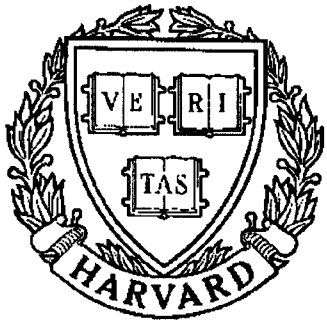
THESIS REPORT
Ph.D.

SYSTEMS
RESEARCH
CENTER



Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Multiple Frequency Estimation in Mixed-Spectrum Time Series by Parametric Filtering

by T-H. Li
Advisor: B. Kedem

Ph.D. 92-7

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1992		**00-00-1992 to 00-00-1992**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multiple Frequency Estimaiton in Mixed-Spectrum Time Series by Parametric Filtering	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,The Graduate School,2123 Lee Building,College Park,MD,20742**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**155**	

Abstract

Title of Dissertation: MULTIPLE FREQUENCY ESTIMATION IN MIXED-SPECTRUM TIME SERIES BY PARAMETRIC FILTERING

Ta-Hsin Li, Doctor of Philosophy, 1992

Dissertation directed by: Dr. Benjamin Kedem
Mathematics Department and
Systems Research Center

A general parametric filtering procedure (the PF method) is proposed for the problem of multiple frequency estimation in mixed-spectrum times series (i.e., superimposed sinusoids in additive noise). The method is based on the fact that a sum of sinusoids satisfies an homogeneous autoregressive (AR) equation. The gist of the method is to parametrize a linear filter so that it possesses a certain parametrization property as suggested by the particular form of the bias encountered by Prony's (least squares) estimator. For any parametric filter with this property, in addition to some mild regularity conditions, the least squares estimator from the filtered data, as a function of the filter parameter, constitutes a contractive mapping — whose multivariate fixed-point serves as a consistent AR estimator. The chronic bias of Prony's estimator is thus eliminated. Coupled with the all-pole (AR) filter endowed with an extra bandwidth parameter, the PF method can achieve the accuracy of nonlinear least squares

by a simple iterative procedure consisting of linear least squares estimation followed by linear recursive filtering. Crude initial guesses such as those from Prony's estimator are sufficient to initiate the iteration. The method is also capable of resolving closely-spaced frequencies which are unresolvable by periodogram analysis or DFT.

To analyze the statistical properties of the PF method, some classical asymptotic results concerning the sample autocovariances are extended to accommodate mixed-spectrum time series and parametric filtering. In particular, under regularity conditions, uniform strong consistency and asymptotic normality are proved for the sample autocovariances of a mixed-spectrum time series after parametric filtering. Equipped with these results, some statistical properties of the PF method itself are investigated. These include the existence of the PF estimator as a fixed-point of the parametric least squares mapping, the convergence of an iterative algorithm that calculates the PF estimator, as well as the strong consistency and asymptotic normality of the PF estimator.

Computer simulations are also presented to demonstrate the effectiveness of the PF method. Directions for future research are briefly discussed at the end of the dissertation.

MULTIPLE FREQUENCY ESTIMATION IN MIXED-SPECTRUM TIME SERIES BY PARAMETRIC FILTERING

by

Ta-Hsin Li

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1992

Advisory Committee:

Dr. Benjamin Kedem, Chairman/Advisor
Dr. Eric V. Slud
Dr. Piotr W. Mikulski
Dr. Harland M. Glaz
Dr. Prakash Narayan

Dedication

To the memory of my grandmother

Acknowledgements

I would like to express my sincere gratitude and appreciation to my advisor, Professor Benjamin Kedem, for his encouragement and guidance. His rigorous scholarship and optimistic attitude toward life influenced and stimulated me throughout my studies at the University of Maryland.

My gratitude also extends to Professors Piotr Mikulski, Eric Slud, C. Z. Wei, Ivo Babuska, and Prakash Narayan. From them I learned many things about statistics, time series, numerical analysis, and signal processing, which not only have contributed to this dissertation, but also shall benefit the rest of my career. I would also like to thank Professor Harland Glaz for his review of this dissertation. Special thanks are due to Professor Sid Yakowitz of the University of Arizona for his stimulating lectures which helped the development of this research.

I am very grateful to my wife, Zeping, my parents, and my family. Without their constant love and unlimited support, this dissertation would have been impossible.

I am also indebted to Professor Xuelong Zhu and other friends of Tsinghua University. Their support and assistance made my studying in the United States possible.

Thanks are also due to Silvia Lopes, John Barnnet, and Felix Santos for their helpful comments and suggestions.

This research was supported by grants AFOSR-89-0049, ONR N00014-89-J-1051, and NSF CDR-88-03012.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The estimation of frequency from superimposed noisy sinusoids is an important subject in time series analysis and signal processing. It has applications in various disciplines of science and technology, such as geophysics, seismology, meteorology, rotating machinery, radar, sonar, and communications. The survey paper by Kay and Marple (1981) serves as an excellent reference of modern approaches to this problem.

1.1 Overview

The problem of frequency estimation has attracted a great deal of attention in the statistics and engineering literature, and is still of prime concern at the present time. Traditional approaches to this problem are based on the Fourier transform. A typical example is periodogram analysis (Whittle, 1952). This method is able to provide satisfactory results in many cases. In particular, when the frequencies of the sinusoids are well separated, periodogram analysis, which locates the local maxima in the periodogram as a continuous function of a frequency variable, is capable of producing consistent frequency estimates with asymptotic standard deviation of order $n^{-3/2}$ (Walker, 1971), where n is the data length. Despite its high accuracy, periodogram analysis is computationally intensive, since it requires not only an iterative routine to obtain the optimal frequency estimates, but also a certain exhaustive search on a local

mesh in order to provide initial guesses of accuracy $o(n^{-1})$, needed for the iterative routine to converge to the optimal solution (Rice and Rosenblatt, 1988). Furthermore, in order to achieve consistency, the sinusoidal frequencies are required to be separated from each other by a distance greater than $O(n^{-1})$ (Walker 1971). In other words, periodogram analysis is unable to resolve two frequencies closer than the reciprocal of the data length. This resolution limit happens to be the most serious problem of the Fourier approach (Kay and Marple, 1981).

An alternative way to avoid the nonlinear optimization in periodogram analysis is to evaluate the periodogram only at a finite number of locations known as Fourier frequencies. The resulting procedure, which we refer to as the DFT method [1], is equivalent to transforming the data by discrete Fourier transform and seeking the Fourier frequencies that correspond to the largest absolute magnitudes of the DFT. The DFT method is computationally efficient, as compared to periodogram analysis, since FFT (fast Fourier transform) algorithms can be employed to compute the transform. However, this reduction in computational complexity is not achieved without any cost. In fact, the estimation accuracy of the DFT method deteriorates from $n^{-3/2}$ to n^{-1}, due to the focus on Fourier frequencies. The resolution limit clearly persists in DFT.

The so-called nonlinear least squares (NLS) method (or MLE in the case of white Gaussian noise) is a procedure that has a higher resolution than periodogram analysis. This procedure fits a sum of sinusoids to the data by minimizing the sum of squared errors with respect to the amplitudes, phases, as well as frequencies of the sinusoids. It turns out that the NLS method not only has a higher resolution than periodogram analysis, but also achieves similar estimation accuracy (Hannan and Quinn, 1989). Unfortunately, the NLS method suffers from the same, maybe more, problems of high computational complexity as periodogram analysis. Indeed, the NLS method also requires a certain iterative routine that must be started with initial guesses of

[1] It is sometimes referred to as periodogram analysis in the statistics literature.

accuracy $o(n^{-1})$, in order to obtain the optimal solution (Rice and Rosenblatt, 1988; Hannan and Quinn, 1989).

The trade-off between high accuracy/resolution and low computational complexity has been recognized in the literature of frequency estimation for many years. In applications, as long as the computational burden is not a major problem, satisfactory frequency estimates with high accuracy and high resolution can almost always be obtained by certain nonlinear procedures such as the NLS method. This requires some rough initial estimation via, for example, the DFT method, and an exhaustive search over a fine grid around the rough estimates, followed by an iterative algorithm starting with the refined estimates (Abatzoglou, 1985; Stoica, Moses, Friedlander, and Söderström, 1989). The number of frequencies q can also be determined by a certain goodness-of-fit test, comparing the fitting errors for different values of q. However, when "real-time" algorithms are required for the frequency estimation, these computationally burdensome procedures are obviously out of the question. In these cases, one has to rely on simple algorithms that can be implemented easily. Therefore, the ultimate objective of frequency estimation in these applications is to *seek high accuracy and high resolution procedures that require low computational burden.*

Most of these so-called modern approaches can be roughly divided into two broad categories. In the first category, procedures are based upon the fact that the sinusoidal signal satisfies a homogeneous autoregressive (AR) equation whose coefficients are uniquely determined by the frequencies of the sinusoids. Using this fact, one is able to transform the frequency estimation problem into an AR estimation problem to which many computationally simple linear methods can be applied. Unlike the Fourier approach in which the data are implicitly assumed to be zero outside the observation interval — an unrealistic assumption responsible for the resolution limit of the Fourier approach (Kay and Marple, 1981) — the modern procedures extrapolate the data beyond the observation interval by fitting parametric models to the measured data, and

thus provide higher resolution than the Fourier approach does. A widely-used procedure in this category is Prony's (spectral line) estimator (Hildebrand, 1956; Kay and Marple, 1981) that fits an AR model to the noisy data by the least squares technique, yielding a system of linear equations for the AR coefficients. In addition to providing higher resolution, many procedures in this category can also be implemented recursively, so that the frequency estimates can be easily updated when new data become available. This property is extremely attractive in applications where instantaneous tracking of time-varying frequencies is required, rather than batch processing.

The other category consists of those procedures that are based upon the eigenvalue decomposition of certain data matrices, of which the nonzero eigenvalues or the associated eigenvectors determine the sinusoidal frequencies. To obtain the decomposition, the singular-value decomposition (SVD) algorithms are usually employed, and in some procedures the noisy data matrices are also cleaned up by annihilating small eigenvalues in the decomposition. Details about these procedures can be found, for example, in a book by Kay (1988) and a survey paper by Li (1991).

In this dissertation, we shall concentrate on the procedures in the first category by following the ideas of AR modeling. In particular, we shall propose a general approach of parametric filtering, which we refer to as PF method, that improves Prony's estimator by eliminating its bias and increasing its accuracy, while inheriting its high resolution property and computational simplicity. In the statistics and engineering literature, the idea of parametric filtering has been applied to the problem of frequency estimation (see, for example, Kay, 1984; Kedem, 1986; Dragošević and Stanković, 1989; He and Kedem, 1989; Truong-Van, 1990; Yakowitz, 1991; Quinn and Fernandes, 1991). In this dissertation, all these methods are unified under the framework of AR estimation and extended to provide consistent and efficient frequency estimates which require simple computations.

This dissertation is organized as follows. In the remaining part of Chapter 1, we

shall formulate more precisely the frequency estimation problem and review in detail some of the existing procedures, such as the NLS method mentioned earlier, that are closely related to the development and evaluation of the PF method. In Chapter 2, we shall summarize the AR approach and discuss its asymptotic properties. As we shall see, the asymptotic bias inherent in Prony's estimator motivates the development of the PF method. Chapter 3 provides some new limit theorems, including (i) the strong uniform consistency of sample autocovariances after parametric filtering, and (ii) the asymptotic normality of sample autocovariances. These results extend in different aspects the well-known asymptotic theory of sample autocovariances for continuous-spectrum processes, and lay the theoretical foundation for the statistical analysis of the PF method in a later chapter. In Chapter 4, we shall present the PF method and discuss its statistical properties, concerning (i) the existence of the PF estimator, (ii) the convergence of an iterative algorithm that computes the PF estimator, (iii) the strong consistency, and (iv) the asymptotic normality of the PF estimator. This analysis points to the fact that the PF method is a highly effective procedure for frequency estimation. In Chapter 5, we shall specialize the PF method by considering a useful parametric filter, which we refer to as the AR filter (also known as all-pole filter). Variations of this filter have been considered before in the engineering literature (Mataušek, *et al.* 1983; Kay, 1984; Dragošević and Stanković, 1989), but no statistical analysis has been done, especially for multiple sinusoids. In this work, we apply the general principle of the PF method in connection with the AR filter and investigate statistical properties of the resulting frequency estimates. We shall show that significant improvements over the existing methods using similar filters can be achieved by the PF estimator in terms of the sensitivity to initial guesses, estimation accuracy, and resolution, especially for closely-spaced frequencies. Some simulation results are provided at the end of this chapter to demonstrate the effectiveness of the PF method.

1.2 Problem Formulation

The problem of frequency estimation is very well formulated in the literature. Suppose that a time series $\{y_1, \ldots, y_n\}$ of length n is observed from a random process $\{y_t\}$ which consists of q superimposed real sinusoids $\{x_t\}$ contaminated by additive noise $\{\epsilon_t\}$, namely,

$$y_t = x_t + \epsilon_t \quad \text{and} \quad x_t = \sum_{k=1}^{q} \beta_k \cos(\omega_k t + \phi_k) \qquad (t = 0, \pm 1, \pm 2, \ldots). \tag{1.1}$$

Assume in this expression that the number of sinusoids $q > 0$ is a known integer and that the amplitudes β_k and the frequencies ω_k are unknown constants, satisfying

$$\beta_k > 0 \quad \text{and} \quad 0 < \omega_1 < \cdots < \omega_q < \pi.$$

For convenience, we assume that the phases ϕ_k are independent and identically distributed (i.i.d.) random variables with uniform distribution on the interval $[0, 2\pi)$. In some literature (e.g., Walker 1971) the phases are also assumed to be constants instead of random variables. We find it convenient to assume that the phases are i.i.d. and uniform random variables, and note that the asymptotic theory is not altered by this assumption. The signal $\{x_t\}$ under this assumption becomes a zero-mean strictly stationary process (Grenander and Rosenblatt, 1957, p. 30).

The noise $\{\epsilon_t\}$ is assumed to be independent of $\{\phi_k\}$ and hence of the signal $\{x_t\}$. Moreover, in this dissertation, $\{\epsilon_t\}$ is modeled as a linear process of the form

$$\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \, \xi_{t-j}, \qquad \{\xi_t\} \sim \mathrm{IID}(0, \sigma_\xi^2), \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty. \tag{1.2}$$

In some literature (e.g., Hannan, 1973; Quinn and Fernandes, 1991), $\{\xi_t\}$ is assumed to be a martingale difference sequence. The i.i.d. assumption is made here only for simplicity, so that $\{\epsilon_t\}$ is strictly stationary with continuous spectrum.

It is easy to verify that under the above assumptions about the signal and the noise, the process $\{y_t\}$ becomes strictly stationary with mean zero and autocovariance

function

$$r_\tau^y := E(y_{t+\tau}y_t) = r_\tau^x + r_\tau^\epsilon \qquad (\tau = 0, \pm 1, \pm 2, \ldots) \tag{1.3}$$

where $r_\tau^x := E(x_{t+\tau}x_t)$ and $r_\tau^\epsilon := E(\epsilon_{t+\tau}\epsilon_t)$ are the autocovariance functions of the signal and the noise, respectively, and can be written as

$$r_\tau^x = \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \cos(\omega_k \tau) \quad \text{and} \quad r_\tau^\epsilon = \sigma_\xi^2 \sum_{j=-\infty}^{\infty} \psi_{j+\tau}\psi_j. \tag{1.4}$$

These expressions will be used frequently in this dissertation.

The objective of frequency estimation is to find estimators of the sinusoidal frequencies ω_k on the basis of the data set $\{y_1, \ldots, y_n\}$.

Notice that the estimation of the amplitudes and phases (when fixed) of the sinusoids is also desirable in some applications. Clearly, estimating these quantities is relatively easier as compared to the frequency estimation. In fact, by representing each sinusoid as a linear combination of both cosine and sine waves with the same frequency, the estimation of its coefficients becomes a linear problem, provided that the frequency estimates are available (see, e.g., Bresler and Markovski, 1986). For this reason, we shall only concentrate on the frequency estimation problem in this work.

As remarked earlier, the number of frequency q is assumed to be known *a priori*. When it is unknown, there are several methods available in the literature that can be used to estimate this number. Some of these methods are based on AIC-like criteria and others on the eigenvalue decomposition of covariance matrix of the data. Details concerning this matter can be found, for example, in Kay (1988) and Fuchs (1988).

1.3 Periodogram Analysis and DFT

As briefly mentioned in Section 1.1, one of the traditional Fourier-transform-based procedures is *periodogram analysis*, proposed by Whittle (1952) as an approximation to the nonlinear least squares method (Walker, 1971).

The *periodogram* of the time series $\{y_1, \ldots, y_n\}$ is defined by

$$P_n(\omega) := \frac{1}{n}\left|\sum_{t=1}^{n} y_t \exp(-it\omega)\right|^2 \tag{1.5}$$

as a continuous function of the frequency variable $\omega \in [0, \pi]$. Periodogram analysis is a method of frequency estimation that seeks q extremum points $0 < \hat{\omega}_1 < \cdots < \hat{\omega}_q < \pi$ that correspond to the largest values in the periodogram $P_n(\omega)$. This can be done by maximizing the sum

$$S_n := \sum_{k=1}^{q} P_n(\omega_k) \tag{1.6}$$

with respect to $\{\omega_k\}$. Clearly, if no other restrictions are imposed on this problem, the q maxima would cluster at the global maximum of $P_n(\omega)$, yielding incorrect frequency estimates. To remedy this problem, Walker (1971) introduced the following *separation condition* [2]

$$\begin{cases} \omega_k - \omega_{k'} \geq c_n/n \quad \forall\ 1 \leq k' < k \leq q, \\ c_n/n \leq \omega_k \leq \pi - c_n/n \quad \forall\ k, \quad \text{where } c_n \to \infty,\ c_n/n \to 0. \end{cases} \tag{1.7}$$

This condition simply says that the sinusoidal frequencies must be separated from each other and from $\{0, \pi\}$ by a distance greater than $O(n^{-1})$. In particular, when two frequencies are closer than n^{-1}, the separation condition is clearly violated, and hence the estimation accuracy by periodogram analysis is no longer guaranteed. This is known as the *resolution limit* of periodogram analysis.

Large sample properties of periodogram analysis have been investigated by Walker (1971, 1973) and Hannan (1973). The following theorem summarizes their results regarding the frequency estimates.

Theorem 1.1 *Let* $\{y_1, \ldots, y_n\}$ *be a time series observed from* (1.1) *and* $P_n(\omega)$ *be its periodogram given by* (1.5). *Suppose that the frequency estimates* $\hat{\omega}_k$ *maximize* S_n *in*

[2] In Walker (1971), the separation condition was given in a slightly different form, and the frequencies were not explicitly required to stay away from 0 and π.

(1.6) *under the separation condition* (1.7). *Then, the* $\hat{\omega}_k$ *are consistent in the sense that* $\hat{\omega}_k - \omega_k \xrightarrow{a.s.} 0$, $(k = 1, \ldots, q)$, *as* $n \to \infty$. *They are also asymptotically jointly independent and normally distributed such that* $n^{3/2}(\hat{\omega}_k - \omega_k) \xrightarrow{D} \mathrm{N}(0, 12/\gamma_k)$, *where*

$$\gamma_k := \frac{\beta_k^2}{2\sigma_\xi^2} \left| \sum \psi_j \exp(ij\omega_k) \right|^{-2}$$

is the signal-to-noise ratio (SNR) *of the kth sinusoid.*

In this theorem the asymptotic standard deviation of the frequency estimates is shown to be of order $n^{-3/2}$, as compared to $n^{-1/2}$ for many standard estimators. This high accuracy, however, is not surprising. Notice that in the frequency domain the power of the sinusoidal signal concentrates only at a finite number of locations and all other power components are due to the noise. It is therefore possible to clean up the noise considerably by using a certain bandpass filter that passes the sinusoids and suppresses the noise.

This, in fact, explains the high accuracy of periodogram analysis. Indeed, for any fixed ω_0, let us consider a linear filter — the *complex exponential filter* — whose impulse response $\{h_j\}$ is given by $h_j := \exp(i(j+1)\omega_0)$ for $j = 0, 1, \ldots, n-1$, and $h_j := 0$ elsewhere. Applying this filter to the data yields the output sequence

$$y_t(\omega_0) := \sum_{j=1}^{t} y_j h_{t-j} = \sum_{j=1}^{t} y_j \exp(i(t-j+1)\omega_0) \qquad (t = 1, \ldots, n). \tag{1.8}$$

Note that y_t is assumed to be zero for $t \leq 0$ in the filtering. From (1.5) and (1.8), it is easy to verify that

$$P_n(\omega_0) = n^{-1} |y_n(\omega_0)|^2. \tag{1.9}$$

That is, the periodogram at ω_0 is proportional to the squared magnitude of the output $y_t(\omega_0)$ when $t = n$.

Notice that the squared gain of the complex exponential filter $\{h_j\}$ can be written

as

$$G_n(\omega) = \begin{cases} \dfrac{\sin^2(n(\omega-\omega_0)/2)}{\sin^2((\omega-\omega_0)/2)} & \text{for } \omega \neq \omega_0 \\ n^2 & \text{for } \omega = \omega_0. \end{cases} \tag{1.10}$$

Given any $\delta > 0$, it is easy to see that $G_n(\omega) = o(n^\delta)$ for all $\omega \neq \omega_0$ (but of course this "o" is not uniform in $\omega \neq \omega_0$). Figure 1.1 shows the normalized squared gain $n^{-2}G_n(\omega)$ with $\omega_0 = 0.5\pi$ as a function of the *normalized* frequency $f := \omega/\pi$. Clearly, for sufficiently large n, $\{h_j\}$ is a bandpass filter with a narrow pass band around ω_0. Therefore, if the filter locks on one of the sinusoids, the locked sinusoid will be significantly enhanced and the periodogram is expected to take a large value of order $O(n)$, as can be seen from (1.9) and (1.10). The output sequence $\{y_1(\omega_0), \ldots, y_n(\omega_0)\}$ in this case consists basically of the locked sinusoid plus some filtered noise which has a considerably smaller variance than in the original data. Being the sum of the variances of the locked sinusoid and the filtered noise, the *variance* of $y_n(\omega_0)$, which is proportional to the expected value of the periodogram, is therefore significantly larger than in the case where no sinusoids are captured by the filter, typically $O(n)$ versus $O(1)$ (see, e.g., Priestley, 1981). For this reason, the locations where $P_n(\omega)$ assumes the largest values produce very accurate frequency estimates.

Furthermore, the behavior of the squared gain $G_n(\omega)$ also explains the resolution limit of periodogram analysis. Notice that the effective bandwidth of $G_n(\omega)$ is roughly of order $O(n^{-1})$ (see Figure 1.1). It is therefore very difficult for the complex exponential filter to distinguish two frequencies that are closer than the bandwidth, since in this case the filter tends to either enhance or suppress *both* frequencies simultaneously. The separation condition (1.7) can be interpreted as a requirement that prevents this situation from happening. Figure 1.2 illustrates the resolution limit of the periodogram. In this example, the periodogram $P_n(\omega)$, plotted again as a funtion of the normalized frequency f, was computed for a time series of length $n = 100$ which consists of four sinusoids in additive white Gaussian noise. Two frequencies of the

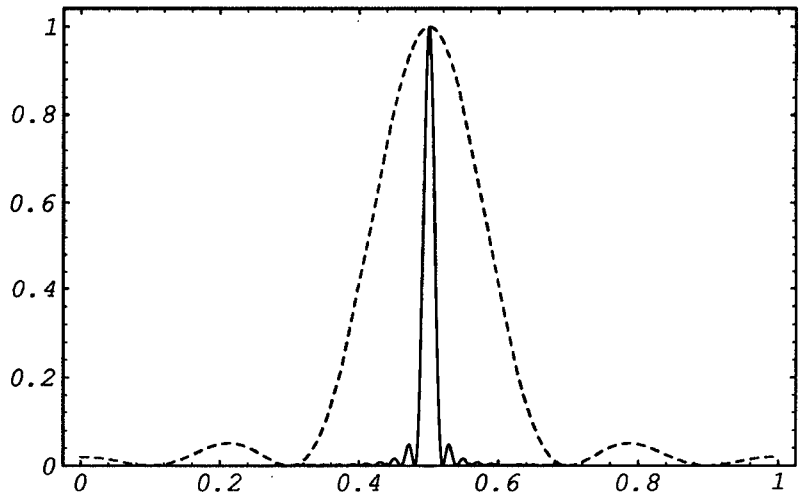


Figure 1.1: Plot of the normalized squared gain $n^{-2}G_n(\omega)$ of the complex exponential filter with $\omega_0 = 0.5\pi$ for $n = 10$ (dashed curve) and $n = 100$ (solid curve).

sinusoids ($\omega_1 = 0.31\pi$ and $\omega_2 = 0.35\pi$) are well-separated, while others ($\omega_3 = 0.713\pi$ and $\omega_4 = 0.725\pi$) are closely-spaced with a distance 0.012π ($\approx \pi/n$). The phase of each sinusoid is zero and the SNR is 3 dB. Clearly, the closely-spaced frequencies are not resolved by the periodogram.

From the computational point of view, periodogram analysis is a highly nonlinear procedure. Certain iterative routines, such as the Newton-Raphson algorithm, are therefore necessary in order to obtain the optimal estimates that maximize $P_n(\omega)$. As pointed out by Rice and Rosenblatt (1988), finding the global maxima in the periodogram is not an easy job. Since there are many local maxima with a separation in frequency about $O(n^{-1})$, as can be seen in Figure 1.2, one must start the iterative routines with initial guesses that are very close to the true frequencies, in order to converge to the desired solution. Initial guesses of accuracy $o(n^{-1})$ are typically required for the convergence. It is usually suggested that the DFT method be used first to yield some rough frequency estimates, and then followed by an exhaustive search on a fine grid around the rough estimates to produce the desired initial guesses (Abatzogolou,

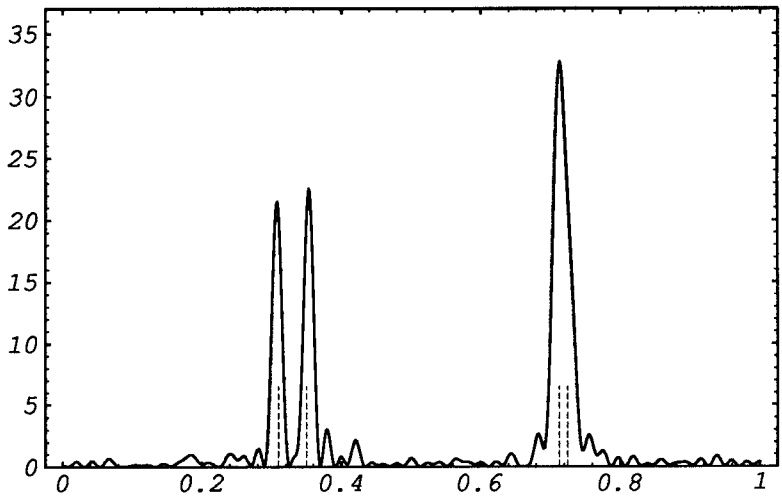


Figure 1.2: Periodogram of four sinusoids in white noise with $n = 100$. The SNR is 3 dB per sinusoid, and the true frequencies are indicated by dashed lines.

1985; Rice and Rosenblatt, 1988; Stoica, Moses, Friedlander, and Söderström, 1989). Using a large number of random initial values was also suggested in the literature (Rice and Rosenblatt, 1988).

Of course, the computational complexity can be reduced considerably by evaluating the periodogram only at *Fourier frequencies* of the form $2\pi j/n$ for $j = 0, 1, \ldots, n-1$. In fact, $P_n(2\pi j/n)$ is proportional to the magnitude of *discrete Fourier transform* (DFT) of the data and can be efficiently computed with the help of FFT (fast Fourier transform) algorithms. In so doing, periodogram analysis reduces to the *DFT method* that seeks the Fourier frequencies which correspond to the largest values of $P_n(2\pi j/n)$.

Despite its computational simplicity, the DFT method does not have the same estimation accuracy as periodogram analysis. In fact, its accuracy is of order n^{-1}, instead of $n^{-3/2}$, since only Fourier frequencies, which are discrete and separated by $2\pi/n$, are considered. Moreover, the resolution limit of the DFT method is similar to that of periodogram analysis. Figure 1.3 presents the DFT magnitude for the same time series as used in Figure 1.2.

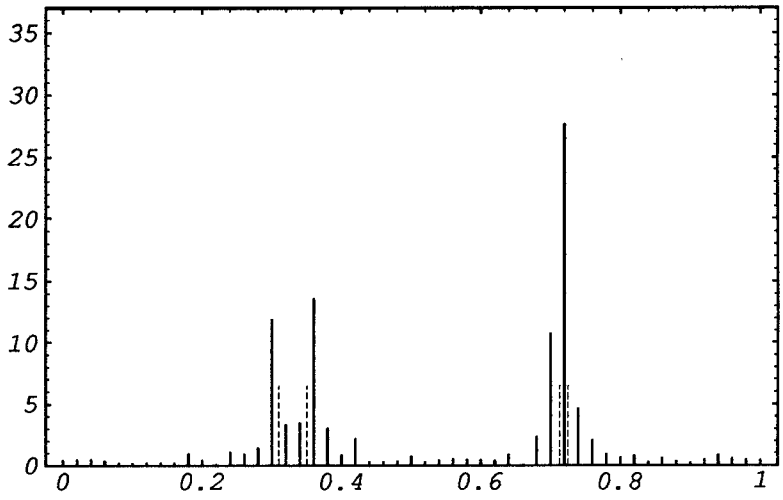


Figure 1.3: Plot of DFT magnitude for the same data as in Figure 1.2.

1.4 Nonlinear Least Squares

Closely related to periodogram analysis is the *nonlinear least squares* (NLS) method that has been considered from different aspects by many researchers (see, for example, Walker, 1971; Hannan, 1973; Rife and Boorstyn, 1974, 1976; Rice and Rosenblatt, 1988; Stoica and Nehorai, 1989).

The NLS method is conceptually very simple — it fits a sum of q sinusoids to the data $\{y_1, \ldots, y_n\}$ by minimizing the sum of squared errors with respect to the amplitudes, phases, as well as frequencies of the sinusoids. More precisely, the NLS method minimizes the criterion

$$J_n := \sum_{t=1}^{n} \left| y_t - \sum_{k=1}^{q} \beta_k \cos(\omega_k t + \phi_k) \right|^2 \tag{1.11}$$

with respect to $\{\beta_k, \phi_k, \omega_k\}$. Clearly, it is a highly nonlinear optimization problem.

It is interesting to note that once the NLS frequency estimates are available, one can easily obtain the NLS estimates of the amplitudes and phases. In fact, if we define $A_k := \beta_k \cos\phi_k$ and $B_k := -\beta_k \sin\phi_k$, then $\cos(\omega_k t + \phi_k)$ can be written as $A_k \cos(\omega_k t) + B_k \sin(\omega_k t)$. By this reparametrization, one can minimize J_n with respect

to $\{A_k, B_k, \omega_k\}$ instead of $\{\beta_k, \phi_k, \omega_k\}$. The resulting estimates of A_k and B_k can be transformed to yield the NLS estimates of β_k and ϕ_k, by using $\beta_k = \sqrt{A_k^2 + B_k^2}$ and $\phi_k = -\arctan(B_k/A_k)$. The NLS estimates of A_k and B_k can be easily obtained by linear methods. In fact, it is readily shown that for any fixed ω_k the minimizer of J_n with respect to $[A_1, \cdots, A_q, B_1, \cdots, B_q]^T$ is given by $(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\tilde{\mathbf{y}}$, where $\tilde{\mathbf{y}} := [y_1, \cdots, y_n]^T$ is the data vector and

$$\mathbf{G} := \begin{bmatrix} \cos\omega_1 & \cdots & \cos\omega_q & \sin\omega_1 & \cdots & \sin\omega_q \\ \vdots & & \vdots & \vdots & & \vdots \\ \cos n\omega_1 & \cdots & \cos n\omega_q & \sin n\omega_1 & \cdots & \sin n\omega_q \end{bmatrix}. \tag{1.12}$$

Substituting these optimal A_k and B_k in J_n yields

$$J_n' := \|\tilde{\mathbf{y}} - \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\tilde{\mathbf{y}}\|^2 = \tilde{\mathbf{y}}^T\{\mathbf{I} - \mathbf{G}\,(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\}\tilde{\mathbf{y}}. \tag{1.13}$$

Therefore, to obtain the NLS frequency estimates, it remains to minimize J_n' with respect to the frequencies ω_k. A similar argument for complex sinusoids can be found in Bresler and Macovski (1986). It is worth pointing out that the NLS estimator of $[A_1, \cdots, A_q, B_1, \cdots, B_q]^T$ obtained by substituting the NLS frequency estimates in $\mathbf{G}$ is consistent and asymptotically normal with the usual normalizing factor $n^{1/2}$ (see, e.g., Stoica and Nehorai, 1989).

1.4.1 For Well-Separated Frequencies

When the separation condition (1.7) holds, the NLS frequency estimates, denoted by $\hat{\omega}_k$, can be shown to possess the same asymptotic properties as those from periodogram analysis (Walker, 1971). In other words, Theorem 1.1 remains true (with the same variances) for the $\hat{\omega}_k$ that minimize J_n in (1.11).

In fact, periodogram analysis was proposed by Whittle (1952) as an approximation to NLS (Walker, 1971). To see this, let us expand J_n in (1.11) and write $J_n = U_n + R_n$,

where

$$U_n := \sum_{t=1}^{n} y_t^2 + \frac{n}{2}\sum_{k=1}^{q}\beta_k^2 - 2\sum_{t=1}^{n}\sum_{k=1}^{q} y_t\beta_k\cos(\omega_k t + \phi_k)$$

$$R_n := \sum_{t=1}^{n}\sum_{k,\,k'=1}^{q} \beta_k\beta_{k'}\cos(\omega_k t + \phi_k)\cos(\omega_{k'} t + \phi_{k'}) - \frac{n}{2}\sum_{k=1}^{q}\beta_k^2.$$

For any fixed ω_k, suppose that U_n is minimized by $\{\hat{\beta}_k, \hat{\phi}_k\}$. Then, it is easy to verify by differentiating U_n with respect to $\{\beta_k, \phi_k\}$ that the following equations must be satisfied, i.e.,

$$\sum_{t=1}^{n} y_t \sin(\omega_k t + \hat{\phi}_k) = 0 \quad \text{and} \quad \sum_{t=1}^{n} y_t \cos(\omega_k t + \hat{\phi}_k) = \frac{n}{2}\hat{\beta}_k.$$

Using these results, the value of U_n corresponding to $\{\hat{\beta}_k, \hat{\phi}_k\}$ can be written as

$$U_n' = \sum_{t=1}^{n} y_t^2 - \tfrac{1}{2}S_n$$

where S_n is the sum defined by (1.6). Clearly, the *frequency* estimates produced by minimizing U_n with respect to $\{\beta_k, \phi_k, \omega_k\}$ can be obtained directly by minimizing U_n', or, equivalently, by maximizing S_n. This is to say that minimizing U_n yields the same frequency estimates as periodogram analysis.

It is now sufficient to show that R_n is asymptotically negligible, so that minimizing J_n will be asymptotically equivalent to minimizing U_n, and hence the frequency estimates by NLS will have the same asymptotic properties as those by periodogram analysis. This can be done under the separation condition (1.7). To do so, we first need the following results.

Lemma 1.1 *Let* $\{\lambda_n\}$ *be a sequence defined in the interval* $(0, \pi)$ *such that* $c_n/n \le \lambda_n \le \pi - c_n/n$, *where* $c_n \to \infty$ *and* $c_n/n \to 0$. *Then, we have* $\sin(n\lambda_n)/\sin\lambda_n = o(n)$. *Moreover, the separation condition* (1.7) *implies that*

$$\frac{\sin(n(\omega_k + \omega_{k'})/2)}{\sin((\omega_k + \omega_{k'})/2)} = o(n) \quad \textit{and} \quad \frac{\sin(n(\omega_k - \omega_{k'})/2)}{\sin((\omega_k - \omega_{k'})/2)} = o(n)$$

where the first equality holds for all k *and* k', *and the second for all* $k \neq k'$.

PROOF. Let $Q_n := |\sin(n\lambda_n)/(n\sin\lambda_n)|$, then it suffices to show that $Q_n = o(1)$. Notice that $Q_n \leq (n\sin\lambda_n)^{-1}$ and hence $\limsup Q_n \leq \limsup (n\sin\lambda_n)^{-1}$. On the other hand, since $c_n/n \leq \lambda_n \leq \pi - c_n/n$ and $c_n/n \to 0$, we obtain $\sin\lambda_n \geq \sin(c_n/n) > 0$ for large n and $\sin(c_n/n)/(c_n/n) \to 1$. Therefore, it follows that

$$\limsup (n\sin\lambda_n)^{-1} \leq \limsup \{n\sin(c_n/n)\}^{-1} = \lim c_n^{-1} = 0.$$

Combining these results gives $\limsup Q_n = 0$, i.e., $Q_n = o(1)$. The remaining proof can be established by considering $\lambda_n = \omega_k - \omega_{k'}$ and $\lambda_n = \omega_k + \omega_{k'}$, respectively. $\Diamond$

Now, using the trigonometric identity

$$\cos\lambda\cos\lambda' = \tfrac{1}{2}\{\cos(\lambda - \lambda') + \cos(\lambda + \lambda')\} \tag{1.14}$$

we can write

$$\begin{aligned} 2R_n &= \sum_{\substack{k,k'=1 \\ k\neq k'}}^{q} \beta_k\beta_{k'} \sum_{t=1}^{n} \cos((\omega_k - \omega_{k'})t + (\phi_k - \phi_{k'})) \\ &\quad + \sum_{k,k'=1}^{q} \beta_k\beta_{k'} \sum_{t=1}^{n} \cos((\omega_k + \omega_{k'})t + (\phi_k + \phi_{k'})). \end{aligned}$$

It is easy to verify, upon noting that $\cos\omega = \{\exp(i\omega) + \exp(-i\omega)\}/2$, that

$$\left|\sum_{t=1}^{n} \cos((\omega_k \pm \omega_{k'})t + (\phi_k \pm \phi_{k'}))\right|^2 \leq \frac{\sin^2(n(\omega_k \pm \omega_{k'})/2)}{\sin^2((\omega_k \pm \omega_{k'})/2)}.$$

Therefore, by Lemma 1.1, we obtain $R_n = o(n)$. On the other hand, it can be shown that $U_n = O(n)$. This implies that R_n is asymptotically negligible as compared with U_n, and hence indicates that NLS can be replaced by periodogram analysis without loss of asymptotic accuracy. A derivation of the asymptotic covariance matrix of the NLS estimator can be found in Stoica and Nehorai (1989).

1.4.2 For Closely-Spaced Frequencies

As seen above, the negligibility of R_n in J_n is based entirely upon the separation condition (1.7) that guarantees Lemma 1.1. When this condition fails, R_n is no longer

negligible. As an example, let us consider in this subsection the case of $q = 2$, where ω_1 is a constant while $\omega_2 = \omega_1 + \delta/n$ for some constant $\delta > 0$ (Hannan and Quinn, 1989). In this case $\omega_2 - \omega_1 = O(n^{-1})$ and hence the separation condition (1.7) is not satisfied. Moreover, it can be shown that R_n has the same order of magnitude as U_n, i.e., $R_n = O(n)$, instead of $o(n)$, since

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \cos(\omega_1 t)\cos(\omega_2 t) = \frac{\sin\delta}{2\delta}.$$

Consequently, periodogram analysis is no longer a good approximation to NLS. In fact, for closely-spaced frequencies, as in our example, the NLS method provides better frequency estimates than periodogram analysis does. In a recent paper by Hannan and Quinn (1989), it was shown that, in our example of two closely-spaced frequencies, Theorem 1.1 still holds for the NLS estimates, obtained by minimizing J_n subject to a relaxed separation condition, except that asymptotic variances are different from (in fact larger than) those for well-separated frequencies. The relaxed separation condition was obtained by replacing the quantity c_n in (1.7) with a less restrictive one of the form

$$c_n := \kappa_n \sqrt{\log n/n}, \quad \text{where } \kappa_n \to \infty \text{ and } \kappa_n \sqrt{\log n/n} \to 0.$$

Notice that in this case c_n tends to zero, instead of infinity, so that frequencies are allowed to stay within a distance closer than or equal to $O(n^{-1})$. In this sense, the NLS method provides a higher resolution than periodogram analysis for frequency estimation.

1.4.3 Computational Considerations

The nonlinear least squares method, as its name implies, is a highly nonlinear optimization problem. It requires not only iterative routines to obtain the optimal frequency estimates, but also very accurate initial guesses for the iterative routines to converge to the optimal solution.

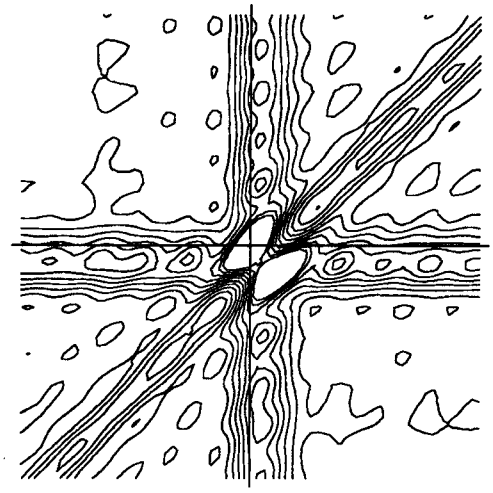


Figure 1.4: Contour plot of J_n for closely-spaced sinusoids, centered at true frequency.

As pointed out by Rice and Rosenblatt (1988), the criterion J_n has many local minima separated in frequency by $O(n^{-1})$, similar to the situation in periodogram analysis. Therefore, in order for the iterative routines, such as the Newton-Raphson algorithm, to converge to the global minimum point of J_n, initial frequency estimates of accuracy $o(n^{-1})$ are typically required.

For well-separated frequencies, periodogram analysis is the preferred method for frequency estimation since it reduces the *multi*-dimensional optimization problem of NLS to several independent *one*-dimensional optimization problems. However, as discussed in the previous section, maximizing the periodogram $P_n(\omega)$ instead of minimizing J_n is still an uneasy task. In fact, an iterative routine that starts with a certain exhaustive search is unavoidable in order to obtain the optimal solution. For closely-spaced frequencies, one has to return to the use of J_n since the periodogram does not resolve closely-spaced frequencies. This, again, requires an iterative routine plus an exhaustive search procedure (Hannan and Quinn, 1989).

Figure 1.4 illustrates the situation by an example of two sinusoids in additive white Gaussian noise with closely-spaced frequencies. The data length is $n = 100$, and the

Table 1.1: Summary of Estimation Methods

	Accuracy	Resolution	Complexity	Initial Guesses
PA	$O(n^{-3/2})$	$O(n^{-1})$	nonlinear	$o(n^{-1})$
NLS	$O(n^{-3/2})$	$O\left(n^{-1}\sqrt{\log n/n}\right)$	nonlinear	$o(n^{-1})$
DFT	$O(n^{-1})$	$O(n^{-1})$	$O(n \log n)$	*

SNR is 3 dB per sinusoid. The contour plot of $-\log J_n$ is shown in Figure 1.4 as a function of frequency variable (ω_1, ω_2) over the region $[0.613\pi, 0.813\pi] \times [0.625\pi, 0.825\pi]$. In calculating J_n, the *exact* values, instead of estimates, are used for the amplitudes and phase. The true frequencies are $\omega_1 = 0.713\pi$ and $\omega_2 = 0.725\pi$, shown as the center of the plot. In this plot, a peak of width of $O(n^{-1})$ is clearly seen near the center, indicating that very accurate estimates can be obtain by minimizing J_n. On the other hand, many local maxima with a separation of $O(n^{-1})$ are also seen in the plot. Therefore, initial guesses of accuracy $o(n^{-1})$ must be used in order to guarantee that gradient-based algorithms will converge to the optimal solution.

1.5 Summary

As seen in the previous sections, nonlinear methods, such as periodogram analysis (PA) and NLS, are able to provide very accurate frequency estimates but at the expense of high computational complexity. On the other hand, the DFT method is computationally simple but its accuracy and resolution cannot match the NLS method. These facts are briefly summarized in Table 1.1.

Since there are applications (e.g., radar and sonar) where frequency estimators with high accuracy and high resolution are demanded while the computation resources are limited, many researchers have been motivated to seek alternative approaches

suitable for these situations. To improve the resolution of periodogram analysis and the DFT method, which is crucial to many applications, it has been understood that the main reason for the resolution limit of the periodogram is that the data outside the observation interval are implicitly assumed to be *zero* in the computation of the periodogram. This is well reflected in the definition (1.8) of the output sequence $\{y_t(\omega_0)\}$ whose last value with $t = n$ defines the periodogram. Due to this practical assumption, the energy of a sinusoid — which theoretically concentrates at a single frequency — will spread all over the nearby frequency components in the periodogram, making it difficult for the periodogram to resolve closely-spaced frequencies. This phenomenon is known as the *spectrum leakage* (Key and Marple, 1981).

In the next chapter, we shall review and analyze an alternative approach, called AR method, that extrapolates the data beyond the observation interval, and thus hopefully provides a higher resolution, by fitting an autoregressive (AR) model to the data.

Chapter 2

Autoregressive Estimation

In this chapter, we shall discuss the autoregressive (AR) approach for frequency estimation. This method is widely used in spectral analysis because of its computational simplicity and high resolution property. As we shall see, however, the AR approach leads to biased and hence inconsistent frequency estimates.

2.1 AR Approach of Frequency Estimation

The AR approach of frequency estimation is based upon the following observations. Suppose that $\{x_t\}$ is the sinusoidal signal in (1.1) which consists of q superimposed sinusoids with frequencies $\omega_1, \ldots, \omega_q$. Let us denote by z^{-1} the backward-shift operator so that $z^{-1}x_t = x_{t-1}$, and consider the polynomial $A(z^{-1})$ — the *AR polynomial* — defined by

$$A(z^{-1}) := \prod_{k=1}^{q}(1 - z_k z^{-1})(1 - \bar{z}_k z^{-1}) = \sum_{j=0}^{2q} a_j z^{-j} \tag{2.1}$$

where $z_k := \exp(i\omega_k)$, and $\bar{z}_k := \exp(-i\omega_k)$ is the complex conjugate of z_k. The following results can be obtained immediately from the definition of $A(z^{-1})$.

Lemma 2.1 *Let $A(z^{-1})$ be the polynomial in* (2.1). *Then,* (a) *the $2q$ zeros of $A(z^{-1})$ are $z = \exp(\pm i\omega_k)$, $(k = 1, \ldots, q)$;* (b) *the $2q+1$ coefficients a_j of $A(z^{-1})$ are real and*

symmetric in the sense that

$$a_0 = 1 \quad \textit{and} \quad a_{2q-j} = a_j \qquad (j = 0, 1, \ldots, q-1); \tag{2.2}$$

and (c) *the a_j uniquely determine, and are determined by, the frequencies ω_k.*

PROOF. Part (a) and Part (c) are trivial. In Part (b), the a_j are real because the zeros of $A(z^{-1})$ are complex conjugate pairs, and the symmetry of a_j follows from the identity $z^{2q}A(z^{-1}) = A(z)$, since both z_k and z_k^{-1} are zeros of $A(z^{-1})$. ♢

Since $\exp(\pm i\omega_k)$ are zeros of $A(z^{-1})$, it is easy to verify that

$$A(z^{-1}) \exp(\pm i\omega_k t) = \exp(\pm i\omega_k t)\, A(\exp(\mp i\omega_k)) = 0$$

for all t. Notice that x_t can be written as a linear combination of $\exp(\pm i\omega_k t)$. Therefore, it follows that $A(z^{-1})\, x_t = 0$, namely,

$$\sum_{j=0}^{2q} a_j\, x_{t-j} = 0 \qquad (t = 0, \pm 1, \pm 2, \ldots). \tag{2.3}$$

This is to say that the sinusoidal signal $\{x_t\}$ satisfies a homogeneous autoregressive equation of order $2q$, with the AR coefficients being identical to the coefficients a_j as defined by (2.1).

According to these results, the original problem of frequency estimation can be equivalently stated as that of estimating the AR parameter

$$\mathbf{a} := [a_1, \cdots, a_q]^T.$$

This *reparametrization* enables us to employ many well-studied *linear* methods that usually end up with solving systems of linear equations. Once an estimate of $\mathbf{a}$ becomes available, the frequency estimates can be obtained from the zeros of $A(z^{-1})$ in (2.1), with the a_j replaced by their estimates. We refer to this method as the *AR approach* of frequency estimation.

2.2 Estimation of AR Coefficients

A widely-used procedure for estimating the AR parameter **a** is *Prony's estimator* (Hildebrand, 1956; Kay and Marple, 1981), also known as the *least squares* (LS) estimator, which can be summarized as follows. From (2.3), we obtain $A(z^{-1})\, y_t = A(z^{-1})\, x_t + A(z^{-1})\, \epsilon_t = A(z^{-1})\, \epsilon_t$. This implies that $\{y_t\}$ satisfies the equation

$$\sum_{j=0}^{2q} a_j\, y_{t-j} = e_t \qquad (t = 0, \pm 1, \pm 2, \ldots) \tag{2.4}$$

where $e_t := A(z^{-1})\, \epsilon_t$ depends on the noise. For the time series $\{y_1, \ldots, y_n\}$ of length $n > 2q$, the AR equation (2.4), together with (2.2), yields the following multivariate linear regression model

$$\mathbf{y} = -\mathbf{Y}\mathbf{Q}\mathbf{a} + \mathbf{e}. \tag{2.5}$$

In this expression, $\mathbf{y}$ and $\mathbf{Y}$ are data matrices as defined by

$$\mathbf{y} := \begin{bmatrix} y_{2q+1} + y_1 \\ \vdots \\ y_n + y_{n-2q} \end{bmatrix} \quad \text{and} \quad \mathbf{Y} := \begin{bmatrix} y_{2q} & \cdots & y_2 \\ \vdots & & \vdots \\ y_{n-1} & \cdots & y_{n-2q+1} \end{bmatrix} \tag{2.6}$$

respectively, and $\mathbf{e}$ the error term given by

$$\mathbf{e} := [e_{2q+1}, \cdots, e_n]^T. \tag{2.7}$$

The $(2q-1)$-by-q matrix $\mathbf{Q}$ takes care of the symmetry of the AR coefficients and admits the following form

$$\mathbf{Q} := \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & 1 \\ \tilde{\mathbf{I}} & \mathbf{0} \end{bmatrix}$$

where $\mathbf{I}$ stands for the $(q-1)$-by-$(q-1)$ identity matrix, $\tilde{\mathbf{I}}$ the $(q-1)$-by-$(q-1)$ reverse permutation matrix, with 1's on the anti-diagonal and 0's elsewhere, and $\mathbf{0}$ the zero vector of dimension $q-1$.

Prony's estimator, or the LS estimator, of the AR parameter $\mathbf{a}$, denoted by $\mathbf{a}_{\mathrm{LS}}$, is defined as the minimizer of the sum of squared errors $\|\mathbf{e}\|^2 = \|\mathbf{y} + \mathbf{YQa}\|^2$. A simple calculation shows that $\mathbf{a}_{\mathrm{LS}}$ satisfies the following *normal equations*

$$\mathbf{Q}^T\mathbf{Y}^T\mathbf{YQ}\,\mathbf{a}_{\mathrm{LS}} = -\mathbf{Q}^T\mathbf{Y}^T\mathbf{y}. \tag{2.8}$$

As will be seen shortly, the q-by-q matrix $\mathbf{Q}^T\mathbf{Y}^T\mathbf{YQ}$ is almost surely nonsingular if n is sufficiently large. Therefore, Prony's estimator $\mathbf{a}_{\mathrm{LS}}$ can be explicitly written as

$$\mathbf{a}_{\mathrm{LS}} = -(\mathbf{Q}^T\mathbf{Y}^T\mathbf{YQ})^{-1}\mathbf{Q}^T\mathbf{Y}^T\mathbf{y}. \tag{2.9}$$

As compared to the nonlinear optimization required by NLS, the estimator $\mathbf{a}_{\mathrm{LS}}$ is relatively easy to compute. The computational simplicity is one of the reasons that the AR approach is preferred in many applications where fast algorithms are desired for frequency estimation.

In the literature, there is an alternative method, known as the *forward-backward linear prediction* (FBLP), which is claimed to work better than straightforward least squares for AR estimation (Kay and Marple, 1981). Instead of minimizing the sum of squared forward prediction errors

$$\|\mathbf{e}\|^2 = \sum_{t=2q+1}^{n} \left| \sum_{j=0}^{2q} a_j\, y_{t-j} \right|^2$$

the FBLP method minimizes the sum of squared forward *and* backward prediction errors, namely,

$$\sum_{t=2q+1}^{n} \left| \sum_{j=0}^{2q} a_j\, y_{t-j} \right|^2 + \sum_{t=2q+1}^{n} \left| \sum_{j=0}^{2q} a_j\, y_{t-2q+j} \right|^2 .$$

In our case, however, since the AR coefficients a_j are symmetric, the backward prediction error coincides with the forward prediction error, i.e.,

$$\sum_{j=0}^{2q} a_j\, y_{t-2q+j} = \sum_{j=0}^{2q} a_j\, y_{t-j}.$$

This can be easily verified upon changing the variable j in the first summation to $2q-j$ and then using the symmetry of a_j. Consequently, the FBLP method produces the *same* AR estimator $\mathbf{a}_{\mathrm{LS}}$ as given by (2.9).

2.3 Inconsistency of Prony's Estimator

The computational simplicity of Prony's estimator makes it attractive in many applications. However, the estimator has been found to be inconsistent for frequency estimation (Kay and Marple, 1981; Dragošević and Stanković, 1989), namely, as the data length n grows, $\mathbf{a}_{\mathrm{LS}}$ does not converge to the desired AR parameter $\mathbf{a}$.

According to the large sample theory which we shall present in Chapter 3 (see also Li and Kedem, 1992), for any s and τ, we have (see Chapter 3, Remark 3.8)

$$n^{-1}\sum_{t=2q+1}^{n} y_{t+\tau}y_{t+s} \overset{a.s.}{\to} E(y_{t+\tau}y_{t+s}) = r^y_{\tau-s}$$

as $n \to \infty$. Since the process is real-valued, we also have $r^y_\tau = r^y_{-\tau}$ for any τ. Using these results, it is easy to verify that, as $n \to \infty$,

$$n^{-1}\mathbf{Y}^T\mathbf{Y} \overset{a.s.}{\to} \bar{\mathbf{R}}_y \quad \text{and} \quad n^{-1}\mathbf{Y}^T\mathbf{y} \overset{a.s.}{\to} \bar{\mathbf{r}}_y + \bar{\mathbf{r}}^B_y$$

where $\bar{\mathbf{R}}_y$ and $\bar{\mathbf{r}}_y$ are autocovariance matrices of $\{y_t\}$ as defined by

$$\bar{\mathbf{R}}_y = \begin{bmatrix} r^y_0 & r^y_1 & \cdots & r^y_{2q-2} \\ r^y_{-1} & r^y_0 & \cdots & r^y_{2q-3} \\ \vdots & \vdots & \ddots & \vdots \\ r^y_{-2q+2} & r^y_{-2q+3} & \cdots & r^y_0 \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{r}}_y = \begin{bmatrix} r^y_{-1} \\ r^y_{-2} \\ \vdots \\ r^y_{-2q+1} \end{bmatrix} \tag{2.10}$$

respectively, and $\bar{\mathbf{r}}^B_y := [r^y_{2q-1}, r^y_{2q-2}, \cdots, r^y_1]^T$ is the backward rearrangement of $\bar{\mathbf{r}}_y$. Notice that $\bar{\mathbf{R}}_y = \bar{\mathbf{R}}_x + \bar{\mathbf{R}}_\epsilon$, where $\bar{\mathbf{R}}_x$ and $\bar{\mathbf{R}}_\epsilon$ are autocovariance matrices of $\{x_t\}$ and $\{\epsilon_t\}$, respectively, with the same structure as $\bar{\mathbf{R}}_y$. It follows immediately that $\bar{\mathbf{R}}_y$ is

nonsingular, since $\bar{\mathbf{R}}_\epsilon$ is positive definite, by Proposition 5.1.1 in Brockwell and Davis (1987), and $\bar{\mathbf{R}}_x$ is at least non-negative definite[1].

As a consequence, Prony's estimator $\mathbf{a}_{\rm LS}$ can be written as (2.9) almost surely for sufficiently large n, and, as $n \to \infty$, it converges almost surely to a deterministic limit as specified by

$$\mathbf{a}_{\rm LS} \stackrel{a.s.}{\to} -\mathbf{R}_y^{-1}\mathbf{r}_y \tag{2.11}$$

where

$$\mathbf{R}_y := \mathbf{Q}^T\bar{\mathbf{R}}_y\mathbf{Q} \quad \text{and} \quad \mathbf{r}_y := \mathbf{Q}^T(\bar{\mathbf{r}}_y + \bar{\mathbf{r}}_y^B) = 2\,\mathbf{Q}^T\bar{\mathbf{r}}_y. \tag{2.12}$$

Notice that the q-by-q matrix $\mathbf{R}_y$ is also nonsingular since $\bar{\mathbf{R}}_y$ is nonsingular and $\mathbf{Q}$ is of full column rank q.

The limit of $\mathbf{a}_{\rm LS}$ in (2.11) can be more conveniently expressed in terms of the AR parameter $\mathbf{a}$ so that the bias of $\mathbf{a}_{\rm LS}$ could be easily identified. This result is presented in the following lemma.

Lemma 2.2 *Let* $\mathbf{a}_{\rm LS}$ *be Prony's estimator of* $\mathbf{a}$ *as defined in* (2.9). *Then, as* $n \to \infty$,

$$\mathbf{a}_{\rm LS} \stackrel{a.s.}{\to} \mathbf{a} - \mathbf{R}_y^{-1}(\mathbf{R}_\epsilon\mathbf{a} + \mathbf{r}_\epsilon) \tag{2.13}$$

where $\mathbf{R}_\epsilon$ *and* $\mathbf{r}_\epsilon$ *are defined from* r_τ^ϵ *in the same way as* $\mathbf{R}_y$ *and* $\mathbf{r}_y$ *in* (2.12).

PROOF. Denote by $\mathbf{X}$ and $\mathbf{Z}$ the matrices defined from $\{x_t\}$ and $\{\epsilon_t\}$, respectively, in the same way as $\mathbf{Y}$ in (2.6). Then, we can write $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Moreover, since $e_t = A(z^{-1})\,\epsilon_t$, namely,

$$e_t = \sum_{j=0}^{2q} a_j\,\epsilon_t \qquad (t = 0, \pm 1, \pm 2, \ldots), \tag{2.14}$$

it is easy to show that the error term $\mathbf{e}$ in (2.7) can be represented as

$$\mathbf{e} = \mathbf{Z}\mathbf{Q}\mathbf{a} + \mathbf{z}$$

[1]In fact, $\bar{\mathbf{R}}_x$ is positive definite, as we shall see later.

where $\mathbf{z}$ is defined from $\{\epsilon_t\}$ in the same way as $\mathbf{y}$ in (2.6). As a result,

$$E(\mathbf{Y}^T\mathbf{e}) = E(\mathbf{Y}^T\mathbf{Z})\,\mathbf{Q}\mathbf{a} + E(\mathbf{Y}^T\mathbf{z}). \tag{2.15}$$

Since $\{x_t\}$ and $\{\epsilon_t\}$ are independent, it follows that $E(\mathbf{X}^T\mathbf{Z}) = \mathbf{0}$ and $E(\mathbf{X}^T\mathbf{z}) = \mathbf{0}$. Therefore, we have

$$\begin{aligned} E(\mathbf{Y}^T\mathbf{Z}) &= E(\mathbf{Z}^T\mathbf{Z}) = (n-2q)\,\bar{\mathbf{R}}_\epsilon \\ E(\mathbf{Y}^T\mathbf{z}) &= E(\mathbf{Z}^T\mathbf{z}) = (n-2q)\,(\bar{\mathbf{r}}_\epsilon + \bar{\mathbf{r}}_\epsilon^B) \end{aligned}$$

where $\bar{\mathbf{R}}_\epsilon$ and $\bar{\mathbf{r}}_\epsilon$ are autocovariance matrices of $\{\epsilon_t\}$ defined in the same way as $\bar{\mathbf{R}}_y$ and $\bar{\mathbf{r}}_y$, respectively. Substituting these results in (2.15) yields

$$E(\mathbf{Y}^T\mathbf{e}) = (n-2q)\,(\bar{\mathbf{R}}_\epsilon\mathbf{Q}\mathbf{a} + \bar{\mathbf{r}}_\epsilon + \bar{\mathbf{r}}_\epsilon^B).$$

This, together with (2.5), gives

$$\begin{aligned} E(\mathbf{Y}^T\mathbf{y}) &= -E(\mathbf{Y}^T\mathbf{Y})\,\mathbf{Q}\mathbf{a} + E(\mathbf{Y}^T\mathbf{e}) \\ &= (n-2q)\,(-\bar{\mathbf{R}}_y\mathbf{Q}\mathbf{a} + \bar{\mathbf{R}}_\epsilon\mathbf{Q}\mathbf{a} + \bar{\mathbf{r}}_\epsilon + \bar{\mathbf{r}}_\epsilon^B). \end{aligned}$$

On the other hand, a straightforward calculation shows that $E(\mathbf{Y}^T\mathbf{y}) = (n-2q)\,(\bar{\mathbf{r}}_y + \bar{\mathbf{r}}_y^B)$. Thus, we obtain

$$\bar{\mathbf{r}}_y + \bar{\mathbf{r}}_y^B = -\bar{\mathbf{R}}_y\mathbf{Q}\mathbf{a} + \bar{\mathbf{R}}_\epsilon\mathbf{Q}\mathbf{a} + \bar{\mathbf{r}}_\epsilon + \bar{\mathbf{r}}_\epsilon^B.$$

By definition, $\mathbf{r}_y = \mathbf{Q}^T(\bar{\mathbf{r}}_y + \bar{\mathbf{r}}_y^B) = -\mathbf{R}_y\mathbf{a} + \mathbf{R}_\epsilon\mathbf{a} + \mathbf{r}_\epsilon$. Substituting this expression in (2.11) proves the lemma. ◊

Lemma 2.2 tells us that the almost sure limit of $\mathbf{a}_{\mathrm{LS}}$ is in general different from the AR parameter $\mathbf{a}$ that we intend to estimate. It can be shown that the bias $-\mathbf{R}_y^{-1}(\mathbf{R}_\epsilon\mathbf{a} + \mathbf{r}_\epsilon)$ is more pronounced when the signal-to-noise ratio is not sufficiently high. For example, let us consider (2.13) in the cases of a single sinusoid and of two sinusoids, respectively.

Example 2.1 For $q = 1$, we have $a_1 = -2\cos\omega_1$, and (2.13) reduces to

$$\mathbf{a}_{\mathrm{LS}} \overset{a.s.}{\to} a_1 - \frac{a_1 + 2\rho_1^\epsilon}{1+\gamma}$$

where $\rho_\tau^\epsilon := r_\tau^\epsilon / r_0^\epsilon$ is the autocorrelation function of $\{\epsilon_t\}$, and $\gamma := r_0^x / r_0^\epsilon$ the signal-to-noise ratio. Clearly, the bias $-(a_1 + 2\rho_1^\epsilon)/(1+\gamma)$ does not vanish unless $\gamma \to \infty$ or in the unusual case of $\rho_1^\epsilon = \cos\omega_1$. In general the bias is inversely related to the signal-to-noise ratio γ. Therefore, in order for Prony's estimator to provide satisfactory results, the signal-to-noise ratio γ should be very high (e.g., 30 dB), as reported in many papers (see Kay and Marple, 1981, and references therein).

Example 2.2 For $q = 2$, it can be shown from (2.1) that

$$a_1 = -2(\cos\omega_1 + \cos\omega_2) \quad \text{and} \quad a_2 = 2(1 + 2\cos\omega_1\cos\omega_2). \tag{2.16}$$

When the noise is white, we have $\mathbf{r}_\epsilon = \mathbf{0}$, and thus the bias in (2.13) becomes

$$\frac{-1}{(1-\rho_1^2+\rho_2)\gamma^2 + (2+\rho_2)\gamma + 1} \begin{bmatrix} \gamma+1 & -\rho_1\gamma \\ -\rho_1\gamma & (1+\rho_2)\gamma+1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

where $\rho_\tau := \rho_\tau^x$ is the autocorrelation function of $\{x_t\}$. As we can see again, the bias vanishes if $\gamma \to \infty$. On the other hand, when $\gamma \to 0$, the bias tends to $-\mathbf{a}$ and hence the limit of Prony's estimator equals $\mathbf{0}$, which, by (2.16), corresponds to $\cos\omega_1 = -\cos\omega_2 = \sqrt{2}/2$, or, equivalently, $\omega_1 = \pi/4$ and $\omega_2 = 3\pi/4$. This explains why Prony's frequency estimates tend to appear around $\omega_1 = \pi/4$ and $\omega_2 = 3\pi/4$ when the signal-to-noise ratio is low.

Finally, it should be pointed out that the above asymptotic analysis applies not only to Prony's estimator $\mathbf{a}_{\mathrm{LS}}$, but also to any AR estimator of the form $-\hat{\mathbf{R}}^{-1}\hat{\mathbf{r}}$, where $\hat{\mathbf{R}}$ and $\hat{\mathbf{r}}$ are some consistent estimators of $\mathbf{R}_y$ and $\mathbf{r}_y$, respectively. In these cases, the limiting expressions (2.11) and (2.13) always hold, and hence the same inconsistency persists.

2.4 High-Order AR Method

A way to reduce the bias of Prony's estimator is to increase the order of the AR model. In fact, for any $m \geq 2q$, let $B(z^{-1}) = \sum_{j=0}^{m-2q} b_j z^{-j}$ be an arbitrary polynomial in z^{-1} of degree $m - 2q$ with $b_0 = 1$. Then, it follows from (2.3) that $B(z^{-1})\, A(z^{-1})\, x_t = 0$ for all t. Clearly, the product

$$C(z^{-1}) := B(z^{-1})\, A(z^{-1}) \tag{2.17}$$

is a polynomial in z^{-1} of degree m, i.e., $C(z^{-1}) = \sum_{j=0}^{m} c_j z^{-j}$, and $c_0 = 1$. This implies that $\{x_t\}$ satisfies the following *high-order* AR equation

$$\sum_{j=0}^{m} c_j\, x_{t-j} = 0 \qquad (t = 0, \pm 1, \pm 2, \ldots) \tag{2.18}$$

where $m \geq 2q$. Notice that (2.18) reduces to (2.3) when $m = 2q$. In general, when $m > 2q$, the AR coefficients c_j in (2.18) are no longer symmetric, since $B(z^{-1})$ is arbitrary, and hence the zeros of the AR polynomial $C(z^{-1})$ are not necessarily reciprocal pairs as in the case of $m = 2q$. Nevertheless, $C(z^{-1})$ has at least $2q$ zeros *on* the unit circle $|z| = 1$ which coincide with the zeros of $A(z^{-1})$ that determine the frequencies ω_k. This gave rise to the idea of estimating the AR coefficients c_j without the restriction of symmetry. When estimates of c_j become available, the frequency estimates can be obtained either from the $2q$ zeros of the (estimated) AR polynomial $C(z^{-1})$ which are on or closest to the unit circle, or from the $2q$ maxima in the (estimated) *AR spectrum* $|C(\exp(-i\omega))|^{-2}$ (Lang and McClellan, 1980; Kay and Marple, 1981).

Strictly speaking, the c_j should be restricted to those which admit the factorization (2.17) in order to guarantee that the resulting $C(z^{-1})$ has at least $2q$ zeros on the unit circle. In the existing high-order AR methods, however, an *unconstrained* AR(m) model is usually used, perhaps due to the lack of a convenient description of this restriction in terms of the c_j.

Now consider the LS estimator $\mathbf{c}_{\mathrm{LS}}$ that minimizes the criterion

$$\sum_{t=m+1}^{n} \left| \sum_{j=0}^{m} c_j y_{t-j} \right|^2 .$$

Owing to the strong ergodicity of $\{y_t\}$, it can be shown as before that $\mathbf{c}_{\mathrm{LS}}$ converges almost surely to a deterministic limit. For white noise, the limit can be written as (Stoica, Friedlander, and Söderström, 1987)

$$\mathbf{c} + O(m^{-2})$$

where the m-vector $\mathbf{c}$ is the *minimum-norm* solution of the Yule-Walker equations corresponding the autocovariances of $\{x_t\}$. The minimum-norm solution $\mathbf{c}$ has some very interesting properties. For example, it was shown (Tufts and Kumaresan, 1982; Stoica, Friedlander, and Söderström, 1987) that the corresponding AR polynomial $C(z^{-1})$ *can* be factorized as (2.17) with all the zeros of $B(z^{-1})$ appear strictly *inside* the unit circle. This, together with the fact that the difference between the limit of $\mathbf{c}_{\mathrm{LS}}$ and $\mathbf{c}$ vanishes as m increases without bound, indicates that for sufficiently large n and m the high-order AR method produces satisfactory frequency estimates by locating the $2q$ zeros on or closest to the unit circle in the AR polynomial corresponding to the estimator $\mathbf{c}_{\mathrm{LS}}$.

In a recent paper by Makisack and Poskitt (1989), it was shown for the simplest case of a single sinusoid ($q = 1$) in additive white noise that the high-order AR method leads to a consistent frequency estimate as the order m of the AR model increases at a certain rate (e.g., $m^6/n \to \infty$ and $m^2/n \to 0$) along with the data length n. For short data records, it was suggested that m be chosen between $n/3$ and $n/2$ in order to obtain satisfactory results (Kay and Marple, 1981, and references therein).

The weakness of the high-order AR method is that many spurious zeros (or spurious peaks in the AR spectrum) are introduced as the order of the AR model increases (Kay and Marple, 1981, and references therein). In some cases, especially when the signal-to-noise ratio is low, it could be difficult to identify the zeros (peaks) corresponding

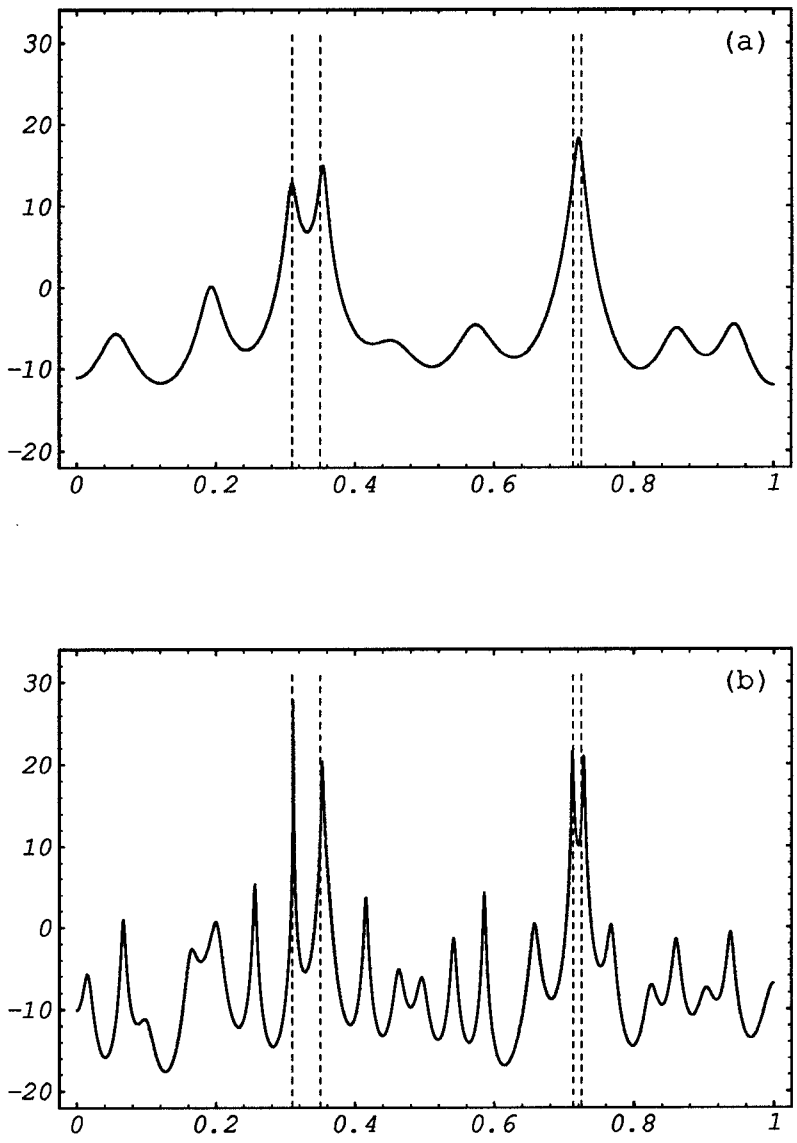


Figure 2.1: Plot of high-order AR log-spectrum for the same data with $n = 100$ as in Figure 1.2. The SNR is 3 dB per sinusoid. (a) For $m = 25$, the well-separated frequencies are resolved but the closely-spaced frequencies are not. (b) For $m = 50$, all the frequencies are resolved, but a large number of spurious peaks may cause difficulties in the calculation of the frequency estimates.

to the sinusoidal signal from a large number of zeros (peaks) in the AR polynomial (spectrum). Figure 2.1 illustrates the performance of the high-order AR method. In this figure, the AR log-spectrum from the Burg estimator (Kay and Marple, 1981) was plotted, as a function of the normalized frequency f, for the same data of four sinusoids in additive white noise with $n = 100$ as in Figure 1.2. It shows that closely-spaced frequencies can be resolved and satisfactory frequency estimates be obtained when the order m is sufficiently high. On the other hand, it is also evident that the spurious peaks presented in the AR spectrum may cause difficulties in the calculation of the global maxima in order to obtain the frequency estimates.

Notice that $\mathbf{c}_{\mathrm{LS}}$ can be written as $\mathbf{c}_{\mathrm{LS}} = -\hat{\mathbf{R}}^{-1}\hat{\mathbf{r}}$ where $\hat{\mathbf{R}}$ is the m-by-m sample autocovariance matrix of $\{y_t\}$. In the noiseless case, the corresponding sample autocovariance matrix is defined from $\{x_t\}$ and can be shown to be of rank q. Making use of this property, one can apply the principal component (PC) analysis and approximate the "noisy" matrix $\hat{\mathbf{R}}$ from $\{y_t\}$ by a rank-q matrix $\check{\mathbf{R}}$ corresponding to the principal eigenvalues (Tufts and Kumaresan, 1982; Kay, 1988). The resulting AR estimator is given by $\mathbf{c}_{\mathrm{PC}} = -\check{\mathbf{R}}^{\dagger}\hat{\mathbf{r}}$ where $\check{\mathbf{R}}^{\dagger}$ is the pseudo-inverse of $\check{\mathbf{R}}$. Simulations have shown that the PC method is able to improve the original least squares estimator $\mathbf{c}_{\mathrm{LS}}$ and to produce very good estimates when the signal-to-noise ratio is high or moderate (Kay, 1988). This is not surprising since the PC method cleans the noise with the help of principal component analysis, while the least squares merely tries to explain the noise with extra poles in the AR spectrum. The difficulty of the PC method, however, is that a singular-value decomposition has to be used for principal component analysis and hence the computational burden could be too high for some applications, especially when m is large, as required for efficient estimates (Kay and Shaw, 1988). Moreover, in the case of low signal-to-noise ratio, the PC method is not as efficient as other procedures which employ linear filters for noise-cleaning (Kay, 1988; Dragošević and Stanković, 1989).

2.5 Nonsingularity of Autocovariance Matrix

Before ending this chapter, we would like to show that the autocovariance matrix $\bar{\mathbf{R}}_x$ of the signal $\{x_t\}$ is positive definite and hence nonsingular. This result is presented in the following lemma.

Lemma 2.3 *Let $\bar{\mathbf{R}}_x$ be the autocovariance matrix of $\{x_t\}$ with the same structure as $\bar{\mathbf{R}}_y$ in (2.10). Then, $\bar{\mathbf{R}}_x$ has full rank $2q-1$ and can be decomposed as*

$$\bar{\mathbf{R}}_x = \mathbf{S}\mathbf{P}\mathbf{S}^H$$

where $\mathbf{P}$ is a $2q$-by-$2q$ diagonal matrix of full rank $2q$ and $\mathbf{S}$ a $(2q-1)$-by-$2q$ Vandermonde matrix of full rank $2q-1$. The superscript H stands for the Hermitian transposition.

PROOF. Let us first extend the notation β_k and ω_k for $k = q+1, \ldots, 2q$ by defining

$$\beta_k := \beta_{2q-k+1} \quad \text{and} \quad \omega_k := -\omega_{2q-k+1} \qquad (k = q+1, \ldots, 2q).$$

Then, the autocovariance function r_τ^x can be written as

$$r_\tau^x = \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \cos(\omega_k \tau) = \sum_{k=1}^{2q} \tfrac{1}{4}\beta_k^2 \, z_k^\tau \tag{2.19}$$

where $z_k := \exp(i\omega_k)$, $(k = 1, \ldots, 2q)$. Consider the $2q$-by-$2q$ diagonal matrix

$$\mathbf{P} := \tfrac{1}{4}\,\mathrm{diag}(\beta_1^2, \ldots, \beta_{2q}^2). \tag{2.20}$$

Clearly, $\mathbf{P}$ is nonsingular since $\beta_k > 0$ for all k. It is easy to verify from (2.19) and the definition of $\bar{\mathbf{R}}_x$ in (2.10) that $\bar{\mathbf{R}}_x = \mathbf{S}\mathbf{P}\mathbf{S}^H$ where

$$\mathbf{S} := [\mathbf{s}_1, \cdots, \mathbf{s}_{2q}] \quad \text{and} \quad \mathbf{s}_k := [1, z_k, \cdots, z_k^{2q-2}]^T. \tag{2.21}$$

To show that $\mathbf{S}$ has full (row) rank $2q-1$, we first note that $\mathbf{S}^H$ can be written as

$$\mathbf{S}^H = [\mathbf{c}_0, \cdots, \mathbf{c}_{2q-2}]$$

where $\mathbf{c}_j := [z_1^{-j}, \cdots, z_{2q}^{-j}]$, $(j = 0, \ldots, 2q-2)$. Suppose that there exist constants p_j such that $\sum_{j=0}^{2q-2} p_j \mathbf{c}_j = \mathbf{0}$. Then, it implies that $\Phi(z_k^{-1}) = \sum_{j=0}^{2q-2} p_j z_k^{-j} = 0$ for all $k = 1, \ldots, 2q$, namely, the polynomial $\Phi(z) := \sum_{j=0}^{2q-2} p_j z^j$ has $2q$ distinct zeros z_k^{-1} while its degree is at most $2q-2$. It is only possible when $p_j = 0$ for all j. As a result, the $\mathbf{c}_j$ are linear independent and hence the rank of $\mathbf{S}$ is equal to $2q-1$. The nonsingularity of $\bar{\mathbf{R}}_x$ follows immediately. $\diamondsuit$

Remark 2.1 Since $\mathbf{Q}$ has full column rank q, Lemma 2.3 implies that $\mathbf{R}_x$ is nonsingular and can be decomposed as $\mathbf{R}_x = \mathbf{Q}^T \mathbf{S} \mathbf{P} \mathbf{S}^H \mathbf{Q}$.

Chapter 3

Limit Theorems of Sample Autocovariance Function

In this chapter, we investigate some limiting properties of the sample autocovariance function (SACF) of the process $\{y_t\}$ in (1.1) concerning its consistency and asymptotic distribution. In particular, we provide a central limit theorem for the sample autocovariance function, and consider the *uniform* consistency of the sample autocovariance function after parametric filtering. Since $\{y_t\}$ has a mixed spectrum — its spectrum consists of a discrete part corresponding to the sinusoids and a continuous part corresponding to the noise — the central limit theorem in this chapter extends the classical results for a time series with continuous spectrum. The consistency of the sample autocovariance function of $\{y_t\}$ has been used in the engineering literature for a long time without rigorous proof. Therefore, the results on uniform consistency of the sample autocovariance function after parametric filtering not only fill this gap but also provide some insight into the effect of parametric filtering on the sample autocovariance function of a time series with mixed spectrum. The limit theorems developed in this chapter lay the foundation of the asymptotic analysis for the parametric filtering method of frequency estimation which we shall present in the next chapter.

3.1 Asymptotic Normality of SACF

Given a finite sample $\{y_1, \ldots, y_n\}$ from the random process in (1.1), let us consider the sample autocovariance function (SACF) as defined by

$$\hat{r}_j := n^{-1} \sum_{t=1}^{n-j} y_{t+j} y_t \qquad (j = 0, 1, \ldots, p) \tag{3.1}$$

where p is a fixed integer with $0 \leq p < n$. In this section, we would like to show that the $\hat{r}_j$ are asymptotically jointly normal as $n \to \infty$.

For simplicity, let us first establish the normality for the similar quantities

$$\tilde{r}_j := n^{-1} \sum_{t=1}^{n} y_{t+j} y_t \qquad (j = 0, 1, \ldots, p), \tag{3.2}$$

and then prove the asymptotic negligibility of the differences $\hat{r}_j - \tilde{r}_j$.

We assume for the time being that the phases ϕ_k of the sinusoids are *constants*. This assumption, however, does not alter the following asymptotic theory. In fact, assuming constant ϕ_k is equivalent to considering the *conditional* properties of the SACF given ϕ_k. As we shall see, the conditional asymptotic distribution of $\hat{r}_j$ given ϕ_k does not depend on ϕ_k, and therefore coincides with the the unconditional distribution.

The following lemma shows that the covariance of $\tilde{r}_i$ and $\tilde{r}_j$ converges to a finite limit at the rate of n^{-1}.

Lemma 3.1 *Suppose that* $E(\xi_t^4) = \kappa \sigma_\xi^4 < \infty$ *where* $\{\xi_t\}$ *is the i.i.d. sequence in* (1.2). *Then, it follows that*

$$\lim_{n \to \infty} n \operatorname{cov}(\tilde{r}_i, \tilde{r}_j) = \lim_{n \to \infty} n E\{(\tilde{r}_i - r_i^y)(\tilde{r}_j - r_j^y)\} = \sigma_{ij}$$

where σ_{ij} *is finite and can be written as*

$$\begin{aligned} \sigma_{ij} \;:=\; & \sum_{k=1}^{q} 2\beta_k^2 \cos(\omega_k i) \cos(\omega_k j) \sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega_k \tau) \\ & + (\kappa - 3) r_i^\epsilon r_j^\epsilon + \sum_{\tau=-\infty}^{\infty} (r_\tau^\epsilon r_{\tau+i-j}^\epsilon + r_{\tau+i}^\epsilon r_{\tau-j}^\epsilon). \end{aligned} \tag{3.3}$$

PROOF. Using the trigonometric identity (1.14), it is easy to show that

$$\cos(\omega_k t + \phi_k)\cos(\omega_k(t+j)+\phi_k) = \tfrac{1}{2}\{\cos(\omega_k j) + \cos(\omega_k(2t+j)+2\phi_k)\}.$$

Therefore, from (1.1), we can write $\tilde{r}_j$ as

$$\begin{aligned}\tilde{r}_j &= r_j^y + n^{-1}\sum_{t=1}^{n}\zeta_{jt} + (2n)^{-1}\sum_{k=1}^{q}\beta_k^2\sum_{t=1}^{n}\cos(\omega_k(2t+j)+2\phi_k) \\ &\quad + n^{-1}\sum_{\substack{k,k'=1\\k\neq k'}}^{q}\beta_k\beta_{k'}\sum_{t=1}^{n}\cos(\omega_k t+\phi_k)\cos(\omega_{k'}(t+j)+\phi_{k'}) \qquad (3.4)\end{aligned}$$

where r_j^y is given by (1.3) and (1.4), and ζ_{jt} is defined by

$$\zeta_{jt} := x_t\epsilon_{t+j} + x_{t+j}\epsilon_t + \epsilon_{t+j}\epsilon_t - r_j^{\epsilon}. \qquad (3.5)$$

Since $\cos\omega = \{exp(i\omega) + \exp(-i\omega)\}/2$, it is not difficult to show that

$$\left|\sum_{t=1}^{n}\cos(\omega_k(2t+j)+2\phi_k)\right| \leq |\sin\omega_k|^{-1}.$$

Similarly, using the trigonometric identity (1.14), we obtain

$$\begin{aligned}&\left|\sum_{t=1}^{n}\cos(\omega_k t+\phi_k)\cos(\omega_{k'}(t+j)+\phi_{k'})\right| \\ &\quad \leq \tfrac{1}{2}\{|\sin((\omega_k+\omega_{k'})/2)|^{-1} + |\sin((\omega_k-\omega_{k'})/2)|^{-1}\}\end{aligned}$$

for any ϕ_k, $\phi_{k'}$, j, and $k \neq k'$. Since $\omega_k \in (0,\pi)$ for all k, there exists a constant $K > 0$ such that $|\sin\omega_k|^{-1} \leq K$ and $|\sin((\omega_k \pm \omega_{k'})/2)|^{-1} \leq K$ for all $k \neq k'$. This, together with (3.4), implies that

$$\tilde{r}_j = r_j^y + n^{-1}\sum_{t=1}^{n}\zeta_{jt} + O(n^{-1}). \qquad (3.6)$$

Moreover, since $E(\zeta_{jt}) = 0$, we also have

$$E(\tilde{r}_j) = r_j^y + O(n^{-1}).$$

Combining these results yields

$$\begin{aligned}\lim_{n\to\infty} n\,\mathrm{cov}(\tilde{r}_i,\tilde{r}_j) &= \lim_{n\to\infty} n\,E\{(\tilde{r}_i - r_i^y)(\tilde{r}_j - r_j^y)\} \\ &= \lim_{n\to\infty} n^{-1}\,\mathrm{cov}\left(\sum_{t=1}^{n}\zeta_{it},\sum_{t=1}^{n}\zeta_{jt}\right).\end{aligned}$$

To proceed with the proof, we note that $\{x_t\}$ is deterministic under the constant phase assumption. Therefore, from (3.5), we obtain

$$n^{-1}\operatorname{cov}\left(\sum_{t=1}^{n}\zeta_{it},\sum_{t=1}^{n}\zeta_{jt}\right)=n^{-1}\sum_{t,s=1}^{n}E(\zeta_{it}\zeta_{js})=I_1+I_2+I_3$$

where

$$\begin{aligned}
I_1 &:= n^{-1}\sum_{t,s=1}^{n}(x_t x_s r^{\epsilon}_{t-s+i-j}+x_t x_{s+j} r^{\epsilon}_{t-s+i}\\
&\qquad + x_{t+i}x_s r^{\epsilon}_{t-s-j}+x_{t_i}x_{s+j}r^{\epsilon}_{t-s})\\
&:= T_1+T_2+T_3+T_4\\
I_2 &:= n^{-1}\sum_{t,s=1}^{n}\{x_t c(t-s+i,j)+x_{t+i}c(t-s,j)\\
&\qquad + x_s c(s-t+j,i)+x_{s+j}c(s-t,i)\}\\
I_3 &:= n^{-1}\operatorname{cov}\left(\sum_{t=1}^{n}\epsilon_{t+i}\epsilon_t,\sum_{t=1}^{n}\epsilon_{t+j}\epsilon_t\right),
\end{aligned}$$

and $c(u,v):=E(\epsilon_{t+u}\epsilon_{t+v}\epsilon_t)$ is the third-order cumulant function of $\{\epsilon_t\}$. Using the substitution $\tau:=t-s$ and the trigonometric identity (1.14), we can write T_1 as

$$\begin{aligned}
T_1 = &\sum_{k=1}^{q}\tfrac{1}{2}\beta_k^2\sum_{|\tau|<n}r^{\epsilon}_{\tau+i-j}\left\{\left(1-\frac{|\tau|}{n}\right)\cos(\omega_k\tau)+n^{-1}\sum_{t\in D}\cos(\omega_k(2t-\tau)+2\phi_k)\right\}\\
&+\sum_{\substack{k,k'=1\\k\neq k'}}^{q}\beta_k\beta_{k'}\sum_{|\tau|<n}r^{\epsilon}_{\tau+i-j}\;n^{-1}\sum_{t\in D}\cos(\omega_k t+\phi_k)\cos(\omega_{k'}(t-\tau)+\phi_{k'})
\end{aligned}$$

where $D:=\{t:\max(1,\tau+1)\le t\le\min(n,\tau+n)\}$. Clearly, for any τ and $k\neq k'$, the two summations over $t\in D$ are bounded in absolute value by n. In addition, it is easy to show that

$$n^{-1}\sum_{t\in D}\cos(\omega_k(2t-\tau)+2\phi_k)\to 0$$

and

$$n^{-1}\sum_{t\in D}\cos(\omega_k t+\phi_k)\cos(\omega_{k'}(t-\tau)+\phi_{k'})\to 0$$

as $n \to \infty$ for any τ and $k \neq k'$. Since $\sum |r_\tau^\epsilon| < \infty$, it follows from the bounded convergence theorem that

$$\begin{aligned} T_1 &\to \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \sum_{\tau=-\infty}^{\infty} r_{\tau+i-j}^\epsilon \cos(\omega_k \tau) \\ &= \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega_k(\tau - i + j)) \end{aligned}$$

as $n \to \infty$. Similarly, we obtain

$$\begin{aligned} T_2 &\to \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega_k(\tau - i - j)) \\ T_3 &\to \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega_k(\tau + i + j)) \\ T_4 &\to \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega_k(\tau + i - j)). \end{aligned}$$

Since the symmetry of r_τ^ϵ implies that $\sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \sin(\omega_k \tau) = 0$, adding up these expressions, followed by an application of (1.14), yields

$$I_1 \to \sum_{k=1}^{q} 2\beta_k^2 \cos(\omega_k i)\cos(\omega_k j) \sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega_k \tau).$$

Furthermore, it is easy to verify from (1.2) that

$$c(\tau + u, v) = E(\epsilon_{\tau+u}\epsilon_v\epsilon_0) = E(\xi_0^3) \sum_{j=-\infty}^{\infty} \psi_{j+\tau+u}\psi_{j+v}\psi_j.$$

and hence $\sum |c(\tau+u, v)| < \infty$ for any fixed u and v. Following an argument similar to the proof of T_1 leads to $I_2 \to 0$. Finally, according to the classical results (Brockwell and Davis, 1987, Proposition 7.3.1), we obtain

$$I_3 \to (\kappa - 3) r_i^\epsilon r_j^\epsilon + \sum_{\tau=-\infty}^{\infty} (r_\tau^\epsilon r_{\tau+i-j}^\epsilon + r_{\tau+i}^\epsilon r_{\tau-j}^\epsilon).$$

The assertion follows immediately upon combining the limits of I_1, I_2, and I_3. $\Diamond$

Remark 3.1 As can be seen from (3.3), the asymptotic covariance σ_{ij} consists of two parts. The first part — namely, the first term in (3.3) — is completely due to the presence of sinusoids in $\{y_t\}$, while the second part, i.e., the last two terms in (3.3),

comes from the classical results for a time series with continuous spectrum (Brockwell and Davis, 1987, Proposition 7.3.1).

Using Lemma 3.1, we now present the central limit theorem (CLT) for $\tilde{r}_j$ in (3.2).

Theorem 3.1 *Assume that* $E(\xi_t^4) = \kappa\sigma_\xi^4 < \infty$. *Then,* $n^{1/2}(\tilde{r}_j - r_j^y)$, $(j = 0, 1, \ldots, p)$, *are asymptotically jointly normal with mean zero and covariance matrix* $[\sigma_{ij}]$, $(i, j = 0, 1, \ldots, p)$, *where* σ_{ij} *is defined by* (3.3).

Since the proof of Theorem 3.1 is long and tedious, we break it up into a series of lemmas. For this purpose, we note that equipped with the Cramér-Wold device (Brockwell and Davis, 1987, Proposition 6.3.1), all we need is to show that for any $\{\lambda_j, j = 0, 1, \ldots, p\} \neq \{0\}$, it holds that $n^{1/2}\sum \lambda_j(\tilde{r}_j - r_j^y) \xrightarrow{D} \mathrm{N}(0, \sigma^2)$, where

$$\sigma^2 := \sum_{i,j=0}^{p} \lambda_i \lambda_j \sigma_{ij} > 0. \tag{3.7}$$

This, according to (3.6), can be accomplished by showing that

$$n^{-1/2} \sum_{t=1}^{n} \zeta_t \xrightarrow{D} \mathrm{N}(0, \sigma^2) \tag{3.8}$$

where

$$\zeta_t := \sum_{j=0}^{p} \lambda_j \zeta_{jt} \tag{3.9}$$

and ζ_{jt} is defined by (3.5).

We first consider the case where only finite many ψ_j in (1.2) are nonzero. The following lemma shows that in this case the sum in (3.8) is asymptotically equivalent to a martingale (i.e., a sum of martingale differences).

Lemma 3.2 *Suppose that* $E(\xi_t^4) = \kappa\sigma_\xi^4 < \infty$, *and that the sequence* $\{\psi_j\}$ *in* (1.2) *has a finite length, i.e.,* $\psi_j \equiv 0$ *for all* $|j| > m$, *with* m *being a positive integer. Then,*

$$\sum_{t=1}^{n} \zeta_t = \sum_{t=1}^{n} M_t + O_P(1)$$

where $\{M_t\}$ *is a martingale difference sequence with respect to the filtration* $\mathcal{F}_t$ *generated by* $\{\xi_s, s \leq t\}$ *for* $t \geq 0$.

PROOF. Notice that from (3.5) we can write

$$\sum_{t=1}^{n} \zeta_{jt} = J_1 + J_2 + J_3 \tag{3.10}$$

where

$$J_1 := \sum_{t=1}^{n} x_t \epsilon_{t+j}, \quad J_2 := \sum_{t=1}^{n} x_{t+j} \epsilon_t, \quad \text{and} \quad J_3 := \sum_{t=1}^{n} (\epsilon_{t+j}\epsilon_t - r_j^{\epsilon}).$$

By definition, $\epsilon_t = \sum \psi_u \xi_{t-u}$. Therefore, we have

$$\begin{aligned} J_3 &= \sum_{u,v=-m}^{m} \psi_u \psi_v \sum_{t=1}^{n} (\xi_{t+j-u}\xi_{t-v} - \sigma_\xi^2 \delta_{u-v-j}) \\ &= \sum_{u,v=-m}^{m} \psi_u \psi_v \sum_{t=j-u+1}^{j-u+n} (\xi_t \xi_{t+u-v-j} - \sigma_\xi^2 \delta_{u-v-j}) \end{aligned}$$

where δ_u is Kronecker's delta function, i.e., $\delta_u = 0$ for $u \neq 0$ and $\delta_0 = 1$, and the second equality is obtained by replacing $t+j-u$ with t in the first expression. Given u and v, the variance of $\xi_t \xi_{t+u-v-j} - \sigma_\xi^2 \delta_{u-v-j}$ is finite and independent of t. Therefore, any weighted sum of these quantities over a *finite* region[1] of (t, u, v) can be written as $O_P(1)$. Armed with this fact, we add and subtract a finite number of $\xi_t \xi_{t+u-v-j} - \sigma_\xi^2 \delta_{u-v-j}$ in the last expression of J_3 and obtain

$$J_3 = \sum_{u,v=-m}^{m} \psi_u \psi_v \sum_{t=1}^{n} (\xi_t \xi_{t+u-v-j} - \sigma_\xi^2 \delta_{u-v-j}) + O_P(1).$$

Moreover, using the substitution $\tau = -(u-v-j)$ and the fact (see equation (1.4)) that $r_{\tau-j}^{\epsilon} = \sigma_\xi^2 \sum \psi_{v-\tau+j}\psi_v$, we can also write

$$\begin{aligned} J_3 &= \sigma_\xi^{-2} \sum_{\tau=-2m+j}^{2m+j} r_{\tau-j}^{\epsilon} \sum_{t=1}^{n} (\xi_t \xi_{t-\tau} - \sigma_\xi^2 \delta_\tau) + O_P(1) \\ &= \sigma_\xi^{-2} \left(\sum_{\tau=-2m+j}^{-1} + \sum_{\tau=0}^{2m+j} \right) r_{\tau-j}^{\epsilon} \sum_{t=1}^{n} (\xi_t \xi_{t-\tau} - \sigma_\xi^2 \delta_\tau) + O_P(1). \end{aligned} \tag{3.11}$$

Replacing τ with $-\tau$ and then substituting $t+\tau$ by t, the first term in (3.11) becomes

$$\sigma_\xi^{-2} \sum_{\tau=1}^{2m-j} r_{\tau+j}^{\epsilon} \sum_{t=1}^{n} \xi_t \xi_{t+\tau} = \sigma_\xi^{-2} \sum_{\tau=1}^{2m-j} r_{\tau+j}^{\epsilon} \sum_{t=\tau+1}^{\tau+n} \xi_t \xi_{t-\tau}.$$

[1]Namely, the size of the region is finite and independent of n.

Therefore, by adding and subtracting a finite number of $\xi_t\xi_{t-\tau}$ in the last expression, we can write the first term in (3.11) as

$$\sigma_\xi^{-2} \sum_{\tau=1}^{2m-j} r_{\tau+j}^{\epsilon} \sum_{t=1}^{n} \xi_t \xi_{t-\tau} + O_P(1). \tag{3.12}$$

Notice that $r_{\tau-j}^{\epsilon} \equiv 0$ for $\tau > 2m + j$ and $r_{\tau+j}^{\epsilon} \equiv 0$ for $\tau > 2m - j$. As a consequence, we can extend the summations in (3.11) and (3.12) to $0 \leq \tau < \infty$ and write

$$J_3 = \sum_{t=1}^{n} \sum_{\tau=0}^{\infty} B_{j\tau}(\xi_t \xi_{t-\tau} - \sigma_\xi^2 \delta_\tau) + O_P(1) \tag{3.13}$$

where

$$B_{j0} := r_j^{\epsilon}/\sigma_\xi^2, \quad \text{and} \quad B_{j\tau} := \ (r_{\tau+j}^{\epsilon} + r_{\tau-j}^{\epsilon})/\sigma_\xi^2 \quad \text{for } \tau > 0. \tag{3.14}$$

Note that $B_{j\tau} \equiv 0$ for $\tau > 2m + p$. Similarly, J_1 can be expressed as

$$J_1 = \sum_{u=-m}^{m} \psi_u \sum_{t=1}^{n} \xi_{t+j-u} x_t = \sum_{u=-m}^{m} \psi_u \sum_{t=j-u+1}^{j-u+n} \xi_t x_{t+u-j}$$

where the second equality is obtained by substituting $t + j - u$ with t. Since the variance of $\xi_t x_{t+u-j}$ can be bounded by a constant, an argument analogous to the one we employed earlier leads to the representation

$$\begin{aligned} J_1 &= \sum_{u=-m}^{m} \psi_u \sum_{t=1}^{n} \xi_t x_{t+u-j} + O_P(1) \\ &= \sum_{t=1}^{n} \xi_t \sum_{u=-m}^{m} \psi_u x_{t+u-j} + O_P(1). \end{aligned} \tag{3.15}$$

In a similar way, we can write J_2 as

$$J_2 = \sum_{t=1}^{n} \xi_t \sum_{u=-m}^{m} \psi_u x_{t+u+j} + O_P(1).$$

This, in connection with (3.13), (3.15), and (3.10), gives the following representation

$$\sum_{t=1}^{n} \zeta_{jt} = \sum_{t=1}^{n} \left\{ A_{jt}\xi_t + \sum_{\tau=0}^{\infty} B_{j\tau}(\xi_t \xi_{t-\tau} - \sigma_\xi^2 \delta_\tau) \right\} + O_P(1) \tag{3.16}$$

where

$$A_{jt} := \sum_{u=-m}^{m} \psi_u (x_{t+u+j} + x_{t+u-j}). \tag{3.17}$$

Now let us define

$$M_t := A_t\xi_t + \sum_{\tau=0}^{\infty} B_\tau(\xi_t\xi_{t-\tau} - \sigma_\xi^2\delta_\tau) \tag{3.18}$$

with $A_t := \sum \lambda_j A_{jt}$ and $B_\tau := \sum \lambda_j B_{j\tau}$. Then, from (3.9) and (3.16), we obtain

$$\sum_{t=1}^{n} \zeta_t = \sum_{j=0}^{p} \lambda_j \sum_{t=1}^{n} \zeta_{jt} = \sum_{t=1}^{n} M_t + O_P(1).$$

Since the A's and B's are constants, it can be easily shown that $\{M_t\}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_t$ generated by $\{\xi_s, s \leq t\}$. Indeed, the measurability of M_t with respect to $\mathcal{F}_t$ is obvious from the definition (3.18). Furthermore, we have

$$E(M_t|\mathcal{F}_{t-1}) = A_t E(\xi_t) + B_0 E(\xi_t^2 - \sigma_\xi^2) + \sum_{\tau=1}^{\infty} B_\tau \xi_{t-\tau} E(\xi_t) = 0.$$

The lemma is thus proved. ◊

As indicated by Lemma 3.2, the central limit theorem (3.8) can be established if we can prove the same result for the martingale difference sequence $\{M_t\}$. To this end, we consider the quantities

$$V_n^2 := \sum_{t=1}^{n} E(M_t^2|\mathcal{F}_{t-1}) \quad \text{and} \quad s_n^2 := E(V_n^2). \tag{3.19}$$

The following lemma shows that both $n^{-1}V_n^2$ and $n^{-1}s_n^2$ converge to the same limit σ^2 as $n \to \infty$.

Lemma 3.3 *Assume that the conditions in Lemma* 3.2 *are satisfied, and let* V_n^2 *and* s_n^2 *be defined by* (3.19). *Then, as* $n \to \infty$, $n^{-1}V_n^2 \xrightarrow{P} \sigma^2$, $n^{-1}s_n^2 \to \sigma^2$, *and hence* $V_n^2/s_n^2 \xrightarrow{P} 1$, *where* σ^2 *is given by* (3.7).

PROOF. From (3.18), it is easy to show by straightforward computations that

$$\begin{aligned} E(M_t^2|\mathcal{F}_{t-1}) \;=\; & \sigma_\xi^2 A_t^2 + 2B_0 E(\xi^3)\, A_t + 2\sigma_\xi^2 \sum_{\tau=1}^{\infty} B_\tau A_t \xi_{t-\tau} + B_0^2 E(\xi^2 - \sigma_\xi^2)^2 \\ & + 2B_0 E(\xi^3) \sum_{\tau=1}^{\infty} B_\tau \xi_{t-\tau} + \sigma_\xi^2 \sum_{s,\tau=1}^{\infty} B_s B_\tau \xi_{t-s}\xi_{t-\tau}. \end{aligned} \tag{3.20}$$

To verify that $n^{-1}V_n^2$ converges to σ^2, we first note that for any fixed τ and s

$$n^{-1}\sum_{t=1}^{n} x_{t+\tau}x_{t+s} \to \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \cos(\omega_k(\tau - s)) = r_{\tau-s}^x \tag{3.21}$$

as $n \to \infty$. Therefore, from (3.17) and (3.21), we can write

$$\begin{aligned} & n^{-1}\sum_{t=1}^{n} A_t^2 \\ &= \sum_{i,j}\lambda_i\lambda_j \sum_{u,v}\psi_u\psi_v \; n^{-1}\sum_{t=1}^{n}(x_{t+u+i} + x_{t+u-i})(x_{t+u+j} + x_{t+u-j}) \\ &\to \sum_{i,j}\lambda_i\lambda_j \sum_{u,v}\psi_u\psi_v(r_{u-v+i-j}^x + r_{u-v+i+j}^x + r_{u-v-i-j}^x + r_{u-v-i+j}^x). \end{aligned} \tag{3.22}$$

For any ω, it is easy to verify that $\sum\sum \psi_u\psi_v \sin(\omega(u-v)) = \Im\{|\sum \psi_u \exp(i\omega u)|^2\} = 0$ and, from (1.4),

$$\sum_{u,v}\psi_u\psi_v \cos(\omega(u-v)) = \sigma_\xi^{-2}\sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega\tau).$$

Therefore, it follows that for any s,

$$\begin{aligned} \sum_{u,v}\psi_u\psi_v r_{u-v+s}^x &= \sum_{k=1}^{q}\tfrac{1}{2}\beta_k^2 \sum_{u,v}\psi_u\psi_v \cos(\omega_k(u-v+s)) \\ &= \sigma_\xi^{-2}\sum_{k=1}^{q}\tfrac{1}{2}\beta_k^2\cos(\omega_k s)\sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon \cos(\omega_k\tau). \end{aligned}$$

Using this expression in (3.22) yields

$$n^{-1}\sum_{t=1}^{n} A_t^2 \to \sigma_\xi^{-2}\sum_{i,j=0}^{p}\lambda_i\lambda_j\left\{\sum_{k=1}^{q} 2\beta_k^2\cos(\omega_k i)\cos(\omega_k j)\sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon\cos(\omega_k\tau)\right\}. \tag{3.23}$$

Similarly, since $n^{-1}\sum_{t=1}^{n} x_{t+s} \to 0$ for any s, it follows from (3.17) that

$$n^{-1}\sum_{t=1}^{n} A_t = \sum_j \lambda_j \sum_u \psi_u \; n^{-1}\sum_{t=1}^{n}(x_{t+u+j} + x_{t+u-j}) \to 0. \tag{3.24}$$

In addition, it can be shown (An, *et al.*, 1983) that

$$n^{-1}\left|\sum_{t=1}^{n}\cos(\omega(t+s)+\phi)\xi_{t-\tau}\right| = O_P\left(\sqrt{n^{-1}\log n}\right) \xrightarrow{P} 0$$

for any fixed ω, ϕ, s, and τ. This implies that for any s and τ,

$$n^{-1}\sum_{t=1}^{n} x_{t+s}\xi_{t-\tau} \xrightarrow{P} 0$$

as $n \to \infty$. Therefore, for any τ, we obtain

$$n^{-1}\sum_{t=1}^{n} A_t\xi_{t-\tau} = \sum_j \lambda_j \sum_u \psi_u \; n^{-1}\sum_{t=1}^{n}(x_{t+u+j} + x_{t+u-j})\xi_{t-\tau} \xrightarrow{P} 0. \tag{3.25}$$

Finally, by the law of large numbers, we have

$$n^{-1}\sum_{t=1}^{n}\xi_{t-\tau} \xrightarrow{P} 0 \quad \text{and} \quad n^{-1}\sum_{t=1}^{n}\xi_{t-s}\xi_{t-\tau} \xrightarrow{P} \sigma_\xi^2\delta_{\tau-s}$$

for any s and τ. This, together with (3.23), (3.24), (3.25), and (3.20), implies that

$$\begin{aligned} n^{-1}V_n^2 \xrightarrow{P} & \sum_{i,j=0}^{p}\lambda_i\lambda_j\left\{\sum_{k=1}^{q}2\beta_k^2\cos(\omega_k i)\cos(\omega_k j)\sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon\cos(\omega_k\tau)\right\} \\ & + B_0^2E(\xi^2-\sigma_\xi^2)^2 + \sigma_\xi^4\sum_{\tau=1}^{\infty}B_\tau^2. \end{aligned}$$

Using (3.14), it can be easily verified that

$$\begin{aligned} & B_0^2E(\xi^2-\sigma_\xi^2)^2 + \sigma_\xi^4\sum_{\tau=1}^{\infty}B_\tau^2 \\ & \quad = \sum_{i,j=0}^{p}\lambda_i\lambda_j\left\{(\kappa-3)r_i^\epsilon r_j^\epsilon + \sum_{\tau=-\infty}^{\infty}(r_\tau^\epsilon r_{\tau+i-j}^\epsilon + r_{\tau+i}^\epsilon r_{\tau-j}^\epsilon)\right\}. \end{aligned} \tag{3.26}$$

Thus we have proved that $n^{-1}V_n^2 \xrightarrow{P} \sigma^2$. To show that $n^{-1}s_n^2 \to \sigma^2$, we note from (3.20) that

$$\begin{aligned} n^{-1}s_n^2 &= n^{-1}\sum_{t=1}^{n}E\{E(M_t^2|\mathcal{F}_{t-1})\} \\ &= \sigma_\xi^2\; n^{-1}\sum_{t=1}^{n}A_t^2 + B_0^2E(\xi^2-\sigma_\xi^2)^2 + \sigma_\xi^4\sum_{\tau=1}^{\infty}B_\tau^2. \end{aligned}$$

The assertion follows immediately from (3.23) and (3.26). ◊

Armed with these lemmas, we claim in the following lemma that (3.8) holds when $\{\psi_j\}$ has a finite length.

Lemma 3.4 *Suppose that* $E(\xi_t^4) = \kappa\sigma_\xi^4 < \infty$, *and that the sequence* $\{\psi_j\}$ *in* (1.2) *has a finite length. Then, the central limit theorem* (3.8) *holds as* $n \to \infty$.

PROOF. By Lemma 3.2, it suffices to show that

$$n^{-1/2}\sum_{t=1}^{n} M_t \xrightarrow{D} \mathrm{N}(0,\sigma^2). \tag{3.27}$$

As Lemma 3.3 has been proved, all we need, according to the central limit theorem for martingales (Brown, 1971), is to verify the Lindeberg condition (Brown, 1971, eq. 2)

$$s_n^{-2}\sum_{t=1}^{n} E\{M_t^2 I(|M_t| \geq \varepsilon s_t)\} \to 0$$

as $n \to \infty$, where $I(\cdot)$ is the indicator. Since $n^{-1}s_n^2 \to \sigma^2$, as shown in Lemma 3.3, the Lindeberg condition is equivalent to

$$n^{-1}\sum_{t=1}^{n} E\{M_t^2 I(|M_t| \geq \varepsilon s_t)\} \to 0.$$

Therefore, it suffices to verify that $E\{M_t^2 I(|M_t| \geq \varepsilon s_t)\} \to 0$ as $t \to \infty$ for any $\varepsilon > 0$. To this end, we note that $|A_t| \leq A$ for some $A > 0$ and all t. Therefore,

$$|M_t| \leq U_t := A|\xi_t| + \left|\sum_{\tau=0}^{\infty} B_\tau(\xi_t\xi_{t-\tau} - \sigma_\xi^2\delta_\tau)\right|.$$

Moreover, the convergence of $t^{-1}s_t^2$ to $\sigma^2 > 0$ as $t \to \infty$ implies that $s_t \geq \varepsilon\, t^{1/2}$ for small $\varepsilon > 0$ and large t. Combining these results, we can write

$$\begin{aligned} E\{M_t^2 I(|M_t| \geq \varepsilon s_t)\} &\leq E\{U_t^2 I(U_t \geq \varepsilon^2 t^{1/2})\} \\ &= E\{U_0^2 I(U_0 \geq \varepsilon^2 t^{1/2})\} \to 0 \end{aligned}$$

as $t \to \infty$, where the equality is due to the stationarity of $\{U_t\}$ and the limit to the finiteness of $E(U_0^2)$. Applying Theorem 2 of Brown (1971) proves (3.27), and hence the lemma. ◊

We now complete the proof of Theorem 3.1 by showing that (3.8) also holds if $\{\psi_j\}$ has an infinite length.

PROOF OF THEOREM 3.1. When $\{\psi_j\}$ is of infinite length, the central limit theorem (3.8) can be obtained by following the proof of Proposition 7.3.3 (Brockwell and Davis,

1987). In fact, for any $m > 0$, let us define

$$\epsilon_t^m := \sum_{j=-m}^{m} \psi_j \xi_{t-j}, \quad r_j^m := E(\epsilon_{t+j}^m \epsilon_t^m)$$

$$\zeta_{jt}^m := x_t \epsilon_{t_j}^m + x_{t+j} \epsilon_t^m + \epsilon_{t+j}^m \epsilon_t^m - r_j^m, \quad \text{and} \quad \zeta_t := \sum_{j=0}^{p} \lambda_j \zeta_{jt}.$$

Lemma 3.4 guarantees that for any fixed m,

$$n^{-1/2} \sum_{t=1}^{n} \zeta_t^m \xrightarrow{D} S_m \sim \mathrm{N}(0, \sigma_m^2).$$

as $n \to \infty$, where $\sigma_m^2 := \sum\sum \lambda_i \lambda_j \sigma_{ij}^m$ and σ_{ij}^m is defined by (3.3) with the autocovariance function r_τ^ϵ replaced by r_τ^m. It is not difficult to verify that $\sigma_m^2 \to \sigma^2$ as $m \to \infty$. This implies that $S_m \xrightarrow{D} \mathrm{N}(0, \sigma^2)$. Moreover, a straightforward calculation shows that

$$\lim_{m \to \infty} \limsup_{n \to \infty} \; n^{-1} E \left| \sum_{t=1}^{n} (\zeta_t^m - \zeta_t) \right|^2 = 0.$$

Using Chebychev's inequality, we obtain

$$\lim_{m \to \infty} \limsup_{n \to \infty} \; \mathrm{pr} \left(n^{-1/2} \left| \sum_{t=1}^{n} (\zeta_t^m - \zeta_t) \right| > \varepsilon \right) = 0$$

for any $\varepsilon > 0$. The proof is then completed by applying Proposition 6.3.9 (Brockwell and Davis, 1987). ◊

With the help of Theorem 3.1, we now claim the asymptotic normality of the sample autocovariances $\hat{r}_j$ in (3.1).

Theorem 3.2 *Assume that* $E(\xi_t^4) = \kappa \sigma_\xi^4 < \infty$. *Then,* $n^{1/2}(\hat{r}_j - r_j^y)$, $(j = 0, 1, \ldots, p)$, *are asymptotically jointly normal with mean zero and covariance matrix* $[\sigma_{ij}]$, $(i, j = 0, 1, \ldots, p)$, *where* σ_{ij} *is defined by* (3.3).

PROOF. Since the asymptotic normality has been proved in Theorem 3.1 for $\tilde{r}_j$, it suffices to show that $n^{1/2}(\tilde{r}_j - \hat{r}_j) = o_P(1)$ for $j = 1, \ldots, p$. To this end, we note that

$$n^{1/2}(\tilde{r}_j - \hat{r}_j) = n^{-1/2} \sum_{t=n-j+1}^{n} (x_{t+j} x_t + x_{t+j} \epsilon_t + x_t \epsilon_{t+j} + \epsilon_{t+j} \epsilon_t). \tag{3.28}$$

Since $|x_t| \leq \beta := \sum \beta_k$ for all t, the first term in (3.28) is $o(1)$. For the same reason,

$$E\left|\sum_{t=n-j+1}^{n} x_{t+j}\epsilon_t\right| \leq \beta \sum_{t=n-j+1}^{n} E|\epsilon_t| \leq p\beta E|\epsilon_0| < \infty.$$

By Markov's inequality, this implies that the second term in (3.28) is $o_P(1)$. The same conclusion applies to the third term in (3.28). Finally, since

$$E\left|\sum_{t=n-j+1}^{n} \epsilon_{t+j}\epsilon_t\right| \leq jE|\epsilon_j\epsilon_0| \leq p\, r_0^{\epsilon} < \infty,$$

the last term in (3.28) is also $o_P(1)$. The theorem is thus proved. ◊

As a direct consequence of Theorem 3.2, the asymptotic normality of the sample autocorrelation $\hat{\rho}_j := \hat{r}_j/\hat{r}_0$ can also be established as follows.

Corollary 3.1 *Suppose that the conditions in Theorem* 3.2 *are satisfied. Then, for any fixed* $j \geq 1$, $n^{1/2}(\hat{\rho}_j - \rho_j)$ *is asymptotically normal with mean zero and variance* v_j, *where* $\rho_j := r_j^y/r_0^y$,

$$v_j := (\rho_j^2\sigma_{00} - 2\rho_j\sigma_{0j} + \sigma_{jj})/(r_0^y)^2, \tag{3.29}$$

and σ_{ij} *is given by* (3.3).

PROOF. The assertion follows immediately from Theorem 3.1 and the "delta method" (Brockwell and Davis, 1987, Proposition 6.4.3). ◊

We would like to end this section by making the following remarks.

Remark 3.2 The asymptotic normality of SACF has been proved by Mackisack and Poskitt (1989) for the simplest case where $q = 1$ and $\{\epsilon_t\}$ is an i.i.d. random sequence (a single sinusoid in white noise). Therefore, Theorem 3.1 and Theorem 3.2 extend this result to the case of *multiple* sinusoids in *colored* noise.

Remark 3.3 As we remarked earlier, the proof of Theorem 3.1 and Theorem 3.2 is based upon the assumption that the phases ϕ_k are constants. Since the asymptotic

distribution of $\tilde{r}_j$ and $\hat{r}_j$ does not depend on ϕ_k, the same conclusion also holds for random phases.

Remark 3.4 The proof of Theorem 3.2 indicates that the same central limit theorem holds for any sample autocovariances of the form

$$\check{r}_j = n^{-1} \sum_{t=u}^{n+v} y_{t+j} y_t$$

where the integers $u = u(j)$ and $v = v(j)$ are independent of n.

3.2 CLT for SACF from Filtered Process

In this section, we consider the sample autocovariance function from a filtered time series. We show that the asymptotic normality remains valid after filtering, provided that the filter is *strictly stable.* The definition of a strictly stable filter is as follows (Ljung, 1987).

Definition 3.1 *A linear time-invariant causal filter* $\{h_j\}$, $(j = 0, 1, \ldots)$, *is said to be* strictly stable *if* $\sum j|h_j| < \infty$.

Suppose that $\{h_j\}$ is a strictly stable filter, and denote its transfer function by

$$H(\omega) := \sum_{j=0}^{\infty} h_j e^{-ij\omega}.$$

For a given time series $\{y_1, \ldots, y_n\}$ from (1.1), define the filtered time series by

$$\hat{y}_t(h) := \sum_{j=0}^{t-1} h_j y_{t-j} \qquad (t = 1, \ldots, n).$$

The sample autocovariances of the filtered times series are defined by

$$\hat{r}_j(h) := n^{-1} \sum_{t=1}^{n-j} \hat{y}_{t+j}(h)\, \hat{y}_t(h) \qquad (j = 0, 1, \ldots, p). \tag{3.30}$$

Note that the filtered time series $\{\hat{y}_1(h), \ldots, \hat{y}_n(h)\}$ depends completely on the given data record $\{y_1, \ldots, y_n\}$. For convenience, we introduce $\{y_t(h)\}$ where

$$y_t(h) := \sum_{j=0}^{\infty} h_j y_{t-j}.$$

It is clear that $y_t(h)$ requires the entire history of the process $\{y_t\}$ up to time t. From (1.1), we can rewrite $\{y_t(h)\}$ as

$$y_t(h) = \sum_{k=1}^{q} \beta_k(h) \cos(\omega_k t + \phi_k(h)) + \epsilon_t(h) \tag{3.31}$$

where $\beta_k(h) := \beta_k |H(\omega_k)|$ and $\phi_k(h) := \phi_k + \arg\{H(\omega_k)\}$ are the amplitudes and phases of the filtered sinusoids, and $\{\epsilon_t(h)\}$ is the filtered noise as specified by

$$\epsilon_t(h) := \sum_{j=0}^{\infty} h_j \epsilon_{t-j} = \sum_{j=-\infty}^{\infty} \mu_j \xi_{t-j}$$

where $\mu_j := \sum \psi_u h_{j-u}$. Clearly, the filtered signal remains a sum of q sinusoids with the same frequencies, while the filtered noise $\{\epsilon_t(h)\}$ is still a linear process.

Let $r_\tau^y(h)$ and $r_\tau^\epsilon(h)$ denote the autocovariance functions of $\{y_t(h)\}$ and $\{\epsilon_t(h)\}$, respectively. Then, we have the following theorem regarding the asymptotic normality of the sample autocovariances $\hat{r}_j(h)$ in (3.30).

Theorem 3.3 *Suppose that the filter $\{h_j\}$ is strictly stable, and that $E(\xi_t^4) = \kappa\sigma_\xi^4 < \infty$. Then, as n tends to infinity, $n^{1/2}(\hat{r}_j(h) - r_j^y(h))$, $(j = 0, 1, \ldots, p)$, are asymptotically jointly normal with mean zero and covariance matrix $[\sigma_{ij}(h)]$, $(i, j = 0, 1, \ldots, p)$, where $\sigma_{ij}(h)$ can be represented by (3.3), except that the β_k and r_τ^ϵ are replaced by $\beta_k(h)$ and $r_\tau^\epsilon(h)$, respectively.*

PROOF. Because of the representation (3.31) and Theorem 3.2, the asymptotic normality can be established for

$$\tilde{r}_j(h) := n^{-1} \sum_{t=1}^{n-j} y_{t+j}(h)\, y_t(h) \qquad (j = 0, 1, \ldots, p).$$

Therefore, it suffices to show that $n^{1/2}(\tilde{r}_j(h) - \hat{r}_j(h)) = o_P(1)$. Let us define

$$\tilde{y}_t(h) := \sum_{u=t}^{\infty} h_u y_{t-u}.$$

Since $y_t(h) = \hat{y}_t(h) + \tilde{y}_t(h)$, a simple calculation shows that

$$\begin{aligned} n^{1/2}(\tilde{r}_j(h) - \hat{r}_j(h)) &= n^{-1/2} \sum_{t=1}^{n-j} \{\hat{y}_{t+j}(h)\, \tilde{y}_t(h) + \tilde{y}_{t+j}(h)\, \hat{y}_t(h) + \tilde{y}_{t+j}(h)\, \tilde{y}_t(h)\} \\ &:= I_1 + I_2 + I_3. \end{aligned}$$

Using (3.31), we can write I_1 as

$$\begin{aligned} I_1 &= n^{-1/2}\sum_{t=1}^{n-j}(\hat{x}_{t+j}\tilde{x}_t + \hat{x}_{t+j}\tilde{\epsilon}_t + \tilde{x}_{t+j}\hat{\epsilon}_t + \hat{\epsilon}_{t+j}\hat{\epsilon}_t) \\ &:= T_1 + T_2 + T_3 + T_4 \end{aligned}$$

where $\hat{x}_t$, $\hat{\epsilon}_t$, $\tilde{x}_t$, and $\tilde{\epsilon}_t$ are similarly defined (with the argument h being omitted for brevity) as $\hat{y}_t(h)$ and $\tilde{y}_t(h)$. By definition, we have

$$\begin{aligned} T_1 &= n^{-1/2}\sum_{t=1}^{n-j}\hat{x}_{t+j}\tilde{x}_t \\ &= n^{-1/2}\sum_{t=1}^{n-j}\sum_{u=0}^{t+j-1}\sum_{v=t}^{\infty}h_u h_v x_{t+j-u}x_{t-v}. \end{aligned}$$

Since $|x_t|$ is bounded by $\beta = \sum\beta_k$ and $H := \sum|h_u|$ is finite, it follows that

$$|T_1| \le n^{-1/2}H\beta^2\sum_{t=1}^{n}\sum_{v=t}^{\infty}|h_v| \le n^{-1/2}H\beta^2\sum_{v=1}^{\infty}v|h_v|$$

and hence $T_1 = o(1)$. It can also be shown from (3.31) that

$$E|T_2| \le n^{-1/2}H\beta E|\epsilon_0|\sum_{v=1}^{\infty}v|h_v|.$$

Applying Markov's inequality yields $T_2 = o_P(1)$. In a similar way, we can show that $T_3 = o_P(1)$ and $T_4 = o_P(1)$. Combining these results gives $I_1 = o_P(1)$. The same results can be obtained for I_2 and I_3 by an analogous argument. ♢

Remark 3.5 The asymptotic normality of the sample autocorrelation

$$\hat{\rho}_j(h) := \hat{r}_j(h)/\hat{r}_0(h)$$

can be obtained in the same way as Corollary 3.1, except that ρ_j and σ_{ij}, and r_0^y are replaced by $\rho_j(h) := r_j^y(h)/r_0^y(h)$, $\sigma_{ij}(h)$, and $r_0^y(h)$, respectively.

Remark 3.6 As claimed in Remark 3.4, the same central limit theorem holds for any sample autocovariances of the form

$$\check{r}_j(h) = n^{-1}\sum_{t=u}^{n+v}y_{t+j}(h)\,y_t(h)$$

where the integers $u = u(j)$ and $v = v(j)$ are independent of n.

3.3 Uniform Strong Consistency of SACF

In this section, we consider the uniform consistency of SACF after parametric filtering. More precisely, let $\{h_j(\alpha)\}$, $(j = 0, 1, \ldots)$, be a parametric causal linear time-invariant filter, where α is a parameter (possibly a vector) that takes on values in $\mathcal{A}$. Following the notation in Section 3.2, we define the filtered time series by

$$\hat{y}_t(\alpha) := \sum_{j=0}^{t-1} h_j(\alpha)\, y_{t-j} \qquad (t = 1, \ldots, n) \tag{3.32}$$

and the sample autocovariances of the filtered time series by

$$\hat{r}_\tau(\alpha) := n^{-1} \sum_{t=1}^{n-\tau} \hat{y}_{t+\tau}(\alpha)\, \hat{y}_t(\alpha). \tag{3.33}$$

We would like to show that under certain conditions $\hat{r}_\tau(\alpha)$ converges to $r_\tau^y(\alpha)$ almost surely as $n \to \infty$, and uniformly in $\alpha \in \mathcal{A}$, where $r_\tau^y(\alpha)$ is the autocovariance function of the process

$$y_t(\alpha) := \sum_{j=0}^{\infty} h_j(\alpha)\, y_{t-j}. \tag{3.34}$$

Recall that the noise $\{\epsilon_t\}$ is a linear process as defined in (1.2). In addition, we assume that the filter $\{h_j(\alpha)\}$ is uniformly strictly stable according to the following definition.

Definition 3.2 *A causal linear filter* $\{h_j(\alpha)\}$, $(j = 0, 1, \ldots)$, *is said to be* uniformly strictly stable *if there exist constants* $c_j > 0$ *such that* $\sum j c_j < \infty$ *and* $|h_j(\alpha)| \leq c_j$ *for all* $j = 0, 1, \ldots$, *and uniformly for all* $\alpha \in \mathcal{A}$.

We first provide a general result concerning the uniform strong consistency of the sample cross-covariances between two differently filtered time series.

Theorem 3.4 *Let* $\{h_j(\alpha)\}$ *and* $\{g_j(\alpha)\}$ *be uniformly strictly stable filters with transfer functions* $H(\omega; \alpha)$ *and* $G(\omega; \alpha)$, *respectively. Let* $F(\omega)$ *be the spectral distribution*

function of $\{\epsilon_t\}$. *Then, uniformly in* $\alpha \in \mathcal{A}$,

$$n^{-1}\sum_{t=1}^{n}\left(\sum_{j=0}^{t+\tau-1} h_j(\alpha)\, y_{t+\tau-j}\right)\left(\sum_{j=0}^{t-1} g_j(\alpha)\, y_{t-j}\right)$$
$$\stackrel{a.s.}{\to} \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2\, \Re\{H(\omega_k;\alpha)\,\overline{G(\omega_k;\alpha)}\, e^{i\tau\omega_k}\} + \int_{-\pi}^{\pi} H(\omega;\alpha)\,\overline{G(\omega;\alpha)}\, e^{i\tau\omega}\, dF(\omega)$$

as $n \to \infty$ *for any* $\tau \geq 0$, *where* $\Re\{\cdot\}$ *stands for the real part and the overbar for the complex conjugate of a complex number.*

Before we prove this theorem, let us introduce the following lemma concerning the consistency of the sample covariance of $\{x_t\}$ and $\{\epsilon_t\}$.

Lemma 3.5 *For any fixed* u, v, *and* w, *it is true that*

$$n^{-1}\sum_{t=w}^{n} x_{t-u}x_{t-v} \stackrel{a.s.}{\to} r^x_{v-u} \quad \textit{and} \quad n^{-1}\sum_{t=w}^{n} \epsilon_{t-u}\epsilon_{t-v} \stackrel{a.s.}{\to} r^\epsilon_{v-u}$$

as $n \to \infty$, *where* r^x_τ *and* r^ϵ_τ *are the autocovariance function of* $\{x_t\}$ *and* $\{\epsilon_t\}$, *respectively. It is also true that*

$$n^{-1}\sum_{t=w}^{n} \epsilon_{t-u}x_{t-v} \stackrel{a.s.}{\to} 0$$

for any fixed u, v, *and* w.

PROOF. The first limit can be proved upon using the trigonometric identity (1.14) in connection with the fact that

$$\lim_{n\to\infty} n^{-1}\sum_{t=w}^{n} \cos(\omega_k(2t-u-v)+2\phi_k) = 0 \qquad a.s.$$

for any k and

$$\lim_{n\to\infty} n^{-1}\sum_{t=w}^{n} \cos(\omega_k(t-u)+\phi_k)\cos(\omega_{k'}(t-v)+\phi_{k'}) = 0 \qquad a.s.$$

for any $k \neq k'$. The second limit in the lemma is due to the strong ergodicity of $\{\epsilon_t\}$ (Hannan, 1970, pp. 203–204; Karlin and Taylor, 1975). Finally, to prove the last limit in the lemma, we note that

$$\left|\sum_{t=w}^{n} \epsilon_{t-u}x_{t-v}\right| \leq \sum_{k=1}^{q}\beta_k\left\{\left|\sum_{t=w}^{n}\epsilon_{t-u}\cos(\omega_k t)\right| + \left|\sum_{t=w}^{n}\epsilon_{t-u}\sin(\omega_k t)\right|\right\}.$$

The assertion follows from the fact that (An, *et al.*, 1983)

$$\lim_{n\to\infty} n^{-1}\left|\sum_{t=w}^{n}\epsilon_{t-u}\cos(\omega t)\right| = \lim_{n\to\infty} n^{-1}\left|\sum_{t=w}^{n}\epsilon_{t-u}\sin(\omega t)\right| = 0 \qquad a.s.$$

for any ω, u, and w. $\Diamond$

Equipped with this lemma, we now prove Theorem 3.4. Throughout the proof we drop out the argument α for brevity.

PROOF OF THEOREM 3.4. For any sequence $\{w_t\}$, let us define $w_t^+ := w_t$ for $t > 0$ and $w_t^+ := 0$ for $t \le 0$. Then, it is readily shown that

$$\left(\sum_{j=0}^{t+\tau-1} h_j y_{t+\tau-j}\right)\left(\sum_{j=0}^{t-1} g_j y_{t-j}\right) = I_1(t) + I_2(t) + I_3(t) + I_4(t)$$

where

$$\begin{aligned} I_1(t) &:= \sum_{u,v=0}^{\infty} h_u g_v x_{t+\tau-u}^+ x_{t-v}^+ \qquad I_2(t) := \sum_{u,v=0}^{\infty} h_u g_v \epsilon_{t+\tau-u}^+ x_{t-v}^+ \\ I_3(t) &:= \sum_{u,v=0}^{\infty} h_u g_v \epsilon_{t-v}^+ x_{t+\tau-u}^+ \quad \text{and} \quad I_4(t) := \sum_{u,v=0}^{\infty} h_u g_v \epsilon_{t+\tau-u}^+ \epsilon_{t-v}^+. \end{aligned}$$

For any fixed $N > 0$, we break up the double sum for u and v into four terms and obtain

$$\begin{aligned} \sum_{t=1}^{n} I_1(t) &= \sum_{t=1}^{n}\left(\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} + \sum_{u=N}^{\infty}\sum_{v=0}^{N-1} \right. \\ &\qquad \left. + \sum_{u=0}^{N-1}\sum_{v=N}^{\infty} + \sum_{u=N}^{\infty}\sum_{v=N}^{\infty}\right) h_u g_v x_{t+\tau-u}^+ x_{t-v}^+ \\ &:= U_{11}^N + U_{12}^N + U_{13}^N + U_{14}^N. \end{aligned} \tag{3.35}$$

(Here the index n is omitted in the U's for brevity.) Given u and v, let $w = w(u,v) := \max(u-\tau, v) + 1$. Then, according to Lemma 3.5, we obtain

$$n^{-1}\sum_{t=1}^{n} x_{t+\tau-u}^+ x_{t-v}^+ = n^{-1}\sum_{t=w}^{n} x_{t+\tau-u} x_{t-v} \xrightarrow{a.s.} r_{\tau-u+v}^x$$

as $n \to \infty$ for any $0 \le u, v \le N-1$. Moreover, let c_u^h and c_v^g be the constants in

Definition 3.2 associated with $\{h_u\}$ and $\{g_v\}$. Then, it is readily shown that

$$\left| n^{-1}U_{11}^N - \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} h_u g_v r_{\tau-u+v}^x \right| \leq \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} c_u^h c_v^g \left| n^{-1}\sum_{t=1}^{n} x_{t+\tau-u}^+ x_{t-v}^+ - r_{\tau-u+v}^x \right|$$

for any $\alpha \in \mathcal{A}$. Therefore, we obtain

$$\lim_{n\to\infty} n^{-1}U_{11}^N = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} h_u g_v r_{\tau-u+v}^x \qquad a.s.$$

uniformly in $\alpha \in \mathcal{A}$ for any N. This result, in coupled with the uniform stability of the filters, implies that uniformly in $\alpha \in \mathcal{A}$,

$$\begin{aligned}
\lim_{N\to\infty}\lim_{n\to\infty} n^{-1}U_{11}^N &= \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \sum_{u=0}^{\infty}\sum_{v=0}^{\infty} h_u g_v \cos(\omega_k(\tau-u+v)) \\
&= \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \Re\left\{ \sum_{u=0}^{\infty}\sum_{v=0}^{\infty} h_u g_v e^{i(\tau-u+v)\omega_k} \right\} \\
&= \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \Re\{H(\omega_k;\alpha)\,\overline{G(\omega_k;\alpha)}\, e^{i\tau\omega_k}\}.
\end{aligned}$$

In addition, since $|x_t| \leq \beta$ almost surely, we have

$$n^{-1}|U_{12}^N| \leq \beta^2 C_g \sum_{u=N}^{\infty} c_u^h$$

for all $\alpha \in \mathcal{A}$ and all n, where $C_g := \sum c_v^g$. Therefore, it follows that

$$\lim_{N\to\infty}\lim_{n\to\infty} n^{-1}\sup_{\alpha\in\mathcal{A}}|U_{12}^N| = 0 \qquad a.s.$$

The same result can be proved for U_{13}^N and U_{14}^N by similar arguments. Combining these results yields

$$n^{-1}\sum_{t=1}^{n} I_1(t) \overset{a.s.}{\to} \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2 \Re\{H(\omega_k;\alpha)\,\overline{G(\omega_k;\alpha)}\, e^{i\tau\omega_k}\} \tag{3.36}$$

uniformly in $\alpha \in \mathcal{A}$ as $n \to \infty$.

Now let us consider $I_2(t)$. Using the same technique as in (3.35), we write

$$\begin{aligned}
\sum_{t=1}^{n} I_2(t) &= \sum_{t=1}^{n}\sum_{u=0}^{\infty}\sum_{v=0}^{\infty} h_u g_v x_{t+\tau-u}^+ \epsilon_{t-v}^+ \\
&= U_{21}^N + U_{22}^N + U_{23}^N + U_{24}^N.
\end{aligned}$$

For any $\alpha \in \mathcal{A}$, it is easy to verify that

$$|U_{21}^N| \leq \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c_u^h c_v^g \left| \sum_{t=1}^{n} \epsilon_{t+\tau-u}^+ x_{t-v}^+ \right|.$$

Moreover, for any $0 \leq u, v \leq N-1$, Lemma 3.5 guarantees that

$$n^{-1} \left| \sum_{t=1}^{n} \epsilon_{t+\tau-u}^+ x_{t-v}^+ \right| = n^{-1} \left| \sum_{t=w}^{n} \epsilon_{t+\tau-u} x_{t-v} \right| \stackrel{a.s.}{\to} 0$$

as $n \to \infty$. Therefore, we obtain

$$\lim_{N\to\infty} \lim_{n\to\infty} n^{-1} \sup_{\alpha\in\mathcal{A}} |U_{21}^N| = 0 \qquad a.s.$$

Furthermore, let us define the random process

$$\zeta_N(t) := \beta C_g \sum_{u=N}^{\infty} c_u^h |\epsilon_{t+\tau-u}| \tag{3.37}$$

for any fixed N. Since $\zeta_N(t) \geq 0$ and, by monotone convergence theorem,

$$E\{\zeta_N(t)\} = \beta C_g E|\epsilon_0| \sum_{u=N}^{\infty} c_u^h < \infty,$$

the infinite sum in (3.37) converges almost surely so that $\zeta_N(t)$ is well-defined. Moreover, it is easy to show that

$$|U_{22}^N| \leq \sum_{t=1}^{n} \zeta_N(t)$$

for all $\alpha \in \mathcal{A}$. Since $\{\zeta_N(t\}$ is strictly stationary for each fixed N, according to the strong ergodic theorem (Karlin and Taylor, 1975), there exists a random variable ζ_N such that

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \zeta_N(t) = \zeta_N \qquad a.s.$$

By the same theorem, the expected value of ζ_N can be written as

$$E(\zeta_N) = E\{\zeta_N(0)\} = \beta C_g E|\epsilon_0| \sum_{u=N}^{\infty} c_u^h$$

for any fixed N. Since $\zeta_N \geq 0$ almost surely and

$$\sum_{N=0}^{\infty} E(\zeta_N) = \beta C_g E|\epsilon_0| \sum_{N=0}^{\infty} \sum_{u=N}^{\infty} c_u^h = \beta C_g E|\epsilon_0| \sum_{u=0}^{\infty} (u+1) c_u^h < \infty$$

it is readily shown by using Markov's inequality that

$$\mathrm{pr}\left\{ \bigcup_{N=N'}^{\infty} (\zeta_N > \varepsilon) \right\} \leq \sum_{N=N'}^{\infty} \mathrm{pr}(\zeta_N > \varepsilon) \leq \varepsilon^{-1} \sum_{N=N'}^{\infty} E(\zeta_N) \to 0$$

as $N' \to \infty$ for any $\varepsilon > 0$. According to the Borel-Cantelli lemma, this implies that $\mathrm{pr}\{(\zeta_N > \varepsilon) \text{ i.o.}\} = 0$ for any $\varepsilon > 0$ and hence that $\zeta_N \overset{a.s.}{\to} 0$ as $N \to \infty$. Consequently, we obtain

$$\lim_{N \to \infty} \limsup_{n \to \infty} n^{-1} \sup_{\alpha \in \mathcal{A}} |U_{22}^N| = 0 \qquad a.s. \tag{3.38}$$

By a similar argument, the same result can also be proved for U_{23}^N and U_{24}^N. It is therefore concluded that

$$n^{-1} \sum_{t=1}^{n} I_2(t) \overset{a.s.}{\to} 0 \tag{3.39}$$

uniformly in $\alpha \in \mathcal{A}$ as $n \to \infty$. Upon noting the similarity between $I_2(t)$ and $I_3(t)$, it is not surprising that the same result also holds for $I_3(t)$, namely,

$$n^{-1} \sum_{t=1}^{n} I_3(t) \overset{a.s.}{\to} 0 \tag{3.40}$$

uniformly in $\alpha \in \mathcal{A}$ as $n \to \infty$.

Finally let us consider $I_4(t)$ for which we can write

$$\begin{aligned} \sum_{t=1}^{n} I_4(t) &= \sum_{t=1}^{n} \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} h_u g_v \epsilon_{t+\tau-u}^{+} \epsilon_{t-v}^{+} \\ &= U_{41}^N + U_{42}^N + U_{43}^N + U_{44}^N \end{aligned}$$

in the same way as (3.35). Given $0 \leq u, v \leq N-1$, Lemma 3.5 implies that

$$n^{-1} \sum_{t=1}^{n} \epsilon_{t+\tau-u}^{+} \epsilon_{t-v}^{+} = n^{-1} \sum_{t=w}^{n} \epsilon_{t+\tau-u} \epsilon_{t-v} \overset{a.s.}{\to} r_{\tau-u+v}^{\epsilon}$$

as $n \to \infty$. Moreover, since

$$\left| n^{-1} U_{41}^N - \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} h_u g_v r_{\tau-u+v}^{\epsilon} \right|$$
$$\leq \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c_u^h c_v^g \left| n^{-1} \sum_{t=1}^{n} \epsilon_{t+\tau-u}^{+} \epsilon_{t-v}^{+} - r_{\tau-u+v}^{\epsilon} \right|,$$

we obtain, upon noting the uniform stability of the filters,

$$\begin{aligned} \lim_{N\to\infty} \lim_{n\to\infty} n^{-1} U_{41}^N &= \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} h_u g_v r_{\tau-u+v}^{\epsilon} \\ &= \int_{-\pi}^{\pi} H(\omega;\alpha) \overline{G(\omega;\alpha)} e^{i\tau\omega} \, dF(\omega) \end{aligned}$$

uniformly in $\alpha \in \mathcal{A}$. To show that the remaining terms in $I_4(t)$ are negligible, we observe that for all $\alpha \in \mathcal{A}$,

$$|U_{42}^N| \leq \sum_{t=1}^{n} Z_N(t)$$

where $\{Z_N(t)\}$ is a strictly stationary process as specified by

$$Z_N(t) := \sum_{u=N}^{\infty} \sum_{v=0}^{\infty} c_u^h c_v^g |\epsilon_{t+\tau-u} \epsilon_{t-v}|.$$

The infinite sum in this expression converges almost surely, since $Z_N(t) \geq 0$ and

$$\begin{aligned} E\{Z_N(t)\} &= \sum_{u=N}^{\infty} \sum_{v=0}^{\infty} c_u^h c_v^g \, E|\epsilon_{t+\tau-u} \epsilon_{t-v}| \\ &\leq C_g r_0^{\epsilon} \sum_{u=N}^{\infty} c_u^h < \infty \end{aligned}$$

according to the monotone convergence theorem. By a similar argument as we employed earlier for the proof of (3.38), we obtain

$$\lim_{N\to\infty} \limsup_{n\to\infty} n^{-1} \sup_{\alpha\in\mathcal{A}} |U_{42}^N| = 0 \qquad a.s.$$

The same result is also true for U_{43}^N and U_{44}^N. Therefore, we conclude that uniformly in $\alpha \in \mathcal{A}$,

$$n^{-1} \sum_{t=1}^{n} I_4(t) \overset{a.s.}{\to} \int_{-\pi}^{\pi} H(\omega;\alpha) \overline{G(\omega;\alpha)} e^{i\tau\omega} \, dF(\omega) \tag{3.41}$$

as $n \to \infty$. Notice that the last quantity is a real number because of the symmetry of $G(\omega;\alpha)$, $H(\omega;\alpha)$, and $F(\omega)$ in ω. Collecting (3.36), (3.39), (3.40), and (3.41) completes the proof. $\Diamond$

Remark 3.7 In the proof of Theorem 3.4, we assume that the ϕ_k are random. The same result remains valid if the ϕ_k are constants, since Lemma 3.5 holds under the assumption of constant phases.

As a direct consequence of Theorem 3.4, we now claim the uniform strong consistency of the sample autocovariances $\hat{r}_\tau(\alpha)$ in (3.33).

Theorem 3.5 *Suppose that $\{h_j(\alpha)\}$ is uniformly strictly stable. Then, as $n \to \infty$, $\hat{r}_\tau(\alpha) \stackrel{a.s.}{\to} r_\tau^y(\alpha)$ uniformly in $\alpha \in \mathcal{A}$ for any $\tau \geq 0$.*

PROOF. It follows from Theorem 3.4 with $\{g_j\}$ replaced by $\{h_j\}$ and n by $n - \tau$. $\Diamond$

The uniform strong consistency can also be proved for the autocovariances

$$\check{r}_\tau(\alpha) = n^{-1} \sum_{t=u}^{n+v} \hat{y}_{t+\tau}(\alpha)\, \hat{y}_t(\alpha)$$

where the integer u and v are independent of n.

Corollary 3.2 *If $\{h_j(\alpha)\}$ is uniformly strictly stable, then, as $n \to \infty$, $\check{r}_\tau(\alpha) \stackrel{a.s.}{\to} r_\tau^y(\alpha)$ uniformly in $\alpha \in \mathcal{A}$ for any $\tau \geq 0$.*

PROOF. According to Theorem 3.4, it suffices to show that

$$n^{-1} \sum_{t=1}^{u-1} \hat{y}_{t+\tau}(\alpha)\, \hat{y}_t(\alpha)$$

vanishes almost surely and uniformly in $\alpha \in \mathcal{A}$. (Here we assume $u > 1$ without loss of generality.) To this end, let $\{c_u^h\}$ be the constant sequence associated with $\{h_u(\alpha)\}$ in Definition 3.2, and let $C_h := \sum c_u^h$. Then, it is easy to show that

$$n^{-1} \sum_{t=1}^{u-1} |\hat{y}_{t+\tau}(\alpha)\, \hat{y}_t(\alpha)| \leq n^{-1} \sum_{t=1}^{u-1} \{\beta^2 C_h^2 + \beta\,(W_{t+\tau} + W_t) + W_{t+\tau} W_t\}$$

for all $\alpha \in \mathcal{A}$. The assertion follows immediately since the sum on the right-hand side does not depend on n. $\Diamond$

Remark 3.8 For the original (unfiltered) process $\{y_t\}$ in (1.1), we have

$$\check{r}_\tau = n^{-1} \sum_{t=u}^{n+v} \hat{y}_{t+\tau} \hat{y}_t \xrightarrow{a.s.} r_\tau^y$$

for any fixed τ, u, and v. This corresponds to the trivial case in Corollary 3.2 where $h_0(\alpha) = 1$ and $h_j(\alpha) = 0$ for all $j > 0$.

Suppose that the filter $\{h_j(\alpha)\}$ in (3.32) is differentiable with respect to α and the derivative is also uniformly strictly stable. The following theorem claims that the uniform strong consistency remains for the derivative of $\hat{r}_\tau(\alpha)$.

Theorem 3.6 *Suppose that $\{h_j(\alpha)\}$ is differentiable with respect to α, and that both $\{h_j(\alpha)\}$ and its derivative are uniformly strictly stable. Then, in addition to the uniform strong consistency of $\hat{r}_\tau(\alpha)$ as claimed in Theorem* 3.5, *it is also true that $\partial \hat{r}_\tau(\alpha)/\partial\alpha \xrightarrow{a.s.} \partial r_\tau^y(\alpha)/\partial\alpha$ uniformly in $\alpha \in \mathcal{A}$ as $n \to \infty$.*

PROOF. From (3.32) and (3.33), it is easy to verify that

$$\begin{aligned} \frac{\partial \hat{r}_\tau(\alpha)}{\partial \alpha} &= n^{-1} \sum_{t=1}^{n-\tau} \left(\sum_{j=0}^{t+\tau} h_j'(\alpha) y_{t+\tau-j} \right) \left(\sum_{j=0}^{t} h_j(\alpha) y_{t-j} \right) \\ &\quad + n^{-1} \sum_{t=1}^{n-\tau} \left(\sum_{j=0}^{t+\tau} h_j(\alpha) y_{t+\tau-j} \right) \left(\sum_{j=0}^{t} h_j'(\alpha) y_{t-j} \right) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial r_\tau^y(\alpha)}{\partial \alpha} &= \sum_{k=1}^{q} \tfrac{1}{2} \beta_k^2 \, \Re\{(H_k' \bar{H}_k + H_k \bar{H}_k') \, e^{i\tau\omega_k}\} \\ &\quad + \int_{-\pi}^{\pi} (H' \bar{H} + H \bar{H}') \, e^{i\tau\omega} \, dF(\omega) \end{aligned}$$

where H_k, H, H_k', and H' are short-hand notation of $H(\omega_k; \alpha)$, $H(\omega; \alpha)$, $H'(\omega_k; \alpha)$, and $H'(\omega; \alpha)$, respectively, and the prime stands for the differentiation with respect to α. The proof is completed by applying Theorem 3.4 to these quantities followed by an argument similar to the proof of Theorem 3.5. $\Diamond$

As a corollary, we obtain the uniform strong consistency of the sample autocorrelation $\hat{\rho}_\tau(\alpha) := \hat{r}_\tau(\alpha)/\hat{r}_0(\alpha)$ and its derivative $\partial\hat{\rho}_\tau(\alpha)/\partial\alpha$ as follows.

Corollary 3.3 *Under the conditions in Theorem* 3.6, *if* $r_0^y(\alpha) > 0$ *for all* $\alpha \in \mathcal{A}$, *then, as* $n \to \infty$, $\hat{\rho}_\tau(\alpha) \xrightarrow{a.s.} \rho_\tau^y(\alpha)$ *and* $\partial\hat{\rho}_\tau(\alpha)/\partial\alpha \xrightarrow{a.s.} \partial\rho_\tau^y(\alpha)/\partial\alpha$ *uniformly in* $\alpha \in \mathcal{A}$, *where* $\rho_\tau^y(\alpha) := r_\tau^y(\alpha)/r_0^y(\alpha)$ *is the autocorrelation function of* $\{y_t(\alpha)\}$.

PROOF. Since $r_0^y(\alpha) > 0$ for all $\alpha \in \mathcal{A}$, the uniform strong consistency in Theorem 3.5 implies that $\hat{r}_0(\alpha) > 0$ almost surely for all $\alpha \in \mathcal{A}$, provided that n is sufficiently large. Therefore, $\hat{\rho}_\tau(\alpha)$ is well-defined and differentiable for all $\alpha \in \mathcal{A}$ and large n. Moreover, we have

$$\frac{\partial\hat{\rho}_\tau(\alpha)}{\partial\alpha} = \frac{\hat{r}_0(\alpha)\,\partial\hat{r}_\tau(\alpha)/\partial\alpha - \hat{r}_\tau(\alpha)\,\partial\hat{r}_0(\alpha)/\partial\alpha}{(\hat{r}_0(\alpha))^2}.$$

The assertion follows immediately from Theorem 3.5 and Theorem 3.6. ◇

3.4 More Results on Uniform Strong Consistency

The uniform strong consistency results in Theorem 3.4 and Theorem 3.5 can be extended to include some more general signal-plus-noise models.

Let us consider $\{y_t(\alpha)\}$ which obeys the following parametric model

$$y_t(\alpha) = \sum_{j=0}^{\infty} h_j(\alpha)\, x_{t-j} + \sum_{j=0}^{\infty} g_j(\alpha)\, \epsilon_{t-j} \qquad (t = 0, \pm 1, \pm 2, \ldots) \tag{3.42}$$

where $\{x_t\}$ and $\{\epsilon_t\}$ are zero-mean, strictly stationary, and mutually independent, with finite second moments. As we can see from the proof of (3.39), (3.41), and Theorem 3.5, in order to show that the sample autocovariance function of $\{y_t(\alpha)\}$ converges to

$$\begin{aligned} r_\tau^y(\alpha) &:= E\{y_{t+\tau}(\alpha)\, y_t(\alpha)\} \\ &= \sum_{u,v=0}^{\infty} h_u(\alpha)\, h_v(\alpha)\, r_{\tau-u+v}^x + \sum_{u,v=0}^{\infty} g_u(\alpha)\, g_v(\alpha)\, r_{\tau-u+v}^\epsilon \end{aligned} \tag{3.43}$$

almost surely and uniformly in α, it is sufficient to require that the almost sure limits in Lemma 3.5 exist for any u, v, and w, and that both $\{h_j(\alpha)\}$ and $\{g_j(\alpha)\}$ be uniformly strictly stable[2]. Therefore, we immediately obtain the following results.

Theorem 3.7 *Let the parametric process $\{y_t(\alpha)\}$ be defined by (3.42) for $\alpha \in \mathcal{A}$, with $\{x_t\}$ and $\{\epsilon_t\}$ being any zero-mean, strictly stationary, and mutually independent processes for which Lemma 3.5 holds. If $\{h_j(\alpha)\}$ and $\{g_j(\alpha)\}$ are uniformly strictly stable in $\mathcal{A}$, then*

$$n^{-1} \sum_{t=1}^{n-\tau} y_{t+\tau}(\alpha)\, y_t(\alpha) \stackrel{a.s.}{\to} r_\tau^y(\alpha)$$

uniformly in $\alpha \in \mathcal{A}$ as $n \to \infty$ for any $\tau \geq 0$, where $r_\tau^y(\alpha)$ is given by (3.43).

The uniform consistency in Theorem 3.7 is of importance in applications where the inference of the model parameter α relies on the sample autocovariance function of the observed process $\{y_t(\alpha)\}$. Generalization of this result to higher-order sample moments and cumulants is quite straightforward. All we need is to assume that the corresponding higher-order sample moments of $\{x_t\}$ and $\{\epsilon_t\}$ are strongly consistent, giving rise to similar results as in Lemma 3.5. We have noticed a recent attempt[3] to prove the uniform strong consistency of sample cumulants from a process similar to $\{y_t(\alpha)\}$. Unfortunately, the proof was based on a false lemma which was incorrectly cited from a wrong result of Ljung (1987, Appendix 2B).

Under the same assumption about $\{x_t\}$ and $\{\epsilon_t\}$ as in Theorem 3.7, we can also prove the uniform strong consistency for the process $y_t = x_t + \epsilon_t$ after parametric filtering. Define, as before, the filtered data by

$$\hat{y}_t(\alpha) := \sum_{j=0}^{t-1} h_j(\alpha)\, y_{t-j} = \sum_{j=0}^{\infty} h_j(\alpha)\, (x_{t-j}^+ + \epsilon_{t-j}^+). \tag{3.44}$$

[2]The proof goes through with x_t and ϵ_t in place of x_t^+ and ϵ_t^+, respectively.

[3]Giannakis, G. B. and Tsatsanis, M. K. (1992). A unifying maximum-likelihood view of cumulant and polyspectral measures for non-Gaussian signal classification and estimation. *IEEE Trans. Inform. Theory*, vol. 38, no. 2, part I, pp. 386–406.

Following the same proof of (3.39), (3.41), and Theorem 3.5, we obtain the following results.

Theorem 3.8 *Suppose that the conditions about* $\{x_t\}$ *and* $\{\epsilon_t\}$ *in Theorem* 3.7 *are satisfied, and let* $\{\hat{y}_t(\alpha)\}$ *be defined by* (3.44). *If the filter* $\{h_j(\alpha)\}$ *is uniformly strictly stable in* $\mathcal{A}$, *then, as* $n \to \infty$,

$$n^{-1} \sum_{t=1}^{n-\tau} \hat{y}_{t+\tau}(\alpha)\, \hat{y}_t(\alpha) \xrightarrow{a.s.} r_\tau^y(\alpha)$$

uniformly in $\alpha \in \mathcal{A}$ *for any* $\tau \geq 0$, *where* $r_\tau^y(\alpha)$ *is the autocovariance function in* (3.43) *with* g_j *replaced by* h_j.

Remark 3.9 The assumptions in Theorem 3.7 are clearly satisfied if $\{x_t\}$ and $\{\epsilon_t\}$ are both linear processes (Hannan, 1970, p. 204).

Chapter 4

Frequency Estimation by Parametric Filtering

As we have seen from Chapter 2, Prony's estimator leads to inconsistent frequency estimates. To alleviate this predicament, we propose in this chapter a general procedure of iterative parametric filtering — the *parametric filtering* (PF) *method* — that can be shown to produce consistent frequency estimates. The idea is to judiciously parametrize the filter so that it satisfies a certain parametrization property for all parameters in a neighborhood of the true AR parameter $\mathbf{a}$, and the clue for the correct parametrization comes from the particular form of the bias encountered by Prony's estimator. For any filter which satisfies the parametrization property, we define the PF estimator of the AR parameter $\mathbf{a}$ as the multivariate fixed-point of the parametrized least squares estimator from the filtered data. Under certain mild conditions, the least squares estimator, as a function of the filter parameter, is shown to constitute a *contractive mapping* so that the PF estimator exists as its fixed-point. We also consider an iterative algorithm that calculates the PF estimator and show its almost sure convergence to the PF estimator. The fact of the matter is that the PF estimator can be shown to be strongly consistent and asymptotically normal for estimating the AR parameter $\mathbf{a}$. These results solidify the theoretical foundation of the PF method as a promising procedure of frequency estimation.

4.1 Parametric Filtering Method

Our basic idea of eliminating the bias in Prony's estimator rests upon estimating the AR parameter **a** in (2.4), not directly from the original data as done in Chapter 2, but from the *filtered* data — upon using an appropriate parametric filter. As we shall see, the least squares estimator from the filtered data can be made consistent by iteratively "tuning" the filter parameter according to previously obtained AR estimate.

4.1.1 AR Estimation After Parametric Filtering

Let us consider a linear time-invariant causal filter with impulse response sequence $\{h_j(\boldsymbol{\alpha}), j = 0, 1, \ldots\}$, where $\boldsymbol{\alpha} := [\alpha_1, \cdots, \alpha_q]^T$ is the filter parameter which takes on values in an open set $\mathcal{A}$ that contains the AR parameter **a**. Assume that for any $\boldsymbol{\alpha} \in \mathcal{A}$ the filter is stable, i.e., $\sum |h_j(\boldsymbol{\alpha})| < \infty$. In addition, suppose that the filter passes all of the sinusoidal components in $\{x_t\}$, namely,

(A1) $\qquad H(\omega_k; \boldsymbol{\alpha}) \neq 0 \qquad$ for $k = 1, \ldots, q$ and for all $\boldsymbol{\alpha} \in \mathcal{A}$,

where $H(\omega; \boldsymbol{\alpha})$ is the transfer function of the filter. We apply this filter to $\{y_t\}$ and obtain the filtered process

$$y_t(\boldsymbol{\alpha}) := \sum_{j=0}^{\infty} h_j(\boldsymbol{\alpha})\, y_{t-j}.$$

Similarly, we define the filtered signal $\{x_t(\boldsymbol{\alpha})\}$ and the filtered noise $\{\epsilon_t(\boldsymbol{\alpha})\}$.

The key fact to observe is that the filtered signal $\{x_t(\boldsymbol{\alpha})\}$ remains a sum of q sinusoids with the *same* frequencies, and hence a solution to the *same* homogeneous AR equation (2.3). In fact, according to the theory of linear filtering (Brockwell and Davis, 1987), the filtered signal $\{x_t(\boldsymbol{\alpha})\}$ can be expressed as

$$x_t(\boldsymbol{\alpha}) = \sum_{k=1}^{q} \beta_k(\boldsymbol{\alpha}) \cos(\omega_k t + \phi_k(\boldsymbol{\alpha}))$$

where $\beta_k(\boldsymbol{\alpha}) := \beta_k |H(\omega_k; \boldsymbol{\alpha})|$ and $\phi_k(\boldsymbol{\alpha}) := \phi_k + \arg\{H(\omega_k; \boldsymbol{\alpha})\}$ are the amplitudes and phases of the filtered sinusoids. Since the AR equation (2.3) is completely determined

by the unchanged frequencies ω_k, we still obtain $A(z^{-1})\, x_t(\boldsymbol{\alpha}) = 0$ for all t. It is quite clear that the original problem of frequency estimation is not altered by the linear filtering and can be equivalently stated as that of estimating the *same* AR parameter $\mathbf{a}$ in terms of the *filtered* processes.

Let $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha})$ be the least squares estimator of $\mathbf{a}$ on the basis of $\{y_1(\boldsymbol{\alpha}), \ldots, y_n(\boldsymbol{\alpha})\}$. Since $\{y_t(\boldsymbol{\alpha})\}$ still consists of a sinusoidal signal in additive noise that can be modeled as a linear process, the strong consistency remains for $\{y_t(\boldsymbol{\alpha})\}$ as indicated by the large sample theory in Chapter 3 (see Remark 3.9 with y_t replaced by $y_t(\boldsymbol{\alpha})$). Therefore, similarly to (2.11), the least squares estimator $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha})$ converges almost surely to the deterministic vector

$$\mathbf{a}(\boldsymbol{\alpha}) := -\mathbf{R}_y^{-1}(\boldsymbol{\alpha})\, \mathbf{r}_y(\boldsymbol{\alpha}) \tag{4.1}$$

which, as claimed in Lemma 2.2, can be written as

$$\mathbf{a}(\boldsymbol{\alpha}) = \mathbf{a} - \mathbf{R}_y^{-1}(\boldsymbol{\alpha}) \{\mathbf{R}_\epsilon(\boldsymbol{\alpha})\, \mathbf{a} + \mathbf{r}_\epsilon(\boldsymbol{\alpha})\} \tag{4.2}$$

where the autocovariances are defined from the corresponding filtered processes.

Since $\mathbf{R}_y(\boldsymbol{\alpha}) = \mathbf{R}_x(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha})$, multiplying each side of (4.2) by $\mathbf{R}_y(\boldsymbol{\alpha})$ gives

$$\mathbf{R}_x(\boldsymbol{\alpha})\, \mathbf{a}(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha})\, \mathbf{a}(\boldsymbol{\alpha}) = \mathbf{R}_x(\boldsymbol{\alpha})\, \mathbf{a} - \mathbf{r}_\epsilon(\boldsymbol{\alpha}). \tag{4.3}$$

Note that $\mathbf{R}_x(\boldsymbol{\alpha})$ is nonsingular if all of the q sinusoids are retained after the filtering. In fact $\mathbf{R}_x(\boldsymbol{\alpha})$ admits a similar decomposition as in Lemma 2.3 and Remark 2.1 with $\mathbf{P}$ in (2.20) replaced by $\mathbf{P}(\boldsymbol{\alpha}) := \frac{1}{4}\,\mathrm{diag}(\beta_1^2(\boldsymbol{\alpha}), \ldots, \beta_{2q}^2(\boldsymbol{\alpha}))$. Under (A1), $|H(\omega_k; \boldsymbol{\alpha})| > 0$ for all k. It follows that $\beta_k(\boldsymbol{\alpha}) > 0$ for all k, and hence $\mathbf{P}(\boldsymbol{\alpha})$ is nonsingular. A similar argument as in the proof of Lemma 2.3 confirms the nonsingularity of $\mathbf{R}_x(\boldsymbol{\alpha})$. Armed with this fact and (4.3), the bias $\mathbf{a}(\boldsymbol{\alpha}) - \mathbf{a}$ can be written as

$$\mathbf{a}(\boldsymbol{\alpha}) - \mathbf{a} = -\mathbf{R}_x^{-1}(\boldsymbol{\alpha}) \{\mathbf{R}_\epsilon(\boldsymbol{\alpha})\, \mathbf{a}(\boldsymbol{\alpha}) + \mathbf{r}_\epsilon(\boldsymbol{\alpha})\}. \tag{4.4}$$

Unlike what we had before when the least squares (Prony's) estimator was obtained from the original data, the asymptotic bias of $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha})$ is now a function of the filter

parameter $\boldsymbol{\alpha}$. This, as we shall see, makes it possible to fulfill our goal of eliminating the bias via appropriate parametric filtering.

4.1.2 PF Method of Frequency Estimation

As suggested by (4.4), if there exists a filter parameter $\boldsymbol{\alpha}^*$ in $\mathcal{A}$ such that

$$\mathbf{a}(\boldsymbol{\alpha}^*) = -\mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha}^*)\,\mathbf{r}_\epsilon(\boldsymbol{\alpha}^*), \tag{4.5}$$

then from (4.4) we would obtain $\mathbf{a}(\boldsymbol{\alpha}^*) = \mathbf{a}$. This implies that after filtering with $\boldsymbol{\alpha}^*$, the least squares estimator $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha}^*)$ becomes consistent for estimating the AR parameter $\mathbf{a}$.

Suppose now that the *autocorrelation* function of the noise $\{\epsilon_t\}$ is known, and, for convenience, that the filter is parametrized so that

(A2) $$\boldsymbol{\alpha} = -\mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha})\,\mathbf{r}_\epsilon(\boldsymbol{\alpha}) \qquad \text{for all } \boldsymbol{\alpha} \in \mathcal{A}.$$

In this case, the filter is said to possess the *parametrization property.* It is interesting to note the similarity between (A2) and the Yule-Walker equations. In fact, for an AR($2q$) process with AR coefficients satisfying the symmetry condition (2.2), the vector of the first q free coefficients will be the solution of (A2), with $\mathbf{R}_\epsilon(\boldsymbol{\alpha})$ and $\mathbf{r}_\epsilon(\boldsymbol{\alpha})$ being replaced by their counterparts defined from the autocovariances of that AR process.

Equipped with the parametrization property, we have the following theorem.

Theorem 4.1 *Suppose that the filter* $\{h_j(\boldsymbol{\alpha})\}$ *satisfies* (A1) *and* (A2). *Then* $\boldsymbol{\alpha}^* = \mathbf{a}$ *is the unique fixed-point of the deterministic mapping* $\mathbf{a}(\boldsymbol{\alpha})$ *in* $\mathcal{A}$.

PROOF. From (4.2) and (A2), we find that

$$\mathbf{a}(\boldsymbol{\alpha}) - \mathbf{a} = \mathbf{C}(\boldsymbol{\alpha})\,(\boldsymbol{\alpha} - \mathbf{a}) \tag{4.6}$$

where

$$\mathbf{C}(\boldsymbol{\alpha}) := \mathbf{R}_y^{-1}(\boldsymbol{\alpha})\,\mathbf{R}_\epsilon(\boldsymbol{\alpha}). \tag{4.7}$$

Since $\mathbf{C}(\boldsymbol{\alpha})$ is nonsingular and $\mathbf{a} \in \mathcal{A}$, the AR parameter $\mathbf{a}$ is a fixed-point of $\mathbf{a}(\boldsymbol{\alpha})$. On the other hand, since the parametrization property (A2) is satisfied, (4.5) is equivalent to $\mathbf{a}(\boldsymbol{\alpha}^*) = \boldsymbol{\alpha}^*$. This implies that $\boldsymbol{\alpha}^*$, if exists, is a also fixed-point of $\mathbf{a}(\boldsymbol{\alpha})$. It also implies that $\boldsymbol{\alpha}^* = \mathbf{a}$ since $\mathbf{a}(\boldsymbol{\alpha}^*) = \mathbf{a}$. The theorem is thus proved. $\diamondsuit$

According to this theorem, it is no longer necessary to distinguish between the AR parameter $\mathbf{a}$ and the filter parameter $\boldsymbol{\alpha}^*$, and thus the problem of estimating $\mathbf{a}$ becomes identical to that of estimating $\boldsymbol{\alpha}^*$. Given a consistent estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ of the deterministic mapping $\mathbf{a}(\boldsymbol{\alpha})$, Theorem 4.1 gives rise to the idea of *finding a fixed-point in the random mapping* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *as an estimator of the AR parameter* $\mathbf{a}$. We refer to this method as the *parametric filtering* (PF) *method*, and call the fixed-point, denoted by $\hat{\boldsymbol{\alpha}}$, the *PF estimator* of $\mathbf{a}$. The corresponding $\hat{\omega}_k$, obtained from the zeros of the AR polynomial $A(z^{-1})$ in (2.1) with $\hat{\boldsymbol{\alpha}}$ in place of $\mathbf{a}$, are called the PF frequency estimates.

To obtain the PF estimator $\hat{\boldsymbol{\alpha}}$ which, by definition, is a fixed-point of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$, we employ the so-called *fixed-point iteration*. Namely, starting with an initial guess $\hat{\boldsymbol{\alpha}}_0$, we recursively calculate

$$\hat{\boldsymbol{\alpha}}_m = \hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}_{m-1}) \qquad (m = 1, 2, \ldots). \tag{4.8}$$

The fixed-point $\hat{\boldsymbol{\alpha}}$ is obtained, when convergence occurs, as the limiting value of $\hat{\boldsymbol{\alpha}}_m$ as $m \to \infty$. Similar to some existing methods pertaining to special cases (Kay, 1984; Dragošević and Stanković, 1989), this procedure is an iterative filtering algorithm which can be implemented as follows: given the initial guess $\hat{\boldsymbol{\alpha}}_0$ and for $m = 1, 2, \ldots,$

STEP 1. Filter the data with $\{h_j(\boldsymbol{\alpha})\}$ such that $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}_{m-1}$;

STEP 2. Compute the estimate $\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}_{m-1})$ from the filtered data;

STEP 3. Set $\hat{\boldsymbol{\alpha}}_m = \hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}_{m-1})$; and

STEP 4. Go to Step 1 with m replaced by $m + 1$.

Related procedures for the special case of $q = 1$ have been considered by He and Kedem (1989), Yakowitz (1991), Kedem and Yakowitz (1992). See also Li and Kedem (1992), and Li, Kedem, and Yakowitz (1992) for a somewhat more rigorous treatment of this special case.

Remark 4.1 Suppose for the moment that the sinusoids are absent in $\{y_t\}$. From (4.1), it is readily shown that $\mathbf{a}(\boldsymbol{\alpha})$ coincides with $-\mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha})\,\mathbf{r}_\epsilon(\boldsymbol{\alpha})$ and hence becomes the identity mapping $\mathbf{a}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}$ under (A2). Therefore, the parametrization property (A2) simply requires that each and every $\boldsymbol{\alpha}$ in $\mathcal{A}$ be a fixed-point of $\mathbf{a}(\boldsymbol{\alpha})$ in the absence of the signal. Moreover, when the signal is present but the filter $\{h_j(\boldsymbol{\alpha})\}$ fails to capture it, we still obtain $\mathbf{a}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}$ since $\{y_t(\boldsymbol{\alpha})\}$ does not contain a sinusoidal part. In this case, little or no change is expected from the iteration (4.8) when initiated by $\boldsymbol{\alpha}$. In other words, when the signal is not captured, the filter essentially remains where it started. This observation suggests that the presence of a sinusoidal signal in $\{y_t\}$ could be tested based on whether significant changes occur after the iteration (4.8) for a number of randomly selected initial guesses.

Remark 4.2 To parametrize the filter $\{h_j(\boldsymbol{\alpha})\}$ in accordance with (A2) requires the autocorrelation function or the normalized spectrum of the noise $\{\epsilon_t\}$. If this information is available, a linear filter can be designed to whiten the noise by the classical theory of Wiener filtering. A filter $\{h_j(\boldsymbol{\alpha})\}$ with the property (A2) can thus be obtained by cascading the whitening filter with a parametric filter (e.g., the AR filter in Chapter 5) which satisfies (A2) when the noise is white. In the cases where the autocorrelation function of the noise is not available but can be modeled as an AR or MA process, certain iterative procedures can be employed to estimate the noise parameters and the sinusoidal frequencies alternatively (see, e.g., Dragošević and Stanković, 1989). Intuitively, however, the performance of the PF method should not be too sensitive to the noise spectrum if *bandpass* filters are applied *locally to frequency clusters* and the noise spectrum is sufficiently smooth. For instance, if three sinusoids

are present with two closely spaced and the other well separated from the first two, a bandpass filter with $q = 1$ (e.g., AR(2) filter in Chapter 5) can be applied locally to the separated frequency while another bandpass filter with $q = 2$ (e.g., AR(4) filter in Chapter 5) to the two closely-spaced frequencies.

4.1.3 Least Squares Estimator

So far we have not yet specified the estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$. In fact, any estimator would qualify as long as it converges (stochastically) to $\mathbf{a}(\boldsymbol{\alpha})$ when the sample size n tends to infinity. In the following, we specialize the choice of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ by considering the least squares estimator from the filtered data where the filtering is completely based upon the observations $y_1, \ldots, y_n$.

Notice that given the finite data record $\{y_1, \ldots, y_n\}$, the filtered process $\{y_t(\boldsymbol{\alpha})\}$ can only be approximated by

$$\hat{y}_t(\boldsymbol{\alpha}) := \sum_{j=0}^{t-1} h_j(\boldsymbol{\alpha})\, y_{t-j} \qquad (t = 1, \ldots, n). \tag{4.9}$$

Therefore, the estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ should be defined on the basis of $\{\hat{y}_1(\boldsymbol{\alpha}), \ldots, \hat{y}_n(\boldsymbol{\alpha})\}$ in stead of the unavailable data $\{y_1(\boldsymbol{\alpha}), \ldots, y_n(\boldsymbol{\alpha})\}$. Motivated by the estimator $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha})$, which requires the unavailable $\{y_t(\boldsymbol{\alpha})\}$, we may take

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) := -\{\mathbf{Q}^T \hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\, \hat{\mathbf{Y}}(\boldsymbol{\alpha})\, \mathbf{Q}\}^{-1} \mathbf{Q}^T \hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\, \hat{\mathbf{y}}(\boldsymbol{\alpha}) \tag{4.10}$$

where $\hat{\mathbf{Y}}(\boldsymbol{\alpha})$ and $\hat{\mathbf{y}}(\boldsymbol{\alpha})$ are the data matrices defined from $\{\hat{y}_t(\boldsymbol{\alpha})\}$ in the same way as $\mathbf{Y}$ and $\mathbf{y}$ in (2.6). It is readily shown that $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in (4.10) is the least squares estimator that minimizes the criterion $\|\hat{\mathbf{y}}(\boldsymbol{\alpha}) + \hat{\mathbf{Y}}(\boldsymbol{\alpha})\, \mathbf{Q}\, \hat{\mathbf{a}}(\boldsymbol{\alpha})\|^2$. According to the large sample theory developed in Chapter 3 (Theorem 3.5), if the filter is uniformly strictly stable, the least squares estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in (4.10) converges almost surely to $\mathbf{a}(\boldsymbol{\alpha})$ as n tends to infinity. In other words, $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ is a strongly consistent estimator of $\mathbf{a}(\boldsymbol{\alpha})$. The strong consistency is also *uniform* in $\boldsymbol{\alpha}$ so that many properties of $\mathbf{a}(\boldsymbol{\alpha})$, as a deterministic

function of $\boldsymbol{\alpha}$, are retained by its estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$. This observation turns out to be very helpful in the statistical analysis of the PF estimator, as we shall consider shortly.

4.1.4 Relation to the CM Method

The PF method described above happens to be an extension of a procedure recently proposed by He and Kedem (1989) and by Yakowitz (1991) for *single* frequency estimation. A more systematic and rigorous treatment of this procedure and its statistical properties can be found in Li and Kedem (1992), and Li, Kedem, and Yakowitz (1992), where the procedure was referred to as the CM (or contraction mapping) method. Indeed, for $q = 1$, we note that the parametrization property (A2) becomes

$$\alpha = -2\, r_1^{\epsilon}(\alpha)/r_0^{\epsilon}(\alpha) \tag{4.11}$$

where $\alpha := \alpha_1$, and the least squares estimator in (4.10) reduces to

$$\hat{a}(\alpha) = -\sum_{t=3}^{n} \hat{y}_{t-1}(\alpha)\,\{\hat{y}_t(\alpha) + \hat{y}_{t-2}(\alpha)\} \Big/ \sum_{t=3}^{n} \hat{y}_{t-1}^2(\alpha). \tag{4.12}$$

If we reparametrize the filter by $\vartheta := -\alpha/2$, then (4.11) can be written as $\vartheta = \rho_1^{\epsilon}(\vartheta)$ where $\rho_1^{\epsilon}(\vartheta)$ stands for the first-order autocorrelation of $\{\epsilon_t(\alpha)\}$. This relation is readily recognized as being the fundamental property required by the CM method for the parametrization of the filter. Moreover, as we have seen in Chapter 3 (see also Li and Kedem, 1992), $\hat{\rho}(\vartheta) := -\hat{a}(\alpha)/2$ is a uniformly and strongly consistent estimator of the first-order autocorrelation of $\{y_t(\alpha)\}$ if the filter is uniformly strictly stable. With this estimator, the fixed-point iteration in (4.8) becomes

$$\hat{\vartheta}_m = \hat{\rho}(\hat{\vartheta}_{m-1}) \qquad (m = 1, 2, \ldots),$$

which coincides with the iteration of the CM method that produces a sequence $\{\hat{\vartheta}_m\}$ for estimating the parameter $\vartheta^* := -\alpha^*/2 = \cos\omega_1$.

Statistical properties of the CM method have recently been studied by Li and Kedem (1992) and Li, Kedem, and Yakowitz (1992). In these works (see also Chapter 5),

it was proved that under appropriate conditions the CM method provides a strongly consistent estimator of ω_1, and that the estimator is asymptotically normal with a variance inversely related to the signal-to-noise ratio of the filtered data. In the next section, we shall analyze the PF method along the same lines as in these works in order to establish the strong consistency and asymptotic normality of the PF estimator $\hat{\boldsymbol{\alpha}}$.

4.2 Statistical Properties of the PF Estimator

To investigate statistical properties of the PF method presented in the preceding section, we shall answer the following questions:

i) Under what conditions does the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ have a fixed-point $\hat{\boldsymbol{\alpha}}$?

ii) Under what conditions does the fixed-point iteration in (4.8) converge to the fixed-point $\hat{\boldsymbol{\alpha}}$? and

iii) What limit and limiting distribution does the PF estimator $\hat{\boldsymbol{\alpha}}$ possess as the sample size n tends to infinity?

This section provides a set of sufficient conditions under which a unique fixed-point $\hat{\boldsymbol{\alpha}}$ exists and can be found in the vicinity of $\boldsymbol{\alpha}^*$ by the iteration in (4.8) almost surely. Under these conditions, $\hat{\boldsymbol{\alpha}}$ is also shown to be strongly consistent and asymptotically normal for estimating $\boldsymbol{\alpha}^*$. For simplicity, these results are formulated in terms of the least squares estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ defined by (4.10).

4.2.1 Existence and Convergence

Suppose that the filter $\{h_j(\boldsymbol{\alpha})\}$ also satisfies the following regularity conditions:

(A3) The filter $\{h_j(\boldsymbol{\alpha})\}$ is uniformly strictly stable in $\mathcal{A}$.

(A4) The filter $\{h_j(\boldsymbol{\alpha})\}$ is continuously differentiable, and its derivatives are uniformly strictly stable in $\mathcal{A}$.

Under these additional conditions, together with (A1) and (A2), we first show that the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ is *uniformly* consistent for estimating $\mathbf{a}(\boldsymbol{\alpha})$ up to the first derivative, as summarized in the following lemma.

Lemma 4.1 *Suppose that* (A1), (A3), *and* (A4) *are satisfied. Then, both* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *and* $\mathbf{a}(\boldsymbol{\alpha})$ *are continuously differentiable. Moreover, as* $n \to \infty$,

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathbf{a}(\boldsymbol{\alpha}) \quad \textit{and} \quad \hat{\mathbf{a}}'(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathbf{a}'(\boldsymbol{\alpha})$$

uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$, *where* $\hat{\mathbf{a}}'(\boldsymbol{\alpha})$ *and* $\mathbf{a}'(\boldsymbol{\alpha})$ *are Jacobian matrices of* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *and* $\mathbf{a}(\boldsymbol{\alpha})$.

PROOF. According to Theorem 3.5, the sample autocovariances of the filtered data are uniformly strongly consistent. Therefore, under (A3), we obtain

$$n^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\,\hat{\mathbf{Y}}(\boldsymbol{\alpha})\mathbf{Q} \overset{a.s.}{\to} \mathbf{R}_y(\boldsymbol{\alpha}) \quad \text{and} \quad n^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\,\hat{\mathbf{y}}(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathbf{r}_y(\boldsymbol{\alpha})$$

uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$ as $n \to \infty$. Since $\mathbf{R}_y(\boldsymbol{\alpha})$ is nonsingular under (A1), it follows that $\mathbf{Q}^T\hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\,\hat{\mathbf{Y}}(\boldsymbol{\alpha})\mathbf{Q}$ is nonsingular almost surely for sufficiently large n, and that

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) \overset{a.s.}{\to} -\mathbf{R}_y^{-1}(\boldsymbol{\alpha})\,\mathbf{r}_y(\boldsymbol{\alpha}) = \mathbf{a}(\boldsymbol{\alpha})$$

uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$. The nonsingularity of $\mathbf{R}_y(\boldsymbol{\alpha})$ and $\mathbf{Q}^T\hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\,\hat{\mathbf{Y}}(\boldsymbol{\alpha})\mathbf{Q}$, in connection with (A4), also guarantees that $\mathbf{a}(\boldsymbol{\alpha})$ and $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ are continuously differentiable. Applying Theorem 3.6 proves that $\hat{\mathbf{a}}'(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathbf{a}'(\boldsymbol{\alpha})$ uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$. ◊

In the sequel, we shall also make use of the following lemma.

Lemma 4.2 *Let* $\mathbf{C}(\boldsymbol{\alpha})$ *be the matrix defined in* (4.7), *and assume that* (A1) *and* (A2) *are valid. Then the spectral radius of* $\mathbf{C}(\boldsymbol{\alpha}^*)$ *is less than 1. Moreover, if* $\mathbf{C}(\boldsymbol{\alpha})$ *is continuous at* $\boldsymbol{\alpha}^*$, *then* $\mathbf{a}'(\boldsymbol{\alpha})$ *is differentiable at* $\boldsymbol{\alpha}^*$ *and* $\mathbf{a}'(\boldsymbol{\alpha}^*) = \mathbf{C}(\boldsymbol{\alpha}^*)$. *Therefore, being the unique fixed-point of* $\mathbf{a}(\boldsymbol{\alpha})$, $\boldsymbol{\alpha}^*$ *is attractive.*

PROOF. It is easy to verify from (4.7) that $\mathbf{C}(\boldsymbol{\alpha})$ can be written as

$$\mathbf{C}(\boldsymbol{\alpha}) = \{\mathbf{I} + \boldsymbol{\Gamma}(\boldsymbol{\alpha})\}^{-1} \quad \text{with} \quad \boldsymbol{\Gamma}(\boldsymbol{\alpha}) := \mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha})\,\mathbf{R}_x(\boldsymbol{\alpha}). \tag{4.13}$$

Let λ_j and $\mathbf{p}_j$, $(j = 1, \ldots, q)$, be the eigenvalues and corresponding eigenvectors of $\boldsymbol{\Gamma}(\boldsymbol{\alpha}^*)$, then $1/(1 + \lambda_j)$ are eigenvalues of $\mathbf{C}(\boldsymbol{\alpha}^*)$, associated with eigenvectors $\mathbf{p}_j$. By definition (Ortega and Rheinboldt, 1970, p. 43), the spectral radius of the matrix $\mathbf{C}(\boldsymbol{\alpha}^*)$ is given by $\varrho = \max\{1/|1 + \lambda_j|\} > 0$. Therefore, $\varrho < 1$ if and only if $|1+\lambda_j| > 1$ for all j. On the other hand, since $\boldsymbol{\Gamma}(\boldsymbol{\alpha}^*)\,\mathbf{p}_j = \lambda_j \mathbf{p}_j$, it follows from (4.13) that

$$\mathbf{p}_j^H \mathbf{R}_x(\boldsymbol{\alpha}^*)\,\mathbf{p}_j = \lambda_j \{\mathbf{p}_j^H \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{p}_j\}.$$

Note that $\mathbf{R}_x(\boldsymbol{\alpha}^*)$ and $\mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)$ are positive definite under (A1). Therefore, we obtain $\mathbf{p}_j^H \mathbf{R}_x(\boldsymbol{\alpha}^*)\,\mathbf{p}_j > 0$ and $\mathbf{p}_j^H \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{p}_j > 0$ for all j. This, in turn, yields $\lambda_j > 0$ for all j. As a consequence, we obtain $|1 + \lambda_j| = 1 + \lambda_j > 1$ for all j and hence $\varrho < 1$.

To proceed with the proof, we note that $\varrho < 1$ implies the existence of a norm $\|\cdot\|$ such that $\|\mathbf{C}(\boldsymbol{\alpha}^*)\| < 1$ (Ortega and Rheinboldt, 1970, Theorem 2.2.8, p. 44). Moreover, under (A2), it follows from (4.6) that

$$\|\mathbf{a}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}^* - \mathbf{C}(\boldsymbol{\alpha}^*)\,(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\| \leq \|\mathbf{C}(\boldsymbol{\alpha}^*) - \mathbf{C}(\boldsymbol{\alpha})\|\,\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|.$$

Since $\mathbf{C}(\boldsymbol{\alpha})$ is continuous at $\boldsymbol{\alpha}^*$ by assumption, and $\mathbf{a}(\boldsymbol{\alpha}^*) = \boldsymbol{\alpha}^*$, we obtain

$$\|\mathbf{a}(\boldsymbol{\alpha}) - \mathbf{a}(\boldsymbol{\alpha}^*) - \mathbf{C}(\boldsymbol{\alpha}^*)\,(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\| = o(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|).$$

By definition (Ortega and Rheinboldt, 1970), the mapping $\mathbf{a}(\boldsymbol{\alpha})$ is differentiable at $\boldsymbol{\alpha}^*$, and its derivative coincides with $\mathbf{C}(\boldsymbol{\alpha}^*)$. The fixed-point $\boldsymbol{\alpha}^*$ of $\mathbf{a}(\boldsymbol{\alpha})$ is attractive since $\|\mathbf{a}'(\boldsymbol{\alpha}^*)\| = \|\mathbf{C}(\boldsymbol{\alpha}^*)\| < 1$. ◊

Based on these lemmas, the existence of the PF estimator $\hat{\boldsymbol{\alpha}}$ as a fixed-point of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ and the convergence of the iteration in (4.8) to $\hat{\boldsymbol{\alpha}}$ can be established as follows.

Theorem 4.2 *Under* (A1)–(A4), *the following assertions hold almost surely, provided that n is sufficiently large.*

a) *There exists a neighborhood* $S_\Delta(\boldsymbol{\alpha}^*) := \{\boldsymbol{\alpha} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\| \leq \Delta\}$ *of* $\boldsymbol{\alpha}^*$, *with* Δ *being independent of n, in which the random mapping* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *constitutes a contractive mapping, and hence possesses a unique fixed-point* $\hat{\boldsymbol{\alpha}}$.

b) *The sequence* $\{\hat{\boldsymbol{\alpha}}_m\}$ *produced by the fixed-point iteration* (4.8) *converges to* $\hat{\boldsymbol{\alpha}}$ *as* $m \to \infty$ *provided that* $\hat{\boldsymbol{\alpha}}_0 \in S_\delta(\hat{\boldsymbol{\alpha}})$, *where* $S_\delta(\hat{\boldsymbol{\alpha}}) := \{\boldsymbol{\alpha} : \|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}\| \leq \delta\}$ *is a neighborhood of* $\hat{\boldsymbol{\alpha}}$, *with* δ *being independent of* n.

PROOF. Since $\mathbf{C}(\boldsymbol{\alpha})$ is continuous under (A4), it follows from Lemma 4.2 that $\mathbf{a}'(\boldsymbol{\alpha}^*) = \mathbf{C}(\boldsymbol{\alpha}^*)$, and that $\|\mathbf{a}'(\boldsymbol{\alpha}^*)\| < 1$ for some norm $\|\cdot\|$. Furthermore, the continuity of $\mathbf{a}'(\boldsymbol{\alpha})$ and $\mathbf{C}(\boldsymbol{\alpha})$ under (A4) also guarantees the existence of a constant $0 < c < 1$ and a neighborhood $S_{\Delta_0}(\boldsymbol{\alpha}^*) := \{\boldsymbol{\alpha} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\| \leq \Delta_0\} \subseteq \mathcal{A}$ such that

$$\|\mathbf{a}'(\boldsymbol{\alpha})\| \leq c \quad \text{and} \quad \|\mathbf{C}(\boldsymbol{\alpha})\| \leq c \tag{4.14}$$

for all $\boldsymbol{\alpha} \in S_{\Delta_0}(\boldsymbol{\alpha}^*)$. Let $\varsigma := (c+1)/2 < 1$, then the uniform convergence of $\hat{\mathbf{a}}'(\boldsymbol{\alpha})$ to $\mathbf{a}'(\boldsymbol{\alpha})$ implies that $\|\hat{\mathbf{a}}'(\boldsymbol{\alpha})\| \leq \varsigma$ almost surely for all $\boldsymbol{\alpha} \in S_{\Delta_0}(\boldsymbol{\alpha}^*)$ when $n \geq N_0$, where N_0 is independent of $\boldsymbol{\alpha}$. On the other hand, using the mean-value theorem (Ortega and Rheinboldt, 1970, p. 71) we can show that for all $n \geq N_0$

$$\|\hat{\mathbf{a}}(\boldsymbol{\alpha}_1) - \hat{\mathbf{a}}(\boldsymbol{\alpha}_2)\| \leq \varsigma \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| \tag{4.15}$$

almost surely and uniformly in $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in S_{\Delta_0}(\boldsymbol{\alpha}^*)$. By definition (Stoer and Bulirsch, 1980, p. 251), this implies that the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ is *contractive* on $S_{\Delta_0}(\boldsymbol{\alpha}^*)$ almost surely. In particular, the inequality (4.15) is valid almost surely on the smaller neighborhood $S_\Delta(\boldsymbol{\alpha}^*) := \{\boldsymbol{\alpha} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\| \leq \Delta\}$ with $\Delta = \frac{1}{2}\Delta_0$ for all $n \geq N_0$. Furthermore, by Lemma 4.1, the convergence of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ to $\mathbf{a}(\boldsymbol{\alpha})$ guarantees that

$$\|\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) - \boldsymbol{\alpha}^*\| = \|\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) - \mathbf{a}(\boldsymbol{\alpha}^*)\| \leq (1 - \varsigma)\Delta$$

almost surely for all $n \geq N$. Therefore, according to Theorem 5.2.3 of Stoer and Bulirsch (1980), the mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ has a unique fixed-point $\hat{\boldsymbol{\alpha}}$ on $S_\Delta(\boldsymbol{\alpha}^*)$ almost surely whenever $n \geq \max(N, N_0)$. Part a) of the theorem is thus proved. To show Part b), we take a constant δ such that $0 < \delta < \frac{1}{2}\Delta_0$. Then the neighborhood $S_\delta(\hat{\boldsymbol{\alpha}})$ of the fixed-point $\hat{\boldsymbol{\alpha}}$ is contained in $S_{\Delta_0}(\boldsymbol{\alpha}^*)$ and hence the inequality (4.15) remains

valid almost surely and uniformly in $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in S_\delta(\hat{\boldsymbol{\alpha}})$, provided that $n \geq \max(N, N_0)$. By Theorem 5.2.2 of Stoer and Bulirsch (1980), the sequence $\{\hat{\boldsymbol{\alpha}}_m\}$ produced by (4.8) converges to $\hat{\boldsymbol{\alpha}}$ as $m \to \infty$ almost surely, if $n \geq \max(N_1, N_0)$ and the initial value $\hat{\boldsymbol{\alpha}}_0$ is chosen in the neighborhood $S_\delta(\hat{\boldsymbol{\alpha}})$. $\Diamond$

Remark 4.3 The theory of numerical analysis tells us (Stoer and Bulirsch, 1980) that the spectral radius ϱ of $\mathbf{C}(\boldsymbol{\alpha}^*)$ is crucial to the rate of convergence of the fixed-point iteration in (4.8). Indeed, the smaller the spectral radius ϱ is, the faster is the convergence to $\hat{\boldsymbol{\alpha}}$. As seen in the proof of Lemma 4.2, $\varrho = 1/(1 + \lambda_{\min})$, where $\lambda_{\min} := \min\{\lambda_j\}$. Therefore, to accelerate the fixed-point iteration, $\lambda_{\min}$ should be made as large as possible. Notice that

$$\lambda_{\min} = \min_{\mathbf{p} \neq \mathbf{0}} \frac{\mathbf{p}^H \mathbf{R}_x(\boldsymbol{\alpha}^*)\,\mathbf{p}}{\mathbf{p}^H \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{p}}.$$

In the case of $q = 1$, $\lambda_{\min}$ reduces to $r_0^x(\boldsymbol{\alpha}^*)/r_0^\epsilon(\boldsymbol{\alpha}^*)$, which is readily recognized to be the signal-to-noise ratio of the filtered process $\{y_t(\boldsymbol{\alpha}^*)\}$. For $q > 1$, $\lambda_{\min}$ can be regarded as a generalized indicator of the amount of signal relative to the amount of noise in the filtered process $\{y_t(\boldsymbol{\alpha}^*)\}$. This implies that the fixed-point iteration can be accelerated if the parametric filter enhances the sinusoidal signal when $\boldsymbol{\alpha}$ takes on values in a neighborhood of $\boldsymbol{\alpha}^*$.

Remark 4.4 In order to accommodate poor initial guesses, it is required to have a large neighborhood of $\hat{\boldsymbol{\alpha}}$ — the convergence region — in which the fixed-point iteration in (4.8) converges. Theorem 4.2 gives a conservative estimate of this region — namely, the neighborhood $S_\delta(\hat{\boldsymbol{\alpha}})$ — which basically consists of those $\boldsymbol{\alpha}$ for which $\|\hat{\mathbf{a}}'(\boldsymbol{\alpha})\| \leq \varsigma < 1$. Therefore, the filter should be selected to provide a large set of such $\boldsymbol{\alpha}$ in order to obtain a large convergence region.

4.2.2 Strong Consistency and Asymptotic Normality

Suppose that $\hat{\boldsymbol{\alpha}}$ is the fixed-point of the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in the vicinity of $\boldsymbol{\alpha}^*$. In the following, we shall investigate asymptotic properties of $\hat{\boldsymbol{\alpha}}$ as the sample size n tends to infinity. The following theorem claims the strong consistency of the PF estimator $\hat{\boldsymbol{\alpha}}$ for estimating $\boldsymbol{\alpha}^*$.

Theorem 4.3 *Suppose that* (A1)–(A4) *are satisfied, and let* $\hat{\boldsymbol{\alpha}}$ *be the unique fixed-point of* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *in* $S_\Delta(\boldsymbol{\alpha}^*)$, *as given by Theorem* 4.2. *Then* $\hat{\boldsymbol{\alpha}}$ *converges to* $\boldsymbol{\alpha}^*$ *almost surely as* $n \to \infty$.

PROOF. Since $\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) = \hat{\boldsymbol{\alpha}}$ and $\mathbf{a}(\boldsymbol{\alpha}^*) = \boldsymbol{\alpha}^*$, it follows from (4.6) that

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^* = \delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) + \mathbf{a}(\hat{\boldsymbol{\alpha}}) - \mathbf{a}(\boldsymbol{\alpha}^*) = \delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) + \mathbf{C}(\hat{\boldsymbol{\alpha}})\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \tag{4.16}$$

where $\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}) := \hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{a}(\boldsymbol{\alpha})$. By (4.14), we have $\|\mathbf{C}(\boldsymbol{\alpha})\| \le c < 1$ for any $\boldsymbol{\alpha} \in S_\Delta(\boldsymbol{\alpha}^*)$. We find that

$$\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| \le \|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| + \|\mathbf{C}(\hat{\boldsymbol{\alpha}})\|\,\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| \le \|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| + c\,\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|$$

and hence

$$\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| \le (1-c)^{-1}\,\|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| \tag{4.17}$$

almost surely for large n. The uniform convergence of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ to $\mathbf{a}(\boldsymbol{\alpha})$, as claimed in Lemma 4.1, guarantees that $\|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| \overset{a.s.}{\to} 0$ as $n \to \infty$, which, together with (4.17), proves the assertion. ◊

The asymptotic normality can also be established for the PF estimator $\hat{\boldsymbol{\alpha}}$. To this end, we first need the following lemma regarding the asymptotic normality of $\hat{\mathbf{a}}(\boldsymbol{\alpha}^*)$.

Lemma 4.3 *Suppose that* (A1)–(A3) *are satisfied and that* $E(\xi_t^4) = \kappa\sigma_\xi^4 < \infty$. *Then, as* $n \to \infty$, $\sqrt{n}\,\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) = \sqrt{n}\,\{\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) - \mathbf{a}\}$ *converges in distribution to a normal random vector with mean zero and covariance matrix*

$$\mathbf{V} := \mathbf{R}_y^{-1}(\boldsymbol{\alpha}^*)\,\mathbf{Q}^T\mathbf{W}(\boldsymbol{\alpha}^*)\,\mathbf{Q}\,\mathbf{R}_y^{-1}(\boldsymbol{\alpha}^*)$$

where $\mathbf{W}(\boldsymbol{\alpha}^*) := [w_{ij}(\boldsymbol{\alpha}^*)]$, $(i, j = 1, \ldots, 2q-1)$, *and*

$$w_{ij}(\boldsymbol{\alpha}^*) := 4 \sum_{\tau=0}^{\infty} \left\{ \sum_{k=0}^{2q} a_k \, r^{\epsilon}_{\tau+i-k}(\boldsymbol{\alpha}^*) \right\} \left\{ \sum_{k=0}^{2q} a_k \, r^{\epsilon}_{\tau+j-k}(\boldsymbol{\alpha}^*) \right\}. \tag{4.18}$$

PROOF. The proof utilizes the large sample theory in Chapter 3 concerning the asymptotic normality of sample autocovariances of filtered data. First of all, we have

$$\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) = -\{(\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{Y}}\mathbf{Q})^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{y}} + \mathbf{a}\}.$$

Here, as well as in the following, the argument $\boldsymbol{\alpha}^*$ is omitted for the sake of brevity. As in the proof of Theorem 3.3, we can write

$$n^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{Y}}\mathbf{Q} = \hat{\mathbf{R}} + o_P(n^{-1/2}) \quad \text{and} \quad n^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{y}} = \hat{\mathbf{r}} + o_P(n^{-1/2}), \tag{4.19}$$

where $\hat{\mathbf{R}} := \mathbf{Q}^T\check{\mathbf{R}}\mathbf{Q}$, $\hat{\mathbf{r}} := \mathbf{Q}^T(\check{\mathbf{r}} + \check{\mathbf{r}}^B) = 2\,\mathbf{Q}^T\check{\mathbf{r}}$, with

$$\check{\mathbf{R}} := \begin{bmatrix} \hat{r}_0 & \cdots & \hat{r}_{2q-2} \\ \vdots & \ddots & \vdots \\ \hat{r}_{-2q+2} & \cdots & \hat{r}_0 \end{bmatrix} \qquad \check{\mathbf{r}} := \begin{bmatrix} \hat{r}_{-1} \\ \vdots \\ \hat{r}_{-2q+1} \end{bmatrix}$$

and

$$\hat{r}_j := n^{-1} \sum_{t=1}^{n-j} \hat{y}_{t+j}(\boldsymbol{\alpha}^*)\, \hat{y}_t(\boldsymbol{\alpha}^*) \qquad (j = 0, 1, \ldots, 2q-1).$$

Using these results, in connection with the nonsingularity of $\hat{\mathbf{R}}$, we obtain

$$\begin{aligned} \delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) &= -(\hat{\mathbf{R}}^{-1}\hat{\mathbf{r}} + \mathbf{a}) + o_P(n^{-1/2}) \\ &= -\hat{\mathbf{R}}^{-1}(\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a}) + o_P(n^{-1/2}). \end{aligned} \tag{4.20}$$

Let $r_j := r_j^y(\boldsymbol{\alpha}^*)$ for brevity. Then, according to Theorem 3.3, $\sqrt{n}\,(\hat{r}_j - r_j), (j = 0, \ldots, 2q-1)$ are asymptotically jointly normal with mean zero and covariance matrix

$\mathbf{V}_r := [\sigma_{ij}]$, where

$$\begin{aligned}\sigma_{ij} &:= 2\sum_{k=1}^{q}\beta_k^2(\boldsymbol{\alpha}^*)\cos(i\omega_k)\cos(j\omega_k)\sum_{\tau=-\infty}^{\infty}r_\tau^\epsilon(\boldsymbol{\alpha}^*)\cos(\tau\omega_k)\\ &\quad+(\kappa-3)\,r_i^\epsilon(\boldsymbol{\alpha}^*)\,r_j^\epsilon(\boldsymbol{\alpha}^*)\\ &\quad+\sum_{\tau=-\infty}^{\infty}\{r_\tau^\epsilon(\boldsymbol{\alpha}^*)\,r_{\tau+i-j}^\epsilon(\boldsymbol{\alpha}^*)+r_{\tau+i}^\epsilon(\boldsymbol{\alpha}^*)\,r_{\tau-j}^\epsilon(\boldsymbol{\alpha}^*)\}\end{aligned}\tag{4.21}$$

for $i, j = 0, 1, \ldots, 2q-1$. On the other hand, the vector $\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a}$ in (4.20) can be regarded as the value of some function $\mathbf{f}(u_0, \ldots, u_{2q-1})$ at $u_j = \hat{r}_j$, that is, $\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a} = \mathbf{f}(\hat{r}_0, \ldots, \hat{r}_{2q-1})$. For this function, it is also true, by (4.1), that

$$\begin{aligned}\mathbf{f}(r_0, \ldots, r_{2q-1}) &= \mathbf{r}_y(\boldsymbol{\alpha}^*) + \mathbf{R}_y(\boldsymbol{\alpha}^*)\,\mathbf{a}\\ &= \mathbf{r}_y(\boldsymbol{\alpha}^*) + \mathbf{R}_y(\boldsymbol{\alpha}^*)\,\mathbf{a}(\boldsymbol{\alpha}^*) \;=\; \mathbf{0}.\end{aligned}$$

Therefore, invoking Proposition 6.4.3 of Brockwell and Davis (1987) proves that

$$\sqrt{n}\,(\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a}) = \sqrt{n}\,\{\mathbf{f}(\hat{r}_0, \ldots, \hat{r}_{2q-1}) - \mathbf{f}(r_0, \ldots, r_{2q-1})\}$$

converges in distribution to a normal random vector with mean zero and covariance matrix $\mathbf{V}_f := \mathbf{F}\mathbf{V}_r\mathbf{F}^T$, where $\mathbf{F}$ is the Jacobian matrix of $\mathbf{f}$ evaluated at $(r_0, \ldots, r_{2q-1})$. It is easy to verify that

$$\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a} = \mathbf{Q}^T(\check{\mathbf{r}} + \check{\mathbf{r}}^B + \check{\mathbf{R}}\mathbf{Q}\mathbf{a}) = \mathbf{Q}^T[\check{\mathbf{r}} \,\vdots\, \check{\mathbf{R}} \,\vdots\, \check{\mathbf{r}}^B]\,\bar{\mathbf{a}},$$

where $\bar{\mathbf{a}} := [a_0, a_1, \cdots, a_{2q}]^T$. Simple algebra shows $\mathbf{F} = [\mathbf{f}_0, \cdots, \mathbf{f}_{2q-1}]$, where

$$\mathbf{f}_0 = \mathbf{Q}^T\begin{bmatrix} a_1 \\ \vdots \\ a_{2q-1}\end{bmatrix} \quad \text{and} \quad \mathbf{f}_j = \mathbf{Q}^T\begin{bmatrix} a_{1-j} + a_{1+j} \\ \vdots \\ a_{2q-1-j} + a_{2q-1+j}\end{bmatrix}$$

for $j = 1, \ldots, 2q-1$, with $a_k := 0$ for $k < 0$ and $k > 2q$. Upon noting that σ_{ij} given by (4.21) are symmetric in the sense that $\sigma_{-i,j} = \sigma_{i,-j} = \sigma_{-i,-j} = \sigma_{ij}$, we can rewrite $\mathbf{V}_f$ as $\mathbf{V}_f = \mathbf{Q}^T\mathbf{B}\mathbf{Q}$, where $\mathbf{B} := [b_{ij}]$, $(i, j = 1, \ldots, 2q-1)$, with

$$b_{ij} := \sum_{u,v=0}^{2q}\sigma_{i-u,j-v}\;a_u\,a_v\,.$$

As can be seen from (4.21), there are three groups in the expression of σ_{ij}. The first group involves the sinusoidal terms, all of which are cancelled out in the expression of b_{ij}, because $A(z_k^{-1}) = 0$ for $k = 1, \ldots, q$ and hence

$$\sum_{v=0}^{2q} a_v \cos(\omega_k(j-v)) = 0 \qquad (j = 1, \ldots, 2q-1;\ k = 1, \ldots, q).$$

The second group in (4.21) consists of $(\kappa - 3)\, r_i^{\epsilon}(\boldsymbol{\alpha}^*)\, r_j^{\epsilon}(\boldsymbol{\alpha}^*)$. It is easy to see that the corresponding term in $\mathbf{V}_f$ can be written as $(\kappa - 3)\,\mathbf{U}\mathbf{U}^T$, where

$$\begin{aligned} \mathbf{U} &:= \mathbf{Q}^T[\bar{\mathbf{r}}_{\epsilon}(\boldsymbol{\alpha}^*) \,\vdots\, \bar{\mathbf{R}}_{\epsilon}(\boldsymbol{\alpha}^*) \,\vdots\, \bar{\mathbf{r}}_{\epsilon}^{B}(\boldsymbol{\alpha}^*)]\,\bar{\mathbf{a}} \\ &= \mathbf{r}_{\epsilon}(\boldsymbol{\alpha}^*) + \mathbf{R}_{\epsilon}(\boldsymbol{\alpha}^*)\,\mathbf{a} \;=\; \mathbf{r}_{\epsilon}(\boldsymbol{\alpha}^*) + \mathbf{R}_{\epsilon}(\boldsymbol{\alpha}^*)\,\boldsymbol{\alpha}^*. \end{aligned}$$

Since $\mathbf{U} = \mathbf{0}$ under (A2), this term also vanishes in $\mathbf{V}_f$. Combining these results, we obtain $\mathbf{V}_f = \mathbf{Q}^T\mathbf{B}_0\,\mathbf{Q}$ where $\mathbf{B}_0 := [b_{ij}^{(0)}]$ with

$$b_{ij}^{(0)} := \sum_{\tau=-\infty}^{\infty} \sum_{u,\,v=0}^{2q} a_u\, a_v\; r_{\tau+i-u}^{\epsilon}(\boldsymbol{\alpha}^*)\{r_{\tau+j-v}^{\epsilon}(\boldsymbol{\alpha}^*) + r_{\tau-u+v}^{\epsilon}(\boldsymbol{\alpha}^*)\}.$$

Furthermore, it is not difficult to verify that $\mathbf{B}_0$ can be written compactly as

$$\mathbf{B}_0 = \sum_{\tau=-\infty}^{\infty} \mathbf{A}^T\mathbf{r}_{\tau}(\boldsymbol{\alpha}^*)\,\mathbf{r}_{\tau}^T(\boldsymbol{\alpha}^*)\,(\mathbf{A} + \tilde{\mathbf{I}}\mathbf{A}),$$

where $\mathbf{A}$ is a $(4q-1)$-by-$(2q-1)$ matrix of the form

$$\mathbf{A} := \begin{bmatrix} a_{2q} & & 0 \\ \vdots & \ddots & \\ a_0 & & a_{2q} \\ & \ddots & \vdots \\ 0 & & a_0 \end{bmatrix} \tag{4.22}$$

and $\mathbf{r}_{\tau}(\boldsymbol{\alpha}^*) := [r_{\tau-2q+1}^{\epsilon}(\boldsymbol{\alpha}^*), \cdots, r_{\tau+2q-1}^{\epsilon}(\boldsymbol{\alpha}^*)]^T$. Since $\tilde{\mathbf{I}}\mathbf{A} = \mathbf{A}\tilde{\mathbf{I}}$ and $\tilde{\mathbf{I}}\mathbf{Q} = \mathbf{Q}$, we obtain $\tilde{\mathbf{I}}\mathbf{A}\mathbf{Q} = \mathbf{A}\mathbf{Q}$. Therefore,

$$\mathbf{V}_f = 2\sum_{\tau=-\infty}^{\infty} \mathbf{Q}^T\mathbf{A}^T\mathbf{r}_{\tau}(\boldsymbol{\alpha}^*)\,\mathbf{r}_{\tau}^T(\boldsymbol{\alpha}^*)\,\mathbf{A}\mathbf{Q}.$$

This expression can be further simplified as

$$\mathbf{V}_f = 4\sum_{\tau=0}^{\infty} \mathbf{Q}^T\mathbf{A}^T\mathbf{r}_\tau(\boldsymbol{\alpha}^*)\,\mathbf{r}_\tau^T(\boldsymbol{\alpha}^*)\,\mathbf{A}\mathbf{Q} = \mathbf{Q}^T\mathbf{W}(\boldsymbol{\alpha}^*)\,\mathbf{Q},$$

upon noting that $\mathbf{r}_{-\tau}(\boldsymbol{\alpha}) = \tilde{\mathbf{I}}\,\mathbf{r}_\tau(\boldsymbol{\alpha})$, $\mathbf{Q}^T\mathbf{A}^T\mathbf{r}_0(\boldsymbol{\alpha}^*) = \mathbf{U} = \mathbf{0}$, and

$$\mathbf{W}(\boldsymbol{\alpha}^*) = 4\sum_{\tau=0}^{\infty} \mathbf{A}^T\mathbf{r}_\tau(\boldsymbol{\alpha}^*)\,\mathbf{r}_\tau^T(\boldsymbol{\alpha}^*)\,\mathbf{A}.$$

Finally, since $\hat{\mathbf{R}} \overset{a.s.}{\to} \mathbf{R}_y(\boldsymbol{\alpha}^*)$, by Slutsky's theorem (Lehmann, 1982, Lemma 4.1, pp. 432–433) we obtain from (4.20) that $\sqrt{n}\,\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) = -\hat{\mathbf{R}}^{-1}\sqrt{n}\,(\hat{\mathbf{r}}+\hat{\mathbf{R}}\mathbf{a})+o_P(1)$ converges in distribution to $\mathrm{N}(\mathbf{0},\mathbf{V})$. ◊

With the aid of this lemma, we are now able to show the asymptotic normality of the PF estimator $\hat{\boldsymbol{\alpha}}$.

Theorem 4.4 *Under the conditions in Theorem* 4.3, $\sqrt{n}\,(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*)$ *converges in distribution as* $n \to \infty$ *to a normal random vector with mean zero and covariance matrix*

$$\mathbf{V}_\alpha = \mathbf{R}_x^{-1}(\boldsymbol{\alpha}^*)\,\mathbf{Q}^T\mathbf{W}(\boldsymbol{\alpha}^*)\,\mathbf{Q}\,\mathbf{R}_x^{-1}(\boldsymbol{\alpha}^*) \tag{4.23}$$

where $\mathbf{W}(\boldsymbol{\alpha}^*)$ *is defined in Lemma* 4.3.

PROOF. It follows from (4.16) that $\{\mathbf{I}-\mathbf{C}(\hat{\boldsymbol{\alpha}})\}\,(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*) = \delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})$. According to the mean-value theorem (Ortega and Rheinboldt, 1970, p. 71), $\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})$ can be written as

$$\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) = \delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) + \left\{\int_0^1 \delta\hat{\mathbf{a}}'(\boldsymbol{\alpha}^* + \lambda(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*))\,d\lambda\right\}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*)$$

where $\delta\hat{\mathbf{a}}'(\boldsymbol{\alpha})$ is the Jacobian matrix of $\delta\hat{\mathbf{a}}(\boldsymbol{\alpha})$. Since $\delta\hat{\mathbf{a}}'(\boldsymbol{\alpha}) = \hat{\mathbf{a}}'(\boldsymbol{\alpha}) - \mathbf{a}'(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathbf{0}$ uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$, as guaranteed by Lemma 4.1, we have

$$\int_0^1 \delta\hat{\mathbf{a}}'(\boldsymbol{\alpha}^* + \lambda(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*))\,d\lambda \overset{a.s.}{\to} \mathbf{0}.$$

In addition, the consistency of $\hat{\boldsymbol{\alpha}}$, together with the continuity of $\mathbf{C}(\boldsymbol{\alpha})$, implies that $\mathbf{C}(\hat{\boldsymbol{\alpha}}) \overset{a.s.}{\to} \mathbf{C}(\boldsymbol{\alpha}^*)$. Therefore, by Slutsky's theorem, $\sqrt{n}\,(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*)$ has the same asymptotic distribution as $\sqrt{n}\,\{\mathbf{I}-\mathbf{C}(\boldsymbol{\alpha}^*)\}^{-1}\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*)$. Invoking Lemma 4.3 shows that

$\sqrt{n}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ converges in distribution to $\mathrm{N}(\mathbf{0}, \mathbf{V}_\alpha)$, where $\mathbf{V}_\alpha := \{\mathbf{I} - \mathbf{C}(\boldsymbol{\alpha}^*)\}^{-1}\mathbf{V}\,\{\mathbf{I} - \mathbf{C}(\boldsymbol{\alpha}^*)\}^{-1}$. Furthermore, using the expression of $\mathbf{C}(\boldsymbol{\alpha})$ in (4.13) and applying the matrix-inversion formula (Haykin, 1986), we can write

$$\{\mathbf{I} - \mathbf{C}(\boldsymbol{\alpha})\}^{-1} = \mathbf{I} + \boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha}) = \boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha})\,\{\mathbf{I} + \boldsymbol{\Gamma}(\boldsymbol{\alpha})\}.$$

On the other hand, $\mathbf{R}_y(\boldsymbol{\alpha}) = \mathbf{R}_x(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha}) = \mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\{\mathbf{I} + \boldsymbol{\Gamma}(\boldsymbol{\alpha})\}$. Therefore, we obtain $\{\mathbf{I} - \mathbf{C}(\boldsymbol{\alpha})\}^{-1}\mathbf{R}_y^{-1}(\boldsymbol{\alpha}) = \boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha})\,\mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha}) = \mathbf{R}_x^{-1}(\boldsymbol{\alpha})$. Substituting this result in the above expression of $\mathbf{V}_\alpha$ completes the proof. ◊

Remark 4.5 The asymptotic variance $\mathbf{V}_\alpha$ in (4.23) is seen to be inversely related to the signal-to-noise ratio in the filtered process $\{y_t(\boldsymbol{\alpha}^*)\}$. This is compatible with the requirement that the filter should effectively enhance the sinusoids and suppress the noise. According to Remark 4.3, the convergence of the fixed-point iteration in (4.8) can also be accelerated by appropriate filters that improve the SNR.

4.3 Extension to Complex Sinusoids

A parallel theory of the PF method can be easily established for complex sinusoids in additive noise. In fact, if $\{x_t\}$ is a sum of p complex sinusoids given by

$$x_t := \sum_{k=1}^{p} \beta_k\, e^{i(\omega_k t + \phi_k)}$$

with $0 < \omega_1 < \cdots < \omega_p < 2\pi$, then it satisfies a pth-order AR autoregressive equation of the form $\sum_{j=0}^{p} a_j\, x_{t-j} = 0$, where the AR parameter vector $\mathbf{a} := [a_1, \cdots, a_p]^T$ is defined by the coefficients of the polynomial

$$\sum_{j=0}^{p} a_j\, z^{-j} = \prod_{k=1}^{p}(1 - z_k z^{-1})$$

with $z_k := \exp(i\omega_k)$. Given a finite data set $\{y_1, \ldots, y_n\}$ observed from (1.1), one of the widely-used estimators of the AR parameter $\mathbf{a}$ is given by

$$\mathbf{a}_{\mathrm{LS}} := -(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H\mathbf{y} \tag{4.24}$$

where $\mathbf{Y}$ and $\mathbf{y}$ are redefined by

$$\mathbf{Y} := \begin{bmatrix} y_p & \cdots & y_1 \\ \vdots & & \vdots \\ y_{n-1} & \cdots & y_{n-p} \end{bmatrix} \quad \text{and} \quad \mathbf{y} := \begin{bmatrix} y_{p+1} \\ \vdots \\ y_n \end{bmatrix}.$$

This estimator is known as the forward linear prediction (FLP) (Kay and Marple, 1981) which minimizes the criterion $\|\mathbf{y}+\mathbf{Ya}\|^2$. Other procedures such as the forward-backward linear prediction (FBLP) (Kay and Marple, 1981) are also applicable for estimating the AR parameter $\mathbf{a}$ with only a slight modification of $\mathbf{Y}$ and $\mathbf{y}$ in (4.24).

Introducing a parametric filter $\{h_j(\boldsymbol{\alpha})\}$ indexed by $\boldsymbol{\alpha} := [\alpha_1, \cdots, \alpha_p]^T$, an estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ of the AR parameter $\mathbf{a}$ can be obtained according to (4.24), with the data matrices replaced by those of the filtered data $\{\hat{y}_t(\boldsymbol{\alpha})\}$ in (4.9). The parametrization property (A2) retains its form, but $\mathbf{R}_\epsilon(\boldsymbol{\alpha})$ and $\mathbf{r}_\epsilon(\boldsymbol{\alpha})$ are now of the structure

$$\mathbf{R}_\epsilon := \begin{bmatrix} r_0^\epsilon & r_1^\epsilon & \cdots & r_{p-1}^\epsilon \\ r_{-1}^\epsilon & r_0^\epsilon & \cdots & r_{p-2}^\epsilon \\ \vdots & \vdots & \ddots & \vdots \\ r_{-p+1}^\epsilon & r_{-p+2}^\epsilon & \cdots & r_0^\epsilon \end{bmatrix} \quad \mathbf{r}_\epsilon := \begin{bmatrix} r_{-1}^\epsilon \\ r_{-2}^\epsilon \\ \vdots \\ r_{-p}^\epsilon \end{bmatrix} \tag{4.25}$$

with r_τ^ϵ replaced by $E\{\overline{\epsilon_{t+\tau}(\boldsymbol{\alpha})}\,\epsilon_t(\boldsymbol{\alpha})\}$. In this case, (A2) is readily recognized as being the Yule-Walker equations. In other words, for complex sinusoids, (A2) can be interpreted as the requirement that the filter be parametrized so that the parameter $\boldsymbol{\alpha}$ satisfies the Yule-Walker equation for the filtered noise. With this property being fulfilled, the PF estimator $\hat{\boldsymbol{\alpha}}$ is defined as the fixed-point of the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ and can be obtained by the fixed-point iteration in (4.8).

Chapter 5

PF Method with AR (All-Pole) Filter

Although its consistency is guaranteed by the asymptotic theory we developed in Section 4.2, the accuracy of the PF estimator depends on the choice of the parametric filter to be applied to the data (see also Remark 4.5). Intuitively, a "good" filter should be bandpass — so that the sinusoidal signal can be effectively enhanced after filtering. A useful example of such a filter is the AR filter (not to be confused with the AR method) we shall consider in this chapter as an illustration of the PF method.

The AR filter — also known as the all-pole filter — is considered in the literature as a filter that whitens the error term in the AR representation of $\{y_t\}$. In fact, as we have seen in Chapter 2, the reason that Prony's estimator leads to inconsistent frequency estimates is that $\{y_t\}$ in (1.1) does not obey an AR model which requires the error $\{e_t\}$ to be uncorrelated (white) in the AR representation (2.4). Instead, the process $\{e_t\}$ is colored, admitting the moving-average (MA) representation (2.14). Clearly, when $\{\epsilon_t\}$ is an i.i.d. random sequence, $\{e_t\}$ would be an MA process with the MA coefficients being identical to the AR coefficients in (2.4). This observation gave rise to the idea of iteratively whitening the error $\{e_t\}$ by AR filtering based on previously estimated AR coefficients from the filtered data. Along this line is the iterative filtering algorithm (IFA), proposed by Kay (1984), which estimates the AR coefficients from iteratively filtered data by Burg's estimator and whitens the error term by an AR filter with poles

on the unit circle. This procedure updates the filter so that the coefficients of the filter on the $(m+1)$st iteration coincide with the estimated AR coefficients from the previous mth iteration. The procedure is known to provide very good frequency estimates also for relatively low SNR. However, since IFA uses a filter whose poles are *on* the unit circle, the bandwidth is very narrow and the iterative procedure requires a rather precise initial guess. In a recent paper by Dragošević and Stanković (1989), iterative least squares estimation was used in connection with an AR filter which has an extra parameter to force the poles to be *within* the unit circle. This extra feature guarantees the stability, and also controls the bandwidth of the filter. The resulting estimator is referred to as the generalized least squares (GLS) estimator. (GLS without the extra bandwidth parameter can be found in Mataušek, *et al.* (1983) where instability of AR filtering was reported, especially when the SNR is low.) Simulations indicate that the GLS estimator (endowed with the bandwidth parameter) improves on the performance of the IFA estimator for high SNR (Dragošević and Stanković, 1989). In general, however, the GLS estimator is not consistent.

In this chapter, we show that the AR filter used by Dragošević and Stanković (1989) can be readily modified to satisfy the parametrization property (A2). With this modification, the resulting PF estimator outperforms the GLS estimator in terms of mean-squared error, and especially so for closely-spaced frequencies.

5.1 General AR Filter and PF Estimator

The AR($2q$) filter (or simply the AR filter) is a parametric filter defined by[1]

$$y_t(\boldsymbol{\alpha}) + \theta_1(\boldsymbol{\alpha})\,\eta\, y_{t-1}(\boldsymbol{\alpha}) + \cdots + \theta_{2q}(\boldsymbol{\alpha})\,\eta^{2q}\, y_{t-2q}(\boldsymbol{\alpha}) = y_t \tag{5.1}$$

where $\eta \in (0, 1]$ is the bandwidth parameter that attracts, when taking on values less than 1, the poles of the filter inside the unit circle, and controls the bandwidth of the

[1]Assuming zero initial values yields the filtered data $\hat{y}_1(\boldsymbol{\alpha}), \ldots, \hat{y}_n(\boldsymbol{\alpha})$ for the calculation of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$.

filter. The coefficients $\theta_j(\boldsymbol{\alpha})$ of the filter are functions of $\boldsymbol{\alpha}$, and are symmetric in the sense that

$$\theta_0(\boldsymbol{\alpha}) = 1 \quad \text{and} \quad \theta_{2q-j}(\boldsymbol{\alpha}) = \theta_j(\boldsymbol{\alpha}) \qquad (j = 0, 1, \ldots, q).$$

In the GLS method (Dragošević and Stanković, 1989), this filter was employed with a specific choice of the filter coefficients so that $\theta_j(\boldsymbol{\alpha}) = \alpha_j$, that is,

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = \boldsymbol{\alpha} \tag{5.2}$$

where $\boldsymbol{\theta}(\boldsymbol{\alpha})$ is the q-dimensional vector

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) := [\theta_1(\boldsymbol{\alpha}), \cdots, \theta_q(\boldsymbol{\alpha})]^T$$

A similar filter without η (or, equivalently, with $\eta = 1$) was also used by Kay (1984) for estimating complex sinusoids.

Let us assume that the additive noise $\{\epsilon_t\}$ in (1.2) is white with zero mean and finite fourth moment. It will be shown that a very simple relationship between $\boldsymbol{\theta}(\boldsymbol{\alpha})$ and $\boldsymbol{\alpha}$ can be established so that the resulting filter possesses the parametrization property (A2), and that the theoretical results in Chapter 4 can be applied to this filter upon appropriately selecting $\boldsymbol{\alpha}$ in a parameter space.

5.1.1 Parametrization

As we have seen in Chapter 3, the PF method requires the filter to possess the parametrization property (A2). For white noise $\{\epsilon_t\}$, the filtered noise $\{\epsilon_t(\boldsymbol{\alpha})\}$ is readily recognized as being an AR process with the autocovariance function $r_\tau^\epsilon(\boldsymbol{\alpha})$ satisfying the equations

$$\sum_{j=0}^{2q} \eta^j \, \theta_j(\boldsymbol{\alpha}) \, r_{\tau-j}^\epsilon(\boldsymbol{\alpha}) = 0 \qquad (\tau = 1, 2, \ldots). \tag{5.3}$$

In matrix notation, (5.3) can be written as

$$\bar{\mathbf{r}}_\epsilon(\boldsymbol{\alpha}) + \eta^{2q} \, \bar{\mathbf{r}}_\epsilon^B(\boldsymbol{\alpha}) = -\bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha}) \, \tilde{\mathbf{Q}} \, \boldsymbol{\theta}(\boldsymbol{\alpha}) \tag{5.4}$$

where $\tilde{\mathbf{Q}}$ is a $(2q-1)$-by-q matrix of the form

$$\tilde{\mathbf{Q}} := \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0}^T & \eta^q \\ \mathbf{Q}_2 & \mathbf{0} \end{bmatrix}$$

with $\mathbf{Q}_1 := \mathrm{diag}(\eta, \ldots, \eta^{q-1})$ and $\mathbf{Q}_2 := \eta^q \mathbf{Q}_1 \tilde{\mathbf{I}}$. Now, by pre-multiplying each side of (5.4) with $2\,\mathbf{Q}^T$, we obtain

$$(1+\eta^{2q})\,\mathbf{r}_\epsilon(\boldsymbol{\alpha}) = -2\,\mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\,\tilde{\mathbf{Q}}\,\boldsymbol{\theta}(\boldsymbol{\alpha}).$$

Therefore, the parametrization property (A2) requires that

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = \tfrac{1}{2}(1+\eta^{2q})\{\mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\,\tilde{\mathbf{Q}}\}^{-1} \mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\boldsymbol{\alpha}. \tag{5.5}$$

On the other hand, a simple calculation shows that

$$\mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\,\tilde{\mathbf{Q}} = \frac{2}{1+\eta^{2q}}\,\mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\,\mathbf{Q}\,\mathbf{T}_\eta^{-1} = \frac{2}{1+\eta^{2q}}\,\mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\mathbf{T}_\eta^{-1}$$

where $\mathbf{T}_\eta$ is a q-by-q diagonal matrix of the form

$$\mathbf{T}_\eta := \mathrm{diag}\left(\frac{1+\eta^{2q}}{\eta+\eta^{2q-1}}, \ldots, \frac{1+\eta^{2q}}{\eta^{q-1}+\eta^{q+1}}, \frac{1+\eta^{2q}}{2\,\eta^q}\right).$$

As a result, the required parametrization (5.5) is simplified to trivial linear equations

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = \mathbf{T}_\eta\,\boldsymbol{\alpha}. \tag{5.6}$$

With the coefficients given by (5.6), the AR filter in (5.1) satisfies (A2) so that a sequence $\{\hat{\boldsymbol{\alpha}}_m\}$ can be generated by the fixed-point iteration in (4.8) to estimate the AR parameter **a**.

As compared to (5.2), we observe that when $\eta < 1$, the parametrization of the PF method differs from that of the GLS method. It is this difference that makes the PF estimator consistent for *any* bandwidth parameter $\eta < 1$, while the GLS estimator is inconsistent. Note that the PF and GLS methods coincide when $\eta = 1$. Moreover,

the iterative filtering algorithm (IFA) of Kay (1984) is easily seen to correspond to the complex version of the PF method with $\eta = 1$. The difference is that Burg's estimator was employed in IFA, instead of the least squares, in order to guarantee the stability of AR filtering (Kay, 1984).

5.1.2 Parameter Space

For the theoretical analysis in Chapter 4 to hold, the AR filter (5.1), with $\boldsymbol{\theta}(\boldsymbol{\alpha})$ given by (5.6), is required to satisfy the regularity conditions (A3) and (A4). Therefore, the parameter $\boldsymbol{\alpha}$ must take on values within a certain *parameter space* in which these conditions are guaranteed.

For any $\eta < 1$, let $\Theta(\eta)$ be the collection of $\boldsymbol{\theta} := [\theta_1, \cdots, \theta_q]^T$ for which the polynomial $Q(z^{-1}) := \sum_{j=0}^{2q} \theta_j z^{-j}$ with $\theta_0 = 1$ and $\theta_{2q-j} = \theta_j$ has $2q$ strictly complex zeros (with non-zero imaginary part) on or inside the circle $|z| = (1-\delta)/\eta$, where $\delta \in (0,1)$ is a small real number such that $1 - \delta \geq \eta$. Let $\mathcal{A}(\eta)$ be the set of $\boldsymbol{\alpha}$ such that $\boldsymbol{\theta}(\boldsymbol{\alpha}) \in \Theta(\eta)$, namely,

$$\mathcal{A}(\eta) := \{\boldsymbol{\alpha} : \ \boldsymbol{\theta}(\boldsymbol{\alpha}) \in \Theta(\eta)\}. \tag{5.7}$$

Then, for $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$, we have the following theorem which claims the validity of (A3) and (A4) for the AR filter (5.1) with $\boldsymbol{\theta}(\boldsymbol{\alpha})$ given by either (5.2) or (5.6).

Theorem 5.1 *Let $\mathcal{A}(\eta)$ be the parameter space defined by* (5.7). *Then the AR filter* (5.1) *satisfies* (A3) *and* (A4) *for $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$ if $\boldsymbol{\theta}(\boldsymbol{\alpha})$ is given by* (5.2) *or* (5.6).

To prove this theorem, we first need the following lemma.

Lemma 5.1 *For any $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$, the poles of the AR filter* (5.1) *either appear on the circle $|z| = \eta \leq 1-\delta$, or occur in reciprocal pairs within the band $\eta^2/(1-\delta) \leq |z| \leq 1-\delta$. Consequently, the AR filter is stable for all $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$.*

PROOF. By definition, $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$ implies that $\boldsymbol{\theta} := \boldsymbol{\theta}(\boldsymbol{\alpha}) \in \Theta(\eta)$. Therefore, there exist constants $0 < \lambda_1 \leq \cdots \leq \lambda_q < \pi$ and $0 < p_k \leq (1-\delta)/\eta$ such that

$$Q(z^{-1}) := \sum_{j=0}^{2q} \theta_j z^{-j} = \prod_{k=1}^{q} (1 - p_k e^{i\lambda_k} z^{-1})(1 - p_k e^{-i\lambda_k} z^{-1}).$$

Let $\nu_k := \eta p_k \exp(i\lambda_k)$ for $k = 1, \ldots, q$ and $\nu_k := \eta p_{2q-k+1} \exp(-i\lambda_{2q-k+1})$ for $k = q+1, \ldots, 2q$. Since $Q(\eta z^{-1})$ can be factorized as

$$Q(\eta z^{-1}) = \prod_{k=1}^{2q} (1 - \nu_k z^{-1}) \tag{5.8}$$

the zeros of $Q(\eta z^{-1})$ — namely, the poles of the AR filter (5.1) — are readily identified to be the complex conjugate pairs (ν_k, ν_{2q-k+1}), $(k = 1, \ldots, q)$. The symmetry of θ_j implies that the zeros of $Q(z^{-1})$ also constitute reciprocal pairs, and hence appear in the region $\eta/(1-\delta) \leq |z| \leq (1-\delta)/\eta$. As a result, the poles of the AR filter must occur within the band $\eta^2/(1-\delta) \leq |z| \leq 1-\delta$. Moreover, for distinct λ_k — the λ_k with multiplicity 1 — the corresponding zeros of $Q(z^{-1})$ occur on the unit circle and hence the corresponding poles of the AR filter lie on the circle $|z| = \eta$. The AR filter is stable[2] for any $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$ because $1/Q(\eta z^{-1})$ is analytic within a small band containing the unit circle, and hence its Taylor series expansion is absolutely summable on the unit circle. ◊

Equipped with this lemma, we now present the proof of Theorem 5.1.

PROOF OF THEOREM 5.1. Let $H(z; \boldsymbol{\alpha}) := 1/P(z; \boldsymbol{\alpha})$ where

$$P(z; \boldsymbol{\alpha}) := \sum_{j=0}^{2q} \theta_j(\boldsymbol{\alpha}) \eta^j z^j.$$

According to Lemma 5.1, $H(z; \boldsymbol{\alpha})$ is analytic in $|z| < (1-\delta)^{-1}$ since the poles of $H(z; \boldsymbol{\alpha})$ are on or outside the circle $|z| = (1-\delta)^{-1} > 1$ for any $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. Given $\rho \in (1-\delta, 1)$, the circle $|z| = \rho^{-1}$ is contained in the analytic region of $H(z; \boldsymbol{\alpha})$. Therefore, by Cauchy's inequality (Markushevich, 1977, vol. 1, Theorem 14.7, p. 302),

[2]Recall that a filter is stable if its impulse response sequence is absolutely summable.

we obtain

$$|H_j(0;\boldsymbol{\alpha})|/j! \leq \rho^j \max_{|z|=\rho^{-1}} |H(z;\boldsymbol{\alpha})| \qquad (j = 0, 1, \ldots) \tag{5.9}$$

where $H_j(z;\boldsymbol{\alpha})$ stands for the jth derivative of $H(z;\boldsymbol{\alpha})$ with respect to z. Note that $|1-\zeta| \geq 1 - |\zeta|$ for any complex number ζ with $|\zeta| \leq 1$. Note also that $P(z^{-1};\boldsymbol{\alpha})$ admits the factorization (5.8). Since $|\nu_k z| \leq (1-\delta)/\rho < 1$ for any $|z| = \rho^{-1}$, it follows that $|1 - \nu_k z| \geq \varrho_0 := 1 - (1-\delta)/\rho > 0$ for all k. Therefore, from (5.8), we obtain $|P(z;\boldsymbol{\alpha})| \geq \varrho_0^{2q}$ for any $|z| = \rho^{-1}$, and hence

$$\max_{|z|=\rho^{-1}} |H(z;\boldsymbol{\alpha})| \leq \varrho_0^{-2q} \tag{5.10}$$

for any $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. On the other hand, let $\{h_j(\boldsymbol{\alpha})\}$ be the impulse response of the AR filter (5.1). Then, we obtain (Markushevich, 1977, vol. 1, Theorem 16.4, p. 349)

$$H(z;\boldsymbol{\alpha}) = \sum_{j=0}^{\infty} h_j(\boldsymbol{\alpha})\, z^j \quad \text{and} \quad h_j(\boldsymbol{\alpha}) = H_j(0;\boldsymbol{\alpha})/j! \qquad (j = 0, 1, \ldots). \tag{5.11}$$

This result, together with (5.9) and (5.10), implies that $|h_j(\boldsymbol{\alpha})| \leq \varrho_0^{-2q}\rho^j$ for $j \geq 0$ and $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. Hence, the condition (A3) is satisfied by the AR filter.

Furthermore, let $P_j(z;\boldsymbol{\alpha})$ be the jth derivative of $P(z;\boldsymbol{\alpha})$ with respect to z. Then, it is readily shown that $H_1(z;\boldsymbol{\alpha}) = -\{H(z;\boldsymbol{\alpha})\}^2 P_1(z;\boldsymbol{\alpha})$. Using the Leibniz rule of differentiation, we obtain, for $j = 1, 2, \ldots$,

$$H_j(z;\boldsymbol{\alpha}) = -\sum_{u=0}^{j-1} C_{j-1}^u P_{j-u}(z;\boldsymbol{\alpha}) \sum_{v=0}^{u} C_u^v H_v(z;\boldsymbol{\alpha})\, H_{u-v}(z;\boldsymbol{\alpha}).$$

This, together with (5.11), yields the following recursive expression for $h_j(\boldsymbol{\alpha})$, namely,

$$h_j(\boldsymbol{\alpha}) = -\sum_{u=0}^{j-1}(1 - u/j)\,\theta_{j-u}(\boldsymbol{\alpha})\,\eta^{j-u} \sum_{v=0}^{u} h_v(\boldsymbol{\alpha})\, h_{u-v}(\boldsymbol{\alpha}) \tag{5.12}$$

for $j = 1, 2, \ldots$, where $\theta_u(\boldsymbol{\alpha}) := 0$ for $u > 2q$. Note that $h_0(\boldsymbol{\alpha})$ is differentiable since $h_0(\boldsymbol{\alpha}) = 1$ for all $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. By induction on j, we can show from (5.12) that $h_j(\boldsymbol{\alpha})$ is also differentiable for all $j \geq 1$ due to the differentiability of $\{\theta_u(\boldsymbol{\alpha})\}$ when $\boldsymbol{\theta}(\boldsymbol{\alpha})$ is given by (5.2) or (5.6). On the other hand, let the prime $'$ denote the differentiation

with respective any component of $\boldsymbol{\alpha}$. Then we have $H'(z;\boldsymbol{\alpha}) = -\{H(z;\boldsymbol{\alpha})\}^2 P'(z;\boldsymbol{\alpha})$. Clearly, $H'(z;\boldsymbol{\alpha})$ is analytic in $|z| < (1-\delta)^{-1}$, just like $H(z;\boldsymbol{\alpha})$. Moreover, when $\boldsymbol{\theta}(\boldsymbol{\alpha})$ is given by (5.2) or (5.6), the derivatives of $\theta_j(\boldsymbol{\alpha})$ are constants, and hence $|P'(z;\boldsymbol{\alpha})|$ is bounded by a constant c for any $|z| = \rho^{-1}$ and $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. Therefore, according to (5.10), we obtain

$$\max_{|z|=\rho^{-1}} |H'(z;\boldsymbol{\alpha})| \leq c\varrho_0^{-4q}. \tag{5.13}$$

On the other hand, since $H'(z;\boldsymbol{\alpha})$ is analytic in $|z| < (1-\delta)^{-1}$, it admits the Taylor series expansion $H'(z;\boldsymbol{\alpha}) = \sum_{j=0}^{\infty} g_j(\boldsymbol{\alpha})\, z^j$ with $g_j(\boldsymbol{\alpha}) := H_j'(0;\boldsymbol{\alpha})/j!$ for $j \geq 0$, where $H_j'(z;\boldsymbol{\alpha})$ is the jth derivative of $H'(z;\boldsymbol{\alpha})$ with respect to z. Since the order of differentiation with respect to z and to $\boldsymbol{\alpha}$ is exchangeable in differentiating $H(z;\boldsymbol{\alpha})$, this result, in connection with (5.11), yields $h_j'(\boldsymbol{\alpha}) = g_j(\boldsymbol{\alpha})$ and hence

$$H'(z;\boldsymbol{\alpha}) = \sum_{j=0}^{\infty} h_j'(\boldsymbol{\alpha})\, z^j$$

for $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. Using (5.13) and Cauchy's inequality for $H_j'(z;\boldsymbol{\alpha})$, we obtain $|h_j'(\boldsymbol{\alpha})| \leq c\varrho_0^{-4q}\rho^j$ for all $j \geq 0$ and $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. Therefore, the AR filter satisfies the condition (A4). The theorem is thus proved. ◊

In order to effectively control the bandwidth of the AR filter by the parameter η, we may consider another parameter space $\mathcal{A}_0(\eta)$ defined by

$$\mathcal{A}_0(\eta) := \{\boldsymbol{\alpha} :\ \boldsymbol{\theta}(\boldsymbol{\alpha}) \in \Theta_0\} \tag{5.14}$$

where Θ_0 is a subset of $\Theta(\eta)$, consisting of $\boldsymbol{\theta} \in \Theta(\eta)$ for which the zeros of the corresponding polynomial $\sum_{j=0}^{2q} \theta_j z^{-j}$ occur *on the unit circle.* For this parameter space, it is readily shown that the poles of the AR filter are constrained to be on the circle $|z| = \eta$, so that the parameter η has a full control of the filter's bandwidth for *all* $\boldsymbol{\alpha} \in \mathcal{A}_0(\eta)$. Moreover, since $\mathcal{A}_0(\eta)$ is a subset of $\mathcal{A}(\eta)$, the AR filter still satisfies the regularity conditions (A3) and (A4) when $\eta < 1$.

In practice, $\boldsymbol{\theta}(\hat{\boldsymbol{\alpha}}_m)$ does not always appear inside Θ_0 when $\hat{\boldsymbol{\alpha}}_m$ in (4.8) is obtained from a finite data record. In case it falls outside, we may project it back into Θ_0 by the following procedure. Suppose that for a given $\boldsymbol{\alpha}$ the zeros of the polynomial $\sum_{j=0}^{2q} \theta_j(\boldsymbol{\alpha})\, z^{-j}$ are of the form $p_k \exp(\pm i\lambda_k)$, $(k = 1, \ldots, q)$, for some $\mathbf{p}_k > 0$ and $0 < \lambda_1 \leq \cdots \leq \lambda_q < \pi$. Then the projection of $\boldsymbol{\theta}(\boldsymbol{\alpha})$, denoted by $\tilde{\boldsymbol{\theta}} := [\tilde{\theta}_1, \cdots, \tilde{\theta}_q]^T$, can be obtained from the coefficients of the polynomial

$$\sum_{j=0}^{2q} \tilde{\theta}_j z^{-j} = \prod_{k=1}^{q} (1 - e^{i\lambda_k} z^{-1})(1 - e^{-i\lambda_k} z^{-1}).$$

By this projection, the poles of the AR filter are guaranteed to be restricted on the circle $|z| = \eta$ so that the AR filtering is always stable as long as $\eta < 1$.

5.2 Statistical Properties

For the properties in Section 4.2 to hold, the condition (A1) is the only thing left for verification. It is readily shown that the AR filter satisfies this condition, since its transfer function

$$H(\omega; \boldsymbol{\alpha}) = \left\{ \sum_{j=0}^{2q} \theta_j(\boldsymbol{\alpha})\, \eta^j e^{-ij\omega} \right\}^{-1}$$

is nonzero for all $\boldsymbol{\alpha} \in \mathcal{A}_0(\eta)$. In conclusion, the AR filter, defined by (5.1) and (5.6) with $\eta < 1$, satisfies (A1)–(A4). Consequently, the corresponding PF estimator $\hat{\boldsymbol{\alpha}}$, possesses the statistical properties in Section 4.2, regarding the existence, convergence, strong consistency, and asymptotic normality, as can be summarized as follows.

Theorem 5.2 *Suppose that $\{\epsilon_t\}$ is white with finite fourth moment. Then, for the AR filter defined by* (5.1) *and* (5.6) *with $\eta < 1$, the results in Theorem* 4.2, *Theorem* 4.3, *and Theorem* 4.4 *are valid, provided that $\boldsymbol{\alpha}^* \in \mathcal{A}_0(\eta)$.*

As mentioned before, the parameter η plays the role of controlling the bandwidth of the AR filter. Indeed, the closer η is to 1, the narrower is the bandwidth of the AR

filter. Since the sinusoidal signal under investigation concentrates only on extremely narrow bands (spikes) in the frequency domain, it is clear that in order to enhance the signal by the AR filter, η should be chosen as close to 1 as possible. From another point of view, the parameter η also determines the asymptotic behavior of the associated PF estimator in terms of its covariance matrix. As a matter of fact, the asymptotic covariance matrix $\mathbf{V}_\alpha(\eta)$ of the PF estimator can be made arbitrarily small if η is chosen arbitrarily close to 1. More precisely, we have the following theorem concerning the limit $\mathbf{V}_\alpha(\eta)$ as $\eta \to 1$.

Theorem 5.3 *Let $\mathbf{V}_\alpha(\eta)$ be the asymptotic covariance matrix of the PF estimator $\hat{\boldsymbol{\alpha}}$ corresponding to the AR filter defined by* (5.1) *and* (5.6) *with $\eta < 1$. Then $\mathbf{V}_\alpha(\eta) = O((1-\eta)^3)$ as $\eta \to 1$, where $\mathbf{V}_\alpha(\eta)$ is given by* (4.23).

PROOF. To prove this theorem, we first need to rewrite $w_{ij}(\boldsymbol{\alpha}^*)$ defined by (4.18) into a more suitable form with the help of the spectral representation

$$r_\tau^\epsilon(\boldsymbol{\alpha}^*) = \frac{\sigma_\epsilon^2}{2\pi} \int_{-\pi}^{\pi} |H(\omega; \boldsymbol{\alpha}^*)|^2 \, e^{i\tau\omega} \, d\omega. \tag{5.15}$$

To this end, let us define $d_j := \eta^j \theta_j(\boldsymbol{\alpha}^*)/a_j$. Then, from (5.6), we obtain

$$d_j = (\eta^j + \eta^{2q+j})/(\eta^j + \eta^{2q-j}) \qquad (j = 0, 1, \ldots, 2q).$$

With this notation, we can rewrite (5.3) as

$$\sum_{j=0}^{2q} d_j \, a_j \, r_{\tau-j}^\epsilon(\boldsymbol{\alpha}^*) = 0 \qquad (\tau = 1, 2, \ldots).$$

This, in connection with (5.15), gives

$$\begin{aligned} S_\tau \;&:=\; \sum_{j=0}^{2q} a_j \, r_{\tau-j}^\epsilon(\boldsymbol{\alpha}^*) \;=\; \sum_{j=0}^{2q} (1 - d_j) \, a_j \, r_{\tau-j}^\epsilon(\boldsymbol{\alpha}^*) \\ &=\; \frac{\sigma_\epsilon^2}{2\pi} \int_{-\pi}^{\pi} |H(\omega; \boldsymbol{\alpha}^*)|^2 \left\{ \sum_{j=0}^{2q} (1 - d_j) \, a_j \, e^{-ij\omega} \right\} e^{i\tau\omega} \, d\omega. \end{aligned}$$

Introducing the polynomials

$$D(z) := \sum_{j=0}^{2q}(1-d_j)\,a_j z^{2q-j} \quad \text{and} \quad P(z) := \sum_{j=0}^{2q}\theta_j(\boldsymbol{\alpha}^*)\,\eta^j z^j,$$

we can write S_τ as a Cauchy integral of the form

$$S_\tau = \frac{\sigma_\epsilon^2}{2\pi i}\oint_{|z|=1}\frac{D(z)}{P(z)\,Q(z)}\,z^{\tau-1}\,dz \qquad (\tau = 1,2,\ldots)$$

where $Q(z) := z^{2q}P(z^{-1})$. Note that $\boldsymbol{\alpha}^* \in \mathcal{A}_0(\eta)$ implies $\boldsymbol{\theta}(\boldsymbol{\alpha}^*) \in \Theta_0$. This, in turn, guarantees that the $2q$ zeros of $P(z^{-1})$ appear on the circle $|z| = \eta$ and can be expressed as $\nu_k := \eta\exp(i\lambda_k)$, $(k = 1,\ldots,2q)$, with the λ_k satisfying $0 < \lambda_1 \leq \cdots \leq \lambda_q < \pi$ and $\lambda_k := -\lambda_{2q-k+1}$, $(k = q+1,\ldots,2q)$. As a result, we can write

$$P(z) = \prod_{k=1}^{2q}(1-\bar{\nu}_k z) \quad \text{and} \quad Q(z) = \prod_{k=1}^{2q}(z-\nu_k)$$

where $\bar{\nu}_k$ is the complex conjugate of ν_k. It follows from the residue theorem of complex analysis (Markushevich, 1977) that

$$S_\tau = \sigma_\epsilon^2\sum_{k=1}^{2q}\frac{D(\nu_k)}{P(\nu_k)\,Q'(\nu_k)}\,\nu_k^{\tau-1} \qquad (\tau = 1,2,\ldots).$$

Using this formula, $w_{ij}(\boldsymbol{\alpha}^*)$ in (4.18) can be written as

$$\begin{aligned} w_{ij}(\boldsymbol{\alpha}^*) &= 4\sum_{\tau=0}^{\infty}S_{\tau+i}\bar{S}_{\tau+j} \\ &= 4\sigma_\epsilon^4\sum_{k,\,k'=1}^{2q}\frac{D(\nu_k)\,\overline{D(\nu_{k'})}}{P(\nu_k)\,\overline{P(\nu_{k'})}\,Q'(\nu_k)\,\overline{Q'(\nu_{k'})}}\,\frac{\nu_k^{i-1}\bar{\nu}_{k'}^{j-1}}{1-\nu_k\bar{\nu}_{k'}} \end{aligned} \tag{5.16}$$

where the overline denotes the complex conjugation.

With the expression in (5.16), the behavior of $\mathbf{W}(\boldsymbol{\alpha}^*)$ as η tends to 1 becomes easer to investigate. In fact, we first note that $\nu_k \to z_k$ as $\eta \to 1$, where the z_k are zeros of the AR polynomial $A(z^{-1}) = \sum_{j=0}^{2q}a_j z^{-j}$ in (2.1). Moreover, $d_0 = 1$ and $(1-\eta^2)^{-1}(1-d_j) \to j/2$ for $j = 1,\ldots,2q$ as $\eta \to 1$. Therefore, we obtain

$$(1-\eta^2)^{-1}\,D(\nu_k) \to \sum_{j=1}^{2q}\tfrac{1}{2}j a_j\,z_k^{2q-j} = \tfrac{1}{2}\,z_k^{2q-1}A'(z_k^{-1}). \tag{5.17}$$

Since $A(z) = \prod(1 - z_k z)$ and $z_k^{-1} = \bar{z}_k$, it is readily shown that

$$(1-\eta^2)^{-1} P(\nu_k) = \prod_{\substack{k'=1 \\ k' \neq k}}^{2q} (1 - \bar{\nu}_{k'} \nu_k) \to \prod_{\substack{k'=1 \\ k' \neq k}}^{2q} (1 - \bar{z}_{k'} z_k) = -z_k A'(z_k)$$

and

$$Q'(\nu_k) = \prod_{\substack{k'=1 \\ k' \neq k}}^{2q} (\nu_k - \nu_{k'}) \to \prod_{\substack{k'=1 \\ k' \neq k}}^{2q} (z_k - z_{k'}) = z_k^{2q-1} \prod_{\substack{k'=1 \\ k' \neq k}}^{2q} (1 - z_{k'} \bar{z}_k) = -z_k^{2q} \overline{A'(z_k)}.$$

Finally, $(1-\eta^2)/(1 - \nu_k \bar{\nu}_{k'}) = 1$ for $k' = k$, and $(1-\eta^2)/(1 - \nu_k \bar{\nu}_{k'}) \to 0$ as $\eta \to 1$ for $k' \neq k$. Substituting all these limits in (5.16) yields

$$(1-\eta^2)\, w_{ij}(\boldsymbol{\alpha}^*) \to \sigma_\epsilon^4 \sum_{k=1}^{2q} \frac{1}{|A'(z_k)|^2} z_k^{i-1} \bar{z}_k^{j-1} \qquad (i, j = 1, \ldots, 2q-1)$$

as $\eta \to 1$. In matrix notation, we can write

$$(1-\eta^2)\, \mathbf{Q}^T \mathbf{W}(\boldsymbol{\alpha}^*)\, \mathbf{Q} \to 4\sigma_\epsilon^4\, \boldsymbol{\Sigma}_0 \quad \text{with} \quad \boldsymbol{\Sigma}_0 := \mathbf{Q}^T \mathbf{S} \mathbf{D}_0 \mathbf{S}^H \mathbf{Q} \tag{5.18}$$

where $\mathbf{S}$ is the Vandermonde matrix in (2.21) and $\mathbf{D}_0$ is the $2q$-by-$2q$ diagonal matrix defined by $\mathbf{D}_0 := \frac{1}{4} \operatorname{diag}\{|A'(z_1)|^{-2}, \ldots, |A'(z_{2q})|^{-2}\}$.

To complete the proof, it suffices to consider the behavior of $\mathbf{R}_x(\boldsymbol{\alpha}^*)$ as η tends to 1. For this purpose, we notice that $|H(\omega_k; \boldsymbol{\alpha}^*)| = 1/|P(z_k^{-1})|$ and

$$P(z_k^{-1}) = \sum_{j=0}^{2q} \theta_j(\boldsymbol{\alpha}^*)\, \eta^j z_k^{-j} = \sum_{j=0}^{2q} d_j a_j z_k^{-j} = -z_k^{-2q} D(z_k).$$

Therefore, it follows from (5.17) that $(1-\eta^2)^2\, |H(\omega_k; \boldsymbol{\alpha}^*)|^2 \to 4/|A'(z_k)|^2$. This, together with the decomposition of $\mathbf{R}_x(\boldsymbol{\alpha}^*)$ in Remark 2.1, implies that

$$(1-\eta^2)^2\, \mathbf{R}_x(\boldsymbol{\alpha}^*) \to 8\, r_0^x\, \boldsymbol{\Sigma} \quad \text{with} \quad \boldsymbol{\Sigma} := \mathbf{Q}^T \mathbf{S} \mathbf{D} \mathbf{S}^H \mathbf{Q} \tag{5.19}$$

where $\mathbf{D} := \frac{1}{4} \operatorname{diag}\{\sigma_1^2/|A'(z_1)|^2, \ldots, \sigma_{2q}^2/|A'(z_{2q})|^2\}$ with $\sigma_k^2 := \beta_k^2 / \sum_{j=1}^q \beta_j^2$ for $k = 1, \ldots, q$ and $\sigma_k^2 := \sigma_{2q-k+1}^2$ for $k = q+1, \ldots, 2q$. Collecting (5.18) and (5.19), we finally obtain

$$\lim_{\eta \to 1} \left(\frac{1+\eta^2}{1-\eta^2} \right)^3 \mathbf{V}_\alpha(\eta) = \frac{1}{2\gamma^2}\, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{-1} \tag{5.20}$$

where $\mathbf{V}_\alpha(\eta)$ is the covariance matrix given by (4.23) and $\gamma := r_0^x/r_0^\epsilon = r_0^x/\sigma_\epsilon^2$ is the signal-to-noise ratio of the data. The theorem is thus proved. $\Diamond$

Remark 5.1 In the special case where the q sinusoids have the same power, that is, $\beta_k \equiv \beta_1$ for all k, the limit in (5.20) reduces to

$$\lim_{\eta \to 1} \left(\frac{1+\eta^2}{1-\eta^2} \right)^3 \mathbf{V}_\alpha(\eta) = \frac{q^2}{2\gamma^2} \boldsymbol{\Sigma}_0^{-1} \tag{5.21}$$

because $\mathbf{D} = q^{-1}\mathbf{D}_0$ in the case.

Remark 5.2 It is interesting to recognize, by comparing (5.18) with the decomposition of $\mathbf{R}_x$ in Remark 2.1, that $\boldsymbol{\Sigma}_0$ is identical to the autocovariance matrix of a sinusoidal signal with the same q frequencies as $\{x_t\}$, and with the amplitude corresponding to ω_k equal to $1/|A'(z_k)|$. Similarly, $\boldsymbol{\Sigma}$ in (5.19) has the same structure as $\boldsymbol{\Sigma}_0$, except that the amplitude associated with ω_k is replaced by $\sigma_k/|A'(z_k)|$. Therefore, these matrices can be rewritten as

$$\boldsymbol{\Sigma}_0 = \left[\sum_{k=1}^{q} \frac{\cos(\omega_k(i-j))}{2\,|A'(z_k)|^2} \right] \quad \text{and} \quad \boldsymbol{\Sigma} = \left[\sum_{k=1}^{q} \frac{\sigma_k^2 \cos(\omega_k(i-j))}{2\,|A'(z_k)|^2} \right]$$

where $i, j = 1, \ldots, 2q-1$.

5.3 Accuracy of the PF Estimator

As guaranteed by Theorem 4.4, the PF method produces an estimator $\hat{\boldsymbol{\alpha}}$ so that $\sqrt{n}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ is asymptotically normally distributed and thus its estimation accuracy is of order $O(n^{-1/2})$. However, for the AR filter (5.1), it has been shown that the asymptotic covariance matrix of the corresponding PF estimator can be made arbitrarily small as η tends to 1. This indicates that a higher order accuracy could be obtained with $\eta \approx 1$. Indeed, as discussed in Truong-Van (1990), Quinn and Fernandes (1991), and Li, Kedem, and Yakowitz (1992) for the case of a single sinusoid ($q = 1$), the PF estimator with the AR filter is able to achieve the same accuracy of order $O(n^{-3/2})$ as the nonlinear least squares (NLS) approach (Hannan, 1973; Stoica and

Nehorai, 1989; Walker, 1971) in the limiting case of $\eta = 1$. We shall return to this point in the next section where the special case of $q = 1$ is considered in detail.

For the case of multiple sinusoids, the fixed-point iteration in (4.8) with $\eta = 1$ can be regarded as an algorithm that approximately calculates the NLS estimator in an iterative fashion. In fact, as we have seen in Chapter 1, the NLS estimator minimizes J_n' in (1.13). Notice that $\mathbf{I} - \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$ in (1.13) is a projection operator that projects an n-vector onto the orthogonal complement of the $2q$-dimensional column-space of $\mathbf{G}$. On the other hand, if we let $\hat{a}_j$ be the AR coefficients determined by (2.1) for any given $\omega_k := \hat{\omega}_k$, and denote by $\hat{\mathbf{A}}$ the corresponding n-by-$(n-2q)$ matrix of the structure (4.22), it is easy to verify that $\hat{\mathbf{A}}^T\mathbf{G} = \mathbf{0}$. This implies that the $n - 2q$ linearly-independent columns of $\hat{\mathbf{A}}$ are orthogonal to the columns of $\mathbf{G}$ and thus span the $(n-2q)$-dimensional orthogonal complement of the column-space of $\mathbf{G}$. As a consequence, we obtain

$$\mathbf{I} - \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T = \hat{\mathbf{A}}(\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^T$$

and hence $J_n' = \tilde{\mathbf{y}}^T\hat{\mathbf{A}}(\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^T\tilde{\mathbf{y}}$. Let $\mathbf{e}(\hat{\mathbf{a}}) := [e_{2q+1}(\hat{\mathbf{a}}), \cdots, e_n(\hat{\mathbf{a}})]^T$ where

$$e_t(\hat{\mathbf{a}}) := \sum_{j=0}^{2q} \hat{a}_j\, y_{t-j}. \tag{5.22}$$

Simple algebra shows that $\hat{\mathbf{A}}^T\tilde{\mathbf{y}} = \mathbf{e}(\hat{\mathbf{a}})$ where $\tilde{\mathbf{y}} = [y_1, \cdots, y_n]^T$. Therefore, J_n' can be rewritten as

$$J_n'(\hat{\mathbf{a}}) = \mathbf{e}^T(\hat{\mathbf{a}})(\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\mathbf{e}(\hat{\mathbf{a}}). \tag{5.23}$$

The NLS method is thus reduced to the problem of minimizing $J_n'(\hat{\mathbf{a}})$ in (5.23) with respect to $\hat{\mathbf{a}}$. To compute the NLS estimator, an iterative procedure can be employed in accordance with the suggestions of Bresler and Macovski (1986). Indeed, for any estimate $\hat{\boldsymbol{\alpha}}_m$ of the AR parameter $\mathbf{a}$, a new estimate can be obtained by minimizing the criterion $J_n^{(m)}(\hat{\mathbf{a}}) := \mathbf{e}^T(\hat{\mathbf{a}})(\hat{\mathbf{A}}_m^T\hat{\mathbf{A}}_m)^{-1}\mathbf{e}(\hat{\mathbf{a}})$ with respect to $\hat{\mathbf{a}}$, where $\hat{\mathbf{A}}_m$ is the matrix $\hat{\mathbf{A}}$ with $\hat{\boldsymbol{\alpha}}_m$ in place of $\hat{\mathbf{a}}$.

Let $\hat{\mathbf{a}}$ be within a tiny neighborhood of the true AR parameter $\mathbf{a}$. It is easy to see that within such a neighborhood $e_t(\hat{\mathbf{a}})$ can be approximated by

$$e_t(\hat{\mathbf{a}}) = \sum_{j=0}^{2q} \hat{a}_j \, x_{t-j} + \sum_{j=0}^{2q} \hat{a}_j \, \epsilon_{t-j} \approx \sum_{j=0}^{2q} \hat{a}_j \, \epsilon_{t-j}.$$

Since $\{\epsilon_t\}$ is white, it can be shown from this approximation that the covariance matrix of $\mathbf{e}(\hat{\mathbf{a}})$ is approximately equal to $\sigma_\epsilon^2 \hat{\mathbf{A}}^T \hat{\mathbf{A}}$. Therefore, when $\hat{\boldsymbol{\alpha}}_m$ and $\hat{\mathbf{a}}$ are both very close to $\mathbf{a}$, the vector $\tilde{\mathbf{e}}_m(\hat{\mathbf{a}}) := (\hat{\mathbf{A}}_m^T \hat{\mathbf{A}}_m)^{-1/2} \mathbf{e}(\hat{\mathbf{a}})$ can be regarded as a whitened version of $\mathbf{e}(\hat{\mathbf{a}})$. Resorting to the frequency-domain interpretation, this whitening procedure can be approximately performed by applying to $\tilde{e}_t := e_t(\hat{\mathbf{a}})$ an AR filter of the form (5.1) with $\eta \approx 1$ and $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}_m$. Let $\{\tilde{e}_t(\hat{\boldsymbol{\alpha}}_m)\}$ be the output of the filter, then $\tilde{\mathbf{e}}_m(\hat{\mathbf{a}}) \approx [\tilde{e}_{2q+1}(\hat{\boldsymbol{\alpha}}_m), \cdots, \tilde{e}_n(\hat{\boldsymbol{\alpha}}_m)]^T$ and hence $J_n^{(m)}(\hat{\mathbf{a}}) = \|\tilde{\mathbf{e}}_m(\hat{\mathbf{a}})\|^2 \approx \sum \{\tilde{e}_t(\hat{\boldsymbol{\alpha}}_m)\}^2$. On the other hand, by interchanging the order of the AR filtering on $e_t(\hat{\mathbf{a}})$ and the FIR filtering on y_t defined by (5.22), we obtain $\tilde{e}_t(\hat{\boldsymbol{\alpha}}_m) \approx \sum_{j=0}^{2q} \hat{a}_j \, \hat{y}_{t-j}(\hat{\boldsymbol{\alpha}}_m)$. Thus, minimizing $J_n^{(m)}(\hat{\mathbf{a}})$ is approximately equivalent to minimizing $\sum \{\sum_{j=0}^{2q} \hat{a}_j \, \hat{y}_{t-j}(\hat{\boldsymbol{\alpha}}_m)\}^2$. The latter yields the PF estimator $\hat{\boldsymbol{\alpha}}_{m+1}$ given by (4.8).

An advantage of the PF method over the NLS method lies in its computational simplicity inherited from the explicit least squares solution of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$. In fact, a direct calculation of $J_n'(\hat{\mathbf{a}})$ in (5.23) requires the inversion of an $(n-2q)$-by-$(n-2q)$ matrix $\hat{\mathbf{A}}^T \hat{\mathbf{A}}$, as compared to the inversion of a q-by-q matrix $\mathbf{Q}^T \mathbf{Y}^T(\boldsymbol{\alpha}) \, \mathbf{Y}(\boldsymbol{\alpha}) \, \mathbf{Q}$ plus a linear recursive filtering for the computation of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in (4.10). The computational complexity is clearly $O(n^3)$ versus $O(n)$.

Another advantage of the PF method comes from its less stringent requirement of the initial estimates. As pointed out by many researchers, the NLS approach and periodogram analysis (Rice and Rosenblatt, 1988; Stoica, Moses, Friedlander, and Söderström, 1989; see also Chapter 1), as well as the PF estimator in the limiting case of $\eta = 1$ (Truong-Van, 1990; Quinn and Fernandes, 1991; see also the next section), require an initial estimate of accuracy $o(n^{-1})$ in order to obtain the optimal

solution by iterative procedures. On the other hand, with $\eta < 1$, an initial estimate of accuracy $O(1)$ is sufficient for the fixed-point iteration in (4.8) to converge to the corresponding PF estimator, as indicated by Theorem 4.2. This, however, is not the end of the story. In fact, thanks to the flexibility in the choice of η, the estimation accuracy of the PF estimator can be further improved upon increasing the value of η toward 1 as more reliable estimates become available to initiate the iteration. For this purpose, we consider an increasing sequence of η such that $0 < \eta_1 < \eta_2 < \cdots \rightarrow 1$ (or $0 < \eta_1 < \eta_2 < \cdots < \eta_K = 1$). To obtain the PF estimator $\hat{\boldsymbol{\alpha}}(\eta_k)$, the fixed-point iteration (4.8) is initiated not with an arbitrary $\hat{\boldsymbol{\alpha}}_0$ but with the previously obtained PF estimator $\hat{\boldsymbol{\alpha}}(\eta_{k-1})$. In so doing for $k = 1, 2, \ldots$, a sequence of PF estimators $\{\hat{\boldsymbol{\alpha}}(\eta_k)\}$ is produced — each estimator in the sequence serving as the initial value of its successor. As k grows, the accuracy of $\hat{\boldsymbol{\alpha}}(\eta_k)$ approaches that of the NLS estimator, but the whole procedure, with η_1 taking on a relatively small value, is able to accommodate poor initial guesses employed to yield $\hat{\boldsymbol{\alpha}}(\eta_1)$. Another possible way of improving the estimation accuracy is to increase η after *each* iteration instead of carrying on the iteration until convergence (Dragošević and Stanković, 1989; Kedem and Yakowitz, 1991). This strategy simplifies the computation to some extent but may result in converging to a false location if η grows too fast.

In the most interesting case where the frequencies are closely spaced, however, one should be very cautious when increasing η. Simulations show that the bias of the PF estimator based on a finite data record increases as η approaches 1, while the variance decreases. For closely-spaced frequencies, the bias eventually dominates the variance, and hence an appropriate value of $\eta < 1$ should be considered to balance the trade-off between the bias and variance in minimizing the mean-squared error.

5.4 Special Cases of One and Two Sinusoids

In the preceding sections, we have discussed the general properties of the PF method with the AR filter. Now let us concentrate on two special cases where the signal consists of a single sinusoid and two sinusoids, respectively. For these examples, we shall provide detailed analysis in order to gain some further insight into the PF method regarding its parameter space and accuracy.

5.4.1 A Single Sinusoid in White Noise

In the case of $q = 1$, the AR filter (5.1) becomes a second-order filter of the form

$$y_t(\alpha) + \theta(\alpha)\,\eta y_{t-1}(\alpha) + \eta^2 y_{t-2}(\alpha) = y_t \tag{5.24}$$

and the parametrization (5.6) reduces to the simple equation

$$\theta(\alpha) = \frac{1+\eta^2}{2\eta}\,\alpha. \tag{5.25}$$

Since the zeros of the polynomial $1 + \theta z^{-1} + z^{-2}$ are $\frac{1}{2}(-\theta \pm \sqrt{\theta^2 - 4}\,)$, it is necessary that $|\theta| < 2$ in order that the zeros are pure complex. In this case, the magnitude of the conjugate pair of zeros is equal to 1, so that $\Theta_0 = \{\theta : \ |\theta| < 2\}$ and

$$\mathcal{A}_0(\eta) = \{\alpha : \ |\alpha| < 4\eta/(1+\eta^2)\}.$$

Moreover, since $a_1 = -2\cos\omega_1$, the results in Theorem 5.2 holds as long as η is chosen such that $\eta < 1$ and

$$-2\eta/(1+\eta^2) < \cos\omega_1 < 2\eta/(1+\eta^2).$$

Obviously, this can be achieved with a sufficiently large η since $|\cos\omega_1| < 1$ for $\omega_1 \in (0,\pi)$. Given $\alpha \in \mathcal{A}_0(\eta)$, since $\theta(\alpha) \in \Theta_0$, we can write $\theta(\alpha) = -2\cos\omega(\alpha)$ for some $\omega(\alpha) \in (0,\pi)$. The poles of the AR filter are therefore $\nu = \eta\exp(i\,\omega(\alpha))$ and

$\bar{\nu} = \eta \exp(-i\,\omega(\alpha))$. In the cases where $\{\epsilon_t\}$ is white noise, a formula given by He and Kedem (1990, Lemma 2.1, Case iii) leads to

$$\rho_1^\epsilon(\alpha) = \frac{2\,\eta}{1+\eta^2}\cos\omega(\alpha) = -\frac{\eta}{1+\eta^2}\,\theta(\alpha)$$

where $\rho_1^\epsilon(\alpha)$ is the first-order autocorrelation of the filtered noise $\{\epsilon_t(\alpha)\}$. On the other hand, a simple calculation shows that $-\mathbf{R}_\epsilon^{-1}(\alpha)\,\mathbf{r}_\epsilon(\alpha) = -2\rho_1^\epsilon(\alpha)$. Therefore, with $\theta(\alpha)$ given by (5.25), we obtain $-\mathbf{R}_\epsilon^{-1}(\alpha)\,\mathbf{r}_\epsilon(\alpha) = \alpha$ for all $\alpha \in \mathcal{A}_0(\eta)$. This verifies again that the AR filter possesses the parametrization property (A2) if it is parametrized according to (5.25).

To compute the asymptotic variance of the PF estimator $\hat{\alpha}$, we note that with $q = 1$ the matrix $\mathbf{V}_\alpha(\eta)$ in (4.23) reduces to the simple form

$$\sigma_\alpha^2(\eta) = \frac{w_{11}(\alpha^*)}{(r_0^x(\alpha^*))^2}$$

where $\alpha^* := a_1 = -2\cos\omega_1$. Notice that we drop the argument η in $w_{11}(\alpha^*)$ and $r_0^x(\alpha^*)$ for notational brevity. The following corollary claims that the expression (5.21) holds for any η without taking the limit.

Corollary 5.1 *Let $\sigma_\alpha^2(\eta)$ and $\sigma_\omega^2(\eta)$ be the asymptotic variances of $\hat{\alpha}$ and the corresponding $\hat{\omega}_1$ using the AR filter* (5.1). *Then, for any $\eta < 1$,*

$$\sigma_\alpha^2(\eta) = \frac{4}{\gamma^2}\left(\frac{1-\eta^2}{1+\eta^2}\right)^3 \sin^2\omega_1 \quad \textit{and} \quad \sigma_\omega^2(\eta) = \frac{1}{\gamma^2}\left(\frac{1-\eta^2}{1+\eta^2}\right)^3 \tag{5.26}$$

where $\gamma := \frac{1}{2}\beta_1^2/\sigma_\epsilon^2$ is the signal-to-noise ratio of $\{y_t\}$.

PROOF. For $q = 1$, we can write $w_{11}(\alpha^*)$ in (5.16) as

$$w_{11}(\alpha^*) = 4\sigma_\epsilon^4\left(\frac{2|D(\nu)|^2}{|P(\nu)|^2|Q'(\nu)|^2(1-|\nu|^2)} + \Re\left\{\frac{2(D(\nu))^2}{(P(\nu))^2(Q'(\nu))^2(1-\nu^2)}\right\}\right)$$

where $\nu := \eta\exp(i\omega)$ with $\omega := \omega(\alpha^*)$. Since $\theta(\alpha^*) = -2\cos\omega$ and $\alpha^* = a_1$, it follows from (5.25) that

$$\cos\omega = -\frac{1+\eta^2}{4\eta}\,a_1 = \frac{1+\eta^2}{2\eta}\cos\omega_1. \tag{5.27}$$

Moreover, by definition, $D(z) = (1 + \frac{1}{2}a_1 z)(1 - \eta^2)$. Therefore, we have

$$\begin{aligned} D(\nu) &= (1 + \tfrac{1}{2}a_1\nu)(1 - \eta^2) \\ &= (1 - \eta^2(1+\eta^2)^{-1}(e^{i\omega} + e^{-i\omega})\, e^{i\omega})(1 - \eta^2) \\ &= (1 - \eta^2 e^{i2\omega})(1+\eta^2)^{-1}(1-\eta^2) \\ &= (1 - \nu^2)(1+\eta^2)^{-1}(1-\eta^2) \end{aligned}$$

where the second equality is due to the relation between $2\cos\omega = \exp(i\omega) + \exp(-i\omega)$ and a_1 given by (5.27). Since $P(z) = (1 - \nu z)(1 - \bar{\nu} z)$, it is readily shown that

$$P(\nu) = (1 - \nu^2)(1 - \eta^2).$$

In addition, since $Q'(z) = (z - \nu) + (z - \bar{\nu})$, a simple calculation gives

$$(Q'(\nu))^2 = (\nu - \bar{\nu})^2 = -4\eta^2 \sin^2\omega$$

and hence $|Q'(\nu)|^2 = 4\eta^2 \sin^2\omega$. Combining all these results, we obtain

$$w_{11}(\alpha^*) = \frac{4\sigma_\epsilon^4}{2\eta^2(1+\eta^2)^2(1-\eta^2)\sin^2\omega}\left(1 - \Re\left\{\frac{1-\eta^2}{1-\nu^2}\right\}\right). \tag{5.28}$$

To proceed with the calculation, we note that

$$\begin{aligned} |1 - \nu^2|^2 &= 1 - 2\eta^2\cos(2\omega) + \eta^4 \\ &= (1+\eta^2)^2 - 4\eta^2\cos^2\omega \\ &= (1+\eta^2)^2 \sin^2\omega_1 \end{aligned}$$

where the last equality is obtained according to (5.27). Similarly, we have

$$\Re\{1 - \bar{\nu}^2\} = 1 - \eta^2\cos(2\omega) = (1+\eta^2)(1 - \tfrac{1}{2}(1+\eta^2)\cos^2\omega_1).$$

Using these expressions, we can write

$$\begin{aligned} 1 - \Re\left\{\frac{1-\eta^2}{1-\nu^2}\right\} &= 1 - \frac{1-\eta^2}{|1-\nu^2|^2}\Re\{1-\bar{\nu}^2\} \\ &= \frac{2\eta^2(1 - (4\eta^2)^{-1}(1+\eta^2)^2\cos^2\omega_1)}{(1+\eta^2)\sin^2\omega_1} \\ &= \frac{2\eta^2\sin^2\omega_1}{(1+\eta^2)\sin^2\omega_1}. \end{aligned}$$

Substituting this result in (5.28) yields

$$w_{11}(\alpha^*) = \frac{4\sigma_\epsilon^4}{(1+\eta^2)^3(1-\eta^2)\sin^2\omega_1}.$$

Finally, with the aid of (5.25), a straightforward calculation shows that

$$|H(\omega_1;\alpha^*)|^2 = \frac{1}{|1+\theta(\alpha^*)\,\eta e^{-i\omega_1}+\eta^2 e^{-i2\omega_1}|^2} = \frac{1}{(1-\eta^2)^2\sin^2\omega_1}.$$

Since $r_0^x(\alpha^*) = \frac{1}{2}\beta_1^2|H(\omega_1;\alpha^*)|^2$, the expression of $\sigma_\alpha^2(\eta)$ in (5.26) follows immediately upon noting that $\sigma_\alpha^2(\eta) = w_{11}(\alpha^*)/\{r_0^x(\alpha^*)\}^2$. Moreover, using Slutsky's theorem (Brockwell and Davis, 1987, Proposition 6.4.1), we obtain $\sigma_\omega^2(\eta) = \frac{1}{4}\sigma_\alpha^2(\eta)/\sin^2\omega_1$ because $\omega_1 = \arccos(-\frac{1}{2}a_1)$ and hence $\partial\omega_1/\partial a_1 = \frac{1}{2}(1-\frac{1}{4}a_1^2)^{-1/2} = (2\sin\omega_1)^{-1}$. ◊

Remark 5.3 For $q = 1$, the matrix Σ_0 in (5.21) reduces to $\Sigma_0 = \frac{1}{2}|A'(z_1)|^{-2}$. Since $A(z) = (1-z_1z)(1-\bar{z}_1z)$, a simple calculation shows that $|A'(z_1)|^2 = |z_1-\bar{z}_1|^2 = 4\sin^2\omega_1$. Therefore, $\Sigma_0 = (8\sin^2\omega_1)^{-1}$. This, in connection with (5.26), verifies our claim that (5.21) holds for $q = 1$ *without* taking the limit.

To end our discussion on the single sinusoid case, let us consider the limiting situation when $\eta = 1$. As aforementioned, the PF estimator with $q = 1$ and $\eta = 1$ coincides with the procedure discussed by Truong-Van (1990) and by Quinn and Fernandes (1991). The following theorem summarizes their results.[3]

Theorem 5.4 *Let $q = 1$ and $\eta = 1$. Then, for large n, the PF estimator $\hat{\alpha}$ exists almost surely as the unique fixed-point of $\hat{a}(\alpha)$ in $S(\alpha^*) := \{\alpha : |\alpha-\alpha^*| \le cn^{-\delta}\}$ for some fixed $1 < \delta < 3/2$ and $c > 0$. The iteration in (4.8) converges to $\hat{\alpha}$ almost surely for large n if $\hat{\alpha}_0 \in S(\hat{\alpha}) := \{\alpha : |\alpha-\hat{\alpha}| \le c_1n^{-\delta}\}$ for some $c_1 > 0$. Furthermore, as $n \to \infty$, $n^\delta(\hat{\omega}_1-\omega_1) \xrightarrow{a.s.} 0$ for any $\delta < 3/2$ and $n^{3/2}(\hat{\omega}_1-\omega_1) \xrightarrow{D} \mathrm{N}(0, 12/\gamma)$.*

PROOF. We briefly outline the proof of existence and convergence. The remaining proof can be found in Truong-Van (1990), or in Quinn and Fernandes (1991). For the

[3]In Quinn and Fernandes (1991), convergence was proved only for the alternative procedure of the form $\hat{\alpha}_m = 2\hat{a}(\hat{\alpha}_{m-1}) - \hat{\alpha}_{m-1}$.

existence of $\hat{\alpha}$, let us consider $S_0(\alpha^*) := \{\alpha : \ |\alpha - \alpha^*| \leq c_0 n^{-\delta}\}$ for some $c_0 > 0$. It can be shown that (Quinn and Fernandes, 1991) that

$$\hat{a}(\alpha_1) - \hat{a}(\alpha_2) = (\alpha_1 - \alpha_2)\left\{\tfrac{1}{2} + O_{a.s.}\left(\sqrt{\log n/n}\right)\right\}$$

uniformly in α_1, $\alpha_2 \in S_0(\alpha^*)$ when n is sufficiently large. Moreover, it can be shown that (Quinn and Fernandes, 1991)

$$\hat{a}(\alpha^*) - \alpha^* = O_{a.s.}\left(n^{-1}\sqrt{\log n/n}\right)$$

for large n. Therefore, with $1/2 < \varsigma < 1$ and $c := c_0/2$, we obtain

$$|\hat{a}(\alpha_1) - \hat{a}(\alpha_2)| \leq \varsigma|\alpha_1 - \alpha_2| \quad \text{and} \quad |\hat{a}(\alpha^*) - \alpha^*| \leq (1-\varsigma)\, c n^{-\delta}$$

almost surely and uniformly in α_1, $\alpha_2 \in S(\alpha^*)$. This implies that $\hat{a}(\alpha)$ is a contractive mapping in $S(\alpha^*)$ and hence the unique fixed-point $\hat{\alpha}$ exists in $S(\alpha^*)$ almost surely, provided that n is sufficiently large. Following the proof of Theorem 4.2, we can show that the iteration (4.8) converges to $\hat{\alpha}$ with any initial guess taken in $S(\hat{\alpha})$ if c_1 is chosen such that $0 < c_1 < c_0/2$. $\Diamond$

Remark 5.4 Theorem 5.4 remains valid if the noise is colored. In this case, we would have $n^{3/2}(\hat{\omega}_1 - \omega_1) \xrightarrow{D} \mathrm{N}(0, 12/\gamma_1)$ where γ_1 is the signal-to-noise ratio given in Theorem 1.1. This implies that the PF method with $\eta = 1$ achieves the same estimation accuracy of periodogram analysis and nonlinear least squares. As we can see, the asymptotic variance depends solely on the noise spectrum at the frequency of the sinusoid.

Remark 5.5 Theorem 5.4 tells us that only if the initial guess falls within a distance of $o(n^{-1})$ to $\hat{\alpha}$, and hence to the true AR parameter α^*, is the iteration (4.8) guaranteed to converge to the fixed-point $\hat{\alpha}$. When the initial accuracy is poorer than $o(n^{-1})$, the fixed-point iteration may converge to a false location, if it converges at all.

From these remarks, we observe again the phenomenon of high estimation accuracy $O(n^{-3/2})$ versus stringent initial requirement $o(n^{-1})$, just like what we saw earlier in

Chapter 1 when we discussed the properties of periodogram analysis. This connection is by no mean a coincidence. As a matter of fact, the PF method using the AR filter with $\eta = 1$ has a close relation with periodogram analysis. Indeed, when $\eta = 1$, the parametrization (5.25) reduces to the trivial equation $\theta(\alpha) = \alpha$ and the parameter space to $\mathcal{A}_0(1) = \Theta_0$. Let $\alpha = \theta(\alpha) = -2\cos\omega$. This equation defines a one-to-one correspondence between $\alpha \in \mathcal{A}_0(1)$ and $\omega \in (0, \pi)$, and the AR filter (5.24) becomes

$$y_t(\alpha) - 2\cos\omega\, y_{t-1}(\alpha) + y_{t-2}(\alpha) = y_t.$$

Assuming zero initial values, the filtered data $\hat{y}_1(\alpha), \ldots, \hat{y}_n(\alpha)$ can be explicitly written as (Truong-Van, 1990)

$$\hat{y}_t(\alpha) = (\sin\omega)^{-1} \sum_{j=0}^{t-1} \sin((j+1)\omega)\, y_{t-j} \qquad (t = 1, \ldots, n). \tag{5.29}$$

Recall (see Chapter 4) that the least squares mapping $\hat{a}(\alpha)$ admits the representation (4.12). Therefore, if we denote by $\hat{\rho}(\vartheta)$ the first-order sample autocorrelation of the filtered data $\{\hat{y}_t(\alpha)\}$, namely,

$$\hat{\rho}(\vartheta) = \frac{1}{2} \sum_{t=3}^{n} \hat{y}_{t-1}(\alpha)\, \{\hat{y}_t(\alpha) + \hat{y}_{t-2}(\alpha)\} \Big/ \sum_{t=3}^{n} \hat{y}_{t-1}^2(\alpha)$$

where $\vartheta := \alpha/2 = \cos\omega$, then the PF estimator can be obtained from the fixed-point $\hat{\vartheta}$ of the mapping $\hat{\rho}(\vartheta)$ according to $\hat{\alpha} = -2\hat{\vartheta}$ and $\hat{\omega}_1 = \arccos\hat{\vartheta}$.

From (5.29), we clearly observe the resemblance between the output sequence $\{\hat{y}_t(\alpha)\}$ of the AR filter and the output sequence $\{y_t(\omega)\}$ of the complex exponential filter in (1.8). In fact, it is easy to verify that

$$\hat{y}_t(\alpha) = (\sin\omega)^{-1} \Im\{y_t(\omega)\} \qquad (t = 1, \ldots, n)$$

where $\Im\{\cdot\}$ stands for the imaginary part of a complex number. With the help of this identity, the relation between periodogram analysis and the PF method with $\eta = 1$ becomes quite evident: the former seeks to find ω that maximizes the power $|y_n(\omega)|^2$ of the last output ($t = n$) of the filter in (1.8), while the latter looks for ω such that

$\vartheta = \cos\omega$ is a fixed-point of the first-order sample autocorrelation $\hat{\rho}(\vartheta)$ based on the output sequence of the *same* filter.

The impact of this observation is twofold. On the one hand, it explains why the PF method with $\eta = 1$ is able to achieve the same accuracy as periodogram analysis; while on the other, it helps to understand why accurate initial guesses are required for convergence. All of these questions can be answered by the behavior of the squared gain function $G_n(\omega)$ of the complex exponential filter in (1.8) as we have discussed in Chapter 1.

Because the complex exponential filter in (1.8) has an extremely narrow bandwidth of $O(n^{-1})$, the PF method with η close or equal to 1 is able to operate locally in the frequency domain with little or no interference from other frequency components far away from the center of the filter. Therefore, in the case of multiple sinusoids where the frequencies are well separated, it is possible to deal with one sinusoidal frequency at a time, using the PF method for the single sinusoid case (Truong-Van, 1990). In so doing, we simply seek to find *local* attractive fixed-points[4] of the mapping $\hat{\alpha}(\alpha)$, just like we seek to find local maxima of the periodogram in periodogram analysis. However, also like periodogram analysis, this method can not be applied when the frequencies are closer than $O(n^{-1})$ because of the resolution limit of the filter. To cope with the more difficult situation of closely-spaced frequencies, we must turn to the multivariate PF method in which the multiple frequencies are considered simultaneously.

5.4.2 Two Sinusoids in White Noise

Let us now consider the case of two sinusoids ($q = 2$) in additive white noise. In this case, Θ_0 consists of all $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ with

$$\theta_1 = -2(\cos\lambda_1 + \cos\lambda_2) \quad \text{and} \quad \theta_2 = 2(1 + 2\cos\lambda_1 \cos\lambda_2) \tag{5.30}$$

[4]Note that $\hat{\alpha}$ is essentially the unique (global) fixed-point of $\hat{a}(\alpha)$ in the single sinusoid case.

for some $0 < \lambda_1 \leq \lambda_2 < \pi$. Simple algebra shows that

$$\begin{aligned} \cos\lambda_1 &= \tfrac{1}{4}\left(-\theta_1 + \sqrt{\theta_1^2 - 4\theta_2 + 8}\right) \\ \cos\lambda_2 &= \tfrac{1}{4}\left(-\theta_1 - \sqrt{\theta_1^2 - 4\theta_2 + 8}\right). \end{aligned}$$

Therefore θ_1 and θ_2 must satisfy the inequalities

$$\tfrac{1}{4}\left|-\theta_1 \pm \sqrt{\theta_1^2 - 4\theta_2 + 8}\right| < 1 \quad \text{and} \quad \theta_1^2 - 4\theta_2 + 8 \geq 0.$$

It turns out by solving these inequalities that Θ_0 can be expressed as

$$\Theta_0 = \{(\theta_1, \theta_2) : \ 2|\theta_1| - 2 < \theta_2 \leq \tfrac{1}{4}\theta_1^2 + 2\}. \tag{5.31}$$

Moreover, for $q = 2$, the parametrization (5.6) reduces to

$$\theta_1(\boldsymbol{\alpha}) = \frac{1+\eta^4}{\eta+\eta^3}\alpha_1 \quad \text{and} \quad \theta_2(\boldsymbol{\alpha}) = \frac{1+\eta^4}{2\eta^2}\alpha_2.$$

According to (5.31), the parameter space $\mathcal{A}_0(\eta)$ is given by

$$\mathcal{A}_0(\eta) = \left\{(\alpha_1, \alpha_2) : \ \frac{4\eta}{1+\eta^2}|\alpha_1| - \frac{4\eta^2}{1+\eta^4} < \alpha_2 \leq \frac{1+\eta^4}{2(1+\eta^2)^2}\alpha_1^2 + \frac{4\eta^2}{1+\eta^4}\right\}. \tag{5.32}$$

Figure 5.1 shows $\mathcal{A}_0(\eta)$ for $\eta = 0.8$ together with Θ_0 defined by (5.31).

From Figure 5.1, as well as (5.31) and (5.32), we observe that $\mathcal{A}_0(\eta)$ is contained in Θ_0 for any $\eta < 1$, and that $\mathcal{A}_0(\eta)$ increases and eventually coincides with Θ_0 as $\eta \to 1$. Therefore, for any given ω_1 and ω_2 satisfying $0 < \omega_1 < \omega_2 < \pi$, the requirement $\boldsymbol{\alpha}^* \in \mathcal{A}_0(\eta)$ in Theorem 5.2 can always be met by choosing η close enough to 1.

On the other hand, for a given $\eta < 1$, the requirement $\boldsymbol{\alpha}^* \in \mathcal{A}_0(\eta)$ imposes a *separation condition* on the frequencies of the signal. As a matter of fact, in order that $\boldsymbol{\alpha}^* \in \mathcal{A}_0(\eta)$, the frequencies ω_1 and ω_2 should stay away from 0 and π, respectively, and from each other, by a certain amount depending on η. It is not too difficult to verify that a sufficient condition for (5.32) to be fulfilled is that

$$\omega_1, \ \pi - \omega_2 > \arccos\left(\frac{1+\eta^2}{\sqrt{2(1+\eta^4)}}\right) \quad \text{and} \quad \omega_2 - \omega_1 > \arccos\left(\frac{2\eta^2}{1+\eta^4}\right).$$

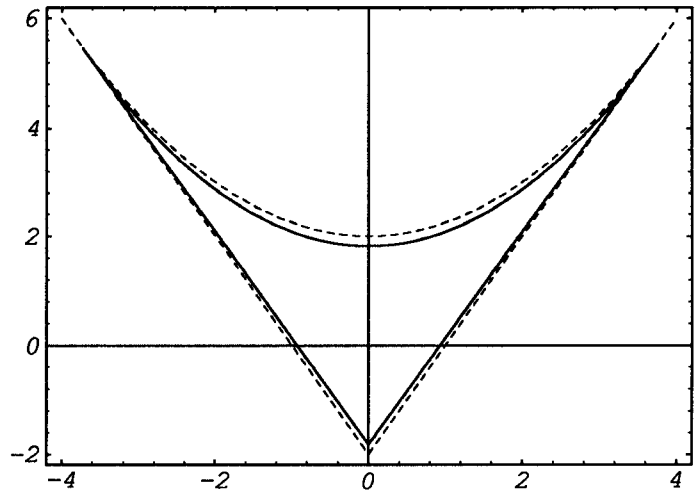


Figure 5.1: Parameter space $\mathcal{A}_0(\eta)$ with $\eta = 0.8$ in the case of two sinusoids. The dotted lines define the region Θ_0.

To provide a complete picture of the separation condition, Figure 5.2 shows for $\eta = 0.8$ and 0.9 the set Ω_η, which we refer to as the *frequency space*, of the frequency pair (ω_1, ω_2) for which $\boldsymbol{\alpha}^* \in \mathcal{A}_0(\eta)$.

Let us now consider the characteristic of the AR filter at $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ which determines the asymptotic behavior of the PF estimator, as indicated by Theorem 4.4. Let us first look at the squared gain function $|H(\omega; \boldsymbol{\alpha}^*)|^2$. In Figure 5.3 (a), $|H(\omega; \boldsymbol{\alpha}^*)|^2$ is plotted for $\eta = 0.92$ as a function of the normalized frequency $f := \omega/\pi$, where $\boldsymbol{\alpha}^* = \mathbf{a}$ is the AR parameter corresponding to $\omega_1 = 0.32\pi$ and $\omega_2 = 0.45\pi$. Clearly, the squared gain function has peaks around the frequencies of the signal, so that the AR filter is able to enhance the sinusoids. On the other hand, the peaks of $|H(\omega; \boldsymbol{\alpha}^*)|^2$ are not exactly located at the sinusoidal frequencies. These slightly biased peaks are required by the parametrization property (A2) in order to make the PF estimator consistent. A further illustration of this point is presented by Figure 5.3 (b), where the poles of the AR filter are plotted in the complex plane. It can be seen that the PF method does not in general force the poles of the AR filter to take the same angular frequencies as

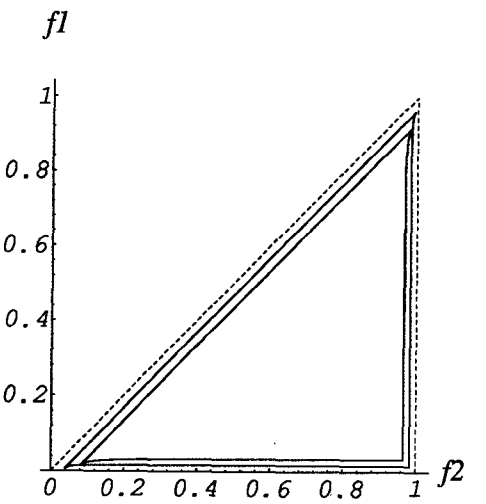


Figure 5.2: Frequency space with $\eta = 0.9$ (smallest) and 0.95 for the normalized frequency pair $(f_1, f_2) = (\omega_1, \omega_2)/\pi$. The region defined by dotted lines corresponds to the case of $\eta = 1$.

the signal in order to produce a consistent estimator, as guaranteed by Theorem 5.2.

Finally, for any given $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ (not necessarily in Θ_0), let $\zeta_1 = p_1 \exp(i\lambda_1)$ and $\zeta_2 = p_2 \exp(i\lambda_2)$ be the zeros of the 4th-degree polynomial

$$1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \theta_1 z^{-3} + z^{-4}$$

such that $0 < \lambda_1 \leq \lambda_2 < \pi$ (or $-\pi < \lambda_2 \leq \lambda_1 < 0$). Then ζ_1 and ζ_2 can be explicitly expressed in closed forms. In fact, it can be shown that when $\Lambda := \theta_1^2 - 4\theta_2 + 8$ is non-negative, we have

$$\zeta_1 = \tfrac{1}{2}\left(s_1 + \sqrt{s_1^2 - 4}\right) \quad \text{and} \quad \zeta_2 = \tfrac{1}{2}\left(s_2 + \sqrt{s_2^2 - 4}\right)$$

where $s_1 = \frac{1}{2}(-\theta_1 + \Lambda^{1/2})$ and $s_2 = \frac{1}{2}(-\theta_1 - \Lambda^{1/2})$; and when $\Lambda < 0$, we have

$$\zeta_1 = \tfrac{1}{2}\left(s_1 + \sqrt{s_1^2 - 4}\right) \quad \text{and} \quad \zeta_2 = \tfrac{1}{2}\left(s_1 - \sqrt{s_1^2 - 4}\right).$$

To obtain $\cos\lambda_1$ and $\cos\lambda_2$ from ζ_1 and ζ_2, we have

$$\begin{aligned} \cos\lambda_1 &= \max(\Re\{\zeta_1/|\zeta_1|\}, \Re\{\zeta_2/|\zeta_2|\}) \\ \cos\lambda_2 &= \min(\Re\{\zeta_1/|\zeta_1|\}, \Re\{\zeta_2/|\zeta_2|\}). \end{aligned}$$

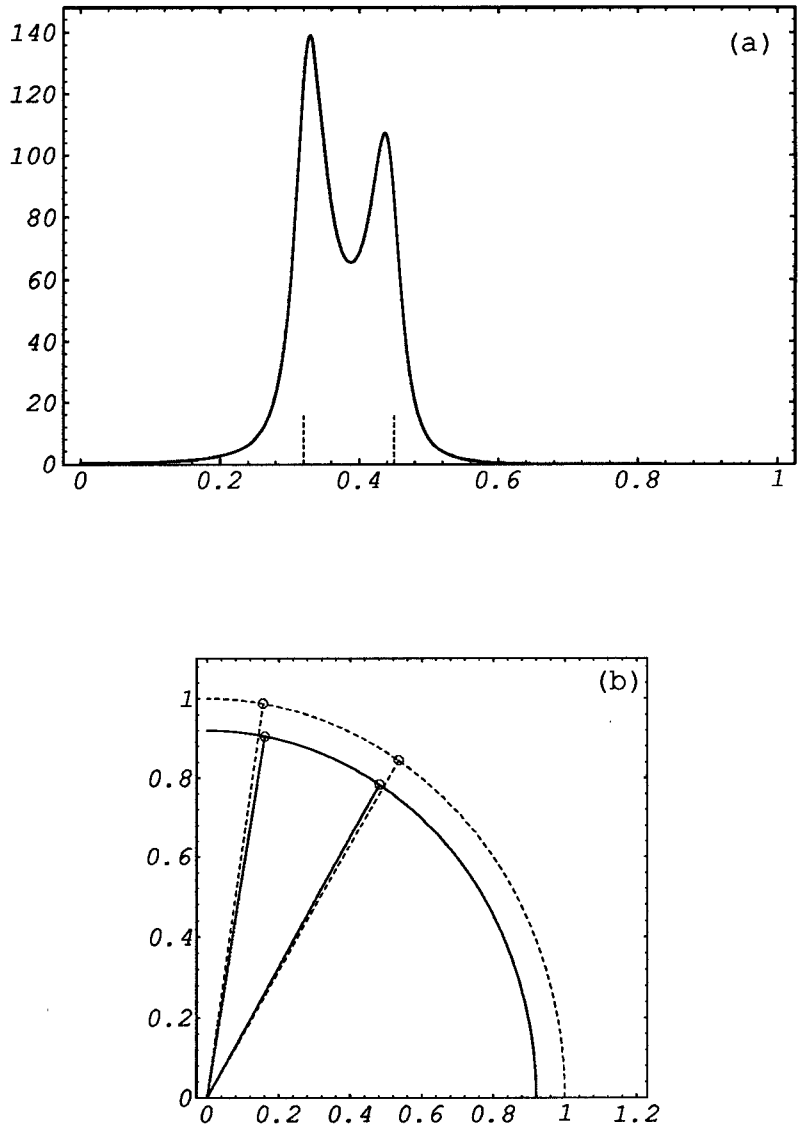


Figure 5.3: (a) Plot of squared gain with $\eta = 0.92$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ as a function of the normalized frequency in the case of two sinusoids with $\omega_1 = 0.32\pi$ and $\omega_2 = 0.45\pi$. (b) Location of poles, with dotted lines indicating true frequencies on the unit circle.

Substituting these results in (5.30) gives the required projection $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ onto Θ_0.

5.5 Experimental Results

In this section, we would like to provide some simulation results to demonstrate the performance of the PF method under various circumstances. For simplicity, our simulations are based on the case of a single sinusoid and the case of two sinusoids, as we have discussed in the preceding section. From these results we shall show the effect of the bandwidth parameter η of the AR filter on the sensitivity of convergence to initial guesses, and on the estimation accuracy of the PF estimator. We shall also gain some insight into the ability of the PF method to resolve close-spaced frequencies which are unresolvable by periodogram analysis.

5.5.1 Univariate PF method

For the univariate PF method corresponding to $q = 1$ in Section 5.4.1, we find it convenient to consider an alternative representation of the mapping $\hat{a}(\alpha)$ in terms of the normalized frequency $f = \omega/\pi$. Note that for any $\alpha \in \mathcal{A}_0(\eta)$ we can always write $\alpha = -2\cos\omega = -2\cos(\pi f)$ with some $f \in (0,1)$. Let us define the mapping

$$\hat{\varphi}(f) := \arccos\{-\tfrac{1}{2}\hat{a}(-2\cos(\pi f))\}/\pi.$$

It is clear that if $\hat{\alpha} = -2\cos(\pi\hat{f})$ is a fixed-point of $\hat{a}(\alpha)$ then $\hat{f}$ is a fixed-point of $\hat{\varphi}(f)$, and vice versa. With the help of this mapping, the algorithm in (4.8) is transformed into a fixed-point iteration in terms of f, namely,

$$\hat{f}_m = \hat{\varphi}(\hat{f}_{m-1}) \qquad (m = 1, 2, \ldots). \tag{5.33}$$

It is therefore sufficient to study the behavior of $\hat{\varphi}(f)$ as a function of f in order to understand the convergence of the iteration. Notice that the derivative of $\hat{\varphi}(f)$ at the fixed-point $\hat{f} = \arccos(-\hat{\alpha}/2)/\pi$ can be easily shown to be $\hat{a}'(\hat{\alpha})$. With $\hat{\alpha}$ being the

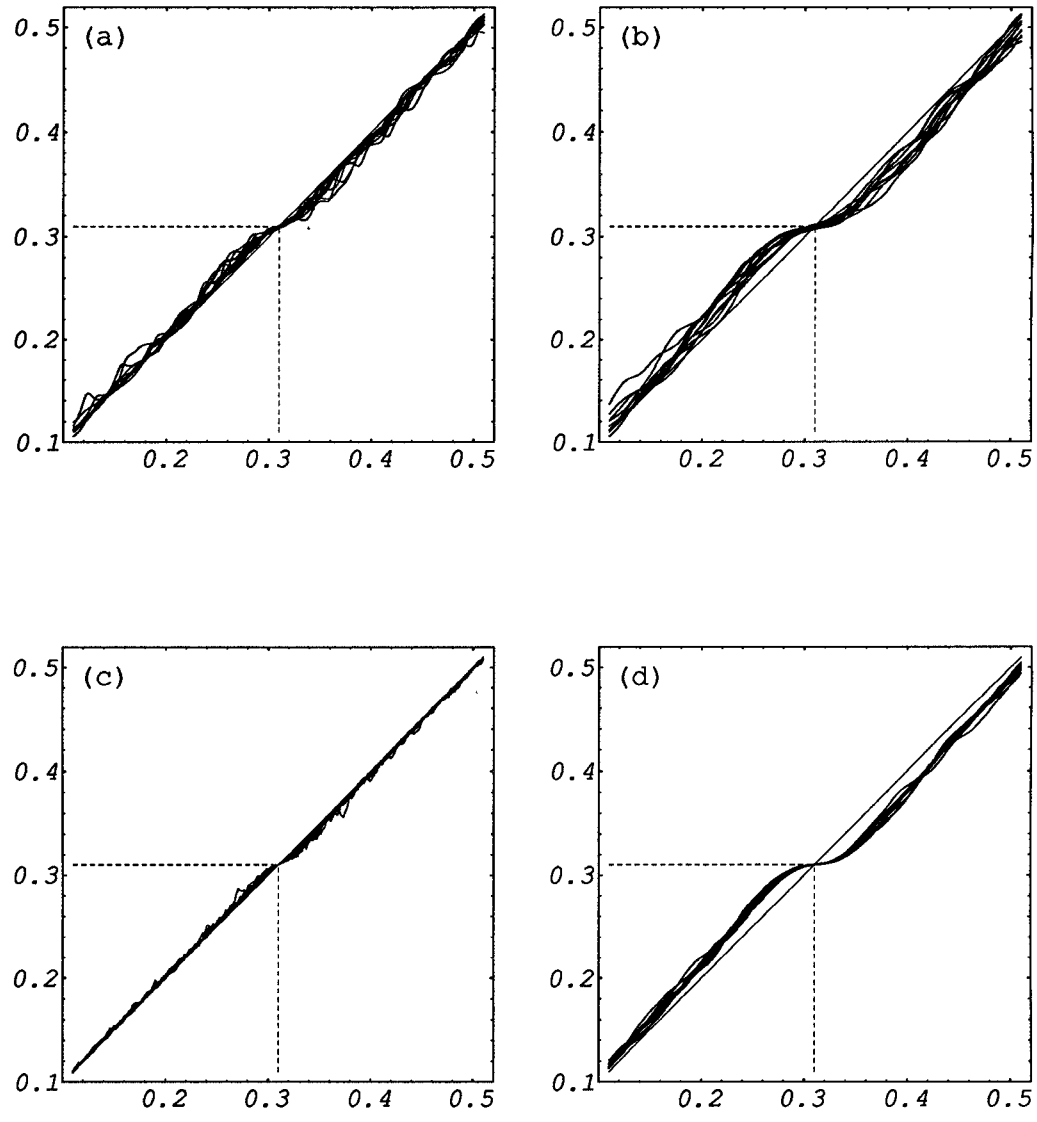


Figure 5.4: Least squares mapping $\hat{\varphi}(f)$ in the case of a single sinusoid with $f_1 = \omega_1/\pi = 0.31$, $\phi_1 = 0$, and SNR $= -3$dB. In (a) and (b), the data length is $n = 100$ for each of the 10 realizations plotted, with $\eta = 1$ for (a) and $\eta = 0.96$ for (b). The data length is increased to $n = 500$ in (c) and (d), with η being 1 and 0.96, respectively.

PF estimator as given in Theorem 4.2 so that $|\hat{a}'(\hat{\alpha})| < 1$, the mapping $\hat{\varphi}(f)$ is also contractive in a neighborhood of the fixed-point $\hat{f}$. Since the least squares technique is used in the calculation of $\hat{a}(\alpha)$ as specified by (4.12), we thus refer to $\hat{\varphi}(f)$ as the least squares mapping of f. In Figure 5.4 are plotted 10 independent realizations of the least squares mapping $\hat{\varphi}(f)$ in the case of a single zero-phase sinusoid with $f_1 = 0.31$ in Gaussian white noise. The SNR is -3 dB, and the data length is $n = 100$ for Figure 5.4(a) and Figure 5.4(b), and $n = 500$ for Figure 5.4(c) and Figure 5.4(d).

It is clear that $\hat{\varphi}(f)$ has always an attractive fixed-point $\hat{f}$ near the true frequency, and that the variation of the fixed-point is directly related to $1 - \eta$. This suggests that the highest accuracy is achieved with $\eta = 1$. On the other hand, by comparing Figure 5.4(a) with Figure 5.4(b), and Figure 5.4(c) with Figure 5.4(d), it can be seen that the basin of attraction — the collection of f with which as initial guesses the iteration (5.33) converges to $\hat{f}$ — is significantly larger when $\eta = 0.96$ than when $\eta = 1$, indicating that the PF method with a relative small η is able to accommodate poor initial guesses for the iteration (5.33) to converge to the desired fixed-point. Moreover, comparing the graphs in Figure 5.4 column-wise reveals that the basin of attraction with $\eta = 0.96$ is basically not affected as the data length n increases, whereas the basin of attraction with $\eta = 1.0$ is inversely related to n. Therefore the initial accuracy of $O(1)$ is sufficient when $\eta < 1$ as compared to $o(n^{-1})$ when $\eta = 1$. In summary, this experiment confirms by studying the behavior of the mapping $\hat{\varphi}(f)$ the necessity of starting with a relatively small value of η to accommodate poor initial guesses and afterwards gradually increasing η toward 1 to improve the estimation accuracy.

To further illustrate this point, Figure 5.5 presents the negative logarithms of mean-squared error (mse) for the normalized frequency estimates $\hat{f} = \arccos(-\hat{\alpha}/2)/\pi$ with various values of η, based on 100 independent realizations of a single sinusoid in Gaussian white noise with different lengths. Here, the true frequency of the sinusoid is $\omega_1 = 0.42\pi$ ($f_1 = 0.42$) and the phase is $\phi_1 = 0.1\pi$. The SNR is fixed at 0 dB for

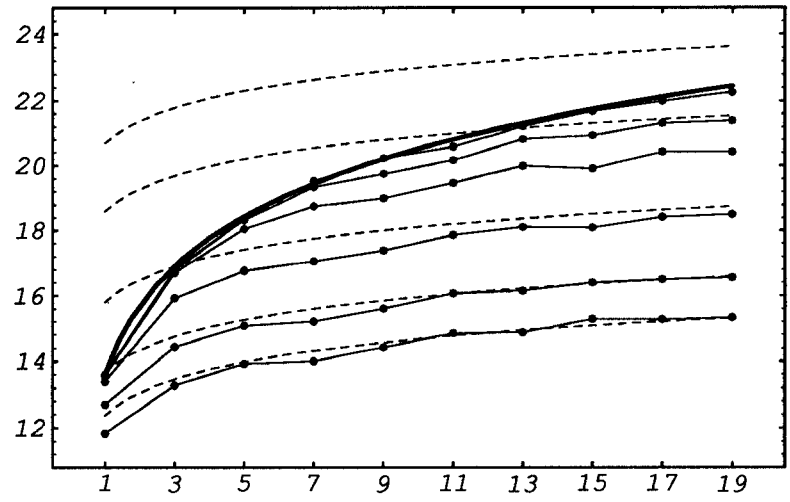


Figure 5.5: Plot of $-\log(\text{mse})$ against the data length n ($\times 100$) with (bottom up) η = 0.85, 0.90, 0.95, 0.98, 0.99, and 0.999. The dashed curves indicate the asymptotic variances of $\hat{f}$ for (bottom up) η = 0.85, 0.90, 0.95, 0.98, and 0.99; and the dark solid curve stands for the asymptotic variance of the NLS estimator.

each realization by adjusting the sample variance of the noise according to the sample variance of the signal. Recall (see Section 5.3) that a sequence of PF estimators $\{\hat{\alpha}(\eta_k)\}$, and hence $\{\hat{f}(\eta_k)\}$, can be obtained in correspondence with an increasing sequence $\{\eta_k\}$ in such a way that $\hat{\alpha}(\eta_k)$ is produced by the iteration (4.8) using $\hat{\alpha}(\eta_{k-1})$ as the initial value. In Figure 5.5, the η-sequence contains η = 0.85, 0.90, 0.95, 0.98, 0.99, and 0.999 for each fixed data length n. The initial guess that generates the PF estimator with $\eta = 0.85$ is fixed at $\hat{\alpha}_0 = -2\cos(0.6\pi)$ as the data length grows from $n = 100$ to $n = 1900$, so that the initial accuracy is merely $O(1)$. For each fixed n and η, the iteration (4.8) is terminated at the mth step if $|\hat{f}_m - \hat{f}_{m-1}| \leq 10^{-5}$.

Two conclusions can be drawn immediately from this figure. First of all, it is clear as before that starting with initial guesses of accuracy $O(1)$ the PF estimator is able to improve the estimation accuracy and eventually achieve the accuracy $O(n^{-3/2})$ of NLS (or periodogram analysis). The key point is that the improvement of accuracy is obtained not by switching from one completely different method to another (e.g., PF

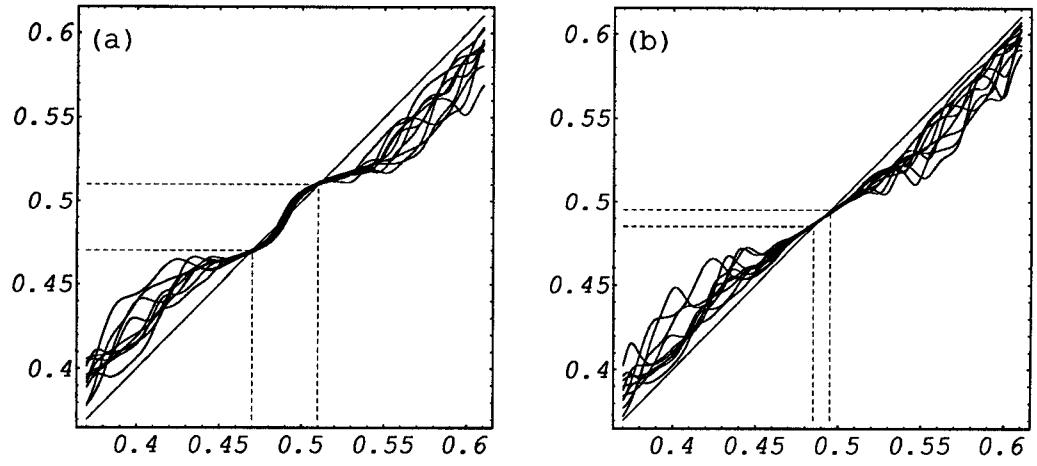


Figure 5.6: Univariate least squares mapping $\hat{\varphi}(f)$ in the case of two sinusoids. The data length is $n = 100$ for each of the 10 realizations plotted, and the SNR is 0 dB per sinusoid. (a) $\eta = 0.99$ for well-separated frequencies with $f_1 = 0.47$ and $f_2 = 0.51$. (b) $\eta = 1$ for closely-spaced frequencies with $f_1 = 0.485$ and $f_2 = 0.495$.

with $\eta = 1$ initiated by DFT, as suggested by Quinn and Fernandes (1991)), but *by an* integrated *method of linear least squares estimation plus linear recursive filtering, with the bandwidth parameter η increasing toward 1.* This simple integration makes it easier for hardware implementation in some applications. Secondly, the mse closely follows the theoretical asymptotic variance of the PF estimator given by (5.26) for relatively small data lengths if η is not too close to 1. When η is close to 1, however, a large sample is required for the theoretical results to be meaningful. This suggests that a more careful analysis be considered when $1 - \eta$ is comparable with n^{-1}.

As aforementioned (see Section 5.4.1), in the case of multiple sinusoids where the frequencies are well separated, the univariate PF method can be applied parallelly to one frequency at a time as if there were only a single sinusoid. This is because the AR filter is bandpass so that the mapping $\hat{\varphi}(f)$ is not significantly affected by the frequency components far away from the center of the filter's effective pass-band. As a result, local attractive fixed-points appear in $\hat{\varphi}(f)$ near the frequencies of the sinusoidal

signal, as illustrated by Figure 5.6(a). In this figure, 10 independent realizations of $\hat{\varphi}(f)$ are plotted, each from a data record of length $n = 100$ containing two zero-phase sinusoids whose (normalized) frequencies, $f_1 = 0.47$ and $f_2 = 0.51$, are separated by two Fourier bins of width $\Delta f := 2/n = 0.02$. The noise is white Gaussian, and the SNR is 0 dB per sinusoid. Clearly, the univariate mapping $\hat{\varphi}(f)$ has two distinct local attractive fixed-points — each corresponding to a sinusoid. This, however, is no longer the case — even with $\eta = 1$ — when the frequencies are closer than a Fourier bin, as shown in Figure 5.6(b) where the true frequencies are $f_1 = 0.485$ and $f_2 = 0.495$ while other conditions remain the same. This phenomenon is due to the resolution limit of the AR filter as we have discussed earlier in Section 5.4.1. To resolve closely-spaced frequencies, we must rely on the multivariate PF method that deals with the frequencies simultaneously.

5.5.2 Multivariate PF method for Two sinusoids

Let us now consider the multivariate PF method using the AR filter for two sinusoids ($q = 2$) in Gaussian white noise. All of the following simulations are based on 100 independent realizations of $\{y_t\}$ with a relatively short length of $n = 100$. The phases of the sinusoids are fixed at zero, and the sample variance of the noise is adjusted in each realization according to the sample variance of the signal in order to achieve the required signal-to-noise ratio.

Furthermore, in both PF and GLS — corresponding to the parametrizations (5.6) and (5.2), respectively — the poles of the AR filter are constrained to be on the circle $|z| = \eta$, by projection if necessary (see Section 5.1.2), so that the parameter η effectively controls the bandwidth of the AR filter and the performance of the estimators. For convenience, the following simulation results are given in regard to the normalized frequencies $f_k = \omega_k/\pi$ again, and the average mean-squared error

$$\text{mse} := \tfrac{1}{2}\{E(\hat{f}_1 - f_1)^2 + E(\hat{f}_2 - f_2)^2\}$$

is employed as an overall performance index. Moreover, we define the average bias and average variance of the frequency estimates by

$$\text{bias} := \tfrac{1}{2}\{(E(\hat{f}_1) - f_1)^2 + (E(\hat{f}_2) - f_2)^2\} \quad \text{and} \quad \text{var} := \tfrac{1}{2}\{\text{var}(\hat{f}_1) + \text{var}(\hat{f}_2)\}$$

respectively. The frequency estimates of both PF and GLS are obtained by the fixed-point iteration (4.8) in connection with (2.16) that provides the relationship between the frequency and AR estimates. The stopping rule of the iteration is given by

$$\sqrt{(\hat{f}_1^{(m)} - \hat{f}_1^{(m-1)})^2 + (\hat{f}_2^{(m)} - \hat{f}_2^{(m-1)})^2} < 10^{-5}.$$

In other words, the iteration is terminated at the mth iteration if this inequality is satisfied.

In our simulations, we first compare the performance of PF and GLS in two cases where the frequencies are separated by four and two Fourier bins, respectively. In both cases the SNR is fixed at 0 dB per sinusoid, while the bandwidth parameter η in the AR filter (5.1) takes on different values. Since η varies, it is convenient to explicitly write the corresponding frequency estimates $(\hat{f}_1(\eta), \hat{f}_2(\eta))$ as functions of η. Table 5.1 and Table 5.2 present some statistics of the frequency estimates for eight *ascending* values of η, that is, $\eta_1 = 0.95$, $\eta_2 = 0.96, \ldots, \eta_8 = 1$. The mean and variance of the stopping time m are also given as "complexity" in the form of "mean $\pm$ variance".

In both PF and GLS, we use Prony's estimator $\mathbf{a}_{\text{LS}}$ in (2.9) as the initial guess of the AR parameter $\mathbf{a}$, corresponding the first value $\eta_1 = 0.95$. When the iteration terminates, the resulting AR estimate, denoted by $\hat{\mathbf{a}}(\eta_1)$, is used not only to obtain the frequency estimates $(\hat{f}_1(\eta_1), \hat{f}_2(\eta_1))$, but also to initiate the iteration for the next value $\eta_2 = 0.96$. In general, as η grows, we employ the previous AR estimate $\hat{\mathbf{a}}(\eta_{k-1})$ to initiate the iteration (4.8) and yield $\hat{\mathbf{a}}(\eta_k)$.

In Table 5.1, the true frequencies are separated by four Fourier bins of width $\Delta f = 0.02$ with $(f_1, f_2) = (0.41, 0.59)$. As we can see, Prony's estimator gives poor frequency estimates, while both PF and GLS significantly improve Prony's estimator in

Table 5.1: PF & GLS Estimates for Well-Separated Frequencies

	PF				GLS			
η	mse	bias	var	complexity	mse	bias	var	complexity
Prony	4.72e-3	4.40e-3	3.23e-4	—	4.72e-3	4.40e-3	3.23e-4	—
0.950	2.20e-6	1.17e-8	2.19e-6	8.0 ± 0.6	2.34e-6	1.48e-7	2.19e-6	8.1 ± 0.8
0.960	1.96e-6	9.14e-9	1.96e-6	3.1 ± 0.2	2.04e-6	9.47e-8	1.95e-6	3.2 ± 0.4
0.970	1.79e-6	9.72e-9	1.78e-6	3.2 ± 0.3	1.84e-6	6.21e-8	1.77e-6	3.3 ± 0.4
0.980	1.67e-6	1.64e-8	1.65e-6	3.3 ± 0.4	1.69e-6	4.96e-8	1.65e-6	3.4 ± 0.4
0.985	1.63e-6	2.39e-8	1.60e-6	2.9 ± 0.4	1.65e-6	4.79e-8	1.60e-6	2.8 ± 0.4
0.990	1.59e-6	3.60e-8	1.55e-6	3.0 ± 0.5	1.60e-6	5.08e-8	1.55e-6	2.7 ± 0.4
0.995	1.55e-6	5.53e-8	1.49e-6	3.4 ± 0.4	1.55e-6	6.08e-8	1.49e-6	2.9 ± 0.6
1.000	1.50e-6	8.27e-8	1.42e-6	3.7 ± 0.6	1.51e-6	8.29e-8	1.42e-6	3.4 ± 0.5

Table 5.2: PF & GLS Estimates for Closely-Spaced Frequencies

	PF				GLS			
η	mse	bias	var	complexity	mse	bias	var	complexity
Prony	1.46e-2	1.40e-2	5.68e-4	—	1.46e-2	1.40e-2	5.68e-4	—
0.950	1.81e-6	2.69e-8	1.78e-6	10.6 ± 3.4	4.68e-6	3.05e-6	1.63e-6	11.0 ± 5.3
0.960	1.67e-6	6.20e-8	1.61e-6	3.3 ± 0.4	3.63e-6	2.17e-6	1.46e-6	3.8 ± 0.2
0.970	1.63e-6	1.63e-7	1.46e-6	3.8 ± 0.3	2.95e-6	1.59e-6	1.36e-6	3.9 ± 0.2
0.980	1.75e-6	4.21e-7	1.32e-6	4.3 ± 0.3	2.56e-6	1.29e-6	1.27e-6	3.7 ± 0.4
0.985	1.90e-6	6.57e-7	1.25e-6	4.2 ± 0.4	2.46e-6	1.25e-6	1.21e-6	2.8 ± 0.4
0.990	2.15e-6	9.83e-7	1.16e-6	4.4 ± 0.3	2.45e-6	1.31e-6	1.15e-6	3.0 ± 0.5
0.995	2.47e-6	1.41e-6	1.06e-6	4.5 ± 0.5	2.56e-6	1.50e-6	1.06e-6	3.7 ± 0.7
1.000	2.85e-6	1.91e-6	9.45e-7	4.5 ± 0.8	2.86e-6	1.91e-6	9.46e-7	4.3 ± 0.8

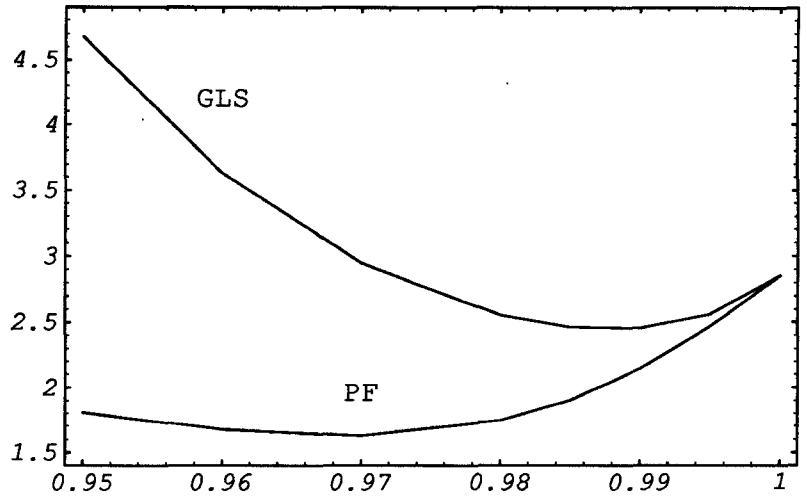


Figure 5.7: Plot of mse($\times 10^{-6}$) against η for closely-spaced frequencies.

terms of mean-squared error, even with a relatively small η. Moreover, the estimation accuracy can be further improved by increasing η toward 1, just like in the single sinusoid case. Table 5.1 shows that as η approaches 1 the PF and GLS estimates achieve a precision (mse) of 1.50×10^{-6} — very close to the asymptotic variance of the NLS estimator which in this case equals 1.22×10^{-6}. Therefore, when the frequencies are well-separated, PF and GLS have the same final performance – which approaches that of the NLS method — as η increases toward 1.

When the frequencies are close to each other, the PF estimator performs better than GLS, as can be seen from Table 5.2 and Figure 5.7. In this experiment, the true frequencies are $(f_1, f_2) = (0.47, 0.51)$ while all other conditions remain the same as in the previous one. Notice that the true frequencies are now separated only by two Fourier bins as compared to four in the previous experiment. It is interesting to observe that as the bandwidth parameter η increases toward 1 the mse of both methods no longer decreases monotonically as in the case of a single sinusoid and in the case where the two frequencies are separated further by four Fourier bins. Instead, it starts increasing after a certain value of η (see also Figure 5.7). The reason is the following.

Table 5.3: Estimation With $\eta = 1$

(f_1, f_2)	mse	$E(\hat{f}_1) \pm \text{var}(\hat{f}_1)$	$E(\hat{f}_2) \pm \text{var}(\hat{f}_2)$	complexity
(0.41,0.59)	4.33e-6	$0.409704 \pm 1.30\text{e-}6$	$0.590529 \pm 6.99\text{e-}6$	17.9 ± 10.7
(0.47,0.51)	1.38e-4	$0.468628 \pm 9.37\text{e-}7$	$0.513672 \pm 2.49\text{e-}4$	24.3 ± 63.3

A closer examination of Table 5.1 and Table 5.2 reveals that as η approaches 1 the bias increases while the variance decreases in both methods. In the first case where the frequencies are well separated (Table 5.1, the bias never dominates the variance, and hence the mse decreases basically along with the decrease of the variance. On the other hand, the bias becomes dominant as η approaches 1 in the second case (see Table 5.2), and a trade-off effect between bias and variance takes place. As we can see from Figure 5.7, the best value of η for the PF estimator lies between 0.96 and 0.98 where the mse achieves the smallest values. The GLS estimator is clearly inferior to the PF estimator in this example because of its relatively higher bias. Indeed, the bias and variance of the GLS estimator play an equal role in the mse, since their magnitudes are of the same order (see Table 5.2).

Table 5.2 illustrates the role of η as a parameter that can be utilized to balance the bias and variance of the PF estimator for minimizing the mean-squared error. Now, in Table 5.3, we illustrate the role of η in the convergence of the fixed-point iteration (4.8) when initial guesses are poor. Instead of gradually increasing η toward 1, as done in Table 5.1 and Table 5.2, the frequency estimates in Table 5.3 were obtained right away with $\eta = 1$, using Prony's estimator as the initial guess. This is equivalent to the iterative procedure that employs the AR filter *without* η and starts with Prony's estimator. As can be seen from Table 5.3, the mean-squared error is higher than the mse reported in Table 5.1 and Table 5.2 corresponding to (gradually achieved) $\eta = 1$, especially for the second case where the frequencies are relatively close to each other.

This indicates that without η in the AR filter the iteration (4.8) may fail to converge to the desired fixed-point when poor initial guesses, such as Prony's estimator, are used. The reason is that the bandwidth of the AR filter without η (or, equivalently, with $\eta = 1$) is extremely narrow. Although it could be helpful to have a narrow bandwidth for the enhancement of the sinusoids if good initial guesses are used, a narrow bandwidth might not be able to capture the sinusoidal signal when tuned according to inaccurate frequency estimates. This experiment verifies once again that the safest way of applying the PF method is to start with a relatively small η, to accommodate even poor initial guesses, and then gradually increase η as improved estimates from previous iterations become available.

To show the performance of the PF estimator under different signal-to-noise ratios when the frequencies are closely spaced *within* a Fourier bin, Figure 5.8(a) presents the negative logarithm of the mse for various values of SNR, with the dotted line indicating the asymptotic variance of NLS as a reference. In this example, the frequencies are $(f_1, f_2) = (0.485, 0.495)$ and the bandwidth parameter η is fixed at 0.985 in both PF and GLS. Prony's estimator again is used to initiate the fixed-point iteration (4.8) for both methods. As can be seen, the mse of the PF estimator closely follows the asymptotic variance of NLS when SNR $\geq$ 2.5 dB, and the performance of both PF and GLS deteriorates rapidly when the SNR is below this threshold. The poor initial accuracy of Prony's estimator is largely responsible for this particular value of threshold. In fact, simulations show that the threshold can be extended to -2 dB if the initial guesses are taken to be the two Fourier frequencies which correspond to the largest absolute values in the FFT of the data. Averages of frequency estimates are plotted against various SNR in Figure 5.8(b). It is clear that both PF and GLS are able to resolve the frequencies which cannot be resolved by periodogram analysis, but the PF method has a smaller bias which allows it to provide more accurate frequency estimates than the GLS method.

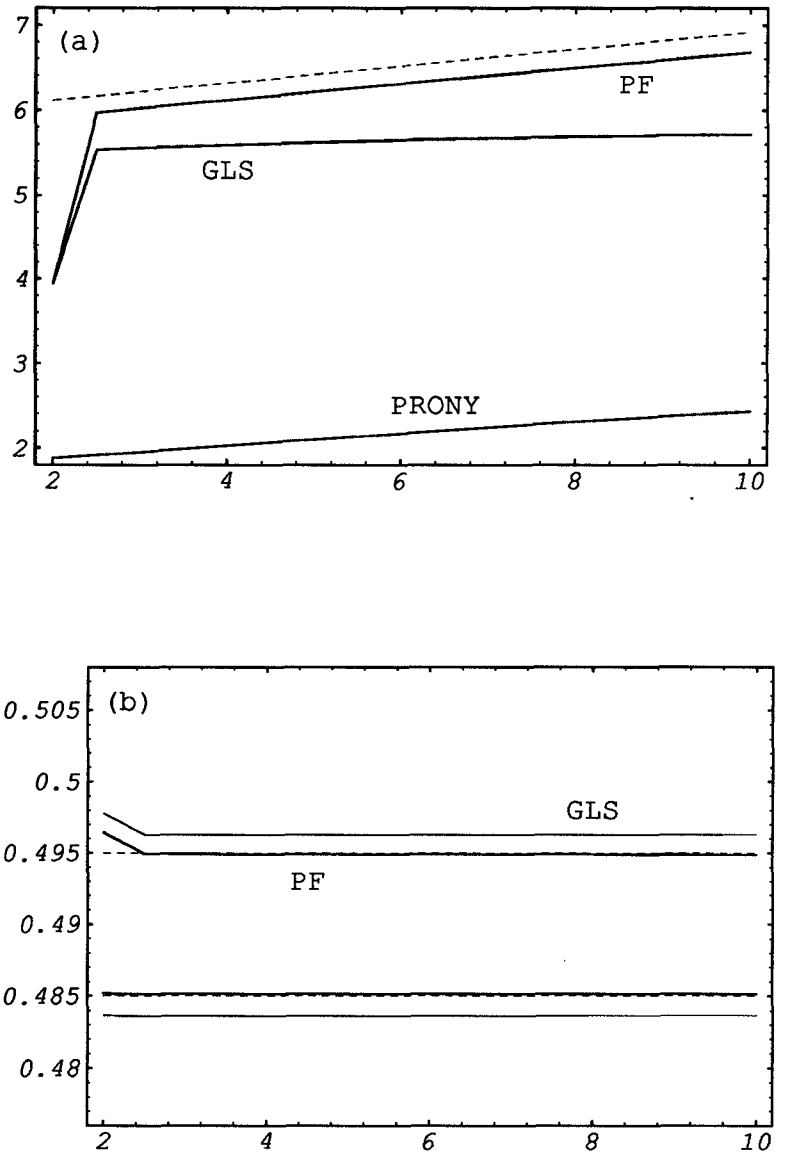
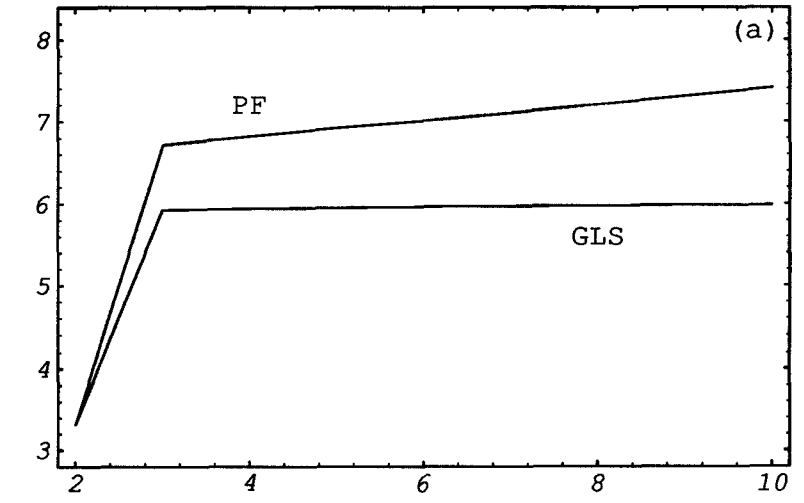


Figure 5.8: Closely-spaced frequencies with $(f_1, f_2) = (0.485, 0.495)$. (a) Plot of $-\log(\text{mse})$ against SNR in dB, with the dotted curve indicating the asymptotic variance of NLS. (b) Plot of averaged frequency estimates against SNR in dB, with dotted lines indicating true frequencies.

We noticed from our intensive simulations that if the distance between the two frequencies is further reduced, it is very likely that the fixed-point $\hat{\boldsymbol{\alpha}}$ will fall outside the parameter space $\mathcal{A}_0(\eta)$, resulting in a single frequency estimate $\hat{f}_1 = \hat{f}_2$ between the true frequencies. If some rough knowledge about the separation is known, more accurate estimates can be obtained upon projecting $\hat{\boldsymbol{\alpha}}$ back into $\mathcal{A}_0(\eta)$. Let $\tilde{\boldsymbol{\theta}}$ be the projection of $\boldsymbol{\theta}(\hat{\boldsymbol{\alpha}})$ into Θ_0. Then, $\tilde{\boldsymbol{\alpha}} := \mathbf{T}_\eta^{-1}\tilde{\boldsymbol{\theta}}$ defines the projection of $\hat{\boldsymbol{\alpha}}$ into $\mathcal{A}_0(\eta)$. By this projection, the separation of $\hat{f}_1$ and $\hat{f}_2$ can be effectively controlled by η, as can be seen in Figure 5.2, and the improvement of estimation accuracy be achieved upon judiciously selecting η. Figure 5.9 shows the improvement of PF over GLS on the estimation accuracy when the frequencies are extremely close. In this experiment, the true frequencies, $(f_1, f_2) = (0.41, 0.412)$, are only 10% apart relative to the width of a Fourier bin. The frequency estimates were obtained with $\eta = 0.997$ and the fixed-point iteration (4.8) was initiated by Prony's estimator. Figure 5.9(a) shows the negative logarithm of the mse for different values of SNR and Figure 5.9(b) presents the averages of the frequency estimates. Compared to the GLS estimator, the PF estimator has a much smaller bias which enables it to achieve a smaller mean-squared error. Notice that the GLS estimator gives essentially a single frequency $f \approx 0.411$ between the two true frequencies. This procedure, however, should not be considered as a method that detects the number of sinusoids, since the projection of $\hat{\boldsymbol{\alpha}}$ into $\mathcal{A}_0(\eta)$ implicitly requires the information about the number of sinusoids.

Finally, we note that in the preceding discussion the phases were fixed at zero. Experience shows, however, that when the phases are chosen at random the mse may worsen somewhat. This is understandable due to the small sample size which cannot explain the addition of extra sources of variability (Kay and Marple, 1981).

To end this section, let us investigate the behavior of the mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in the case of two sinusoids. For convenience, we transform the mapping into the frequency domain, as we have done in the single sinusoid case, and obtain a two-dimensional least



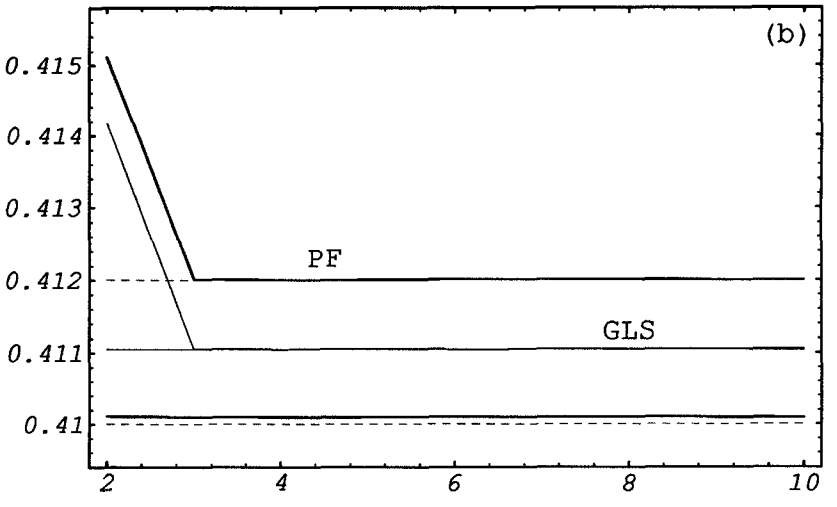


Figure 5.9: Closely-spaced frequencies with $(f_1, f_2) = (0.41, 0.412)$. (a) Plot of $-\log(\text{mse})$ against SNR in dB. (b) Plot of averaged frequency estimates against SNR in dB with dotted lines indicating true frequencies.

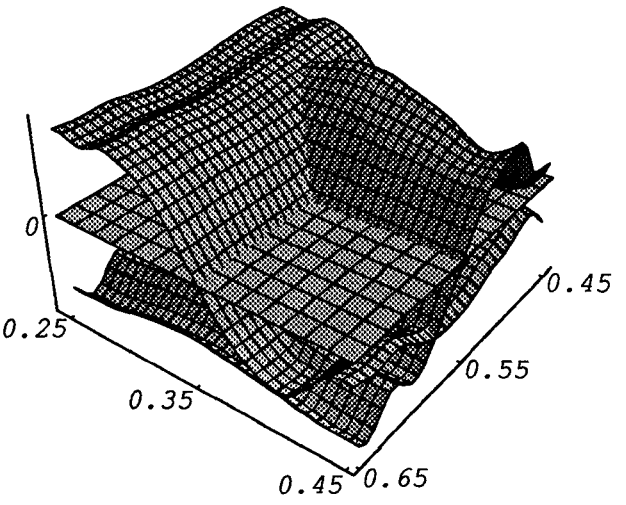


Figure 5.10: Two-dimensional least squares mapping $\hat{\boldsymbol{\psi}}(\boldsymbol{f}) = \hat{\boldsymbol{\varphi}}(\boldsymbol{f}) - \boldsymbol{f}$ in the case of two zero-phase sinusoids with $(\omega_1, \omega_2) = (0.35\pi, 0.55\pi)$. The SNR is 0 dB per sinusoid, and the data length is $n = 100$. A single realization of $\hat{\boldsymbol{\psi}}(\boldsymbol{f})$ with $\eta = 0.96$ is plotted over the region $(f_1, f_2) \in [0.25, 0.45] \times [0.45, 0.65]$.

squares mapping $\hat{\boldsymbol{\varphi}}(\boldsymbol{f}) = [\hat{\varphi}_1(f_1, f_2), \hat{\varphi}_2(f_1, f_2)]^T$, where $\boldsymbol{f} := [f_1, f_2]^T$. More precisely, we define $\hat{\boldsymbol{\varphi}}(\boldsymbol{f})$ as the composition of the following mappings:

$$\hat{\boldsymbol{\varphi}}: \quad \boldsymbol{f} \mapsto \hat{\mathbf{a}}(\boldsymbol{\alpha}) \mapsto \hat{\boldsymbol{f}} = \hat{\boldsymbol{\varphi}}(\boldsymbol{f})$$

where $\boldsymbol{\alpha}$ is determined from the identities in (5.30) with $\pi \boldsymbol{f}$ in place of (λ_1, λ_2), and $\hat{\boldsymbol{f}}$ is obtained from the zeros of the AR polynomial corresponding to $\hat{\mathbf{a}}(\boldsymbol{\alpha})$. With this mapping, the fixed-point iteration (4.8) becomes $\hat{\boldsymbol{f}}_m = \hat{\boldsymbol{\varphi}}(\hat{\boldsymbol{f}}_{m-1})$, $(m = 1, 2, \ldots)$.

Figure 5.10 shows a single realization of the mapping $\hat{\boldsymbol{\psi}}(\boldsymbol{f}) = [\hat{\psi}_1(\boldsymbol{f}), \hat{\psi}_1(\boldsymbol{f})]^T := \hat{\boldsymbol{\varphi}}(\boldsymbol{f}) - \boldsymbol{f}$ together with the zero-plane, where the true frequencies are well separated by ten Fourier bins. The intersection of $\hat{\psi}_1(\boldsymbol{f})$ with the zero-plane defines the *fixed-curve* $\hat{\varphi}_1(f_1, f_2) = f_1$ on which the f_1-coordinate cannot be altered by the mapping $\hat{\boldsymbol{\varphi}}(\boldsymbol{f})$. Similarly, the intersection of $\hat{\psi}_2(\boldsymbol{f})$ with the zero-plane determines the fixed-curve $\hat{\varphi}_2(f_1, f_2) = f_2$ on which the f_2-coordinate cannot be changed by $\hat{\boldsymbol{\varphi}}(\boldsymbol{f})$. The fixed-

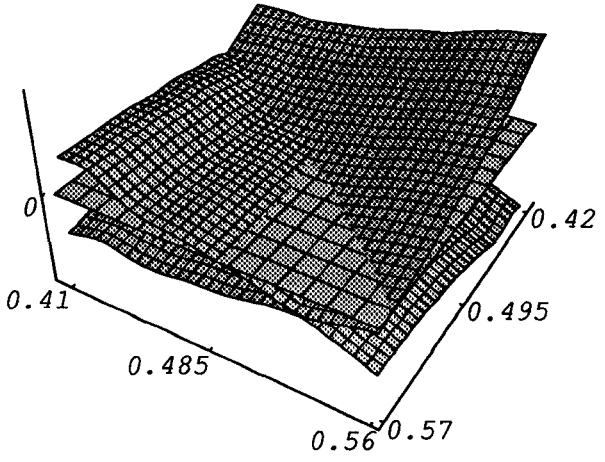


Figure 5.11: Two-dimensional least squares mapping $\hat{\psi}(\boldsymbol{f}) = \hat{\varphi}(\boldsymbol{f}) - \boldsymbol{f}$ in the case of two zero-phase sinusoids with $(\omega_1, \omega_2) = (0.485\pi, 0.495\pi)$. The SNR is 5 dB per sinusoid, and the data length is $n = 100$. A single realization of $\hat{\psi}(\boldsymbol{f})$ with $\eta = 0.985$ is plotted over the region $(f_1, f_2) \in [0.41, 0.56] \times [0.42, 0.57]$.

point of $\hat{\varphi}(\boldsymbol{f})$ is therefore given by the intersection of these fixed-curves. Moreover, the fixed-point iteration can be written in terms of $\hat{\psi}(\boldsymbol{f})$ as follows

$$\hat{\boldsymbol{f}}_m = \hat{\boldsymbol{f}}_{m-1} + \hat{\psi}(\hat{\boldsymbol{f}}_{m-1}) \qquad (m = 1, 2, \ldots).$$

It is clear that $\hat{\psi}_1(\hat{\boldsymbol{f}}_{m-1})$ and $\hat{\psi}_2(\hat{\boldsymbol{f}}_{m-1})$ are the increments in f_1 and f_2, respectively, at the mth iteration.

Figure 5.12(a) and Figure 5.12(b) present the contours of $\hat{\psi}(\boldsymbol{f})$ viewed from above and below the zero-plane. It is interesting to observe that the two fixed-curves roughly coincide with the straight lines $f_1 = 0.35$ and $f_2 = 0.55$, respectively. This implies that the two-dimensional search for the fixed-point of $\hat{\varphi}(\boldsymbol{f})$ is *decoupled* as two independent one-dimensional problems of seeking to find fixed-points along one of the coordinates while keeping the other fixed. This phenomenon is similar to what we have encountered in Chapter 1 where the multi-dimensional optimization problem of non-

linear least squares can be approximated by a number of independent one-dimensional optimization problems (periodogram analysis) when the frequencies are well separated.

For closely-spaced frequencies, however, the behavior of $\hat{\varphi}(\boldsymbol{f})$ is slightly different, as one may have expected: the two-dimensional problem is no longer decoupled when the frequencies occur within a Fourier bin. To illustrate this point, a single realization of $\hat{\psi}(\boldsymbol{f})$ is shown in Figure 5.11 where the true frequencies are separated only by 50% of a Fourier bin. Figure 5.12(c) and Figure 5.12(d) present the contour plots together with the diagonal line $f_1 = f_2$. It is clear that the fixed-curves do not *independently* provide correct frequency estimates any more, since in the region $f_1 < f_2$ they are skewed toward the lines $f_1 = 0.49$ and $f_2 = 0.49$, respectively, tending to yield a single frequency estimate $f_1 = f_2 = 0.49$, i.e., the average of the true frequencies. However, the fixed-curves, *jointly*, provide again the correct frequency estimates by their *intersection* — namely, the (multivariate) fixed-point of the mapping $\hat{\varphi}(\boldsymbol{f})$.

5.6 Concluding Remarks

Given a time series $\{y_1, \ldots, y_n\}$ from a stochastic process $\{y_t\}$ in (1.1), we considered the classical problem of frequency estimation in the presence of additive noise. We proposed the PF method that overcomes the predicament of inconsistency of Prony's estimator and provides consistent frequency estimates. As a general method of parametric filtering, it unifies and extends several existing procedures of frequency estimation in the literature.

Coupled with the AR filter that has an extra bandwidth parameter, the PF method is able to accommodate poor initial guesses of accuracy $O(1)$ and improve them with a simple iterative algorithm of linear least squares estimation plus linear recursive filtering to achieve the same accuracy of $O(n^{-3/2})$ as the computationally cumbersome procedure of nonlinear least squares. In the statistical analysis of the PF method, we

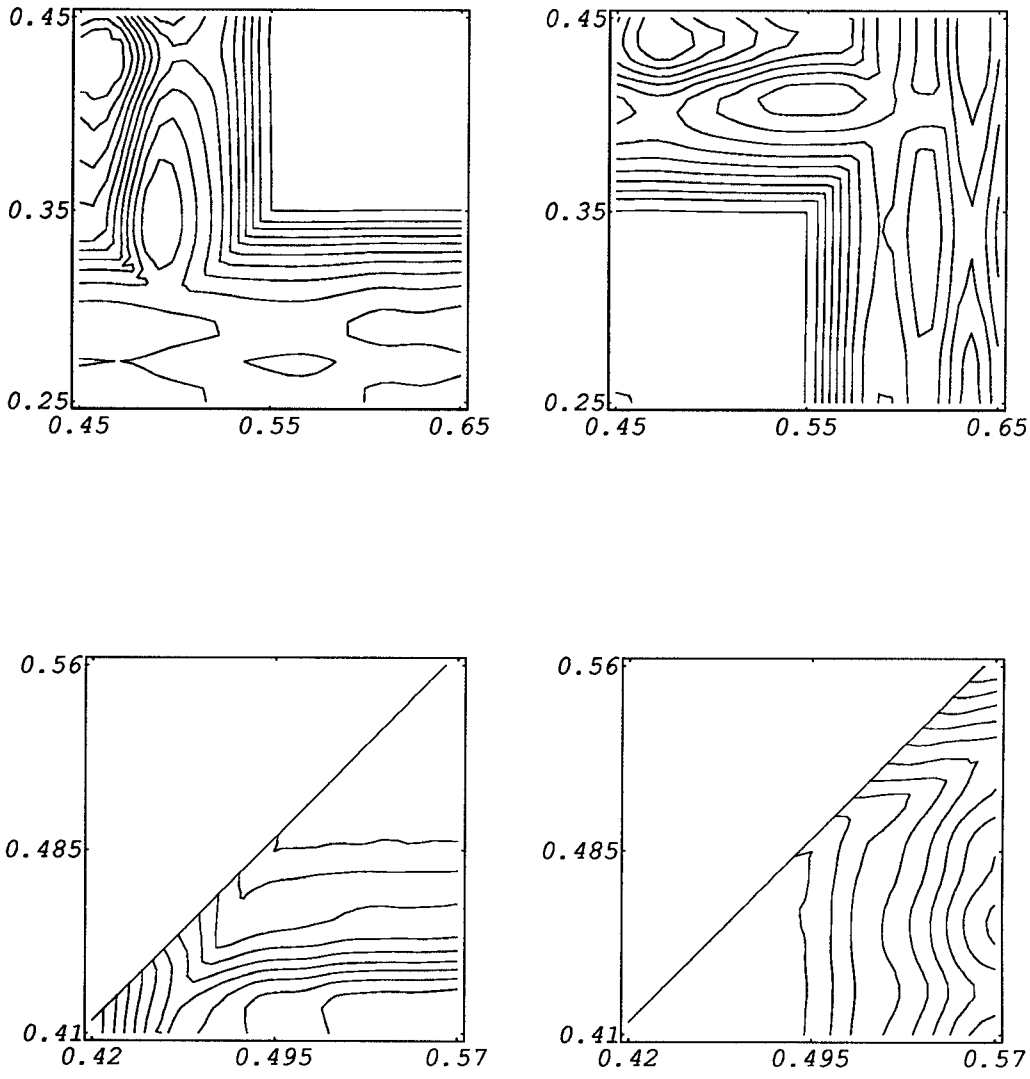


Figure 5.12: Contours of the two-dimensional least squares mapping $\hat{\psi}(\boldsymbol{f})$. First row: (a) Contour plot of $\hat{\psi}(\boldsymbol{f})$ in Figure 5.10 above the zero-plane; (b) Contour plot of $\hat{\psi}(\boldsymbol{f})$ in Figure 5.10 below the zero-plane. Second row: (c) Contour plot of $\hat{\psi}(\boldsymbol{f})$ in Figure 5.11 above the zero-plane with the diagonal line; and (d) Contour plot of $\hat{\psi}(\boldsymbol{f})$ in Figure 5.11 below the zero-plane with the diagonal line.

extended the classical results on the ergodicity and asymptotic normality of sample autocovariances, and proved the uniform strong consistency of sample autocovariances after parametric filtering and the asymptotic normality of sample autocovariances from the mixed-spectrum process $\{y_t\}$ of sinusoids in additive colored noise.

There are many ways of possible continuation and extension of the current work. As we have seen, the AR filter works very well in many cases as a parametric filter in the PF method. Although some other filters are available for single sinusoid estimation (e.g., Kedem, 1990; Lopes, 1991; Kedem and Yakowitz, 1992), it is still interesting to find other parametric filters that provide better results especially in the case of multiple sinusoids where the frequencies are relatively close to 0 or π.

The ability of the PF method to resolve closely-spaced frequencies is inherited from the AR modeling of the sinusoidal signal that allows the data to extrapolate beyond the observation interval without assuming them to be zero. It's ability of producing accurate estimates when the AR filter is employed is due to the noise-cleaning capability of AR filtering. These characteristics are also observed in many other methods of frequency estimation, especially those that employ the principal component analysis (e.g., Tufts and Kumaresam, 1982). For fthe uture research, some comparisons between these procedures are needed in order to understand their advantages and disadvantages, although some results are available in these regards (e.g., Kay, 1988).

Furthermore, since the PF method assumes that the number of sinusoids is known, it is therefore necessary to couple it with a procedure that estimates this number. Some eigenvalue-based procedures are available in the literature (e.g., Fuchs, 1988). Experience shows that the PF method is likely to yield multiple zeros in the AR polynomial when the number of sinusoids is less than the assumed value in the calculation of the PF estimator while the SNR is sufficiently high (see also Kay, 1984). This makes it possible to estimate the number of sinusoids q by a certain *goodness-of-fit test* based on the PF frequency estimates corresponding to a number of assumed values. For in-

stance, to test the hypothesis of $q = 1$ versus $q = 2$, we could first obtain the frequency estimates $\hat{\omega}$ and $(\hat{\omega}_1, \hat{\omega}_2)$ using the PF method with $q = 1$ and $q = 2$, respectively. If $\hat{\omega}_1$ and $\hat{\omega}_2$ are not significantly different, the hypothesis of $q = 1$ would be clearly preferred. Otherwise, we would calculate the error J'_n in (1.13), corresponding to $\hat{\omega}$ with $q = 1$ and $(\hat{\omega}_1, \hat{\omega}_2)$ with $q = 2$, respectively, and obtain $J'_n(1)$ and $J'_n(2)$. The hypothesis of $q = 2$ would be in favor if $J'_n(2)$ is significantly smaller that $J'_n(1)$. An alternative way is to compare the estimated amplitudes of the sinusoids with estimated frequencies and reject the hypothesis of $q = 2$ if one of the amplitudes is significantly small. Primary results along this line seems quite promising but rigorous statistical analysis is still in need.

It is also possible to combine the PF method with principal component analysis and obtain a hybrid procedure of frequency estimation. It may have been noticed that the AR estimation step in the PF method is based on the AR model of exact order $2q$. A high-order AR model is a clear alternative. In fact, we could first estimate the coefficients of a higher order AR model[5] by principal component analysis (see Chapter 2), and obtain the AR parameter $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ for the exact model from the $2q$ zeros of the (estimated) high-order AR polynomial which are in complex conjugate pairs and closest to the unit circle. The AR parameter could in turn be employed in the filtering step. This hybrid procedure cleans up the noise in both eigenvalue and frequency domains, and hopefully would provide better frequency estimates in low SNR cases.

For those who are familiar with adaptive filtering, it is easy to see that the PF method with the AR filter can be readily modified to obtain an adaptive (recursive) algorithm capable of tracking time-varying frequencies in noise. In fact, since the AR filter is already recursive in time, all we need is to employ the recursive least squares algorithm (e.g., Haykin, 1986) to update the least squares estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$. Details in this regard can be found in Li and Kedem (1989) and Dragošević, *et al.* (1982).

[5] Just a slightly higher than $2q$ in order not to increase the computational complexity too much.

It would also be interesting to find connections and possible extensions to other problems in related areas, such as the frequency estimation of damped sinusoids and the estimation of direction of arrival (DOA).

In theoretical aspects, rigorous proof is still needed for multiple sinusoids when the AR filter is used with $\eta = 1$. Especially, a careful study should be carried out to analyze the situation when the frequencies are closely-spaced with respect to n^{-1}. A recent work has been published by Hannan and Quinn (1989) investigating the nonlinear least squares method under this situation. Some more efforts should be paid in this direction to analyze the PF as well as other methods. For the PF method itself, a detailed analysis is needed in order to understand its behavior when $1 - \eta$ is comparable with n^{-1}. This rules out the possibility of using the traditional technique of stationary processes, as we employed in our statistical analysis, since the filter is now a f unction of the data length n.

Finally, as a general idea, parametric filtering has close relations with the filter-bank technique, multiresolution analysis, and wavelet transformation, all of which can be regarded as ways of extracting useful information from the filtered data obtained via parametric filtering. It is therefore not impossible to generalize the basic ideas behind the PF method to other estimation/detection problems, an example of which is the use of parametrized first-order sample autocorrelation function as a tool for discrimination and identification of different signals (Kedem and Li, 1989; 1992).

Bibliography

[1] Abatzoglou, T. J. (1985). A fast maximum likelihood algorithm for frequency estimation of a sinusoid based on Newton's Method. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 1, pp. 77–89.

[2] An, H., Chen, Z., and Hannan, E. J. (1983). The maximum of the periodogram. *J. Multivariate Anal.*, vol. 13, pp. 383–400.

[3] Bienvenu, G. and Kopp, L. (1983). Optimality of high resolution array processing using the eigensystem approach. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 5, pp. 1235–1247.

[4] Blackman, R. B. and Tukey, J. W. (1959). *The Measurement of Power Spetrca From the Point of View of Communications Engineering.* New York: Dover.

[5] Bresler, Y. and Macovski, A. (1986). Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089.

[6] Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods.* New York: Springer-Verlag.

[7] Cadzow, J. A. (1982). Spectral estimation: An overdetermined rational model equation approach. *Proc. IEEE*, vol. 70, no. 9, pp. 907–939.

[8] Cadzow, J. A. and Bronez, T. P. (1983). Time series identification: An annihilation filter approach. *Proc. of IEEE 1983 ASSP Spectrum Estimation Workshop II*, pp. 172–180.

[9] Chan, Y. T., Lavoie, J. M. M., and Plant, J. B. (1981). A parameter estimation approach to estimation of frequencies of sinusoids. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 2, pp. 214–219.

[10] Chan, Y. T. and Langford, R. P. (1982). Spectral estimation via the high-order equations. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 5, pp. 689–698.

[11] Chicharo, J. F. and Ng, T. S. (1990). Gradient-based adaptive IIR notch filtering for frequency estimation. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 769–777.

[12] Dragošević, M. V. and Stanković, S. S. (1989). A generalized least squares method for frequency estimation. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 805–819.

[13] Dragošević, M. V., Stanković, S. S., and Čarapić, M. (1982). An approach to recursive estimation of time-varying spectra. *Proc. ICASSP82*, vol. 3, pp. 2080–3083.

[14] Fuchs, J. (1988). Estimating the number of sinusoids in additive white noise. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 12, pp. 1846–1853.

[15] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series.* New York: Wiley.

[16] Hannan, E. J. (1970). *Multiple Time Series.* New York: Wiley.

[17] Hannan, E. J. (1973). The estimation of frequency. *J. Appl. Prob.*, vol. 10, pp. 510–519.

[18] Hannan, E. J. and Quinn, B. G. (1989). The resolution of closely adjacent spectral lines. *J. Time Series Analysis*, vol. 10, no. 1, pp. 13–31.

[19] Haykin, S. (1986). *Adaptive Filter Theory.* Englewood Cliffs, New Jersey: Prentice-Hall.

[20] He, S. and Kedem, B. (1989). Higher order crossings of an almost periodic random sequences in noise. *IEEE Trans. Inform. Theory*, vol. 35, no. 3, pp. 360–370.

[21] He, S. and Kedem, B. (1990). The zero-crossing rate of autoregressive processes and its link to unit roots. *J. Time Series Analysis*, vol. 11, pp. 201–213.

[22] Hildebrand, F. B. (1956). *Introduction to Numerical Analysis.* New York: McGraw-Hill, Chapter 9.

[23] Hua, Y. and Sarkar, T. K. (1988). Perturbation analysis of TK method for harmonic retrieval problems. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 228–240.

[24] Hua, Y. and Sarkar, T. K. (1990). Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824.

[25] Hua, Y. and Sarkar, T. K. (1991). On SVD for estimating generalized eigenvalues of singular matrix pencil in noise. *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 892–900.

[26] Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes.* New York: Academic.

[27] Kaveh, M. and Barabell, A. (1986). The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 2, pp. 331–341.

[28] Kay, S. M. (1984). Accurate frequency estimation at low signal-to-noise ratio. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 3, pp. 540–547.

[29] Kay, S. M. (1988). *Modern Spectral Estimation: Theory and Application*, Englewood Cliffs, New Jersey: Prentice-Hall.

[30] Kay, S. M. and Marple, S. L. (1981). Spectrum analysis – a modern perspective. *Proc. of IEEE*, vol. 69, no. 11, pp. 1380–1419.

[31] Kay, S. M. and Shaw, A. K. (1988). Frequency estimation by principal component AR spectral estimation method without eigendecomposition. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 95–101.

[32] Kedem, B. (1990). Contraction mappings in mixed spectrum estimation. Presented at Inst. for Mathematics and Its Applications, Univ. of Minnesoda, Minneapolis.

[33] Kedem, B. and Li, T. H. (1989). Higher order crossings from a parametric family of linear filters. Technical Repart TR-89-47, Dept. of Math., Univ. of Maryland, College Park.

[34] Kedem, B. and Li, T. H. (1992). Monotone gain, first-order autocorrelation, and zero-crossing rate. *Ann. Statist.*, vol. 19, no. 3, pp. 1672–1676.

[35] Kedem, B. and Yakowitz, S. (1992). Practical aspects of a fast algorithm for frequency detection. Revised.

[36] Kumaresan, R. and Feng, Y. (1991). FIR prefiltering improves Prony's method. *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 736–741.

[37] Kumaresan, R., Scharf, L. L., and Shaw, A. K. (1986). An algorithm for pole-zero modeling and spectral analysis. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 637–640.

[38] Kumaresan, R., Tufts, D. W., and Scharf, L. L. (1984). A Prony method for noisy data: choosing the signal components and selecting the order in exponential signal models. *Proc. IEEE*, vol. 72, no. 2, pp. 230–233.

[39] Kung, S. Y., Arun, K. S., and Bashkar Rao, D. V. (1983). State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem. *J. Opt. Soc. Amer.*, vol. 73, no. 12, pp. 1799–1811.

[40] Lacoss, R. T. (1971). Data adaptive spectral analysis method. *Geophysics*, vol. 36, no. 4, pp. 661–675.

[41] Lang, S. and McClellan, J. (1980). Frequency estimation with maximum entropy spectral estimator. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, pp. 716–724.

[42] Lehmann, E. L. (1983). *Theory of Point Estimation.* New York: Wiley.

[43] Li, T. H. (1991). On the estimation of sinusoidal signals in additive noise. Unpublished manuscript.

[44] Li, T. H. and Kedem, B. (1991). Adaptive frequency tracking by zero-crossing counts. Technical Report TR91-26, Dept. of Math., Univ. of Maryland, College Park.

[45] Li, T. H. and Kedem, B. (1992). Strong consistency of the contraction mapping method for frequency estimation. Technical Report TR92-22, Systems Research Center, Univ. of Maryland, College Park.

[46] Li, T. H., Kedem, B., and Yakowitz, S. (1992). Asymptotic normality of the contraction mapping estimator for frequency estimation. Technical Report TR92-21, Systems Research Center, Univ. of Maryland, College Park.

[47] Ljung, L. (1987). *System Identification — Theory for the User*, Englewood Cliffs, New Jersey: Prentice-Hall.

[48] Lopes, S. (1991). Spectral analysis in frequency modulated models. Ph.D. dissertation, Dept. of Math., Univ. of Maryland, College Park.

[49] Mackisack, M. S. and Poskitt, D. S. (1989). Autoregressive frequency estimation. *Biometrika*, vol. 76, no. 3, pp. 565–575.

[50] Mackisack, M. S. and Poskitt, D. S. (1990). Some properties of autoregressive estimates for processes with mixed spectra. *J. Time Series Analysis*, vol. 11, no. 4, pp. 325–337.

[51] Markushevich, A. I. (1977). *Theory of Functions of a Complex Variable.* 2nd Ed. New York: Chelsea.

[52] Mataušek, M. R., Stanković, S. S., and Radović, D. V. (1983). Iterative inverse filtering approach to the estimation of frequencies of noisy sinusoids," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 6, pp. 1456–1463.

[53] Ortega, J. M. and W. C. Rheinboldt (1970). *Iterative Solution of Nonlinear Equations in Several Variables.* New York: Academic.

[54] Paliwal, K. K. (1986). Some comments about the iterative filtering algorithm for spectral estimation of sinusoids. *Signal Process.*, vol. 10, pp. 307–310.

[55] Pisarenko, V. F. (1973). The retrieval of harmonics from a covariance function. *Geophys. J. Roy. Astronom. Soc.*, vol. 33, pp. 347–366.

[56] Porat, B. and Friedlander, B. (1988). Analysis of the asymptotic relative efficiency of the MUSIC algorithm. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 4, pp. 532–544.

[57] Priestley, M. B. (1981). *Spectral Analysis and Time Series*, vols. 1 and 2. New York: Academic.

[58] Quinn, B. G. and Fernandes, J. M. (1991). A fast efficient technique for the estimation of frequency. *Biometrika*, vol. 78, no. 3, pp. 489–497.

[59] Rao, B. D. (1988). Perturbation analysis of an SVD-based linear prediction methods for estimating the frequencies of multiple sinusoids. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 1026–1035.

[60] Rice, J. A. and Rosenblatt, M. (1988). On frequency estimation. *Biometrika*, vol. 75, no. 3, pp. 477–484.

[61] Rife, D. C., and Boorstyn, R. R. (1974). Single-tone parameter estimation from discrete-time observations. *IEEE Trans. Inform. Theory*, vol. 20, pp. 591–598.

[62] — (1976). Multiple tone parameter estimation from discrete-time observations. *Bell Syst. Technical J.*, vol. 55, pp. 1389–1410.

[63] Roy, R. and Kailath, T. (1989). ESPRIT – estimation of signal parameters via rotation invariance techniques. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7 pp. 984–995.

[64] Satorius, E. H. and Zeidler, J. R. (1978). Maximum entropy spectral analysis of multiple sinusoids in noise. *Geophysics*, vol. 43, no. 6, pp. 1111–1118.

[65] Stoer, J. and Bulirsch, R. (1980). *Introduction to Numerical Analysis.* New York: Springer-Verlag.

[66] Stoica, P., Friedlander, B., and Söderström, T. (1987). Asymptotic bias of the high-order autoregressive estimates of sinusoidal frequencies. *Circuits, Systems, Signal Processing*, vol. 6, no. 3, pp. 287-298.

[67] Stoica, P., Moses, R. L., Friedlander, B., and Söderström, T. (1989). Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392.

[68] Stoica, P. and Nehorai, A. (1989). Statistical analysis of two nonlinear least-squares estimators of sine wave parameters in the colored-noise case. *Circuits Systems Signal Process.*, vol. 8, no.1, pp. 3–15.

[69] Truong-Van, B. (1990). A new approach to frequency analysis with amplified harmonics. *J. R. Statist. Soc.*, series B, vol. 52, pp. 347–366.

[70] Tufts, D. W. and Kumaresan, R. (1982). Estimation of frequencies of multiple sinusoids: Making linear prediction performance like maximum likelihood. *Proc. IEEE*, vol. 70, no. 9, pp. 975–989.

[71] Ulrych, T. J. and Clayton, R. W. (1976). Time series modeling and maximum entropy. *Phys. Earth Planet. Interiors*, vol. 12, pp. 188–200.

[72] Walker, A. M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika*, vol. 58, no. 1, pp. 21–36.

[73] Walker, A. M. (1973). On the estimation of a harmonic component in a time series with stationary dependent residuals. *Adv. Appl. Prob.*, vol. 5, pp. 217–241.

[74] Whittle, P. (1952). The simultaneous estimation of time series' harmonic components and covariance structure. *Trab. Estad.*, vol. 3, pp. 43–57.

[75] Yakowitz, S. (1991). Some contributions to a frequency location method due to He and Kedem. *IEEE Trans. Inform. Theory*, vol. 37, pp. 1177–1181.

[76] Zoltawski, M. and Stavrinides, D. (1989). Sensor array signal processing via a Procrustes rotation based eigen-analysis of the ESPRIT data pencil. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 832–961.