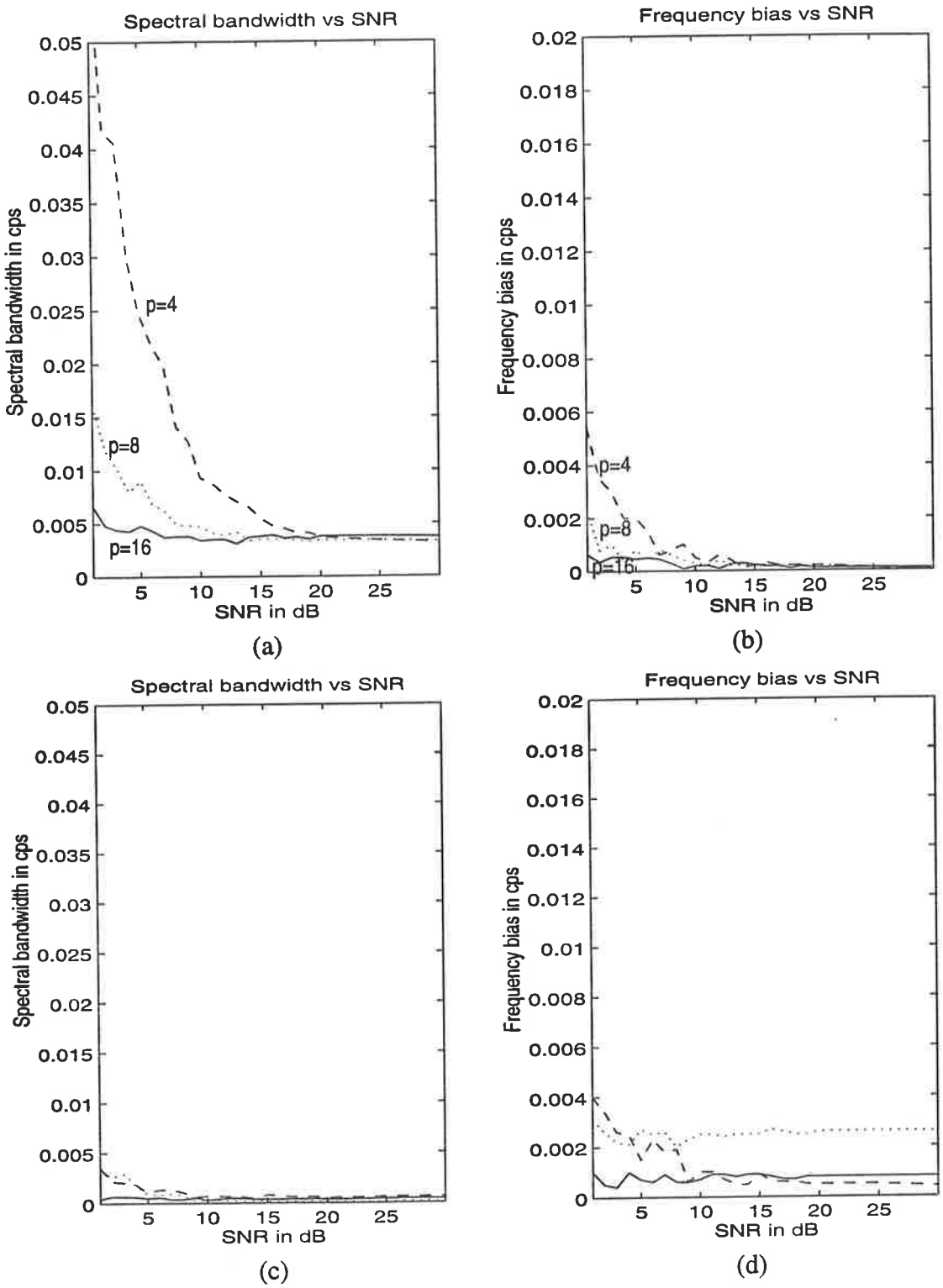

FIGURE 4.2 Spectral bandwidth and frequency bias of the Yule-Walker method, (a) and (b), and the recursive method, (c) and (d): N=100, f_0=0.3 cps, p=4(--), 8(..), and 16(-).

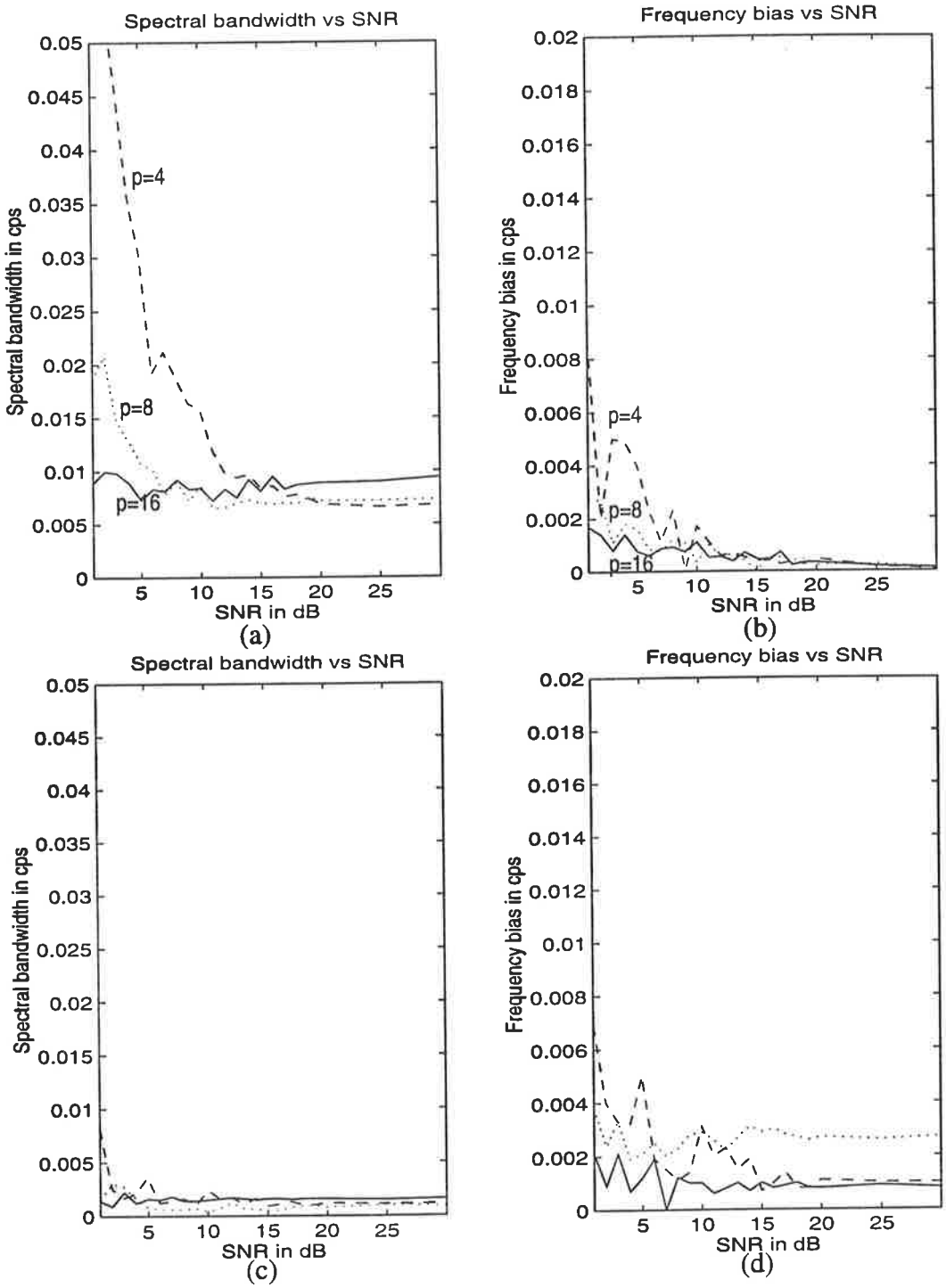

FIGURE 4.3 Spectral bandwidth and frequency bias of the Yule-Walker method, (a) and (b), and the recursive method, (c) and (d): N=50, f_0=0.3 cps, p=4(--), 8(..), and 16(-).

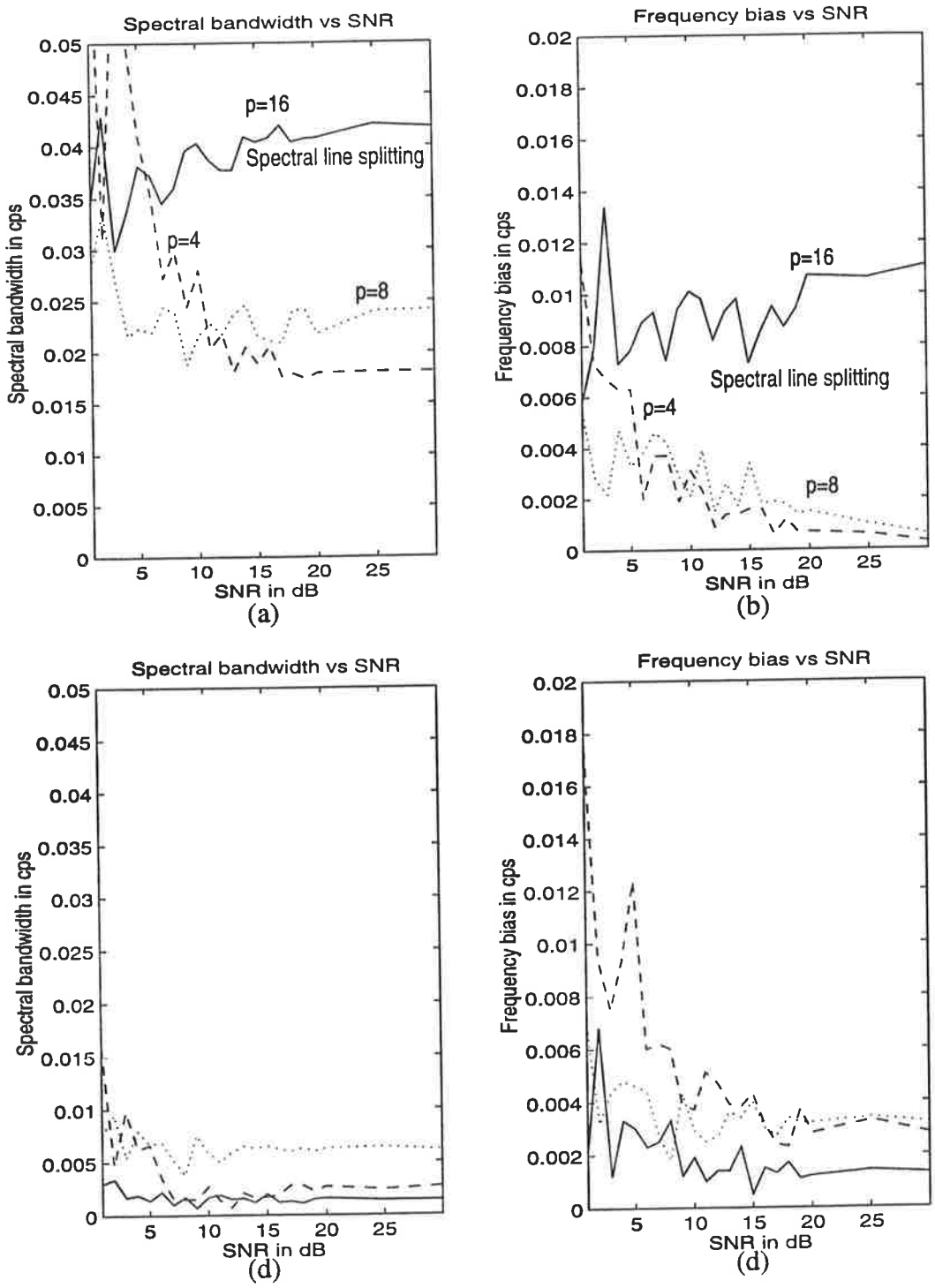

FIGURE 4.4 Spectral bandwidth and frequency bias of the Yule-Walker method, (a) and (b), and the recursive method, (c) and (d): N=20, f_0=0.3 cps, p=4(--), 8(..), and 16(-).

4.2.2 Effect of Low Frequency

In this experiment, the normalised frequency of the sinusoid is reduced to 0.1 cps in order to observe the effect of reducing the number of cycles of the sinusoid on the PSD estimate. The length of the data used is 20 points, hence we have 2 cycles of a sinusoid. In this test, the recursive method is used with one iteration only, and smf is set to 0.6. This value of the smf is chosen because it yields near optimum results.

Fig. 4.5 shows that the Yule-Walker method gives a peak with spectral bandwidth equal to 0.02 cps at the correct frequency position when p=4. Furthermore, when the model order increases the bandwidth resolution becomes worse. When the model order is increased beyond p=10, line splitting occurs, and spurious low level peaks are introduced. This phenomenon agrees with the result in Fig 4.4 (a): for SNR $\geq$ 20 dB, the bandwidth resolution of the peak with p=4 is about 0.02 cps which is the smallest value achieved. With other values of p, the bandwidth resolution increases.

When the model order is p=4, the PSD estimate of the recursive method is smoothed out, and a slight frequency bias is observed. The spectral resolution of the recursive method is slightly better than that of the Yule-Walker method at low model orders. Unlike the Yule-Walker PSD estimate, the spectral resolution of the recursive PSD estimate improves significantly as the model order increases. For p=16, the result is most outstanding: the main lobe has very small frequency bias, very large height and extremely narrow bandwidth. With the recursive method, neither line splitting nor spurious peaks are observed at high model order.

For the conventional AR based PSD estimation techniques such as the Yule-Walker method, the best model order lies in the range N/3 to N/2 to achieve good PSD estimate. This, of course, does not apply to every case, but gives a general idea on how to select the model order. With the recursive method, the appropriate model order should be higher than that of the conventional methods. The range p=12 to p=18 is adequate for a 20-point data record.

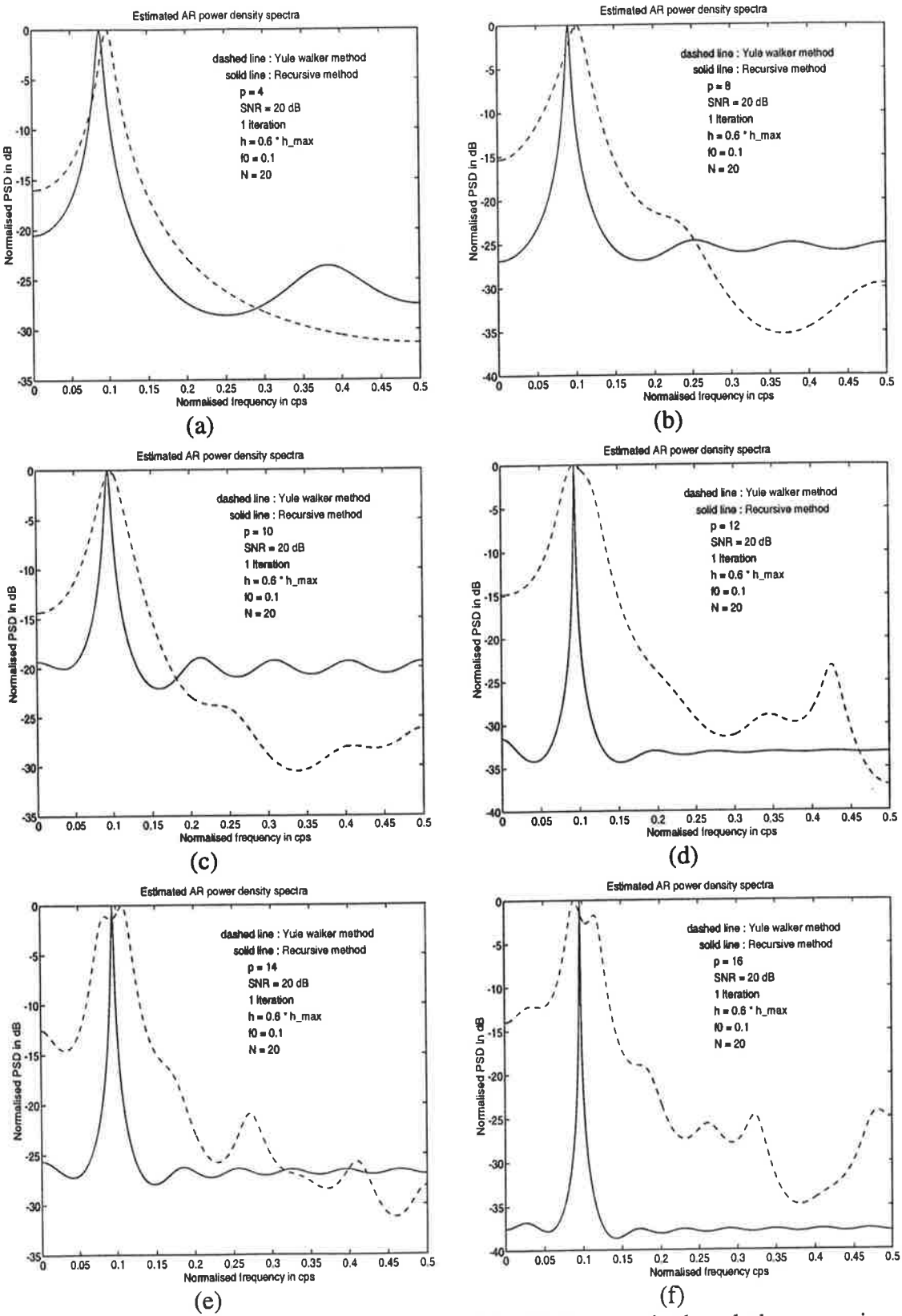

FIGURE 4.5 AR PSD estimates of the Yule Walker method and the recursive method: f_0= 0.1 cps, h = 0.6 * h_{max}, SNR = 20 dB, 1 iteration, p = 4(a), p = 8(b), p = 10(c), p = 12(d), p = 14(e), and p = 16(f).

The frequency was reduced until the Yule-Walker method could not generate a spectral peak. It was found that when the frequency was 0.03 cps, the Yule-Walker method gave a spectrum with a very small peak, which can hardly be recognised, thus, $f_0 = 0.03$ cps was selected for this test.

The model order is set to p=16 because the best result was yielded in the previous test for the recursive method (f_0=0.1 cps), for p=16. Different values of the step multiplication factors were tried, and smf=0.9 appeared to yield good PSD estimate.

Fig. 4.6 (a) shows the PSD of the Yule-Walker method with p=8. With the model order in the range 4 to 8, the Yule-Walker method yields PSD estimate with a noticeable but very small peak; Otherwise, no peak is observed. Also the variation of the shape of the spectrum was relatively large; the PSD estimate plotted in Fig 4.6 (a) has the most frequently observed shape. After one iteration, with the recursive method, no peak is observed (see Fig. 4.6 (b)), hence, no sinusoid was detected. After increasing the number of iterations to 2, a peak at f_0=0.03 cps appeared as shown in Fig. 4.6 (c). Further increase in the number of iterations revealed that the PSD estimate of the recursive method exhibits a peak only for an even number of iterations; no peak is observed for odd numbers of iterations as shown in Fig. 4.6.

Further decreases in the frequency was tried with f_0=0.02 cps and 0.01 cps. In each case, a single peak is observed when the number of iterations is even, and no peak for odd number of iterations. However, the frequency estimate showed a large bias; all peaks occurred at f_0=0.025 cps or slightly greater. There are two interpretations for this misplacement of the peak. One is that the frequency of the sinusoid is too low for the recursive method to make a correct estimate. The other may be due to the existence of a replica of this peak in the negative frequency region. Because the limitation on the resolving power of the recursive method, when the difference in the frequencies of these two peaks is less than 0.05 cps, in other words, when f_0 is less than 0.025 cps, the peaks are pushed away from each other resulting in such a large frequency bias. A test to see whether this is, in fact, the case was carried out with two sinusoids whose frequencies are 0.27 cps and 0.3 cps, respectively. The results found support this reasoning; the details of the discussion are included in Section 4.3.

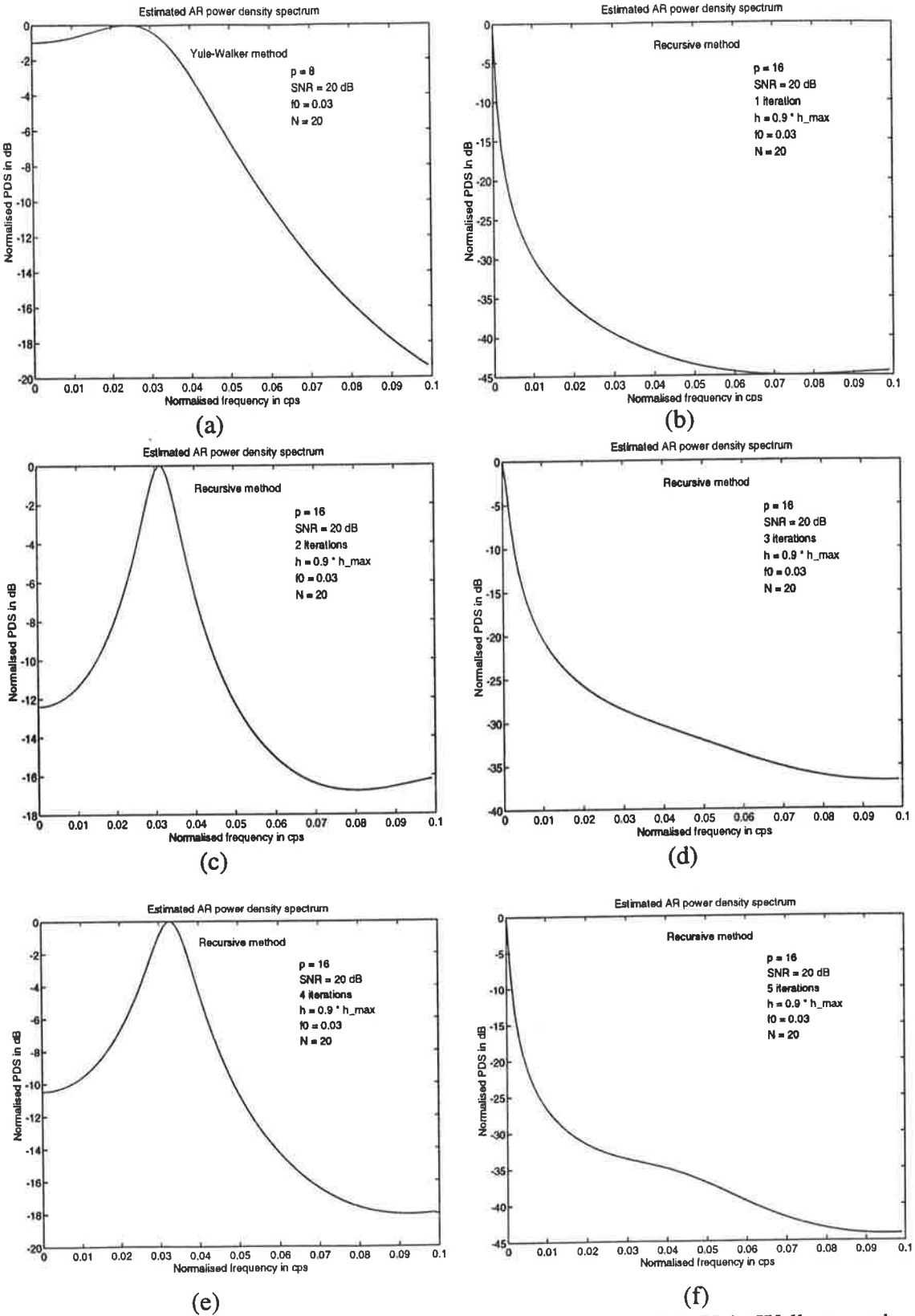

FIGURE 4.6 Power density spectrum estimations using: (a) Yule-Walker method, Recursive method with (b) 1 iteration, (c) 2 iterations, (d) 3 iterations, (e) 4 iterations and (f) 5 iterations (N = 20, p = 16, SNR = 20dB, h = 0.9 * h_max and f_0 = 0.03 cps).

4.2.3 Selection of Step Multiplication Factor

This section deals with the determination of the optimum values of the smf. Since it was observed in Fig. 4.4 (c) that the performance of the recursive method is not affected by the SNR, a value of SNR=20 dB was chosen for this test with $N = 20$. The frequency of the sinusoid varies from 0.05 cps to 0.45 cps with an increment of 0.05 cps, and the model order varies from p=4 to p=18 with an increment of 2. For each pair of these parameters, e.g., f_0=0.3 cps and p=16, seven different values of smf ranging from 0.3 to 0.9, with an increment of 0.1, were applied.

After observing the seven plots of the corresponding PSD for each pair of frequency and model order, the smf resulting in the best estimate, i.e., frequency bias and bandwidth resolution, was selected and recorded in Table 4.2.

Table 4.2 The optimum step multiplication factors(smf).

f_0	p=4	p=6	p=8	p=10	p=12	p=14	p=16	p=18	mean
0.05	-	-	-	0.8**	0.6*	0.6*	0.7**	0.8*	0.7
0.10	0.8	0.7	0.4	0.5*	0.6*	0.6	0.6*	0.7*	0.613
0.15	0.5	0.5*	0.5*	0.5*	0.6**	0.6	0.6*	0.7	0.56
0.20	0.6	0.5*	0.4	0.5	0.5	0.6*	0.6*	0.7	0.55
0.25	0.5	0.3	0.4	0.4	0.5*	0.5	0.6*	0.6	0.48
0.30	0.6*	0.5**	0.4	0.5	0.5	0.6**	0.6*	0.7	0.55
0.35	0.3	0.5**	0.5*	0.5	0.5*	0.6	0.6*	0.7	0.53
0.40	0.9	0.4	0.7	0.5*	0.6*	0.6	0.6*	0.7*	0.625
0.45	-	-	-	0.8*	0.6*	0.6*	0.7*	0.8*	0.7
mean	0.6	0.49	0.47	0.54	0.55	0.59	0.62	0.71	

☛ *→ The height of the peak is greater than 30dB but less than 40dB.

☛ ** → The height of the peak is greater than 40dB.

In order to observe the trend of the smf with respect to the frequency of the sinusoid and the model order, the smf values are averaged both column-wise (frequency-wise) and row-wise (model-order-wise). The frequency-wise averaged smf versus model order and the model-order-wise averaged smf versus frequency are depicted in Fig. 4.7.

In Fig. 4.7 (a), apart from p=4, the frequency-wise averaged smf increases as p increases. In Fig. 4.7 (b), the lowest value of the model-order-wise averaged smf is 0.48 and this

occurs for $f_0 = 0.25$ cps, i.e., the middle frequency. As the frequency moves away from this value, the smf increases in a symmetrical manner. The most ideal situation is that if the optimum smf is constant throughout the entire frequency range. This enables the choice of a single value of smf, after deciding on the model order, to be irrespective of frequency.

By examining the columns of Table 4.2, we can see that there exist relatively large variations in the optimum smf value when the model order is low. The columns associated with high model orders contain optimum smf values which are almost constant over the entire frequency range. In fact the V-shape of the graph shown in Fig. 4.7 (b) is caused mainly by the predominant influence of lower model order smf values. If we plotted the graphs for high model orders only, the plot in Fig. 4.7 (b) would be flat. Therefore, by neglecting the low model orders, a near optimum smf value can be found for each model order by taking the average value of the corresponding column, i.e., the last row in Table 4.2.

The column for p=16 contains consistent values of the smf; i.e., *smf*=0.6 throughout the entire frequency range except at the extreme frequencies such as 0.05 cps and 0.45 cps, where the smf increases slightly to 0.7. Also this column contains the largest number of * and ** symbols, which signify sharp spectral peaks with heights greater than 30 dB in the estimated PSD. This suggests that p=16 is the optimum model order for 20 points data records and the corresponding smf should be 0.6.

So far we have considered, intuitively, how to choose a smf for the best result. However, we must take into account the fact that the selected smf values may not always be optimum. An issue arising from this is the sensitivity of the recursive PSD estimate to variations in the smf; it is important that the PSD estimate is not very sensitive to changes in smf.

A test for sensitivity of 3-dB bandwidth to smf was carried out using 20-point data records; the frequency was 0.3 cps, SNR was 20dB, and the model orders were 4, 8, and 16. The smf was increased from 0.1 to 0.9 in regular intervals of 0.025, and the 3-dB bandwidth for each smf was evaluated. For each smf, the test was carried out 6 times; the results were averaged, and plotted in Fig. 4.8. The curves for p = 4, 8, and 16 have

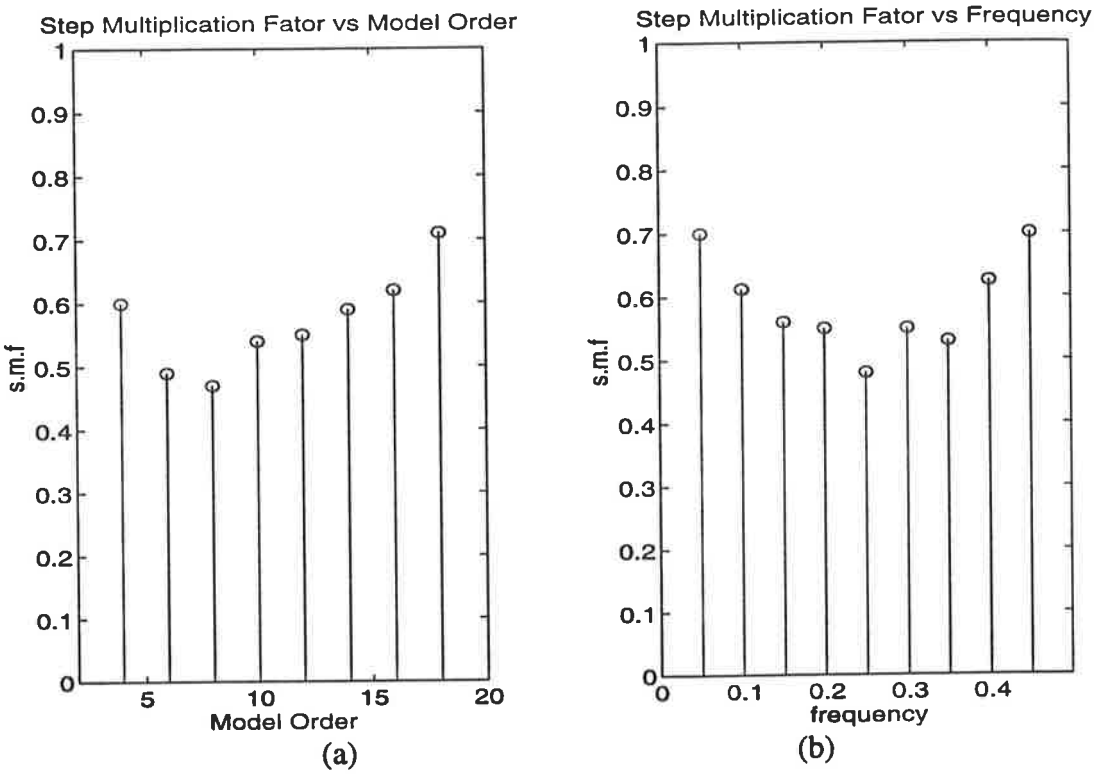

FIGURE 4.7 The averaged step multiplication factors versus (a) the model order p and (b) the frequency f_0, for N=20 and SNR=20dB.

minima at different smf positions: 0.525, 0.45, and 0.625, respectively. The corresponding 3-dB bandwidths are tabulated in Table 4.3. Note that the averaged 3-dB bandwidth decreases with increasing model order.

Table 4.3 Optimum smfs for 20-point data records. SNR = 20dB, and $f_0 = 0.3$ cps.

	***p*=4**	***p*=8**	***p*=16**
smf	0.525	0.45	0.625
3-dB bandwidth (cps)	0.0006	0.0004	0.0002

The 3-dB bandwidths obtained using smf = 0.6 with p = 4, smf = 0.4 with p = 8, smf = 0.6 with p = 16 are 0.0033, 0.0047, and 0.0013 cps, respectively. Therefore, the largest 3-dB bandwidth is obtained with p=8.

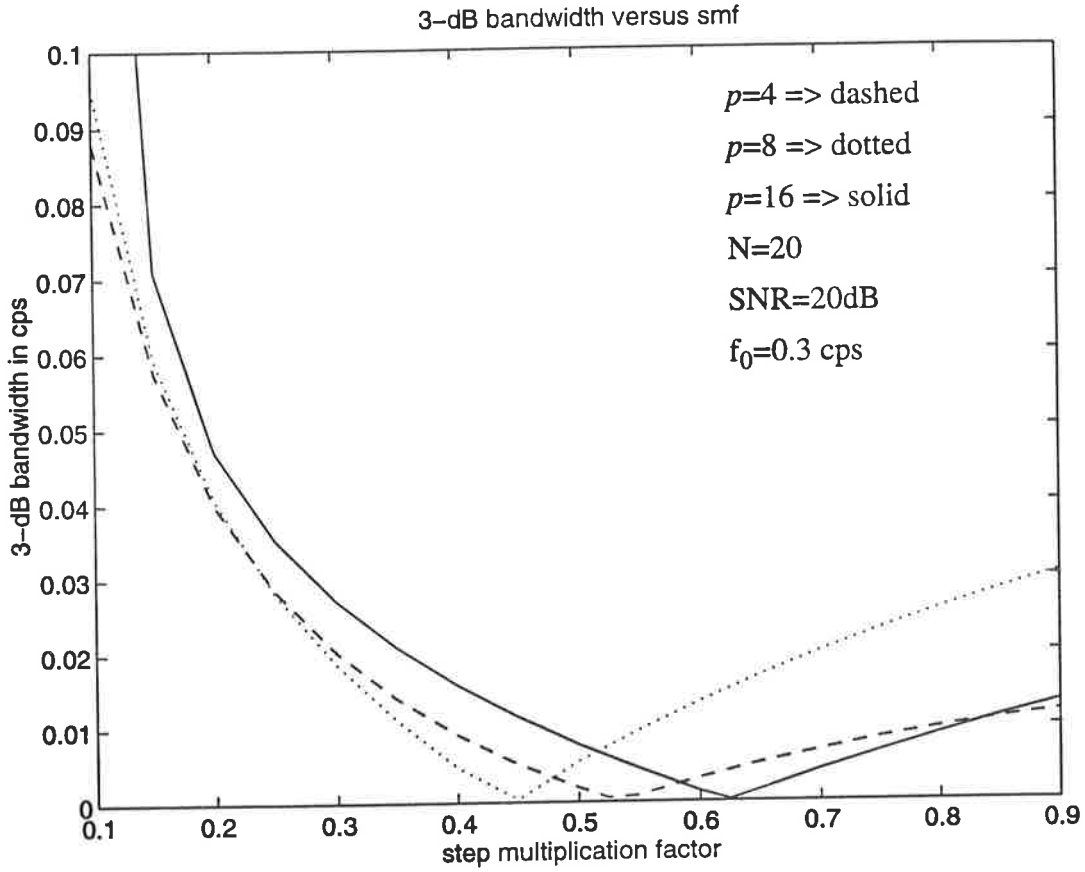

FIGURE 4.8 3-dB bandwidth versus smf. N=20, p=4 (--), 8 (..), 16 (-), f_0=0.3cps, and SNR=20dB.

In Subsection 4.2.1 (see Fig, 4.4), it was found that p=8 gives the largest 3-dB bandwidth. In that test we used smf=0.6, 0.4, 0.6. Therefore the above result is consistent with the result obtained in Subsection 4.2.1.

In Fig. 4.8, the three curves have a similar shape; each of the curves is bowl-shaped, and has one minimum. These curves have small second derivatives; the 3-dB bandwidths increase slowly and monotonously as the smf moves further away from the optima. For smfs smaller than 0.3, the 3-dB increases rapidly as smf decreases. It can be concluded that the recursive PSD estimate is not degraded significantly by choosing near optimum smf, and choosing smf between 0.4 and 0.7, in general, results in a good PSD estimate.

4.2.4 Bandwidth Versus Frequency

In this section, we study the variations of the spectral bandwidth of the peak as a function of the frequency. We use data of length 20 and SNR = 20 dB. By selecting the smf values from the frequency-wise averaged smfs, Fig. 4.7 (a), the spectral bandwidths were evaluated for model orders ranging from p=4 to p=18 at regular intervals of 2, and plotted (see Fig. 4.9).

Fig. 4.9 shows that, when the model order p is greater than N/2, i.e., p=12, 14, 16, and 18, the 3-dB bandwidths are very small, in the order of 10^{-3} cps. The 3-dB bandwidths at f_0=0.2 cps and 0.3 cps are smallest, except for p=18. The graphs exhibit a symmetry about f_0=0.25 cps, at which a local maximum occurs. However the variations throughout the frequency range are relatively small, i.e, less than 5×10^{-3} cps. At the extreme frequencies, f_0=0.05 cps and f_0=0.45 cps, the 3-dB bandwidths are still small and are less than 5×10^{-3} cps.

With the small model orders, p=4 to p=10, the 3-dB bandwidths are larger than with the high model orders. The 3-dB bandwidths remain flat throughout the middle part of the frequency range, and increase at the extreme frequencies are observed. When $p \leq 6$, no spectral peaks at f_0=0.05 cps or f_0=0.45 cps are observed.

The Yule-Walker method exhibits line splitting if the model order is greater than p=10. However, for lower model orders, the results achieved by the Yule-Walker method do not seem to be subject to variation due to the frequency change and are constant about 0.02 cps, throughout the entire frequency range regardless of the model order. In spite of consistency of the bandwidth over the frequency change, the value itself (≈ 0.02 cps) is much greater than the values from the recursive method, except at extreme frequencies.

In summary, if the data is short in duration, i.e N=20, the recursive method yields very small spectral bandwidth at the expense of variations of the spectral bandwidth with frequency. Since the actual values and the variations of the spectral bandwidths are very small, and get even smaller if the model order is high, it can be concluded that the recursive method is better in spite of the variation of the spectral bandwidth.

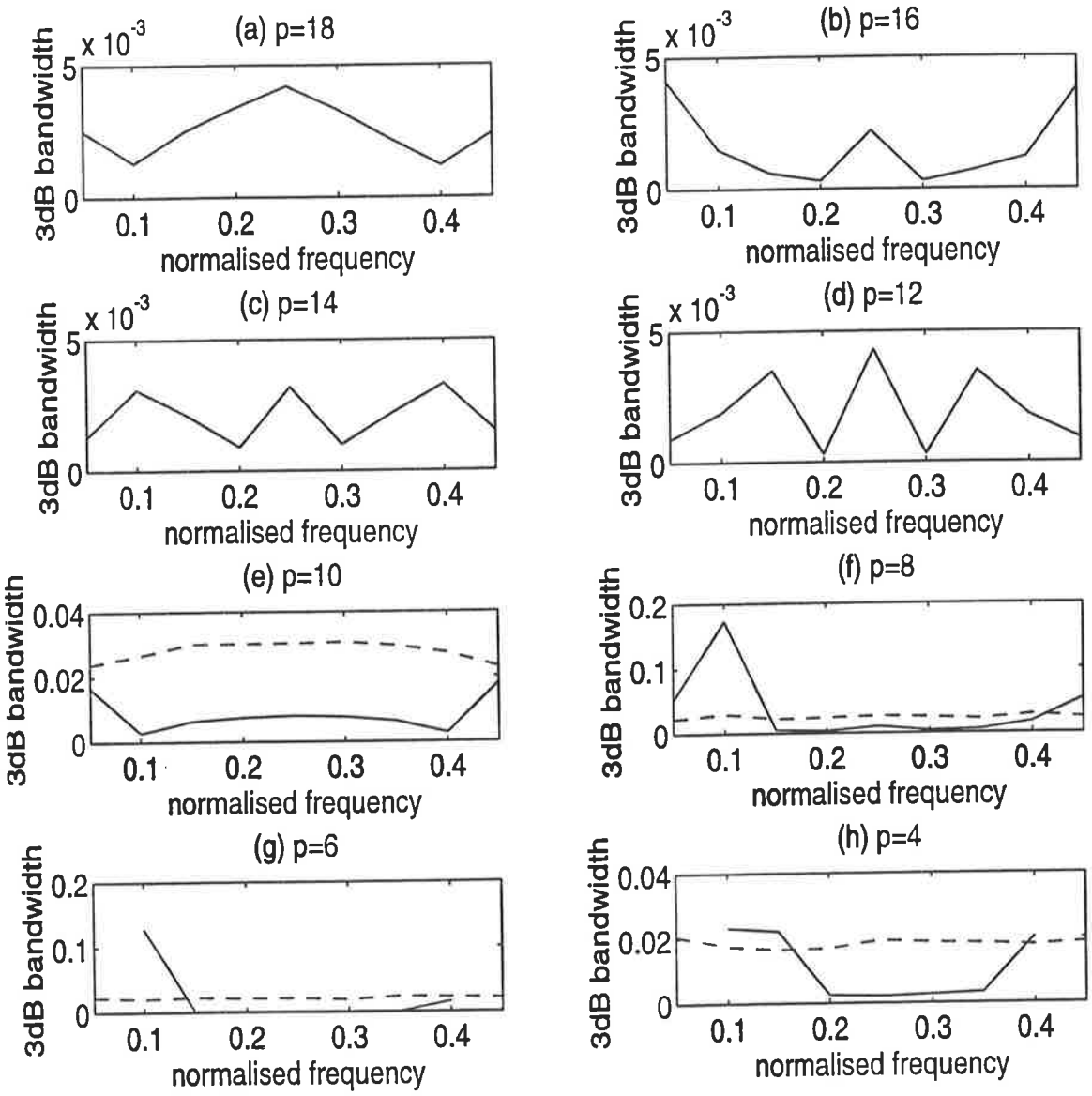

FIGURE 4.9. Variation of 3-dB bandwidth versus relative frequency (cps) of the sinusoid. Dashed and solid lines represent the Yule-Walker and the recursive PSD estimates, respectively.

4.2.5 Effect of Initial Phase

Here, we investigate the effect of initial phase of a sinusoid on the PSD estimate. Again, the Yule-Walker and the recursive methods are tested and the results are compared and analysed.

The data consists of a single sinusoid of frequency $f_0 = 0.3$ cps plus white Gaussian. The tests were performed using signals of length N=20, 50, and 100, and SNR = 20dB with model orders p = 4, 8, and 16. The smfs for the recursive method were chosen from Table 4.2 as follows

- p=4, smf = 0.6
- p=8, smf = 0.4
- p=16, smf = 0.6

We denote the initial phase value in degrees by θ. Seventy-two different initial phase values, ranging from $0°$ to $355°$ in regular intervals of $5°$, are used.

To improve the reliability of the results, for each initial phase the test is carried out twenty times, the 3-dB bandwidth and frequency bias are recorded, and only the averaged results are presented.

Recursive Method

For N=20, 50 and 100, the results obtained with the recursive method are depicted in Figs. 4.10 (a) and (b), 4.11 (a) and (b), and 4.12 (a) and (b), respectively. These figures show that both the 3-dB bandwidth and frequency bias vary as a function of initial phase. Although the amplitude of 3-dB bandwidth fluctuations for p=16 are generally much smaller than for p=4 and p=8, the variations are periodic with a period equal to $180°$. The fluctuations for p=8 and p=16 are sinusoidal, whereas the fluctuations for p=4 are not sinusoidal but periodic with period of about $180°$.

Examination of the frequency bias in Figs. 4.10 (b), 4.11 (b), and 4.12 (b), reveals that the frequency bias fluctuations are periodic with a period equal to $180°$. For N=20, the fluctuations are not sinusoidal. For N=50 and 100, the fluctuations are sinusoidal except for p=4. The amplitude of the frequency bias variations are smallest when p=16. The

variations for p=8 possess the largest DC offset, except for N=20; consequently the frequency bias when p=8 is generally greater than the frequency biases when p=4 and p=16. It is apparent that as the data length is increased the amplitude of fluctuations gets smaller but their shape is preserved.

The maximum value of the frequency bias is approximately 0.0075 cps when the fluctuation with N=20 reaches the crests at phase positions, $\theta \approx 150°$ and $325°$. However, more frequently, we would expect the frequency bias to be less than 0.003 cps as more of the results fall in this region.

Now let us examine the 3-dB bandwidth variations more closely. In Figs. 4.10 (a), 4.11 (a), and 4.12 (a), the variations when p=4 exhibit 2 main lobes centred at $\theta \approx 150°$ and $325°$, and two much smaller lobes centred at $\theta \approx 60°$ and $230°$. Taking a closer look at the shapes of the fluctuations, one would find that the patterns which have four lobes can be thought of as rectified sinusoids. Each one of these sinusoids has a DC offset smaller than amplitude of the sinusoid. Consequently the troughs of the sinusoid lies in the negative region. Since it is meaningless to have negative values for the 3-dB bandwidth and the frequency bias, because they are absolute values as defined in Section 4.1, it can be concluded that any portion of the sinusoids which lies below the threshold limit, which lies below 0.001 cps for all model orders and data lengths, are reflected back upwards about the threshold line. If this is so, all the variations can be regarded as being sinusoidal with period of $180°$; they may, or may not be all in phase, depending on which one of the major and minor lobes, for p=4, is being reflected.

Yule-Walker Method

With the Yule-Walker method, the results are quite apprehensible. In all cases, the results are roughly constant with small variations along the phase axes showing the robustness of this system against the initial phase variations.

With N=20 and p=4, the Yule-Walker method shows the smallest values for both 3-dB bandwidth and frequency bias, but for p=16 they are prominently larger than the others, indicating the advent of spectral line splitting due to excessively high model order with respect to the size of the data record. As the size of the data record increases, the differ-

ence between the three plots fades out; the three plots eventually become almost identical for N=100.

Comparisons

With regard to the 3-dB bandwidth, even though the recursive method shows prominent fluctuations, its results are still superior to those of the Yule-Walker method. The largest values of the 3-dB bandwidth achieved using the recursive method for three data lengths, N=20, N=50 and N=100, are roughly 0.0062 cps (p=8), 0.0021 cps (p=4), and 0.001 cps (p=4), respectively (see Table 4.4). On the other hand, the smallest values for the Yule-Walker method are 0.0176 cps (p=4), 0.0067 cps (p=4), and 0.0032 cps (p=8), for the same signals. This shows that, regardless of the value of initial phase and model order, the recursive method produces smaller spectral bandwidth.

For the frequency bias of the spectral peak, the Yule-Walker method shows overwhelming superiority over the recursive method. Apart from N=20 (notice that line splitting occurs when p=16), the frequency bias values obtained from the Yule-Walker method lie below 1×10^{-3} cps (see Table 4.5). With the recursive method, the frequency bias reaches about 0.008 cps in the worst cases, and the mean value of the frequency bias for all model orders and data lengths falls below 0.004 cps, which is a considerably small value.

Table. 4.4 Max. / min. 3-dB bandwidths of AR PSD estimates with initial phase.

3-dB bandwidth		**Recursive**			**Yule-Walker**		
p		**N=20**	**N=50**	**N=100**	**N=20**	**N=50**	**N=100**
4	max	0.0051	0.0021	0.0010	0.0195	0.0075	0.0039
	min	0.0006	0.0002	0.0001	0.0176	0.0067	0.0036
8	max	0.0062	0.0016	0.00047	0.0253	0.0075	0.0035
	min	0.0028	0.0004	0.00011	0.0224	0.0069	0.0032
16	max	0.0018	0.0017	0.00049	0.0419	0.0095	0.0039
	min	0.0009	0.0014	0.00035	0.0406	0.0082	0.0034

Table 4.5 Max. / min. frequency bias of AR PSD estimates with initial phase.

frequency bias		**Recursive**			**Yule-Walker**		
p		**N=20**	**N=50**	**N=100**	**N=20**	**N=50**	**N=100**
4	max	0.0075	0.0027	0.0013	0.0012	0.00033	0.00017
	min	0.0005	0.0002	0.0001	0.0005	0.00017	0.00006
8	max	0.0049	0.0033	0.0028	0.0019	0.00051	0.00017
	min	0.0003	0.0016	0.0002	0.0009	0.00020	0.00008
16	max	0.0031	0.0013	0.00097	0.0114	0.00059	0.00019
	min	0.0004	0.0002	0.00045	0.0089	0.00027	0.00007

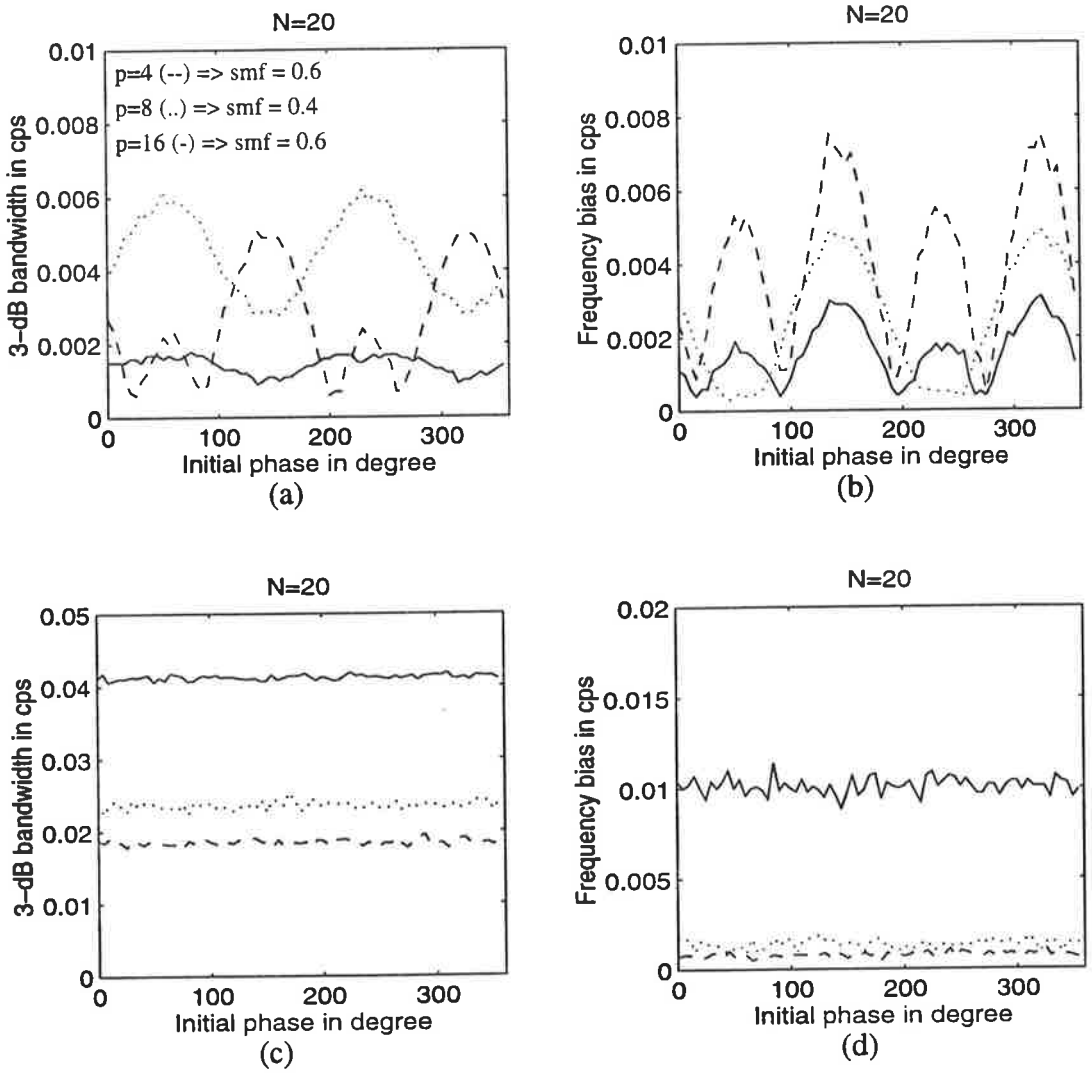

FIGURE 4.10. Variation of 3-dB bandwidth and frequency bias versus initial phase using the recursive method, (a) and (b), and the Yule-Walker method, (c) and (d). smf =0.6, 0.4, 0.6 for p=4(dashed), 8(dotted), 16(solid), respectively, N=20 points, 1 iteration, f_0=0.3 cps, and SNR=20dB.

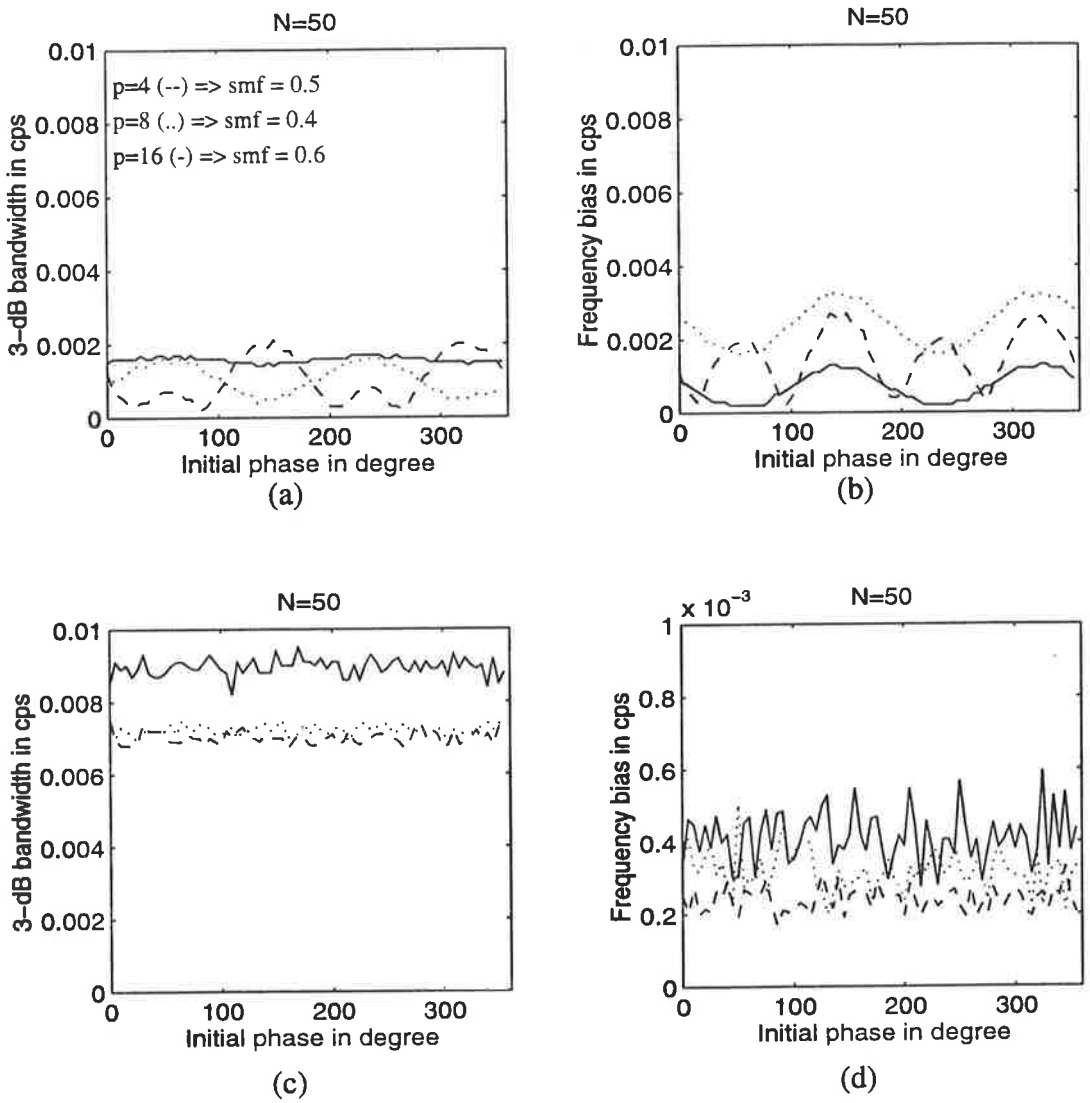

FIGURE 4.11. Variation of 3-dB bandwidth and frequency bias versus initial phase using the recursive method, (a) and (b), and the Yule-Walker method, (c) and (d). smf =0.5, 0.4, 0.6 for p=4(dashed), 8(dotted), 16(solid), respectively, N=50 points, 1 iteration, f_0=0.3 cps, and SNR=20dB.

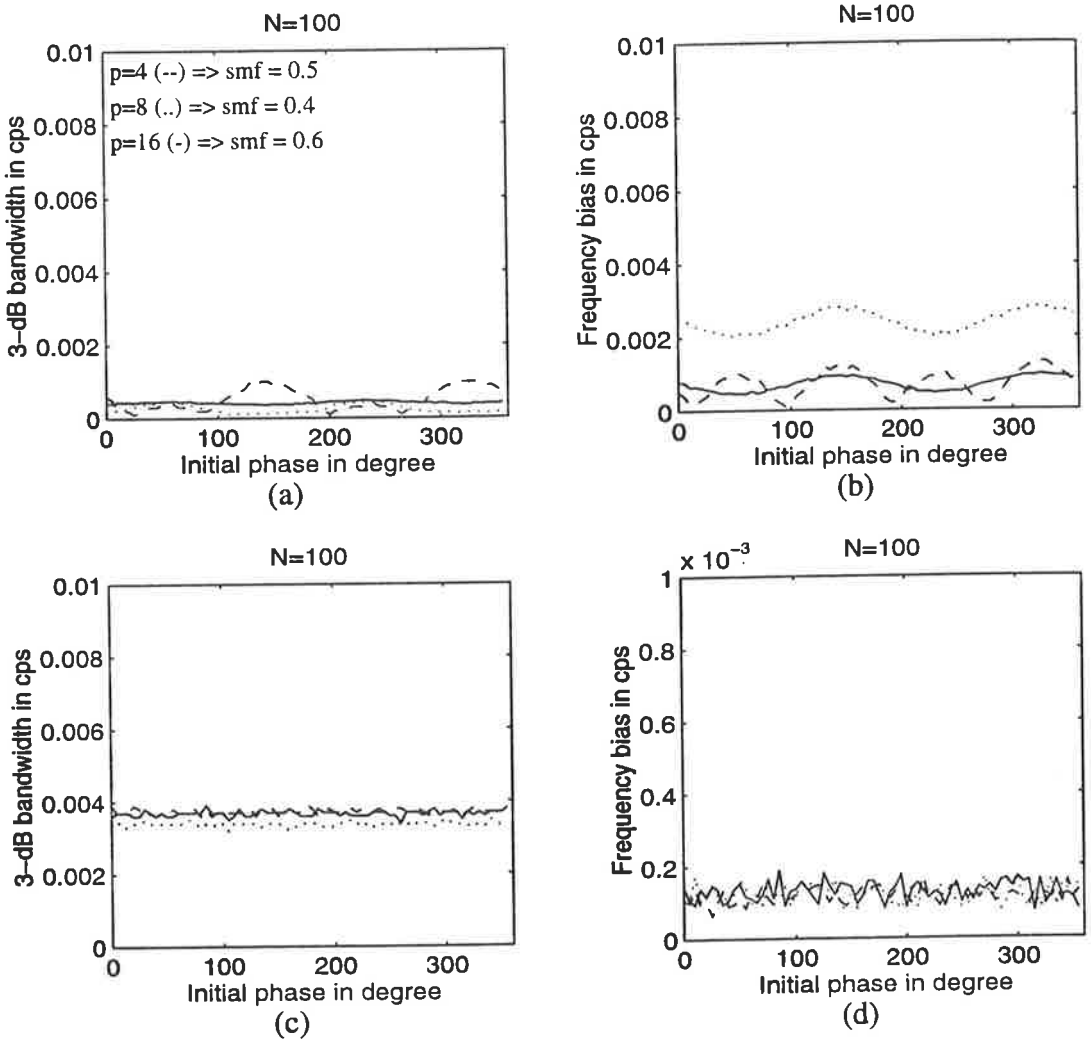

FIGURE 4.12. Variation of 3-dB bandwidth and frequency bias versus initial phase using the recursive method, (a) and (b), and the Yule-Walker method, (c) and (d). smf =0.5, 0.4, 0.6 for p=4(dashed), 8(dotted), 16(solid), respectively, N=100 points, 1 iteration, f_0=0.3 cps, and SNR=20dB.

Remarks

The graphs in Fig. 4.10, 4.11, and 4.12 represent averaged values over twenty tests, hence the smooth nature of these graphs. To see if these graphs depict an accurate picture of the performance of the two methods, similar experiments were conducted using a single test instead of twenty. The results are plotted in Figs. 4.13, 4.15, and 4.15. As expected, the graphs are much coarser but substantially similar to the smoothed ones, indicating that the smoothed graphs in Figs. 4.10, 4.11 and 4.12 are reliable representations of the behaviour of the systems with respect to phase changes.

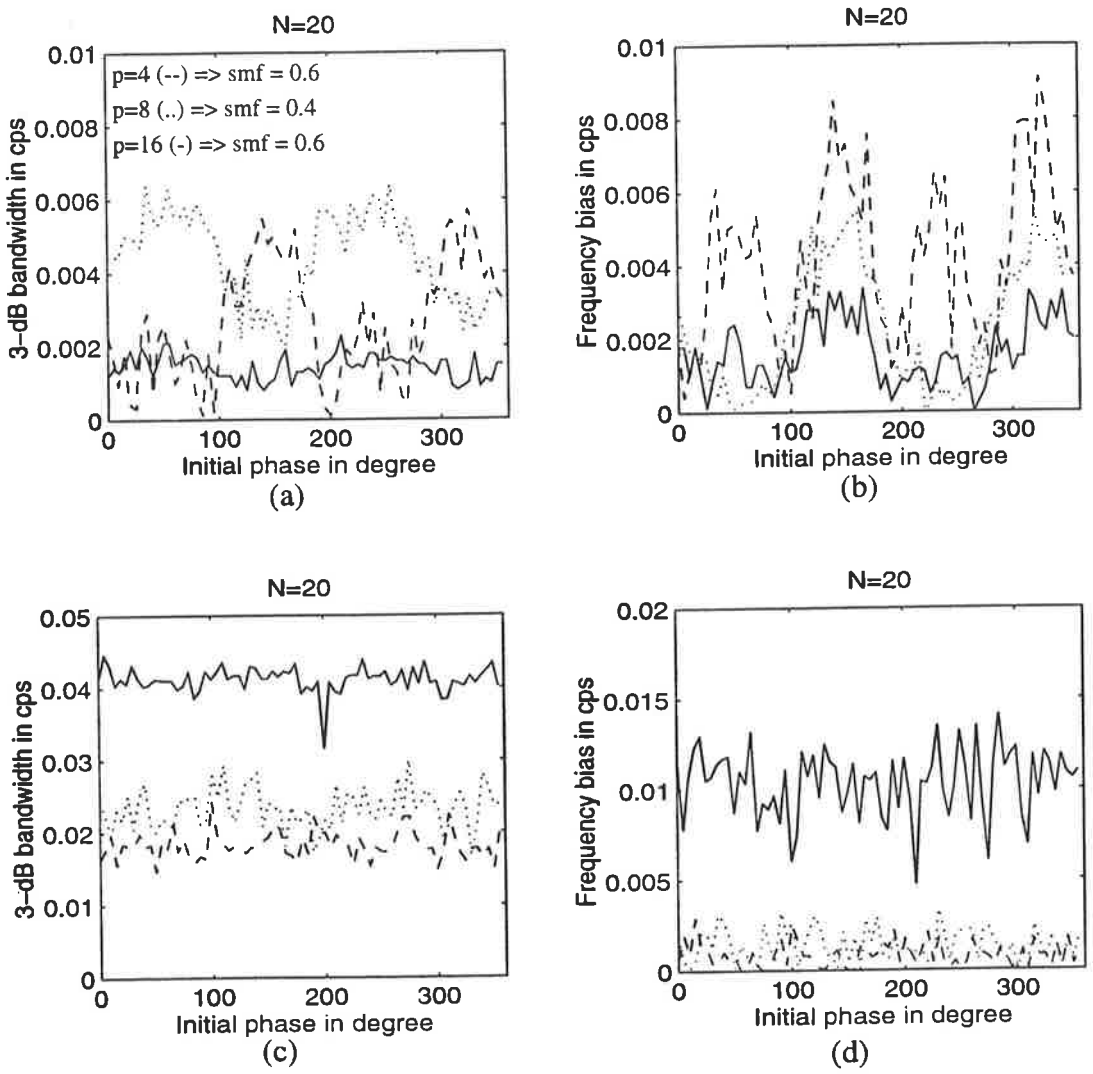

FIGURE 4.13. Variation of 3-dB bandwidth and frequency bias versus initial phase using the recursive method, (a) and (b), and the Yule-Walker method, (c) and (d). smf =0.6, 0.4, 0.6 for p=4(dashed), 8(dotted), 16(solid), respectively, N=20 points, 1 iteration, f_0=0.3 cps, and SNR=20dB.

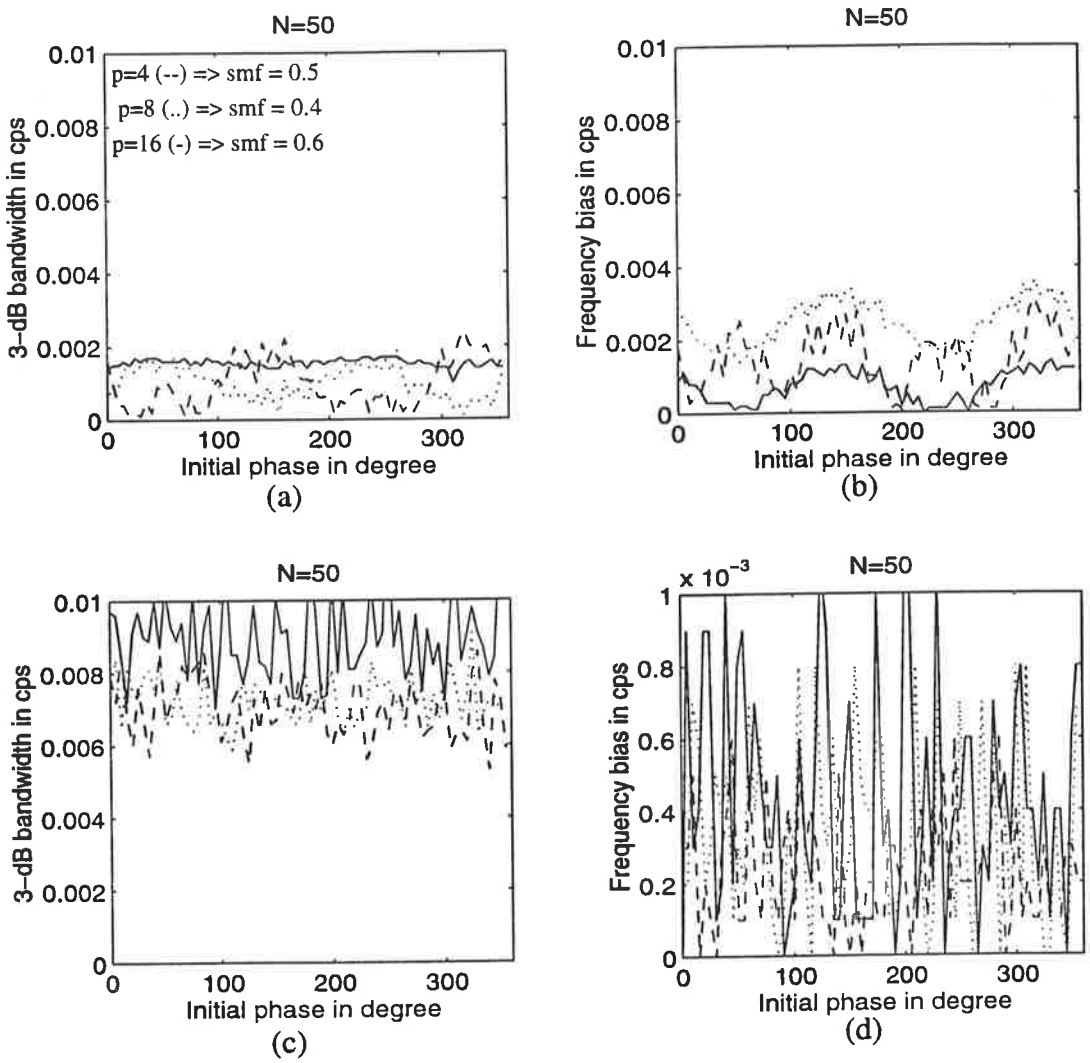

FIGURE 4.14. Variation of 3-dB bandwidth and frequency bias versus initial phase using the recursive method, (a) and (b), and the Yule-Walker method, (c) and (d). smf =0.5, 0.4, 0.5 for p=4(dashed), 8(dotted), 16(solid), respectively, N=50 points, 1 iteration, f_0=0.3 cps, and SNR=20dB.

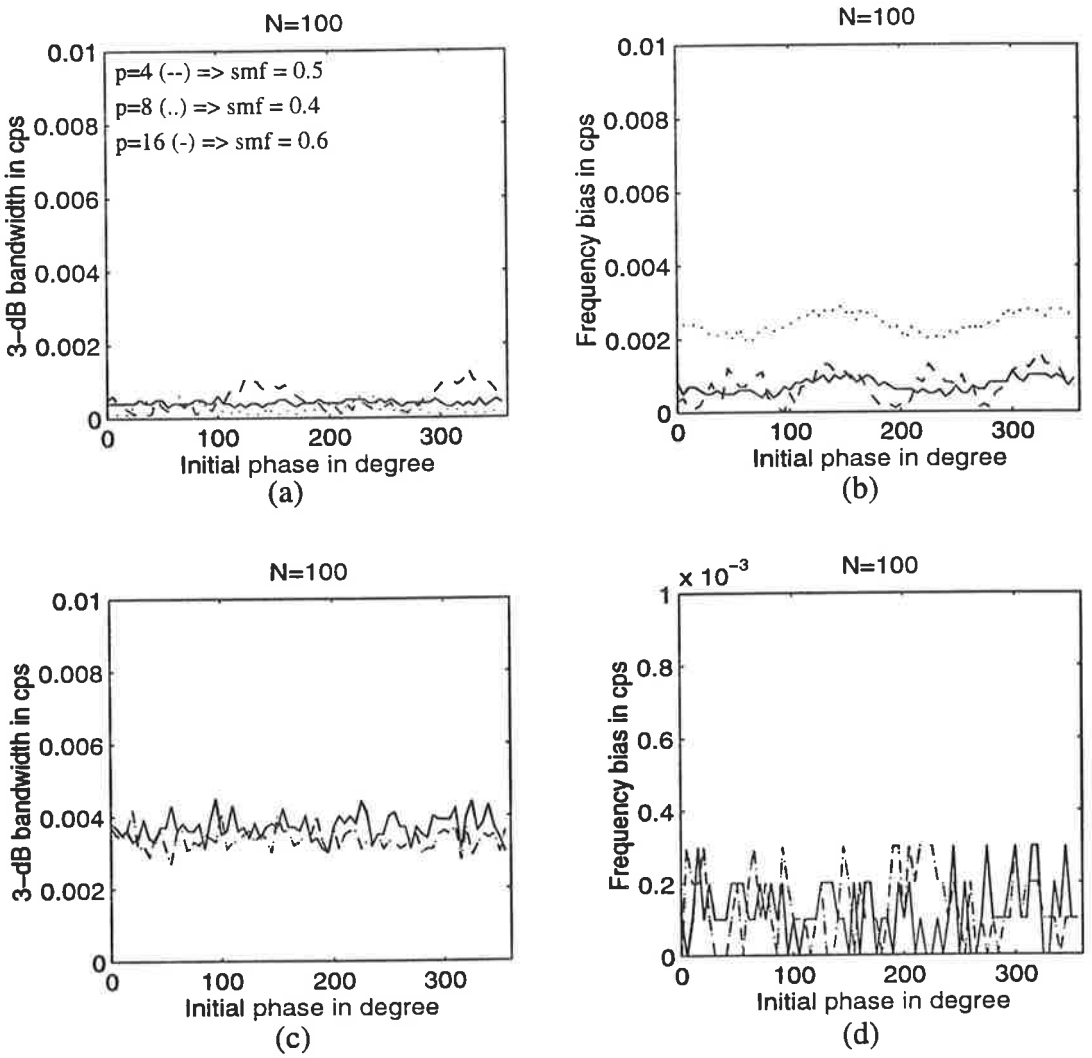

FIGURE 4.15. Variation of 3-dB bandwidth and frequency bias versus initial phase using the recursive method, (a) and (b), and the Yule-Walker method, (c) and (d). smf =0.5, 0.4, 0.5 for p=4(dashed), 8(dotted), 16(solid), respectively, N=100 points, 1 iteration, f_0=0.3 cps, and SNR=20dB.

4.3 Two Sinusoids Test

In this section, we present some experimental results on the AR based PSD estimates obtained from artificially generated data consisting of two sinusoids corrupted with noise, and spaced Δf cps apart. Thus, the underlying process is an ARMA(4,4), but the results presented in this section employ an AR(p) model. The model order p for the AR model is expected to be higher than 4 to accurately approximate the ARMA(4,4) process.

Here, we study the effects of the model order p and the SNR on the quality of the PSD estimates obtained by the Yule-Walker method and the recursive method, and also compare the spectral resolutions of the two PSD estimates.

4.3.1 Effect of Model Order

The test data comprises two sinusoids with relative frequencies $f_0 = 0.2$ cps and $f_1 = 0.3$ cps. The data record has a fixed length of 20 points and SNR=20 dB. The model order is varied from p=4 to p=18 in regular intervals of 2.

The results are depicted in Fig. 4.16. The solid lines represent the PSD estimates of the recursive method, and the dashed lines represent those of the Yule-Walker method. When p=4, the recursive method does not resolve the peaks as illustrated in Fig. 4.16 (a). The Yule-Walker method gives a very smooth (broad) spectral estimate with two small peaks. Some frequency bias is also evident in the Yule-Walker method (i.e. the peaks are pushed apart and away from their true locations). As the model order increases to p=6, Fig. 4.16 (b), the recursive method resolves the peaks putting them at the correct frequency positions. The Yule-Walker method exhibits slight frequency bias, but the resolution is better than that of the recursive estimate. However, for model orders $p \geq 8$ the recursive method exhibits superior spectral resolution and very small frequency bias. According to the graphs in Fig. 4.16, the optimum model orders seem to be p=8 for the Yule-Walker method and p=14 for the recursive method. The peaks, in the recursive method, when p=14 have large heights, about 40dB~45dB, extremely narrow bandwidths, and very small frequency bias. The Yule-Walker method exhibits broadening in peaks as the model order exceeds p=14, and prominent splitting in the peaks when p=18, Fig 4.16 (h).

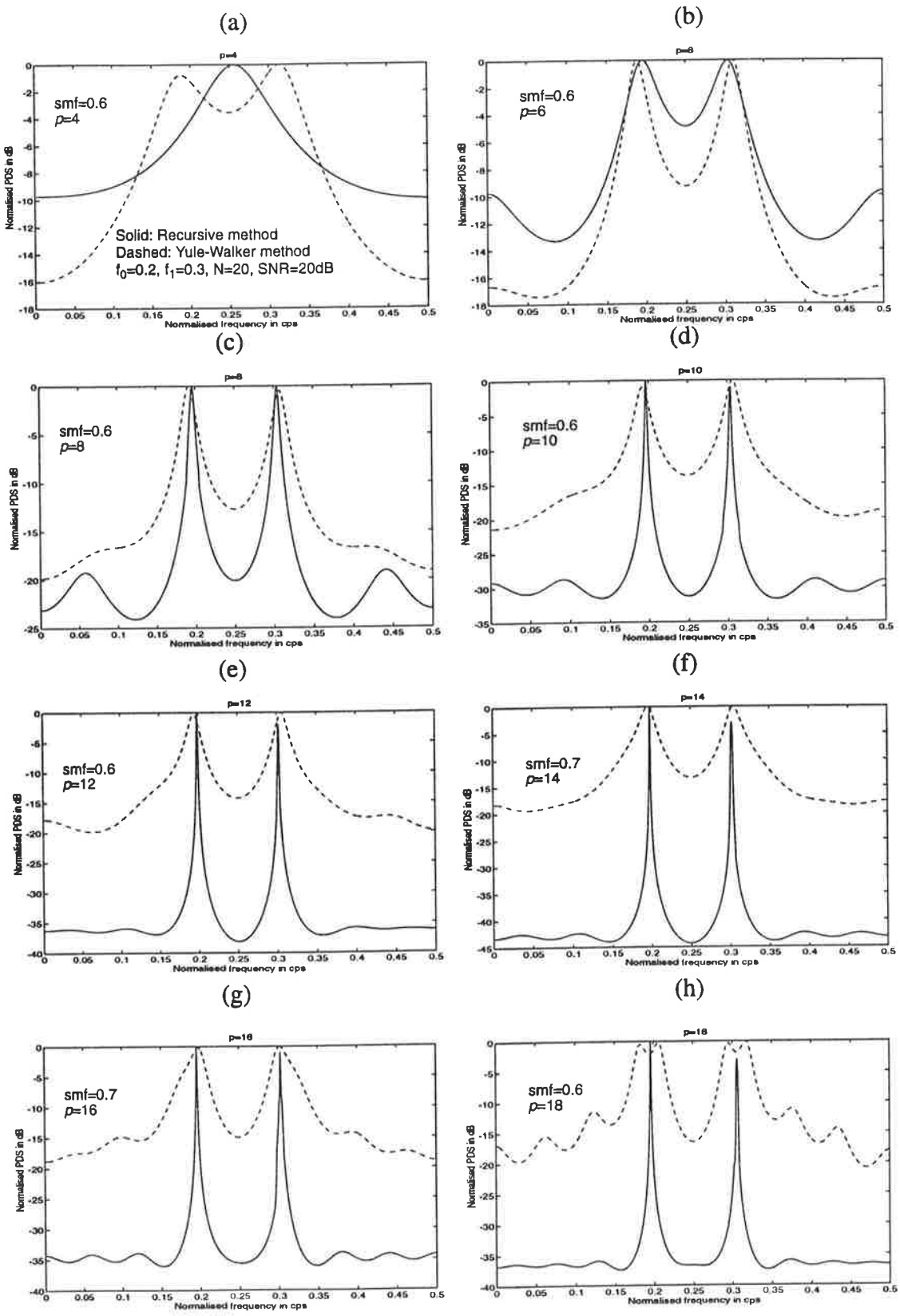

FIGURE 4.16 PSD estimates using the Yule-Walker method (dashed line) and the recursive method (solid line). N=20, SNR=20dB, f_0 =0.2 cps, and f_1 =0.3 cps.

Comparing the best PSD estimates, Figs. 4.16 (c) and (f), it is clearly evident that the recursive method outperforms the Yule-Walker method in this test.

4.3.2 Effect of SNR

It is well known that the resolution of the AR spectral estimate for two equiamplitude sinusoids in white noise decreases as the SNR decreases. The reason for the degradation is that the all-pole model assumed in AR spectral analysis is no longer valid when observation noise is present. The inadequacy of the AR model for a noise corrupted AR process leads to degradation in PSD estimate. The effect of noise is to reduce the dynamic range of PSD of the signal without noise.

To observe the effect of different noise levels on the PSD estimates, the PSD estimates with four different noise levels, SNR= 1dB, 5db, 10dB, and 20dB, are examined. The results using the Yule-Walker method are plotted in Fig. 4.17. When the SNR=1dB, there exist noticeable spurious peaks which have relatively large heights, and the two main spectral peaks have unequal heights. As the SNR increases to 5 dB, the spurious peaks are diminished; however, the difference in height of the main peaks still remains. For larger SNRs, the estimated PSDs become better: no spurious peaks, equal height of the peaks, and consistency against the SNR change (i.e. No distinguishable differences between SNR = 10 dB and SNR = 20dB).

A similar test is conducted using the recursive method, and the results are shown in Fig. 4.18. Similar phenomena as with the Yule-Walker method are observed, except that the heights of the main peaks are much larger than those of the Yule-Walker method. There is also a slight improvement in bandwidth resolution when SNR changes from 1dB to 5dB.

Comparing the SNR=1dB and SNR=20dB, one can easily see that the degradation due to a decrease in SNR is more prominent in the Yule-Walker PSD estimate than it is in the recursive PSD estimate.

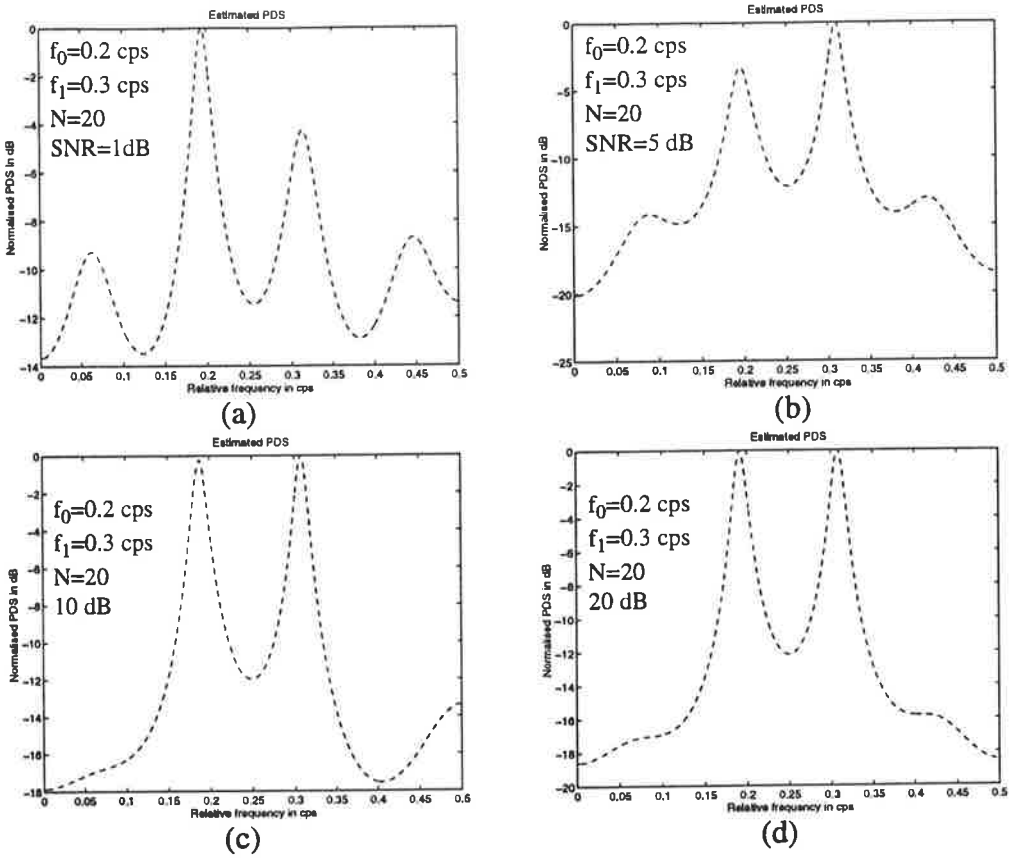

FIGURE 4.17 Effect of SNR on the Yule-Walker PSD estimates

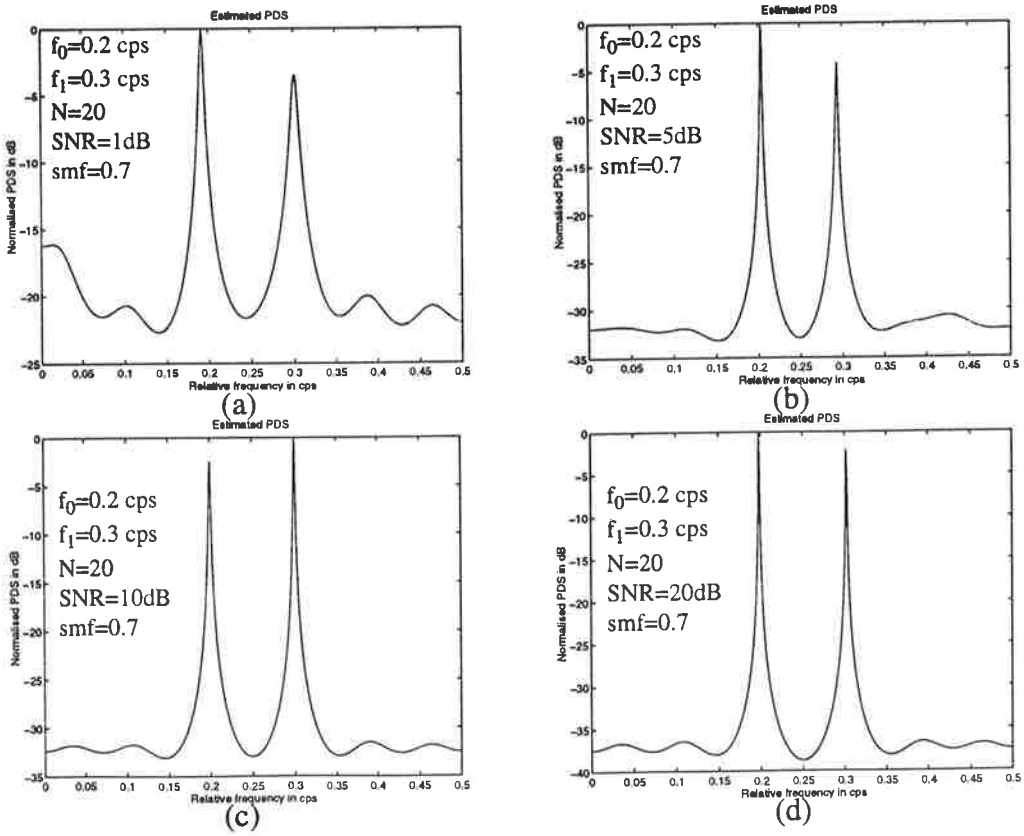

FIGURE 4.18 Effect of SNR on the recursive PSD estimates

4.3.3 Spectral Resolution

In this section, we focus on spectral resolution of the PSD estimates. Spectral resolution is a measure of how closely spaced in the frequency domain two sinusoids can become before they can not be distinguished and are detected as a single broad lobe.

The test data consists of two sinusoids of equal amplitude with frequencies 0.25 cps and 0.3 cps, hence Δf=0.05 cps. The SNR is set to 20dB, and the model order is set to 14. A model order of p = 14 was chosen because in Fig. 4.16 it yielded the best PSD estimate for the recursive method and a near best PSD estimate for the Yule-Walker method. This experiment was repeated using p=8, the optimum model order for the Yule-Walker method in Fig. 4.16; the PSD estimate obtained by the Yule-Walker method using p=8 was poorer than that obtained with p=14.

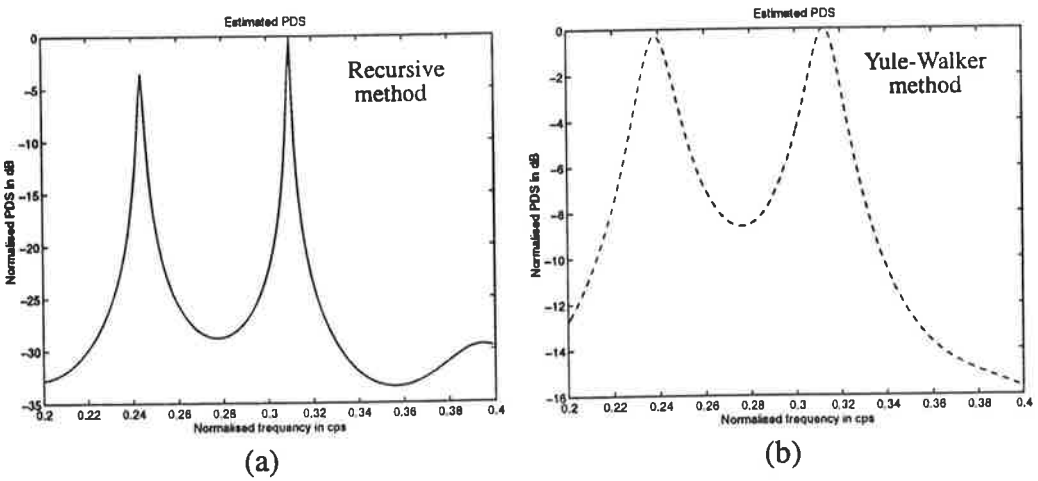

FIGURE 4.19 PSD estimates with Δf = 0.05 cps and p=14 using the recursive method (a) and the Yule-Walker method (b), smf=0.6, f_0=0.25 cps, SNR=20dB, f_1 =0.3 cps, and N=20.

In Fig. 4.19 (a), the recursive method exhibits sharp peaks at f=0.246 cps and f=0.31 cps giving Δf=0.064 cps. In Fig. 4.19 (b), the Yule-Walker method exhibits 2 more broad peaks at f=0.239 cps and f=0.314 cps, giving Δf=0.075 cps. Notice that the peaks obtained with the Yule-Walker method are broader and displaced further from their true positions than those obtained with the recursive method, suggesting that the Yule-Walker

method is inferior to the recursive method with respect to the spectral resolution.

The frequency difference between the two sinusoids is reduced to Δf = 0.03 cps (f_0 = 0.27 cps and f_1 = 0.3 cps). Recalling the low frequency test in Subsection 4.2.2, it was mentioned that when the relative frequency of the peak was below f = 0.025 cps, the peak was displaced from the correct position to where the frequency is greater than 0.03 cps. One reasoning for this result was that it is due to the presence of a replica in the negative frequency region, and to the fact that the resolving power of the recursive method is limited to $\Delta f > 0.05$ cps.

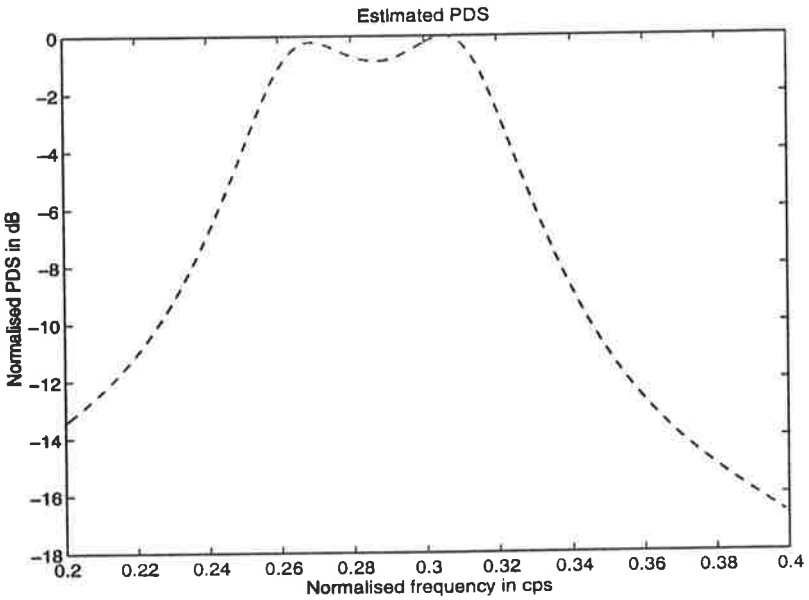

FIGURE 4.20 Yule-Walker PSD estimate with Δf= 0.03 cps. f_0 = 0.27 cps, f_1 = 0.3 cps, SNR=20 dB, N=20, and p=12.

First, the Yule-Walker method is tried and the best spectrum, obtained using p=12, is plotted in Fig. 4.20. Two very small peaks at f = 0.27 cps and f = 0.306 cps are observed giving Δf = 0.036 cps. Of course the spectrum is highly degraded with the peaks being barely perceptible. Peaks are considered to be resolved only if they are separate at -3 dB level. The heights of the small peaks in Fig. 4.20 are about 1 dB; only a single broad lobe is detected, Therefore it is concluded that the Yule-Walker method can not resolve peaks 0.03 cps apart.

Second, the recursive PSD estimate is plotted using the same signal. The best PSD estimates are obtained using smf = 0.9 and p=16. The results are shown in Fig. 4.21. A similar phenomenon as with the low frequency test results is observed. After odd number of iterations, a single peak is observed at a frequency position between the two true frequency positions. However with even number of iterations, two peaks are detected, their frequency positions are $f = 0.258$ cps and $f = 0.317$ cps respectively hence giving $\Delta f \approx 0.059$ cps. Comparing these with their true frequencies, i.e., $f = 0.27$ cps and $f = 0.3$ cps, it can be seen that the peaks are pushed further away from each other giving false frequency estimates.

In conclusion, the Yule-Walker method can not detect two very close peaks; the peaks are not separated at -3 dB level, and only the crests of the peaks are distinguishable. On the other hand, the recursive method detects the presence of two very close peaks, but trades the accuracy of estimated frequencies with the resolution of peaks, i.e., two peaks are clearly separated. Therefore, for even number of iterations, any two peaks with $\Delta f < 0.05$ cps will be detected at frequency locations with $\Delta f \geq 0.05$ cps. This supports the second reasoning of the low frequency test results in Subsection 4.2.2.

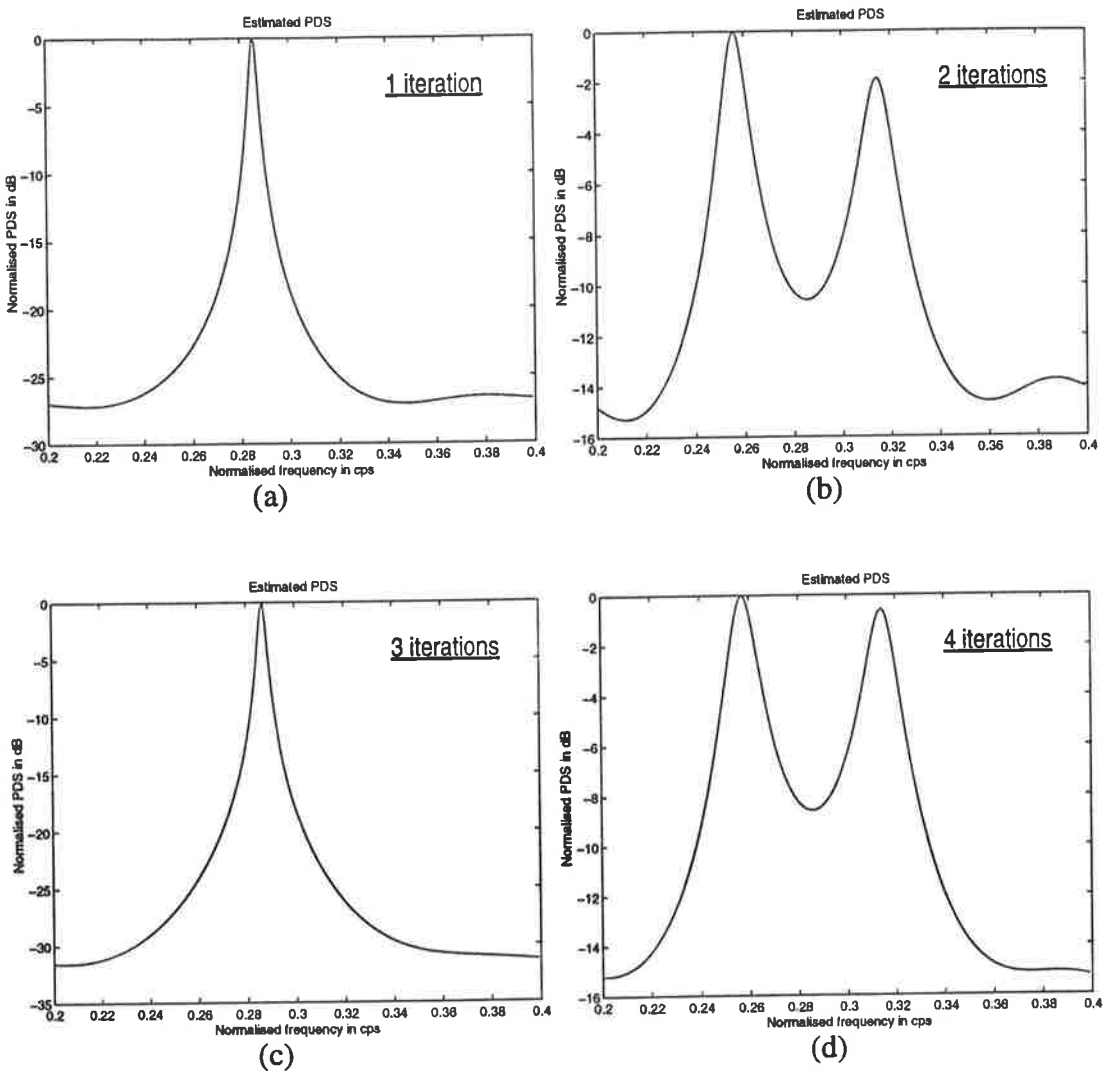

FIGURE 4.21 The recursive PSD estimates for $\Delta f = 0.03$ cps. p=16, $f_0 = 0.27$ cps, $f_1 = 0.3$ cps, smf = 0.9, SNR=20dB, number of iterations=1 (a), 2 (b), 3 (c), and 4 (d).

4.4 Summary of Test Results

This section summarises the attributes of the recursive method based upon the test results presented in the current chapter. For simplicity these attributes are presented in a list form.

- No line splitting.
- No spurious peaks.
- smf value must be chosen optimally.
- Even number of iterations is needed to resolve very closely spaced spectral peaks.
- Very good bandwidth resolution.
- Small bias in frequency estimate.
- Robustness in noisy environment.
- Very small variations of 3-dB bandwidth and frequency bias due to initial phase changes.
- Very small variations of bandwidth resolution due to the frequency changes.
- Higher model orders work better.

The recursive method features many advantages over the Yule-Walker method. Another strength of this method is the simplicity of the recursive equation. Overall, the recursive method outperforms the Yule-Walker method.

CHAPTER 5 *Implementation*

This chapter deals with the implementation issues of the recursive method for PSD estimation. It is shown that the recursive method can be implemented using Fast Fourier Transform (FFT) algorithms, and hence improving its computational speed, especially for large model orders. The implementation of PSD estimation algorithm is crucial and this chapter demonstrates once again the superiority of the new recursive method compared to the other well known techniques. Discussions on how the recursive equation is related to circular convolution is treated in Section 5.1. The implementation of the recursive equation using the FFT algorithms is presented in Section 5.2. Savings on the computational cost of the recursive method, by incorporating the FFT algorithm, and comparisons with the computational complexity of other AR PSD estimation techniques are dealt with in Section 5.3.

5.1 Analogy between Matrix Multiplication and Circular Convolution

In this section we explore the link between the devised recursive equation and circular convolution, which is implementable using the FFT algorithm. To do so, let us recall the recursive equation, Eqn. (3.23),

$$\boldsymbol{u}(n+1) = M \cdot \boldsymbol{u}(n) + y_1 \tag{5.1}$$

where

$$\boldsymbol{u}^T = [u(1)\,u(2)\ldots u(p)]\,,\; y_1^T = \frac{h}{1+hk}[r_{xx}(1)\;r_{xx}(2)\;\ldots\;r_{xx}(p)]\,,\; k>0$$

and $M = \dfrac{I-hC}{1+hk} = \begin{bmatrix} m(0) & \ldots & m(p-1) \\ \ldots & \ldots & \ldots \\ m(p-1) & \ldots & m(0) \end{bmatrix}$

This equation provides a complete description of the system or a mechanism governing the adjustment of the AR model parameters, which are the states of the system. We say this equation provides a state-space description of the system with y_1 and $\boldsymbol{u}(n)$, for $n \geq 0$, being the input and state variables, respectively. Since the matrix M and the input vector y_1 are constant with time, the above equation is called the *linear time-invariant state-space realization* of p^{th} order system. The initial value for $\boldsymbol{u}(n)$ is zero, i.e., $\boldsymbol{u}^T(0) = [0\ldots0]$. This recursive method involves $p^2 \times n$ multiplications where p is the model order and n is the number of iterations.

Let us consider a multiplication of the system matrix M and the state vector $\boldsymbol{u}(n)$

$$\begin{bmatrix} m(0) & \ldots & m(p-1) \\ \ldots & \ldots & \ldots \\ m(p-1) & \ldots & m(0) \end{bmatrix} \times \begin{bmatrix} u(1) \\ \ldots \\ u(p) \end{bmatrix} \tag{5.2}$$

Due to the structure of the matrix M, this multiplication is somewhat analogous to the circular convolution [10] of the two sequences $\{m(n)\}$ and $\{u(n)\}$ given by

$$\{m(n)\} = \{m(0)\;\; m(1)\;\; \ldots\; m(p-1)\;\; m(p-1)\;\; m(p-2)\;\; \ldots\; m(1)\} \tag{5.3}$$

and

$$\{u(n)\} = \{u(1)\;\; u(2)\;\; \ldots\; u(p)\;\; \underbrace{0\;0\;\ldots\;0}_{p\text{-1 zeros}}\;\} \tag{5.4}$$

These two sequences are of equal length, $L = 2p - 1$.

We know that the sequence $\{m(n)\}$ is a fixed quantity and the sequence $\{u(n)\}$ is to be constantly updated at each iteration. The circular convolution of $\{m(n)\}$ and $\{u(n)\}$ is illustrated in Fig. 5.1.

The product of the matrix M and the vector $\boldsymbol{u}$ yields a vector of length p. However the circular convolution of the two sequences $\{m(n)\}$ and $\{u(n)\}$ yields a sequence of length $2p$-1, which is greater than p for $p \geq 2$. This suggests that some kind of elaboration is required.

Let us rewrite the recursive equation (5.1) as a circular convolution of two sequences.

$$\{u_{k+1}(n)\} = \{m(n)\} \circledast \{u_k(n)\} + \{y_1(n)\} \tag{5.5}$$

where $\{y_1(n)\} = \{y_1(1) \ \ldots \ y_1(p) \ \underbrace{0 \ \ldots \ 0}_{\text{-1 zeros}}\}$

The circular convolution of the two sequences $\{m(n)\}$ and $\{u_k(n)\}$ yields a sequence of length $2p$-1. We denote this sequence by $\{b_k(n)\}$

$$\{b_k(n)\} = \{m(n)\} \circledast \{u_k(n)\}$$

$$= \{\underbrace{b_k(1) \ldots b_k(p)}_{= Mu} \ b_k(p+1) \ldots b_k(2p-1)\} \tag{5.6}$$

We notice that the first p elements in the sequence $\{b_k(n)\}$ constitute a sequence equivalent to the multiplication of the matrix M and the vector $\underline{u}$, the rest are redundant terms.

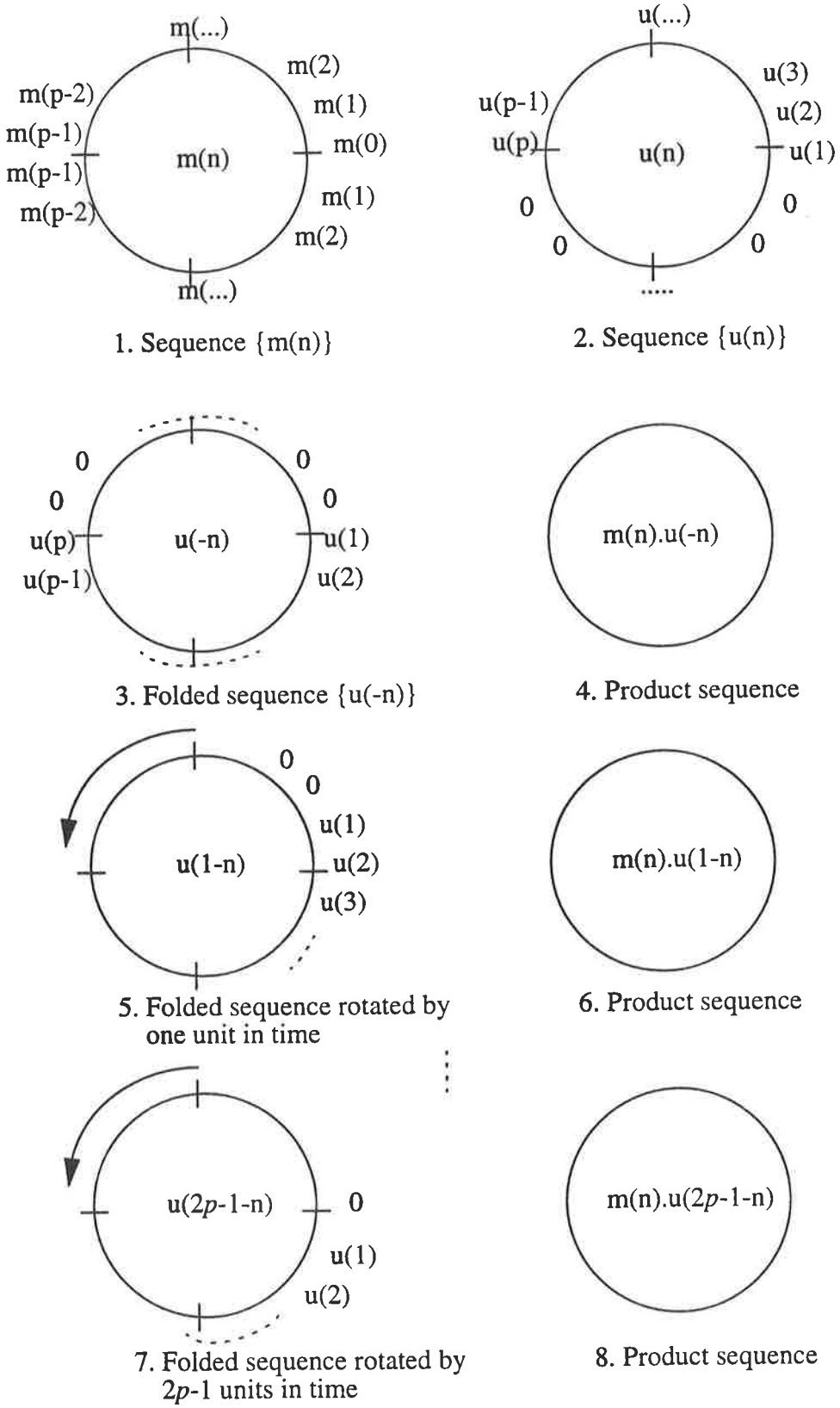

FIGURE 5.1 Circular convolution of two sequences $\{m(n)\}$ and $\{u(n)\}$.

Therefore the sequence $\{u_{k+1}(n)\}$ obtained from Eqn. (5.5) will also contain the redundant elements, $b_k(p+1) \ldots b_k(2p-1)$.

$$\{u_{k+1}(n)\} = \{b_k(n)\} + \{y_1(k)\}$$

$$= \{\underbrace{u_{k+1}(1) \ldots u_{k+1}(p)}_{\text{updated AR coefficients}} \overbrace{b_k(k) \ldots b_k(2p-1)}^{\text{redundant elements}}\} \quad (5.7)$$

The last p-1 elements in the sequence $\{u_{k+1}(n)\}$ need to be replaced by zeros to enter the next iteration; the sequence $\{u_{k+1}(n)\}$ must retain its structure in Eqn. (5.4) to enter the next circular convolution operation, i.e., $\{m(n)\} \circledast \{u_{k+1}(n)\}$.

5.2 Implementation using FFT

If the matrix multiplication were equivalent to full circular convolution, then using the linearity property of the DFT, the recursive equation can be expressed in the frequency domain as follows

$$FFT\left[\{u_{k+1}(k)\}\right] = FFT[\{m(n)\}] \cdot FFT\left[\{u_k(n)\}\right] + FFT\left[\{y_1(n)\}\right] \quad (5.8)$$

Each of these sequences transformed into the frequency domain maintains its length 2p-1. By doing this, the number of multiplications per iteration can be reduced down to 2p-1 which would be a remarkable saving in computation time. However, since this matrix multiplications is equivalent to convolution of the sequences only as indicated by Eqn. (5.6), the above linearity property is no longer valid.

Once $FFT\left[\{u_{k+1}(n)\}\right]$ is obtained from Eqn. (5.8), it needs to be transformed to the time sequence $\{u_{k+1}(n)\}$. However, since the sequence $\{u_{k+1}(n)\}$ contains the redundant elements, $b_k(p+1) \ldots b_k(2p-1)$, these must be replaced by zeros. Then the FFT of the corrected AR coefficient sequence $\{u_{k+1}(n)\}$ is computed and used in Eqn. (5.8) to obtain the $FFT\left[\{u_{k+2}(n)\}\right]$. Note that $FFT\{m(n)\}$ and $FFT\{y_1(n)\}$ are computed once only, since the sequences $\{m(n)\}$ and $\{y_1(n)\}$ do not vary with number of iterations.

If we let L be the length of the sequences $\{m(n)\}$ and $\{u(n)\}$ and assume $L = 2^{\nu}$, then the total number of multiplications in each iteration would be

element by element multiplication

$$\underbrace{2\times\left(\frac{L}{2}\right)\log_2(L)}_{\text{2 FFT operations}} + L = L\times\log_2(2L) \approx 2p\times\log_2(4p) \qquad (5.9)$$

since $L=2p-1 \approx 2p$.

To compare this number with the p^2, the number required for matrix-vector multiplication, the two numbers are plotted against the model order in Fig. 5.2. In Fig. 5.2, It is observed that if p is greater than 11 then the solid line lies below the dotted line. Since p is usually greater than 12 for the recursive method to yield good results, this graph suggests that incorporating FFT reduces the computational complexity of the recursive method.

For illustrative purposes, Fig. 5.3 depicts the computation of an N=8 point DFT. We observe that the computation is performed in three stages, beginning with the computation of four two-point DFTs, then two four-point DFTs, and finally, one eight-point DFT. The combination of the smaller DFTs to form the larger DFTs is also clearly illustrated in Fig. 5.3.

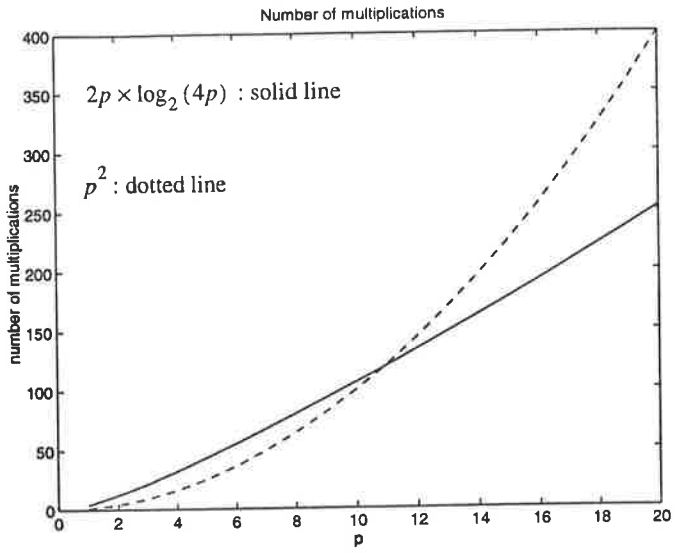

FIGURE 5.2 Number of multiplications per iteration

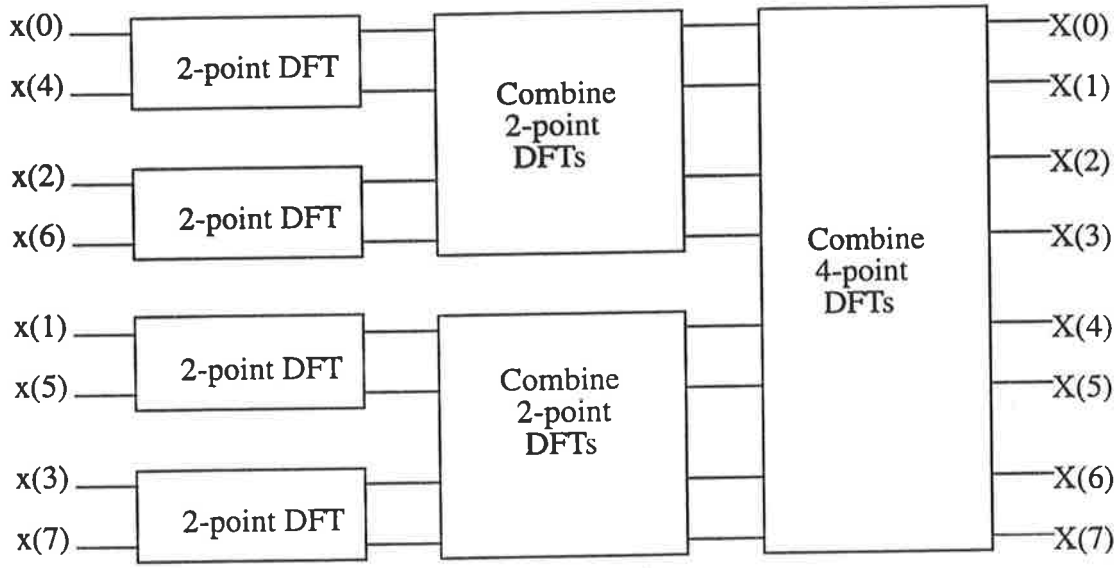

FIGURE 5.3. Three stages in the computation of an 8-point DFT.

The basic computation performed at every stage includes taking two complex numbers, say the pair (a,b), multiplying b by twiddle factor W_L^r, i.e., $W_L^r = e^{-j2\pi r/L}$, and then adding and subtracting the product from a to form two new complex numbers (A,B). This basic computation, depicted in Fig. 5.4, is called a *butterfly computation.* The resulting FFT algorithm is called a *decimation-in-time* FFT algorithm, see Fig. 5.5.

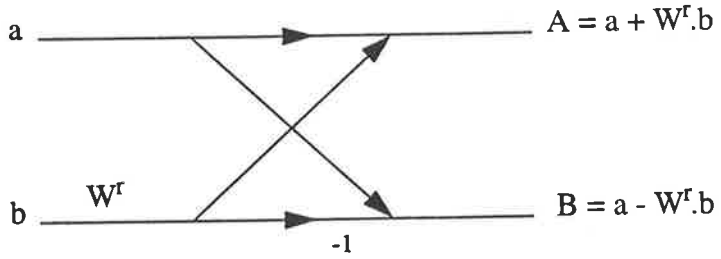

FIGURE 5.4. Basic butterfly computation in the decimation-in-time FFT algorithm.

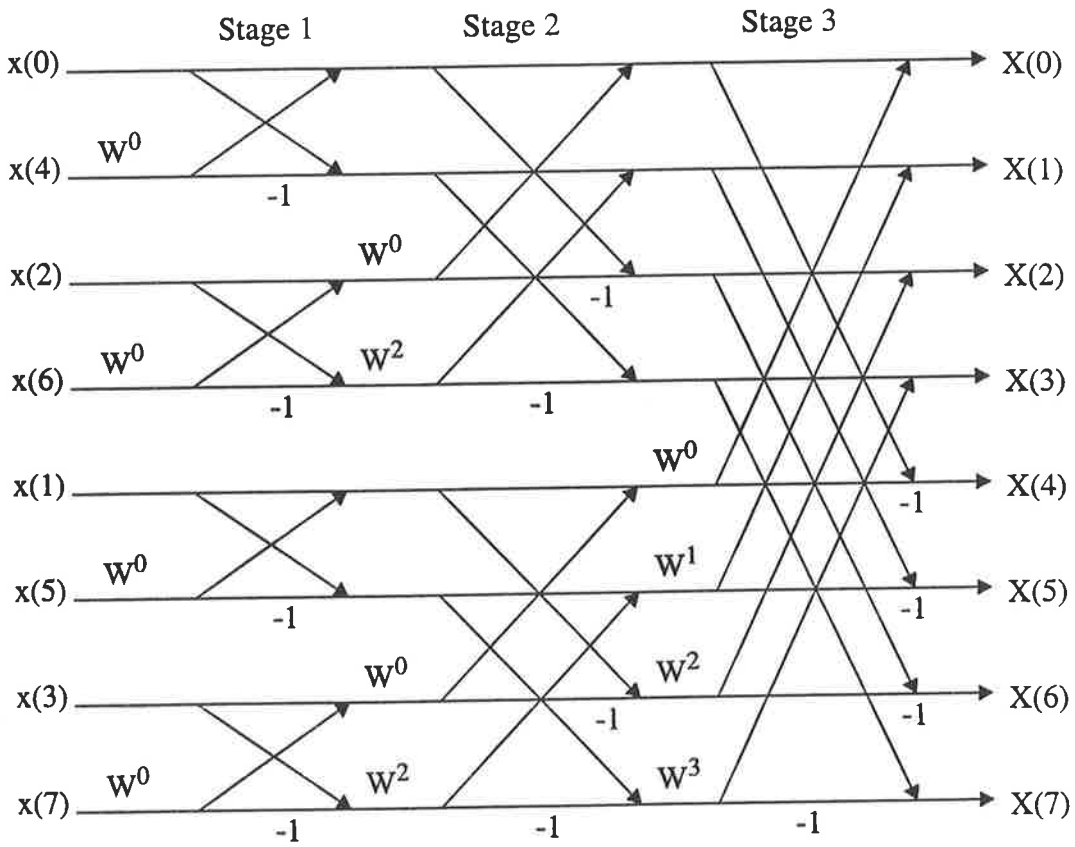

FIGURE 5.5. Eight-point decimation-in-time FFT algorithm.

Another important radix-2 FFT algorithm, called the *decimation-in-frequency* algorithm, is illustrated in Fig. 5.6. We observe that the basic computation in Fig. 5.6 involves the butterfly operation illustrated in Fig. 5.7.

We observe from Fig. 5.6 that a decimation-in-time algorithm has the input occurring in bit-reversed order and the output occurring in natural order. Whereas a decimation-in-frequency algorithm contains the input occurring in natural order and the output occurring in bit-reversed order. Both the decimation-in-time and decimation-in-frequency algorithms require the shuffling of the input and the output data, respectively.

These algorithms are based on the assumption that the lengths of the input and the output sequences of the algorithm are equal and the length is a power of 2.

If the length of the input and the output sequences is not a power of two, extra zeros, to make up for the shortage of the length, must be padded into the sequence in the following manner.

$$\{m(n)\} = \left[m(0)\, m(1) \ldots m(p-1), \overbrace{0 \ldots 0}^{\text{extra padded zeros}}, m(p-1)\, m(p-2) \ldots m(1)\right] \quad (5.10)$$

$$\{u(n)\} = \left[\underbrace{u(1)\, u(2) \ldots u(p), 0 \ldots 0}_{2p\text{-}1}, \overbrace{0 \ldots 0}^{\text{extra padded zeros}}\right] \quad (5.11)$$

The sequences now have a length of 2^{ν}.

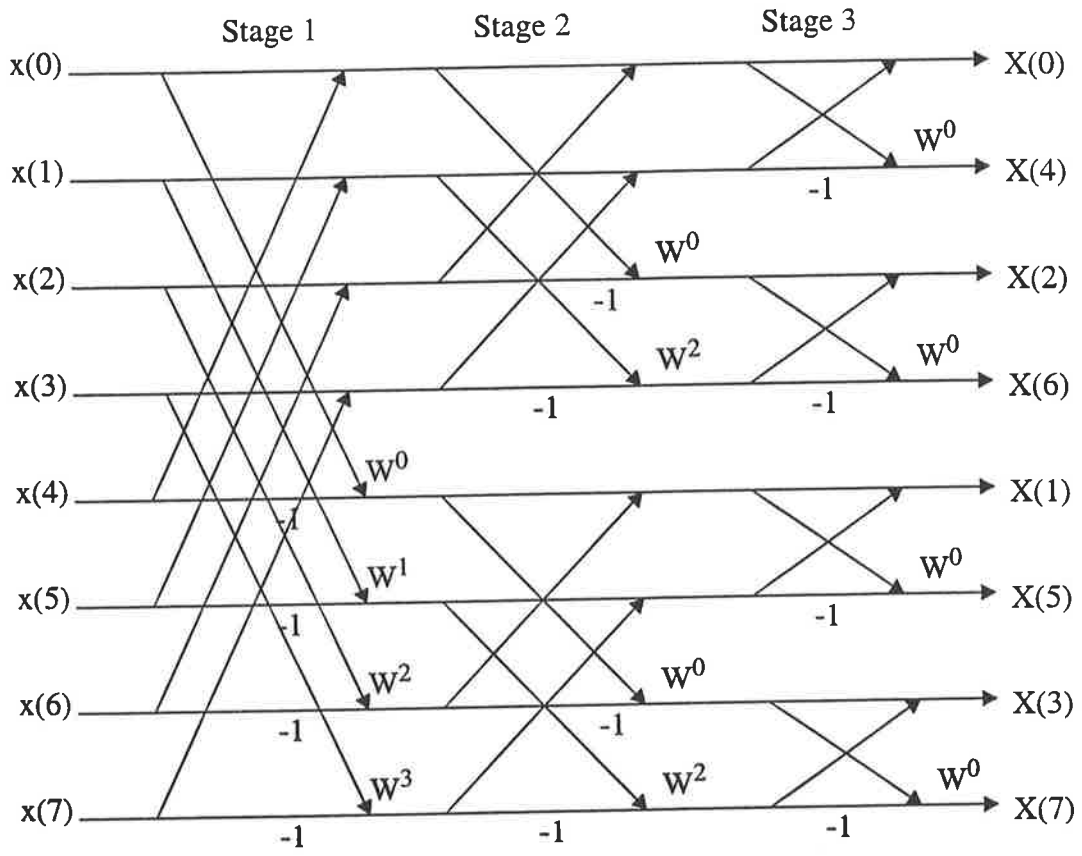

FIGURE 5.6. Eight-point decimation-in-frequency FFT algorithm.

The circular convolution of these two sequences does not violate the matrix multiplication in Eqn. (5.2). For example, if p=6, then the length of the sequence {m(n)} without any extra padded zeros would be $2p$-1 = 11. Since 11 is greater than $2^3 = 8$, L must be chosen to be $2^4 = 16$ and this requires 5 extra zeros. Therefore in the sequence $\{u(n)\}$ we have 16-6=10 padded zeros in total.

$$\{u(n)\} = [u(1)\ u(2)\ u(3)\ \ldots\ u(6)\ \underbrace{0\,0\ldots0\,0}_{10\text{ zeros}}] \tag{5.12}$$

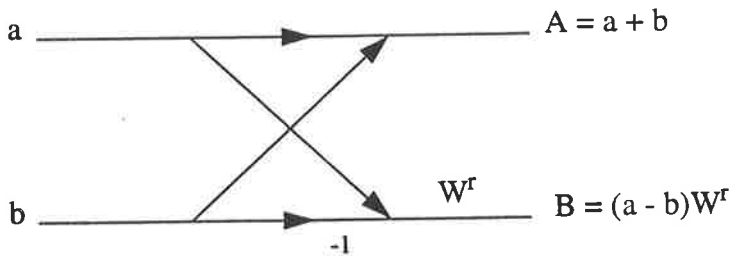

FIGURE 5.7 Basic butterfly computation in the decimation-in-frequency FFT.

Going back to the previous discussion on the recursive equation, by observing the decimation-in-time radix-2 FFT algorithm in Fig. 5.5, we realise that the zeros padded in the sequence $\{u(n)\}$ enable us to skip the first stage and also some of the butterfly computations in the second stage. This is true because, in all cases, the total number of zeros padded in the sequence $\{u(n)\}$ is always greater than half the length of the sequence. Likewise, in the decimation-in-frequency algorithm, every lower part of the butterfly output in the last stage and some of the butterfly computations in the second stage can be skipped since they are to be replaced by zeros immediately after the computations.

When $p = 6$ and $L = 16$ correspondingly, the last ten zeros in the sequence $\{u(n)\}$ simplify the last two stages in the IFFT algorithm and the first two stages in the FFT algorithm. This is illustrated in Fig. 5.8. The number of complex multiplications is reduced by 24 and the total number of multiplication is now 40 instead of 64.

This strategy is called *pruning* with a subset of non-zero data points: first devised by Markel [67] and later improved by Skinner [81], Screenivas and Rao [84], [85], and Nagai [73]. It should be noted that choosing the decimation-in-time algorithm for FFT operation, the decimation-in-frequency algorithm for IFFT operation, and concatenating the IFFT and FFT enable us to complete the first three steps in each iteration of the recursive algorithm together with a reduction in the number of butterfly computations. Furthermore, using this cascaded structure, we can skip the inherent bit reversal steps in FFT and IFFT algorithms alone, leading to further simplification. Recently two very computationally efficient algorithms have been developed:

one is *split-radix* FFT [99] and the other is *transform decomposition* [82], [83] which uses a mixture of a Cooley-Tukey FFT [17] and a computational structure similar to Goertzel's algorithm. The split-radix FFT is very efficient if $2p - 1 \geq \frac{L}{2}$ but since this is not the case here, the saving would not be great. It is documented by Sorensen and Burrus [83] that the transformation decomposition methods are normally better than the pruning methods for any length of non-zero subset in L input points or output points. Incorporating this method is of great interest for further research.

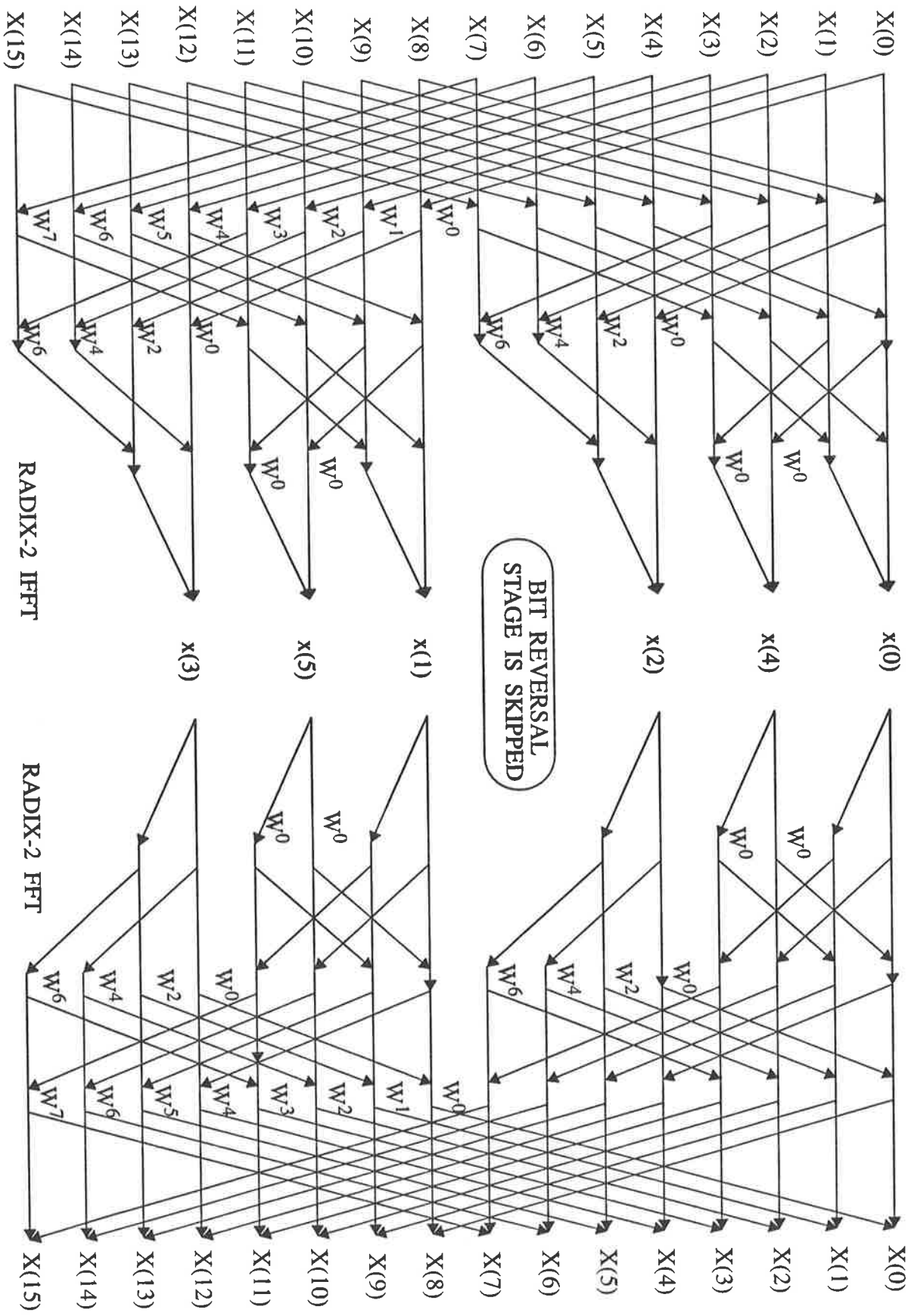

FIGURE 5.8. Concatenation of IFFT and FFT

5.3 Computational complexity

The objective of this section is to analyse the computational complexities of the recursive method with and without the FFT algorithm, and to make comparisons with other conventional methods.

First, the number of FFT points, denoted by L, in a L-point FFT needs to be evaluated. In the previous section, we assumed that $L=2p-1$ and also $L=2^{\nu}$. However this is just a tentative assumption for the sake of simplifying the discussion. It is apparent that $2p-1$ is an odd number, and this can not be equal to 2^{ν}. By taking into account the extra zeros that are padding the sequences $\{m(n)\}$ and $\{u(n)\}$, we can define L as follows:

$$L=2^{\nu} \quad \text{if } 2^{\nu-1} < 2p-1 < 2^{\nu} \tag{5.13}$$

where ν is a positive integer. Then the total number of multiplications per iteration, T, without pruning is

$$T = L \times \log_2(2L) . \tag{5.14}$$

By carefully observing Fig. 5.8, we can realise that $L/2$ multiplications in the last stage of the radix-2 IFFT and the first stage of the radix-2 FFT are skipped; therefore, L multiplications are saved. Also by observing the second last stage of IFFT and the second stage of FFT, we can see that in each stage there is $L-2p$ skipped multiplications, contributing to a total reduction of $2L-4p$ multiplications. In total, we have a saving of $3L-4p$ multiplications. Therefore the total number of multiplications involved in each iteration is

$$T = L \times \log_2(2L) \text{ -}3L+4p. \tag{5.15}$$

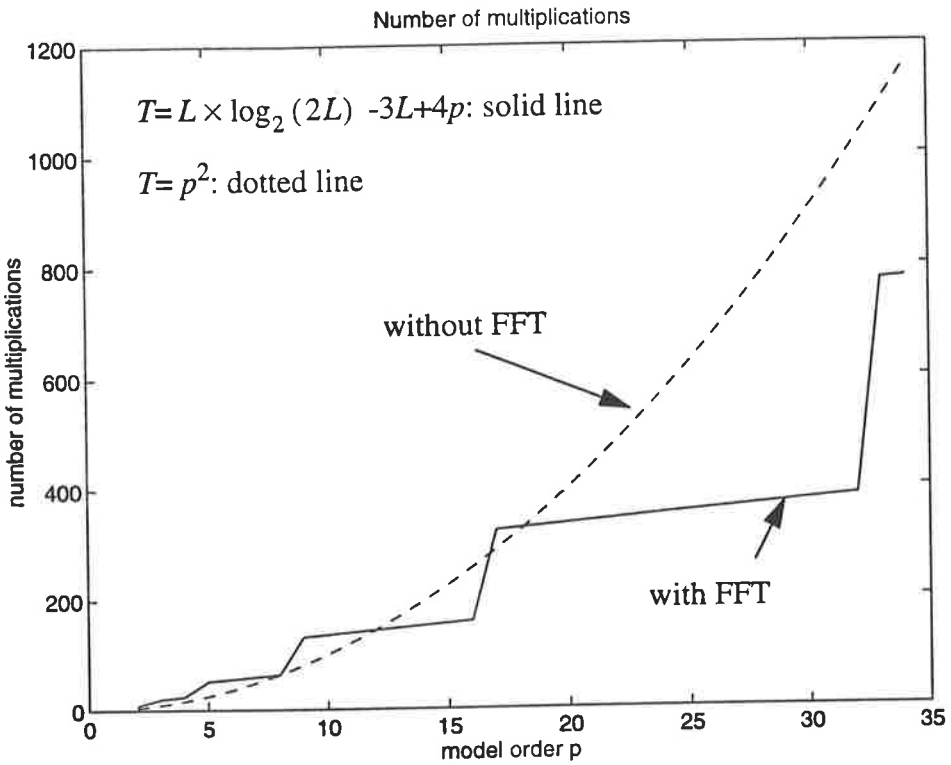

FIGURE 5.9 Number of multiplications per iteration

Figure 5.9 shows both $L \times \log_2(2L)$ -3L+4p and p^2 against p. The dotted graph has a parabolic shape as it is a quadratic function of p, and the solid line is a stair-step function with gradual increase in the step size. For very low p, the number of multiplications without FFT seem to be smaller than with FFT. In fact the recursive equation without FFT has smaller number of multiplications T until p=11 except for the equality when p=8. For $p \geq 12$, the recursive equation without FFT has smaller T except for $16 < p \leq 18$, but the differences are not significant. As p becomes larger and larger, the difference becomes more significant. Since with the recursive method, the good choice of p is $p \geq 12$ as experimentally verified, incorporating the FFT into the recursive equation definitely has an advantage.

Now let us consider a rough comparison with the computational complexity of conventional AR model based PSD estimation methods, [5], [18], [20], [52], [76]. For easy comparison, the number of multiplications is tabulated below.

TABLE 5.1 Computational complexity

Complexity	Recursive	Yule-Walker	Burg	LS
Lag Estimates	Np	Np	NA	$p^2(N-p)$
AR Coefficients	$L\times\log_2(2L)$ -3L+4p*	p^2	Np+p^2	Np+p^2
PSD Estimates	pS	pS	pS	pS

* This is the computational complexity per iteration.

- N = Number of data samples
- p = Order of Model
- L = Length of sequences entering the modified FFT
- S = Number of spectral samples computed
- NA = Not Applicable

It is quite apparent that the other conventional methods have higher computational complexities than the recursive method has. The O(p^2) operations for the Yule-Walker is provided by the Levinson-Durbin recursion [18], [21], [60], [105]. These computations, when performed by p parallel processors, can be accomplished in O($p$log p). The Burg method operates directly on the data; that is, it does not include lag estimates. The O(p^2) operations for the LS method is obtained using an order recursive algorithm developed by Marple [69]; the LS algorithm is almost as computationally efficient as the Burg algorithm requiring typically about 20% more computations than the Burg algorithm does. According to the literature [52], [76], other PSD estimation methods such as the eigenanalysis methods are also subject to high computational complexity. In conclusion, the recursive method has a computational complexity comparable with the conventional AR based PSD estimation techniques.

CHAPTER 6 *Conclusion and Further Studies*

6.1 Conclusion

This thesis has addressed some major issues in the field of PSD estimation: spectral resolution, and computational complexity. A new recursive high-resolution parametric method was devised for PSD estimation. This technique was intended to provide an iterative approach for finding the Yule-Walker solution for PSD estimation. The principal objective of applying this approach to PSD estimation is to obtain a high-resolution PSD estimate, which is free of some inherent limitations imposed on the conventional methods: spurious peaks, spectral line split, sensitivity to initial phase, and bias in estimated frequency.

The formulation of this method traces back to the idea of linear prediction. We showed how the minimization of mean square value of one-step forward prediction error using the steepest descent method led to a simple recursive equation for AR parameter estimation. We discussed the use of a preconditioner matrix and its contribution to improvement in stability. We then related this result to determination of step size parameter; a new parameter called step multiplication factor (smf) was introduced to provide a simple way for generating a value for the step size parameter.

Performance analysis has become a common practice for researchers to provide guide lines or formulae for system design and evaluation. The measures in Chapter 4 reflected

some performance characteristics of the recursive method and also performance comparisons between the recursive and Yule-Walker methods.

We conducted extensive tests using a short length data record comprising single sinusoid corrupted with additive Gaussian noise. The recursive method proved to yield a very high spectral resolution and small bias in estimated frequency. The 3-dB bandwidths obtained by the recursive method were consistently good over a wide SNR range; i.e., 1dB - 30 dB. The recursive method could detect a peak having a frequency value below 0.03 cps. A test was carried out to find the optimum smf values, and to examine the sensitivity of 3-dB bandwidth to smf. The result indicated that the optimum smf was about 0.6 and the sensitivity was moderate. We tested the recursive method and the Yule-Walker method by varying the frequency of sinusoid and by introducing an initial phase. The 3-dB bandwidth versus frequency was symmetric about f=0.25 cps, and the variations of 3-dB bandwidth and frequency bias were periodic. The peak resolving ability of the two methods was examined. The recursive method could resolve peaks closer than $\Delta f = 0.06$ cps while the Yule-Walker method could not. Throughout the various tests, the recursive method held its superiority over the Yule-Walker method.

The recursive method was then implemented using the FFT algorithms; the computational complexity was reduced to a level at least comparable with that of other AR PSD estimation methods.

The significance of this method lies in its high resolution and easy-to-implement structure. Additional merits include no spurious peaks, no spectral line splitting, and small sensitivity to initial phase. This technique can be implemented using p parallel processors to achieve $O(p)$ operations.

6.2 Further Studies

It is clear from the limitations and qualifications concerning the results of this thesis, that there is a need for further work on several issues. A fraction of many ideas that should be pursued to extend the work of this thesis is presented.

- Chapter 4 illustrates the performance comparisons with the Yule-Walker method since it is more related to the recursive method than others are. Since the Yule-Walker method is not outstanding among the conventional AR PSD estimation methods, comparisons with other high-performance PSD estimation methods should be considered.
- Subsection 4.2.3 addresses the selection of optimum smf. More rigorous tests and analysis on selection criteria for optimum smf and optimum model order should be pursued in the hope of establishing more concrete criteria.
- In chapter 5, it is briefly mentioned that the transformation decomposition methods by Sorensen and Burrus [83] might yield greater savings in computational complexity. Further research on this or seeking alternative approaches would be of interest.
- Hardware implementation of the recursive method would be desirable. Development of cost-effective and fast hardware for implementing this method on real data is essential, and exploring its adequacy for real time applications should also be considered.
- To assist the validity of this method, a research to find an analytic explanation to the performance of the recursive method is suggested.

Bibliography

[1] H. Akaike. "Power Spectrum Estimation Through Autoregression Model Fitting." *Ann. Inst. Stat. Math*, vol. 21, pp. 407–419, 1969.

[2] H. Akaike. "A New Look at the Statistical Model Identification." *IEEE Trans. Automatic Control*, vol. AC-19, pp. 716–723, December 1974.

[3] N. O. Andersen. "Comments on the Performance of Maximum Entropy Algorithm." *Proc. IEEE*, vol. AC-19, pp. 1581–1582, November 1978.

[4] B. S. Atal and S. L. Hanauer. "Speech Analysis and synthesis by linear prediction of the speech wave." *J. Acoust. Soc. Amer.*, vol. 50, no. 2, pp. 637–655, 1971.

[5] L. Barrodale and R.E. Erickson. "Algorithms for least squares linear prediction and maximum entropy spectral analysis-Part I: Theory and Part II: FORTRAN Program." *Geophysics*, vol. 45, pp. 420–446, March 1980.

[6] M. S. Bartlett. "Smoothing Periodograms from Time Series with Continuous Spectra." *Nature (London)*, vol. 161, pp. 686–687, May 1948.

[7] J. G. Berryman. "Choice of operator length for maximum entropy spectral analysis." *Geophys.*, vol. 43, pp. 1384–1391, December 1978.

[8] C. Bingham, M. D. Godfrey, and J. W. Tukey. "Modern techniques of power spectrum estimation." *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 56–66, June 1967.

[9] R. B. Blackman and J. W. Tukey. "The Measurement of Power Spectra From the Point of View of Communications Engineering." *New York: Dover*, 1958.

[10] T. Bohlin. "Comparison of two methods of modeling stationary EEG signals." *IBM J. Res. Dev.*, pp. 194–205, May 1973.

[11] A. Bouzerdoum and Tim R. Pattison. "Neural Network for Quadratic Optimazation with Bound Constraints." *IEEE trans. on Neural Networks*, vol. 4, pp. 293–304, March 1993.

[12] P. M. T. Broersen and H. Einar Wensink. "On finite sample theory for autoregressive model order selection." *IEEE Transactions on signal processing*, vol. 41, pp. 194–204, January 1993.

[13] J. P. Burg. "Maximum Entropy Spectral Analysis." *Proc. 37th Meeting Society of Exploration Geophysists*, October 1967.

[14] J. P. Burg. "New concepts in power spectral estimation." *Proc. 40th Annu. Int. Society of Exploration Geophysicists Meeting*, November 1970.

[15] J. P. Burg. *Maximum entropy spectral analysis*. PhD thesis, Dep. Geophysics, Stanford University, Stanford, CA, May 1975.

[16] J. F. Claerbout. *Fundamentals of Geophysical Data Processing*. McGraw-Hill, New York, 1976.

[17] J. W. Cooley and J. W. Tukey. "An algorithm for the machine calculation of complex Fourier Series." *Math. Comput.*, vol. 19, pp. 297–310, April 1965.

[18] 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.

[19] P. T. Daniel. "On the theoretical specification and sampling properties of autocorrelated time-series." *J. R. Stat. Soc., ser. B*, vol. 8, pp. 88–90, 1946.

[20] A. K. Datta. "The complex form of the maximum entropy method for spectral estimation." *Proc. IEEE*, vol. 65, pp. 1219–1220, August 1977.

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

[22] P. B. C. Fenwick, P. Michie, J. Dollimore, and G. W. Fenton. "Mathematical simulation of the electroencephalogram using an autoregressive series." *Bio-Med.*

Comput., vol. 2, pp. 281–307, 1971.

[23] P. F. Fougere, E. J. Zawalic, and H. R. Radoski. "Spontaneous line splitting in maximum entropy power spectrum analysis." *Physics Earth Planetary Interiors*, vol. 12, pp. 210–217, August 1976.

[24] R. F. Fougere. "A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis." *J. Geophysical Res.*, vol. 82, pp. 1051–1054, March 1977.

[25] W. F. Gabriel. "Spectral analysis and adaptive array super-resolutopn technique." *Proc. IEEE*, vol. 68, pp. 654–666, June 1980.

[26] W. A. Gardner. *Statistical Spectral Analysis: A Non-Probabilistic Theory*. Prentice-Hall, Inc., Englewood Cliffs. N. J., 1987.

[27] W. Gersch. "Spectral anaysis of EEG's by autoregressive decomposition of time series." *Math. Biosci*, vol. 7, pp. 205–222, 1970.

[28] W. Gersch and D. R. Sharpe. "Estimation of power spectra with finite-order autoregressive models." *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 367–369, August 1973.

[29] Peter T. Gough. "A Fast Spectral Estimation Algorithm Based on the FFT." *IEEE Trans. on Signal Processing*, vol. 42, pp. 1317–1422, June 1994.

[30] P. M. Grant. "Digital Signal Processing Part 1: Digital filters and DFT." *Electronincs & Commnication Engineering Journal*, February 1993.

[31] P. M. Grant and J. T. E. McDonnel. "Digital Signal Processing Part 2: Spectral Analysis." *Electronics & Commnication Engineering Journal*, pp. 211–217, August 1993.

[32] A. H. Gray, Jr, and D. Y. Wong. "The Burg algorithm for LPC speech analysis/synthesis." *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-28, pp. 609–615, December 1980.

[33] L. J. Griffiths and R. Prieto-Diaz. "Spectral analysis of natural seismic events using autoregressive techniques." *IEEE Tran. Geosci. Electron.*, vol. GE-15, pp. 13–25, January 1977.

[34] Fredric J. Harris. "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transformation." *Proc. of IEEE*, vol. 66, pp. 172–204, January 1978.

[35] J. R. Haskew, J. M. Kelly, R. M. Kelly, Jr., and T. H. Mckinney. "Results of a study of the linear prediction vocoder." *IEEE Trans. Commun.*, vol. COM-21, 1973.

[36] S. S. Haykin. *Nonlinear Methods of Spectral Anaysis*. Springer-Verlag, New York, 1979.

[37] S. Haykin. *Adptive Filter Theory*. Prentice-Hall Information And System Sciences Series. Prentice-Hall A Division of Simon & Schuster, Inc., Englewood Cliffs, New Jersey 07632, 1986.

[38] R. W. Herring. "The cause of line splitting in Burg maximum entropy specral analysis." *IEEE Trans. Acoustics, Speech, Sinal Process.*, vol. ASSP-28, pp. 692–701, December 1980.

[39] S. Holm and J. M. Hovem. "Estimation of scalar ocean wave spectra by the maximum entropy method." *IEEE J. Ocean. Eng.*, vol. OE-4, pp. 76–83, July 1979.

[40] F. M. Hsu and A. A. Giordano. "Line tracking using autoregressive spectral esimates." *IEEE Trans. Acoustics, Speech, Sinal Process.*, vol. ASSP-25, pp. 510–519, December 1977.

[41] M. Huzzii. "On a spectral estimate obtained by an autoregressive model fitting." *Ann. Inst. Statist. Math.*, vol. 29, pp. 415–431, 1977.

[42] F. Itakura and S. Saito. "A statistical method for estimation of speech spectral density and formant frequecies." *Electron. Commun. Japan*, vol. 53-A, no. 1, pp. 36–43, 1970.

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

[44] S. J. Johnsen and N. Andersen. "On power estimation in maximum entropy spectral analysis." *Geophsy.*, vol. 43, pp. 681–690, June 1978.

[45] R. H. Jones. "Identification and autoregressive spectrum estimation." *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 894–898, December 1974.

[46] R. H. Jones. "Autoregression order selection." *Geophys.*, vol. 41, pp. 771–773, Au-

gust 1976.

[47] R. P. Kane. "Maximum entropy analysis of some artificial samples." *J. Geophys. Res.*, vol. 84, pp. 965–966, March 1979.

[48] R. L. Kashyap. "Inconsitency of the AIC rule for estimating order of autoregressive models." *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 996–998, October 1980.

[49] M. Kaveh and G. R. Cooper. "An empirical investigation of the properties of the autoregressive spectral estimator." *IEEE Trans. Infoma. Theory*, vol. IT-22, pp. 313–323, May 1976.

[50] S. M. Kay. "The effects of noise on the autoregressive spectral estimator." *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, pp. 478–485, October 1979.

[51] S. M. Kay and S. L. Marple Jr. "Sources of remedies for spectral line splitting in autoregressive spectrum analysis." *Proc. 1979 ICASSP*, pp. 151–154.

[52] S. M. Kay and S.L. Marple Jr. "Spectrum Analysis-A Modern Perspective." *Proc. of IEEE*, vol. 69, pp. 3–42, November 1981.

[53] S. B. Kesler and S. S. Haykin. "The maximum entropy method applied to the specral anaysis of radar clutter." *IEEE Trans. Informa. Theory*, vol. IT-24, pp. 269–272, March 1978.

[54] S. B. Kesler. "Maximum entropy estimation of radar clutter spectra." *Natl. telecommunications conf. Rec. (Birmingham, AL)*, pp. 18.5.1–18.5.5, December 3-6 1978.

[55] H. Kobatake. "Linear predictive coding of speech signals in a high ambient noise environment." *Conf. Rec 1978 ICASSP*, pp. 472–475.

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

[57] F. Kozin and F. Nakajima. "The order determination problem for linear time- varying AR models." *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 250–257, April 1980.

[58] R. T. Lacoss. "Data adaptive spectral analysis methods." *Geophysics*, vol. 36, pp. 661–675, August 1971.

[59] T. E. Landers and R. T. Lacoss. "Some geophysical applications of autoregressive spectral estimates." *IEEE Trans. Geosci. Electron.*, vol. GE-15, pp. 26–32, January 1977.

[60] N. Levinson. "The Wiener error criterion in filter design and prediction." *J. Math. Phys.*, vol. 25, pp. 262–278, 1947.

[61] J. S. Lim. "All pole modelling of degraded speech." *IEEE Trans. Acoust., Speech, Signal Precess.*, vol. ASSP-29, pp. 197–209, June 1978.

[62] Ja Ling. "Basis-Vector-Decomposition Based Two-Stage Computational Algorithms for DFT and DHT." *IEEE Trans. on Signal Processing*, vol. 41, pp. 1562–1575, April 1993.

[63] J. Makhoul. "Spectral analysis of speech by linear prediction." *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 140–148, June 1973.

[64] J. Makhoul. "Linear Prediction: A tutorial review." *Proc. of IEEE*, vol. 63, pp. 561–580, April 1975.

[65] J. Makhoul. "Spectral Linear Prediction: Properties and Applications." *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-23, pp. 283–296, June 1975.

[66] J. Makhoul. "Stable and Efficient Lattice Methods for Linear Prediction." *IEEE trans. on ASSP*, vol. 25, pp. 423–428, October 77.

[67] J. D. Markel. "FFT pruning." *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 305–311, December 1971.

[68] S. L. Marple. "Frequency resolution of high-resolution spectrum analysis technique." *Proc. 1978 RADC Spectrum Estimation Workshop*, pp. 19–35.

[69] S. L. Marple. "A New Autoregressive Spectrum Analysis Algorithm." *IEEE Trans. on ASSP*, vol. 28, pp. 431–454, August 1980.

[70] S. L. Marple-Jr. *Digital Spectral Analysis with Applications*. Prentice-Hall Signal Procesing Series. Prentice-Hall, Inc, Englewood Cliffs, New Jersey 07632, 1987.

[71] R. N. Mcdonough. "Maximum entropy spatial processing of array data." *Geophysics*, vol. 39, pp. 843–851, December 1974.

[72] W. Murray. *Numerical Methods for Unconstrained Optimization*. Academic Press,

New York, 1972.

[73] K. Nagai. "Pruning the decimation-in-time FFT algorithm with frequency shift." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1008–1010, August 1986.

[74] E. Parzen. "Some recent advances in time series modelling." *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 723–730, 1974.

[75] M. B. Priestley. *Spectral Anaysis and Time Series*, volume 1 and 2. Academic press, New York, 1981.

[76] John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing : Principles, Algorithms, and Applications*. Macmillan Publishing Company & Maxwell Macmillan Canada, Inc, 866 Third Avenue, New York, New York 10022, 1200 Eglinton Avenue East Suite 200 Don MIlls, Ontario M3C 3N1, 2nd edition, 1988.

[77] J. Rissanen. "A Universal Prior for the Integers and Estimation by Minimum Description Length." *Ann. Stat.*, vol. 11, pp. 417–431, 1983.

[78] E.A. Robinson. "Historical Perspective of spectral estimation." *Proc. IEEE*, pp. 885–907, September 1982.

[79] H. Sakai. "Statistical properties of AR spectral analysis." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 402–409, August 1979.

[80] S. Shon and K. Mehrotra. "Performance comparison of autoregressive estimation methods." *Proceedings of the 1984 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 14.3.1–14.3.4, 1984.

[81] D. P. Skinner. "Pruning the decimation-in-time FFT algorithm." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 193–194, 1976 1976.

[82] H. V. Sorensen and C. S. Burrus. "A new efficient algorithm for computing a few DFT points." *Proc. IEEE 1988 Int. Symp. Circuits Syst.*, pp. 1915–1918, June 7-9 1988.

[83] Henrik V. Sorersen and C. Sidney Burrus. "Efficient Computation of the DFT with Only a Subset of Input or Output Points." *IEEE trans. on Signal Processing*, vol. 41, pp. 1184–1199, March 1993.

[84] T. V. Sreenivas and P. V. S. Rao. "FFT algorithm for both input and output pruning." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 291–292, June 1979.

[85] T. V. Sreenivas and P. V. S. Rao. "High Resolution narrow-band spectra by FFT pruning." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 254–257, April 1980.

[86] G. Strang. *Linear Algebra and Its Applications*. Academic Press, New York, 2nd edition edition, 1980.

[87] D. N. Swingler. "A comparison between Burg's maximum entropy method and a nonrecursive method technique for the spectral analysis of deterministic signals." *J. Geophysical Res.*, vol. 84, pp. 679–685, February 1979.

[88] D. N. Swingler. "A modified Burg algorithm for maximum entropy spectral analysis." *Proc. IEEE*, vol. 67, pp. 1368–1369, September 1979.

[89] D. J. Thomson, M. F. Robbins, C. G. Maclennan, and L. J. Lanzerotti. "Spectral and windowing techniques in power spectral analysis of geomagnetic data." *Phys. Earth Planetary interiors*, vol. 12, pp. 217–231, August 1976.

[90] T. Thorvaldsen. "Comparison of the Least Squares method and the Burg method for autoregressive spectral analysis." *IEEE Trans. Antennas Propag.*, vol. AP-29, pp. 675–679, July 198 1.

[91] T. Thorvaldsen, A. T. Waterman-Jr., and R. W. Lee. "Maximum entropy angular response patterns of microwave transhorizon signals." *IEEE Trans. Antennas Propagat.*, vol. AP-28, pp. 722–724, September 1980.

[92] H. Tong. "Autoregressive model fitting with noisy data by Akaikes Information Criterion." *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 476–480, JUly 1975.

[93] H. Tong. "More on Autoregressive model fitting with noisy data by Akaikes Information Criterion." *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 409–410, May 1976.

[94] H. Tong. "Final prediction error and final interpolation error: A paradox?" *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 758–759, November 1979.

[95] T. J. Ulrych. "Maximum entropy power spectrum of long period geomagnetic reversals." *Nature*, vol. 235, pp. 218–219, January 1972.

[96] T. J. Ulrych. "Maximum entropy power spectrum of truncated sinusoids." *J. Geophyical Reserch*, vol. 77, pp. 1396–1400, March 1972.

[97] T. J. Ulrych and T. N. Bishop. "Maximum entropy spectral analysis and autoregressive decomposition." *Rev. Geophysics Space Phys.*, vol. 13, pp. 183–200, February 1975.

[98] T. J. Ulrych and R. W. Clayton. "Time series modelling and maximum entropy." *Phys. Earth Planetary Interiors*, vol. 12, pp. 188–200, August 1976.

[99] M. Vetterli. "Split-Radix Algorithms for Length-p^m DFT's." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 57–64, January 1989.

[100] G. Walker. "On Periocity in Series of Related Terms." *Proc. Roy. Soc. London, Series A*, vol. 131, pp. 518–532, 1931.

[101] P. D. Welch. "The use of fast Fourier transformation for the estimation of power spectra: A model based on time averaging over short, modified periodograms." *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70–73, June 1967.

[102] A. Wennberg and L. H. Zetterberg. "Application of a computer based model for EEG analysis." *Electroencephalogr. Clin. Neurophys.*, vol. 31, no. 5, pp. 457–468, 1971.

[103] S. J. Wernecke and L. R. D'Addario. "Maximum entropy image reconstruction." *IEEE Trans. comput.*, vol. C-26, pp. 351–364, April 1977.

[104] S. J. Wernecke. "Two-dimensional maximum entropy reconstruction of radio brightness." *Radio Sci.*, vol. 12, pp. 831–844, September-October 1977.

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

[106] G.U. Yule. "On a method of Investigating Periodicities in Disturbed Series with Special References to Wolfer's Sunspot Numbers." *Philos. Trans. R. Soc. London, Series A*, vol. 226, pp. 267–298, July 1927.