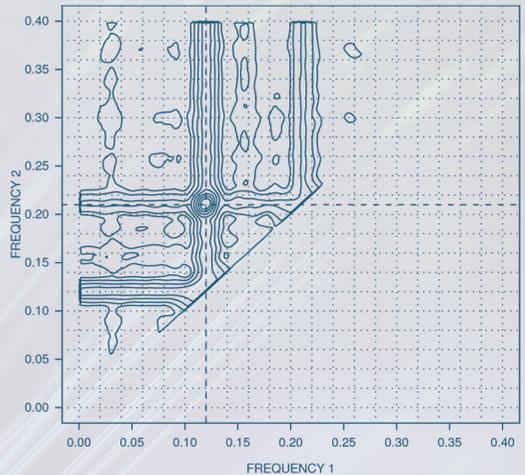
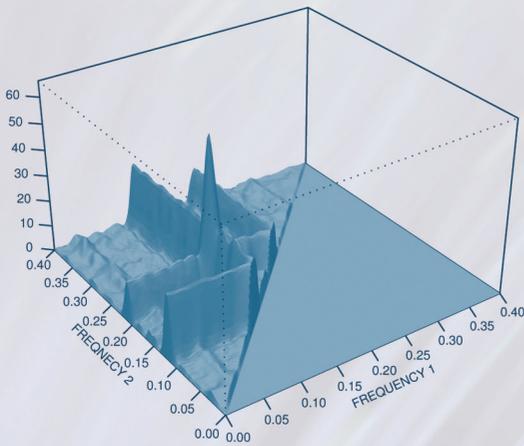


# TIME SERIES WITH MIXED SPECTRA



**Ta-Hsin Li**



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

# TIME SERIES WITH MIXED SPECTRA

**Ta-Hsin Li**



**CRC Press**

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business  
A CHAPMAN & HALL BOOK

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2014 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works  
Version Date: 20130531

International Standard Book Number-13: 978-1-4200-1006-0 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at  
<http://www.taylorandfrancis.com>

and the CRC Press Web site at  
<http://www.crcpress.com>

*To My Family*



---

# Contents

<b>Preface</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Periodicity and Sinusoidal Functions . . . . .	1
1.2 Sampling and Aliasing . . . . .	3
1.3 Time Series with Mixed Spectra . . . . .	4
1.4 Complex Time Series with Mixed Spectra . . . . .	7
<b>2 Basic Concepts</b>	<b>13</b>
2.1 Parameterization of Sinusoids . . . . .	13
2.2 Spectral Analysis of Stationary Processes . . . . .	18
2.3 Gaussian Processes and White Noise . . . . .	23
2.4 Linear Prediction Theory . . . . .	28
2.5 Asymptotic Statistical Theory . . . . .	31
<b>3 Cramér-Rao Lower Bound</b>	<b>37</b>
3.1 Cramér-Rao Inequality . . . . .	37
3.2 CRLB for Sinusoids in Gaussian Noise . . . . .	40
3.3 Asymptotic CRLB for Sinusoids in Gaussian Noise . . . . .	47
3.4 CRLB for Sinusoids in NonGaussian White Noise . . . . .	58
3.5 Proof of Theorems . . . . .	64
<b>4 Autocovariance Function</b>	<b>75</b>
4.1 Autocovariances and Autocorrelation Coefficients . . . . .	75
4.2 Consistency and Asymptotic Unbiasedness . . . . .	77
4.3 Covariances and Asymptotic Normality . . . . .	80
4.4 Autocovariances of Filtered Time Series . . . . .	90
4.5 Proof of Theorems . . . . .	91
<b>5 Linear Regression Analysis</b>	<b>111</b>
5.1 Least Squares Estimation . . . . .	111
5.2 Sensitivity to Frequency Offset . . . . .	120
5.3 Frequency Identification . . . . .	122
5.4 Frequency Selection . . . . .	126

5.5	Least Absolute Deviations Estimation	136
5.6	Proof of Theorems	147
<b>6</b>	<b>Fourier Analysis Approach</b>	<b>167</b>
6.1	Periodogram Analysis	167
6.2	Detection of Hidden Sinusoids	184
6.3	Extension of the Periodogram	194
6.3.1	Sinusoids with NonFourier Frequencies	194
6.3.2	Refined Periodogram	197
6.3.3	Secondary Analysis	201
6.3.4	Interpolation Estimators	204
6.4	Continuous Periodogram	208
6.4.1	Statistical Properties	208
6.4.2	Periodogram Maximization	210
6.4.3	Resolution Limit	223
6.5	Time-Frequency Analysis	225
6.6	Proof of Theorems	229
<b>7</b>	<b>Estimation of Noise Spectrum</b>	<b>253</b>
7.1	Estimation in the Absence of Sinusoids	253
7.1.1	Periodogram Smoother	254
7.1.2	Lag-Window Spectral Estimator	258
7.1.3	Autoregressive Spectral Estimator	261
7.1.4	Numerical Examples	271
7.2	Estimation in the Presence of Sinusoids	276
7.2.1	Modified Periodogram Smoother	277
7.2.2	M Spectral Estimator	279
7.2.3	Modified Autoregressive Spectral Estimator	283
7.2.4	A Comparative Example	287
7.3	Detection of Hidden Sinusoids in Colored Noise	290
7.4	Proof of Theorems	296
<b>8</b>	<b>Maximum Likelihood Approach</b>	<b>311</b>
8.1	Maximum Likelihood Estimation	311
8.2	Maximum Likelihood under Gaussian White Noise	313
8.2.1	Multivariate Periodogram	314
8.2.2	Statistical Properties	322
8.3	Maximum Likelihood under Laplace White Noise	331
8.3.1	Multivariate Laplace Periodogram	332
8.3.2	Statistical Properties	337
8.4	The Case of Gaussian Colored Noise	341
8.5	Determining the Number of Sinusoids	351

8.6	Proof of Theorems . . . . .	355
<b>9</b>	<b>Autoregressive Approach</b>	<b>375</b>
9.1	Linear Prediction Method . . . . .	375
9.1.1	Linear Prediction Estimators . . . . .	376
9.1.2	Statistical Properties . . . . .	385
9.2	Autoregressive Reparameterization . . . . .	392
9.3	Extended Yule-Walker Method . . . . .	396
9.3.1	Extended Yule-Walker Estimators . . . . .	397
9.3.2	Statistical Properties . . . . .	399
9.4	Iterative Filtering Method . . . . .	404
9.4.1	Iterative Filtering Estimator: Complex Case . . . . .	405
9.4.2	Iterative Filtering Estimator: Real Case . . . . .	413
9.5	Iterative Quasi Gaussian Maximum Likelihood Method . . . . .	424
9.5.1	Iterative Generalized Least-Squares Algorithm . . . . .	426
9.5.2	Iterative Least-Eigenvalue Algorithm . . . . .	436
9.5.3	Self Initialization . . . . .	442
9.6	Proof of Theorems . . . . .	444
<b>10</b>	<b>Covariance Analysis Approach</b>	<b>455</b>
10.1	Eigenvalue Decomposition of Covariance Matrix . . . . .	455
10.2	Principal Component Analysis Method . . . . .	458
10.2.1	Reduced Rank Autoregressive Estimators . . . . .	459
10.2.2	Statistical Properties . . . . .	472
10.3	Subspace Projection Method . . . . .	482
10.3.1	MUSIC and Minimum-Norm Estimators . . . . .	482
10.3.2	Statistical Properties . . . . .	489
10.4	Subspace Rotation Method . . . . .	492
10.4.1	Matrix-Pencil and ESPRIT Estimators . . . . .	492
10.4.2	Statistical Properties . . . . .	500
10.5	Estimating the Number of Sinusoids . . . . .	503
10.6	Sensitivity to Colored Noise . . . . .	510
10.7	Proof of Theorems . . . . .	513
<b>11</b>	<b>Further Topics</b>	<b>529</b>
11.1	Single Complex Sinusoid . . . . .	529
11.2	Tracking Time-Varying Frequencies . . . . .	534
11.3	Periodic Functions in Noise . . . . .	540
11.4	Beyond Single Time Series . . . . .	545
11.5	Quantile Periodogram . . . . .	552

<b>12 Appendix</b>	<b>567</b>
12.1 Trigonometric Series . . . . .	567
12.2 Probability Theory . . . . .	573
12.3 Numerical Analysis . . . . .	575
12.4 Matrix Theory . . . . .	577
12.5 Asymptotic Theory . . . . .	585
<b>Bibliography</b>	<b>611</b>
<b>Index</b>	<b>637</b>

---

# Preface

This book focuses on the methods and theory for statistical analysis of time series with mixed spectra. A time series is said to have a mixed spectrum if it comprises a finite number of sinusoids with different frequencies plus random noise. The research on such time series has a long history, and it remains active to this day, especially in the signal processing community where the interest is driven in part by the everlasting desire for fast algorithms to reduce the computational cost. Despite its importance, the subject often receives limited coverage in standard textbooks for understandable reasons. The objective of this book is to provide a more comprehensive and in-depth treatment of the subject. Needless to say, it is impossible to cover every aspect of the subject, not only because of the huge body of literature which keeps growing, but also due to the limited ability and capacity of the author. The topics in this book are selected to reflect what the author thinks are most interesting and relevant.

The intended audience of the book includes graduate students, researchers, engineers, and other professionals who work in the fields of time series analysis and signal processing. For the most part, the book only requires basic knowledge of probability, statistics, and time series analysis. Some theoretical results, especially their proofs, require more advanced knowledge of asymptotic analysis. For this reason, the proofs are deferred to the last section of each chapter in order not to interrupt the flow of intuitive interpretations which are more easily accessible to most readers. To serve the interests of a broader audience, the book deals with both real- and complex-valued time series. Except for the treatment of the proofs, the main style of the book is influenced by the highly successful textbook of Brockwell and David (1991) entitled *Time Series: Theory and Methods*. Other excellent textbooks that influenced the treatment of this book include *Spectral Analysis* by Priestley (1984), *Introduction to Fourier Analysis of Time Series* by Bloomfield (2000), *Spectral Analysis* by Stoica and Moses (1997), and *Modern Spectral Analysis* by Kay (1987).

This book is not merely a survey of the existing literature — it also includes original material which is either unavailable or cannot be easily found in the literature. For example, many results for complex-valued time series are derived in this book by extending the existing results developed in the literature for the real case; some theoretical results, such as the asymptotic theory for closely

spaced frequencies and the proof of the asymptotic normality of the nonlinear least-absolute-deviations frequency estimator, are not available in the literature. The book also contains the author's most recent work on the quantile regression method for spectral analysis.

The infancy of this book was a thesis proposal written in 1990 at the University of Maryland in College Park. The idea of writing a book based on the material came about in 1997 when I was on the faculty of the statistics department at Texas A&M University in College Station. It became a prolonged undertaking as priorities changed with my career moves from Texas A&M to the University of California at Santa Barbara and then to the IBM T. J. Watson Research Center at Yorktown Heights, New York. I am therefore very grateful to Bob Stern and his successor David Grubbs, editors of CRC Press, for their infinite patience.

Over the years with the project, I received the unwavering support from Dr. Benjamin Kedem and Dr. Emanuel Parzen, to whom I am deeply indebted. I am also grateful for the help and encouragement from Dr. H. Joseph Newton, Dr. Gerald R. North, Dr. Jerry D. Gibson, Dr. Keh-Shin Lii, Dr. David V. Hinkley, Dr. Emmanuel Yashchin, Dr. David A. Harville, and Dr. Yasuo Amemiya. My special thanks go to Dr. Kai-Sheng Song for his collaboration on several papers that closed some important gaps in the literature and enriched the content of this book. I would also like to express my gratitude to Dr. Hee-Seok Oh for many helpful comments and suggestions on an earlier draft of the book. Last but not least, I would like to thank Dr. Steve M. Kay for reviewing the proposal and the first draft of the book, and an anonymous reviewer for reviewing the last draft of the book with many thoughtful and constructive suggestions that have been incorporated into the final product.

Ta-Hsin Li  
Yorktown Heights, New York

# Chapter 1

---

## Introduction

A time series is a sequence of data points obtained over successive uniform time intervals. The word “time” can also be interpreted loosely to mean a time-like variable that provides the ordering and spacing for the data points. In this book, we are mainly interested in a particular type of time series, known as time series with mixed spectra, which can be expressed as a sum of sinusoids plus random noise. Time series of this type are abundant in a variety of science and engineering applications, including astronomy, biology, econometrics, geophysics, meteorology, rotating machinery, radar, sonar, and telecommunications. A main objective for spectral analysis of such time series is to detect and estimate the hidden periodicities represented by the sinusoidal functions.

---

### 1.1 Periodicity and Sinusoidal Functions

Periodicity is one of the most important and useful natural phenomena, and widely observable. The earth revolves periodically around its own axis and the sun, giving us different days and seasons. The displacement of a vibrating string or a swinging pendulum exhibits periodic patterns over time.

By definition, a periodic function repeats its values over intervals of a fixed length called the *period*. Sinusoidal, or trigonometric, functions, i.e., sines and cosines, are perfect examples of periodic functions. For any fixed constant  $f > 0$ , the sinusoidal function  $\sin(2\pi f t)$ , defined on the real line  $\mathbb{R} := (-\infty, \infty)$ , has a period  $T := 1/f$ , because

$$\sin(2\pi f(t + T)) = \sin(2\pi f t) \quad \forall t \in \mathbb{R}.$$

The parameter  $f$ , measured in cycles per unit time, is called the *frequency* of the sinusoidal function  $\sin(2\pi f t)$ . The parameter  $\omega := 2\pi f$ , measured in radians per unit time, is called the *angular frequency* of the sinusoid.

Not only the sinusoidal functions are periodic, they can also be superimposed to represent any periodic function. In fact, according to the theory of Fourier

series, any piecewise continuous function  $x(t)$  with period  $T$  can be expressed as a sum of sinusoids with frequencies  $f_k := k/T$  ( $k = 1, 2, \dots$ ), i.e.,

$$x(t) = A_0 + \sum_{k=1}^{\infty} \{A_k \cos(2\pi f_k t) + B_k \sin(2\pi f_k t)\}, \quad (1.1.1)$$

where

$$\begin{aligned} A_0 &:= \frac{1}{T} \int_0^T x(t) dt, \\ A_k &:= \frac{2}{T} \int_0^T x(t) \cos(2\pi f_k t) dt, \\ B_k &:= \frac{2}{T} \int_0^T x(t) \sin(2\pi f_k t) dt. \end{aligned}$$

The convergence of this infinite series takes place at every continuous point of  $x(t)$  for  $t \in \mathbb{R}$ . Observe that the sinusoidal functions in (1.1.1) are orthogonal to each other and to the constant function 1 (i.e., the cosine function with frequency zero) in the sense that

$$\begin{aligned} \int_0^T \cos(2\pi f_k t) \cos(2\pi f_{k'} t) dt &= 0 \quad \forall k \neq k', \\ \int_0^T \sin(2\pi f_k t) \sin(2\pi f_{k'} t) dt &= 0 \quad \forall k \neq k', \\ \int_0^T \cos(2\pi f_k t) \sin(2\pi f_{k'} t) dt &= 0 \quad \forall k, k', \end{aligned}$$

and

$$\int_0^T \cos(2\pi f_k t) dt = \int_0^T \sin(2\pi f_k t) dt = 0 \quad \forall k.$$

Therefore, the sinusoidal functions in (1.1.1), together with the constant function, form an orthogonal basis for  $T$ -periodic functions.

Although the sinusoidal representation (1.1.1) is an infinite sum in general, it can be approximated by a finite sum when  $x(t)$  is sufficiently smooth so that the coefficients decay rapidly as  $k$  grows. Given such an approximation,

$$\tilde{x}(t) := A_0 + \sum_{k=1}^K \{A_k \cos(2\pi f_k t) + B_k \sin(2\pi f_k t)\}, \quad (1.1.2)$$

the total squared error can be expressed as

$$\int_0^T |x(t) - \tilde{x}(t)|^2 dt = \sum_{k=K+1}^{\infty} \frac{1}{2} T (A_k^2 + B_k^2). \quad (1.1.3)$$

An important reason why the sinusoids are the preferred basis for representing periodic functions is that an approximation of the form (1.1.2) is time-invariant:

for any constant  $\tau$ , the function  $\tilde{x}(t + \tau)$  remains a finite sum of sinusoids with the same frequencies, and the error of  $\tilde{x}(t + \tau)$  for approximating  $x(t + \tau)$  is the same as the error of  $\tilde{x}(t)$  for approximating  $x(t)$ . In fact, it is easy to verify that

$$\tilde{x}(t + \tau) = A_0 + \sum_{k=1}^K \{A'_k \cos(2\pi f_k t) + B'_k \sin(2\pi f_k t)\},$$

where

$$\begin{aligned} A'_k &:= A_k \cos(2\pi f_k \tau) + B_k \sin(2\pi f_k \tau), \\ B'_k &:= -A_k \sin(2\pi f_k \tau) + B_k \cos(2\pi f_k \tau). \end{aligned}$$

A similar expression can be obtained for  $x(t + \tau)$ . Because  $A'_k{}^2 + B'_k{}^2 = A_k^2 + B_k^2$  for all  $k$ , the identity (1.1.3) remains true for the error  $\int_0^T |x(t + \tau) - \tilde{x}(t + \tau)|^2 dt$ .

## 1.2 Sampling and Aliasing

When continuous-time functions are observed only at discrete time instants, the problem of aliasing arises. This may lead to difficulties in interpreting the sinusoidal components of the resulting time series.

Consider the periodic function  $x(t)$  in (1.1.1), for example. If samples are taken with sampling interval  $\Delta$  at equally spaced time instants  $\Delta t$  for  $t \in \mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$ , where  $f_s := 1/\Delta$  is known as the *sampling rate* (measured in samples per unit time), then the resulting time series can be expressed as

$$x_t := x(\Delta t) = A_0 + \sum_{k=1}^{\infty} \{A_k \cos(2\pi f_k \Delta t) + B_k \sin(2\pi f_k \Delta t)\} \quad (t \in \mathbb{Z}). \quad (1.2.1)$$

For any  $f_k \in (f_s/2, f_s]$ , define  $f'_k := f_s - f_k \in [0, f_s/2)$ . The  $2\pi$ -periodicity of the sinusoidal functions implies that for all  $t \in \mathbb{Z}$ ,

$$\cos(2\pi f'_k \Delta t) = \cos(2\pi f_k \Delta t), \quad \sin(2\pi f'_k \Delta t) = -\sin(2\pi f_k \Delta t). \quad (1.2.2)$$

Therefore, the frequency  $f_k$  becomes indistinguishable from the frequency  $f'_k$  in the discrete-time representation (1.2.1). For this reason, the frequency  $f'_k$  is called an *alias* of  $f_k$ . Similarly, for any  $f_k > f_s$ , there exists an integer  $u$  such that  $\tilde{f}_k := f_k - u f_s \in [0, f_s)$ . Let  $f'_k := f_s - \tilde{f}_k$  if  $\tilde{f}_k \in (f_s/2, f_s)$ , and let  $f'_k := \tilde{f}_k$  otherwise. In the first case, (1.2.2) is true for all  $t \in \mathbb{Z}$ . In the second case, we have

$$\cos(2\pi f'_k \Delta t) = \cos(2\pi f_k \Delta t), \quad \sin(2\pi f'_k \Delta t) = \sin(2\pi f_k \Delta t).$$

In both cases, the frequency  $f'_k$  is an alias of  $f_k$ . As we can see, the aliasing effect in the time series makes it impossible to correctly identify the original frequencies which are greater than  $f_s/2$ .

If the function  $x(t)$  does not contain sinusoidal components whose frequencies are greater than a known constant  $f_c > 0$ , then an alias-free time series can be obtained by setting the sampling rate  $f_s$  higher than  $2f_c$  so that  $f_c < f_s/2$ . The lower bound  $2f_c$  is known as the Nyquist rate for alias-free sampling, named after the American physicist and electrical engineer Harry Nyquist (1889–1976).

In many applications, the periodic functions of interest are smooth enough to be well approximated by a finite sum of the form (1.1.2) with sufficiently large  $K$  such that the approximation error is comparable to the noise in the measurements. In this case, the aliasing problem can be ignored in the time series obtained with a sampling rate higher than the effective Nyquist rate  $2f_K = 2K/T$ . In other applications, such as radar and telecommunications, where the continuous-time signals are available, the aliasing problem can be mitigated by filtering the continuous-time signals, before sampling at rate  $f_s$ , with an analog device to remove the frequency content higher than  $f_c := f_s/2$ , which is known as the Nyquist frequency. Of course, the sampling rate  $f_s$  needs to be sufficiently high in order to minimize the distortion introduced by anti-alias filtering.

Instead of taking the instantaneous values, which leads to (1.2.1), one can also take the average values of a continuous-time function over regular intervals of length  $\Delta$ . For the periodic function  $x(t)$  in (1.1.1), this sampling technique leads to a time series that can be expressed as

$$x_t := \frac{1}{\Delta} \int_{\Delta(t-1/2)}^{\Delta(t+1/2)} x(s) ds = A_0 + \sum_{k=1}^{\infty} \{\tilde{A}_k \cos(2\pi f_k \Delta t) + \tilde{B}_k \sin(2\pi f_k \Delta t)\},$$

where  $\tilde{A}_k := \text{sinc}(2\pi f_k \Delta/2) A_k$ ,  $\tilde{B}_k := \text{sinc}(2\pi f_k \Delta/2) B_k$ , and  $\text{sinc}(t) := \sin(t)/t$ . As we can see, the time series retains the original form of  $x(t)$  in (1.1.1) as a sum of sinusoids. This is another advantage of the sinusoidal representation of periodic functions. Because the time series takes the same form as that in (1.2.1), the aliasing problem and the alias-free sampling condition remain the same.

### 1.3 Time Series with Mixed Spectra

The sinusoidal representation (1.1.1) can be generalized to include functions that can be expressed as a sum of periodic functions with different periods. If each periodic component is a finite sum of sinusoids, then, with a sufficiently

high sampling rate, the resulting time series can be expressed as

$$y_t = A_0 + \sum_{k=1}^q \{A_k \cos(2\pi f_k t) + B_k \sin(2\pi f_k t)\} + \epsilon_t \quad (t \in \mathbb{Z}),$$

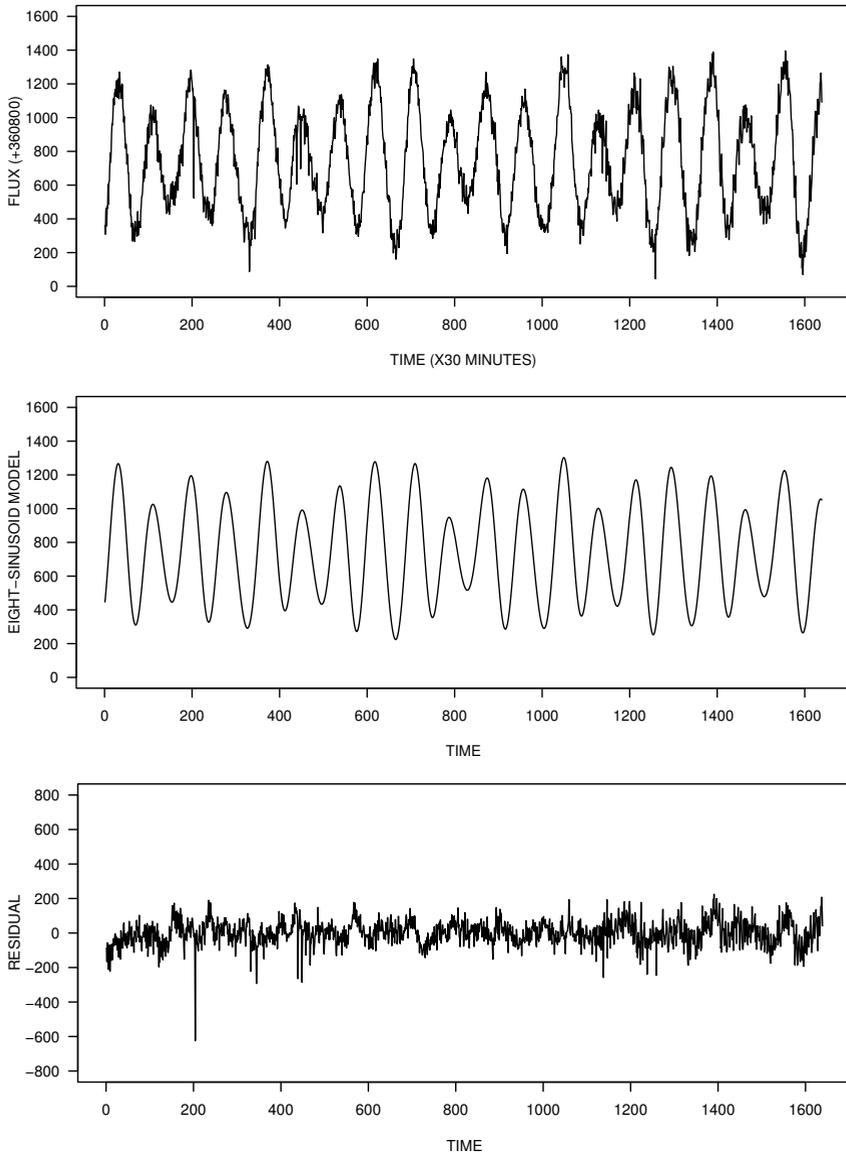
where  $\{\epsilon_t\}$  is a random process representing the noise in the observations and  $\{f_k\}$  is a set of normalized frequencies in  $(0, 1/2)$  which may not be harmonically related as they are in (1.2.1). A time series of this form has a mixed-type spectrum, because the frequency content of the sinusoids concentrates on discrete values in the interval  $(0, 1/2)$  with infinite density, whereas the frequency content of the noise spreads over the interval with finite density.

Figure 1.1 shows an example of the sinusoid-plus-noise model for a real-world time series. The time series, shown in the top panel, is known as the light curve. It comprises 1,639 brightness measurements of a variable star over a period of about 34 days. It is part of a large data set produced by the Kepler Mission\* of the National Aeronautics and Space Administration (NASA). In the Kepler Mission, a spacecraft carrying a simple aperture photometer (SAP) was launched in March 2009 to monitor the brightness of stars in the Milky Way galaxy. The photometer comprises an array of charge-coupled devices (CCDs), which convert light into electrical signals, and measures the average flux of electrons per second over 30-minute intervals (known as long cadence). The raw flux data are corrected for systematic and other errors by a procedure called presearch data conditioning (PDC) [398]. The measurements shown in the top panel of Figure 1.1 are the corrected flux for a variable star with Kepler ID 8073767 (quarter 1).†

As we can see, the time series exhibits a strong sinusoid-like periodic pattern which repeats approximately 19.5 times. A rough estimate for the periodicity is therefore  $1,639/19.5 \approx 84.1$  samples, or  $84.1/2 = 42.05$  hours. However, a single sinusoid would not be able to capture the large variations of the peak values. The middle panel of Figure 1.1 depicts a model which employs eight sinusoids plus a constant. The frequencies in this model are estimated by a periodogram maximization technique to be discussed later in the book (Chapter 6). The revised estimate for the dominating periodicity is 84.7 samples or  $84.7/2 = 42.35$  hours. The coefficients of the resulting sinusoids are estimated by least-squares regression together with the constant term (Chapter 5). As we can see, the model is able to capture the main oscillatory patterns of the time series, leaving just small random-looking variation in the residuals (bottom panel). A further analysis of the residual time series will be given in Chapter 7.

\* <http://kepler.nasa.gov/>.

† Available at <http://archive.stsci.edu/kepler/> and <http://exoplanetarchive.ipac.caltech.edu/>.



**Figure 1.1.** Brightness variation of a variable star obtained from NASA's Kepler Mission. Top to bottom: light curve of the variable star observed over 30-minute intervals, an eight-sinusoid model for the time series, and the residuals.

## 1.4 Complex Time Series with Mixed Spectra

A fundamental and useful property of sinusoidal functions is their invariance to linear time-invariant (LTI) systems. An LTI system is a convolution operator which transforms the input signal  $x(t)$  into  $y(t)$  given by

$$y(t) := \int_{-\infty}^{\infty} h(\tau)x(t - \tau) d\tau,$$

where  $h(t)$  is called the impulse response of the system. If  $x(t)$  is a sinusoidal function of the form  $x(t) = A\sin(2\pi ft) + B\cos(2\pi ft)$ , then

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)\{A\cos(2\pi f(t - \tau)) + B\sin(2\pi f(t - \tau))\} d\tau \\ &= A'(f)\cos(2\pi ft) + B'(f)\sin(2\pi ft), \end{aligned}$$

where

$$\begin{aligned} A'(f) &:= \int_{-\infty}^{\infty} h(\tau)\{A\cos(2\pi f\tau) - B\sin(2\pi f\tau)\} d\tau, \\ B'(f) &:= \int_{-\infty}^{\infty} h(\tau)\{A\sin(2\pi f\tau) + B\cos(2\pi f\tau)\} d\tau. \end{aligned}$$

In other words, the output  $y(t)$  remains to be a sinusoidal function with the same frequency  $f$  and the LTI system only changes the coefficients of the sinusoid from  $A$  and  $B$  to  $A'(f)$  and  $B'(f)$ . Note that the linearity of an LTI system also implies that if the input  $x(t)$  is a sum of sinusoids, the output  $y(t)$  remains to be a sum of sinusoids with unaltered frequencies.

By taking advantage of the invariance to LTI systems, sinusoidal waves, in electrical or acoustical forms, are widely used for transmission of signals in applications such as radar, sonar, biomedical imaging, and telecommunications. Complex sinusoidal functions, which, by Euler's formula, take the form

$$\exp(i\omega t) = \cos(\omega t) + i\sin(\omega t)$$

with  $i := \sqrt{-1}$ , appear naturally in these applications. Complex sinusoidal functions are also known as complex exponentials.

In radar [416], for example, sinusoidal signals of the form

$$A\cos(2\pi f_0 t) + B\sin(2\pi f_0 t)$$

are transmitted as electromagnetic waves which propagate at the speed of light  $c$ . When hitting a target, the wave is reflected with the Doppler effect that shifts its frequency. The echo, or return, received by the radar takes the form

$$r(t) := A'\cos(2\pi(f_0 - f)t) + B'\sin(2\pi(f_0 - f)t), \quad (1.4.1)$$

where  $f := 2v f_0/c$  is the Doppler shift caused by the motion of the target. The parameter  $v$  denotes the radial velocity of the target relative to the radar, which is positive when the target moves toward the radar and negative when it moves away from the radar. Although the Doppler shift is unknown, an upper bound, denoted by  $b > 0$ , can be determined *a priori* for its absolute value. The carrier frequency  $f_0$  is much greater than  $b$ , so the frequency in  $r(t)$  resides in a narrow band between  $f_0 - b$  and  $f_0 + b$ . If  $r(t)$  were sampled at its Nyquist rate  $2(f_0 + b)$ , which is very large, it would produce an enormous amount of unnecessary data. A more efficient way is to bring the frequency in  $r(t)$  down to the neighborhood of zero and sample the resulting baseband function at a much lower rate. This process is known as frequency demodulation.

More precisely, consider the function

$$r_I(t) := \cos(2\pi f_0 t) r(t).$$

It follows from (1.4.1) that

$$\begin{aligned} r_I(t) &= \frac{1}{2}\{A' \cos(2\pi f t) - B' \sin(2\pi f t)\} \\ &\quad + \frac{1}{2}\{A' \cos(2\pi(2f_0 - f) t) + B' \sin(2\pi(2f_0 - f) t)\}. \end{aligned}$$

Observe that the frequency of the second term in  $r_I(t)$ , which equals  $2f_0 - f$ , is far above the frequency of the first term, which equals  $f$  (assume  $f > 0$  for simplicity of discussion). Therefore, the second term can be removed by an analog lowpass filter with impulse response  $h(t)$  and cutoff frequency

$$f_c \in (b, 2f_0 - b).$$

This produces the so-called in-phase signal

$$x_I(t) := \frac{1}{2}\{A'_I \cos(2\pi f t) + B'_I \sin(2\pi f t)\},$$

where

$$\begin{aligned} A'_I &:= \int_{-\infty}^{\infty} h(\tau) \{A' \cos(2\pi f \tau) + B' \sin(2\pi f \tau)\} d\tau, \\ B'_I &:= \int_{-\infty}^{\infty} h(\tau) \{A' \sin(2\pi f \tau) - B' \cos(2\pi f \tau)\} d\tau. \end{aligned}$$

Because  $x_I(t)$  does not have frequency content higher than  $f_c$ , an alias-free time series can be obtained by sampling  $x_I(t)$  at a rate  $f_s \geq 2f_c$ . With instantaneous sampling, the time series takes the form

$$x_I(\Delta t) = \frac{1}{2}\{A'_I \cos(2\pi f \Delta t) + B'_I \sin(2\pi f \Delta t)\} \quad (t \in \mathbb{Z}),$$

where  $\Delta := 1/f_s$ . Similarly, applying the lowpass filter to the function

$$r_Q(t) := \sin(2\pi f_0 t) r(t)$$

followed by instantaneous sampling yields the so-called quadrature signal

$$x_Q(\Delta t) := \frac{1}{2}\{A'_Q \cos(2\pi f \Delta t) + B'_Q \sin(2\pi f \Delta t)\} \quad (t \in \mathbb{Z}),$$

where

$$\begin{aligned} A'_Q &:= \int_{-\infty}^{\infty} h(\tau)\{B' \cos(2\pi f \tau) - A' \sin(2\pi f \tau)\} d\tau, \\ B'_Q &:= \int_{-\infty}^{\infty} h(\tau)\{B' \sin(2\pi f \tau) + A' \cos(2\pi f \tau)\} d\tau. \end{aligned}$$

Combining the in-phase and quadrature signals as the real and imaginary parts of a complex-valued signal leads to the complex sinusoid

$$x_t := x_I(\Delta t) + ix_Q(\Delta t) = \beta \exp(i\omega t) \quad (t \in \mathbb{Z}), \quad (1.4.2)$$

where  $\omega := 2\pi f \Delta \in (0, \pi)$  and  $\beta := \frac{1}{2}(A' + iB') \int h(\tau) \exp(-i\omega \tau) d\tau$ .

The same argument can be used to show that (1.4.2) remains valid when  $f$  is negative because  $\cos(2\pi f t) = \cos(2\pi|f|t)$  and  $\sin(2\pi f t) = -\sin(2\pi|f|t)$ . Therefore, for complex sinusoids, the frequency can be negative. However, due to the  $2\pi$ -periodicity, a negative frequency  $\omega$  in the interval  $(-\pi, 0)$  is an alias of the positive frequency  $\omega' := 2\pi - \omega$  in the interval  $(\pi, 2\pi)$ . Therefore, the frequencies of complex sinusoids can also be restricted to the interval  $(0, 2\pi)$ .

The model (1.4.2) can be extended to include multiple complex sinusoids and the noise. The general model takes the form

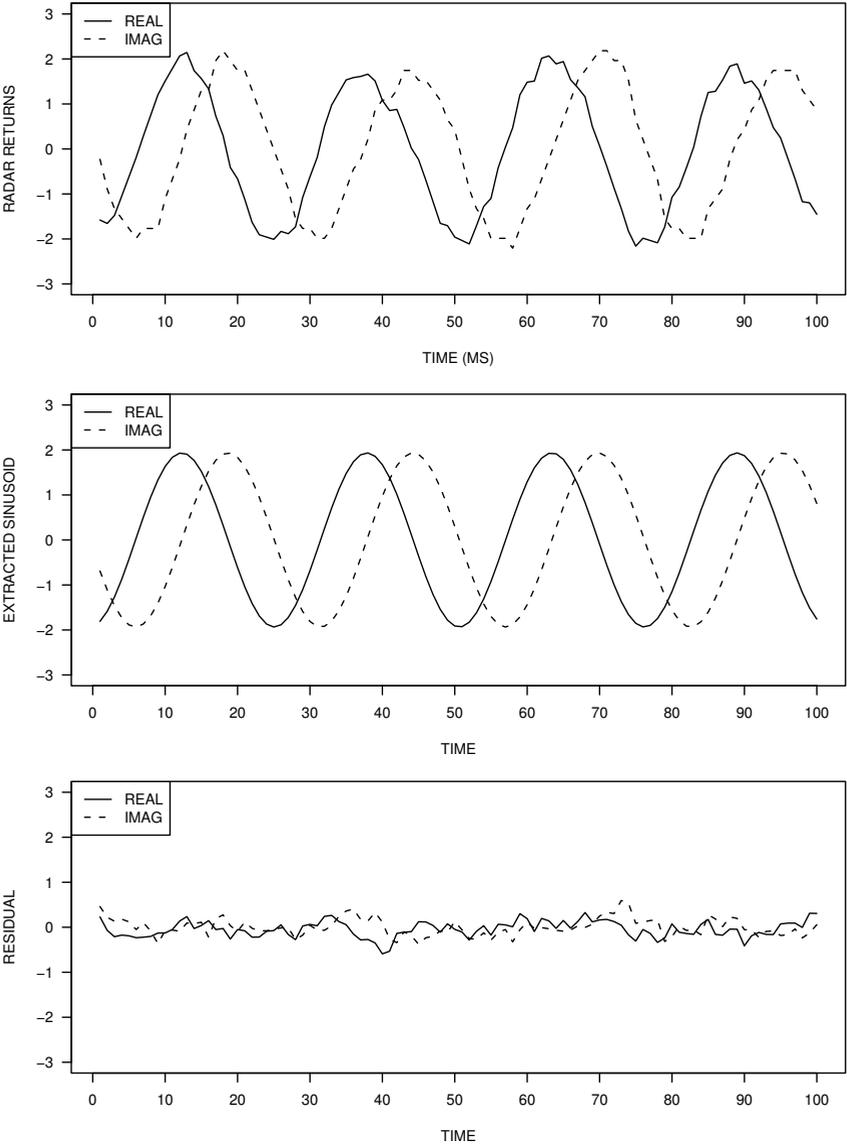
$$y_t = \sum_{k=1}^p \beta_k \exp(i\omega_k t) + \epsilon_t \quad (t \in \mathbb{Z}).$$

In radar applications, this model represents superimposed returns from multiple targets (or scatterers), each moving at a different speed relative to the radar.

Let us consider an example with real-world data. Figure 1.2 shows a 100-sample segment of a radar signal together with the complex sinusoid extracted from it and the residuals. The radar signal is part of a large data set collected in November 1993 by a team of researchers at McMaster University using a high resolution radar overlooking the Atlantic Ocean from a cliff top near Dartmouth, Nova Scotia, Canada.<sup>‡</sup> The target is a spherical block of styrofoam, one meter in diameter and wrapped with wire mesh. The radar signal<sup>§</sup> represents the demodulated returns in the 2,685-meter range bin, sampled at the rate  $f_s = 1,000$  Hertz (hence  $\Delta := 1/f_s = 1$  millisecond).

<sup>‡</sup><http://soma.ece.mcmaster.ca/ipix>.

<sup>§</sup>Available at <http://soma.ece.mcmaster.ca/ipix/dartmouth/datasets.html> as Clutter + Target Data File #283 (range bin 10, vertical polarization).



**Figure 1.2.** Radar returns from a target. Top to bottom: real (solid line) and imaginary (dashed line) parts of the radar data, the extracted complex sinusoid, and the residuals. Time is in milliseconds.

Processing the entire signal requires more sophisticated techniques due to its time-varying nature (see Chapter 6 for details). Figure 1.2 simply shows that for a sufficiently short segment (100 milliseconds in this case) the signal can be well represented by a single complex sinusoid plus random noise.

The frequency of the complex sinusoid is estimated by the periodogram maximization technique to be discussed in Chapter 6, and the complex amplitude is obtained by least-squares regression to be discussed in Chapter 5. The estimated frequency is equal to 0.03912442. It is an estimate for the normalized frequency  $f_n := \Delta f = f/f_s$ . To get an estimate for the velocity of the target, it suffices to observe that  $f = 2vf_0/c$ , where  $f_0 = 9.39 \times 10^9$  Hertz is the carrier frequency of the radar and  $c = 3 \times 10^8$  meters per second is the speed of light. Therefore, the relative velocity of the target is given by  $v = f_n \times f_s c / (2f_0)$ , or approximately 0.62 meters per second toward the radar.



# Chapter 2

---

## Basic Concepts

---

This chapter discusses two types of parameterization for real and complex sinusoids together with some basic assumptions. It also reviews some basic concepts of random processes, linear prediction theory, and asymptotic statistical theory. Most of these results can be found easily in the literature. Therefore, they are stated without proof but with reference to standard textbooks.

---

### 2.1 Parameterization of Sinusoids

As shown in Chapter 1, there are two types of models for sinusoidal functions in practice: the real sinusoid model (RSM) and the complex sinusoid model (CSM). Both models can be further parameterized in two different forms: the Cartesian (or rectangular) form and the polar form.

Consider the real case first. The Cartesian RSM can be expressed as

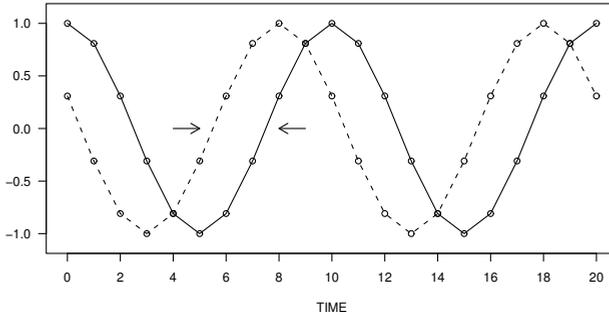
$$x_t = \sum_{k=1}^q \{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\} \quad (t \in \mathbb{Z}), \quad (2.1.1)$$

where the  $\omega_k \in (0, \pi)$  are the frequency parameters, the  $A_k \in \mathbb{R}$  and  $B_k \in \mathbb{R}$  are the amplitude parameters (or coefficients) satisfying  $A_k^2 + B_k^2 > 0$  ( $k = 1, \dots, q$ ). The polar RSM takes the form

$$x_t = \sum_{k=1}^q C_k \cos(\omega_k t + \phi_k) \quad (t \in \mathbb{Z}), \quad (2.1.2)$$

where the  $C_k > 0$  are the amplitude parameters and the  $\phi_k \in (-\pi, \pi]$  are the phase parameters. Note that in both (2.1.1) and (2.1.2) we exclude a possible constant term  $A_0$  for convenience as it can be easily estimated and removed in practice.

While the  $C_k$  in (2.1.2) can be easily interpreted as representing the strength of the sinusoids, the phase parameters  $\phi_k$  represent the advance (if positive) or delay (if negative) of the sinusoidal waves relative to their zero-phase counterparts. This is illustrated in Figure 2.1. In this example, the sinusoidal functions



**Figure 2.1.** Two sinusoids with common frequency  $\omega_0 = 2\pi \times 0.1$  and different phases. Solid line,  $\phi_1 = 0$ ; dashed line,  $\phi_2 = 2\pi \times 0.2$ . The second sinusoid leads the first sinusoid by  $(\phi_2 - \phi_1)/\omega_0 = 2$  samples.

are  $x_1(t) := \cos(\omega_0 t + \phi_1)$  and  $x_2(t) := \cos(\omega_0 t + \phi_2)$ , where  $\omega_0 = 2\pi \times 0.1$ ,  $\phi_1 = 0$ , and  $\phi_2 = 2\pi \times 0.2$ . As we can see, the first sinusoid (solid line) is just a shifted copy of the second sinusoid (dashed line). At  $t = 0$ , the first sinusoid takes the value  $\cos(\phi_1) = 1$ . The second sinusoid takes the same value at  $t = (\phi_1 - \phi_2)/\omega_0 = -2$ . Hence the second sinusoid leads the first sinusoid by 2 time units.

It is straightforward to verify that the Cartesian and polar forms of the RSM are equivalent under the following parameter transformation:

$$\begin{cases} A_k = C_k \cos(\phi_k), & B_k = -C_k \sin(\phi_k), \\ C_k = \sqrt{A_k^2 + B_k^2}, & \phi_k = \arctan(-B_k, A_k), \end{cases} \quad (2.1.3)$$

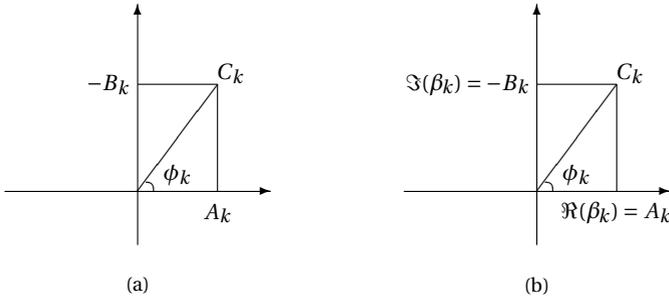
where

$$\arctan(B, A) := \begin{cases} \arctan(B/A) & \text{if } A > 0, \\ \arctan(B/A) + \pi & \text{if } A < 0, B \geq 0, \\ \arctan(B/A) - \pi & \text{if } A < 0, B < 0, \\ \pi/2 & \text{if } A = 0, B > 0, \\ -\pi/2 & \text{if } A = 0, B < 0, \\ 0 & \text{if } A = B = 0. \end{cases}$$

This relationship is illustrated in Figure 2.2(a). The equivalence explains why a linear combination of  $\cos(\omega t)$  and  $\sin(\omega t)$  is considered as a single real sinusoid with frequency  $\omega$  rather than two sinusoids. For this reason, we say that  $\omega_k$ ,  $A_k$ , and  $B_k$  in (2.1.1) are the parameters of the  $k$ th sinusoid.

Without loss of generality, we always assume that the frequencies  $\omega_k$  are arranged in an ascending order such that

$$0 < \omega_1 < \cdots < \omega_q < \pi.$$



**Figure 2.2.** Illustration of Cartesian and polar parameters for real and complex sinusoids. (a) Real case. (b) Complex case.

We also exclude the possibility for the frequency parameter to take value zero because it leads to a constant term which is typically removed (by subtracting the sample mean) prior to spectral analysis. Moreover, instead of the angular frequencies  $\omega_k$  (measured in radians per unit time), it is often more meaningful to consider the normalized frequencies

$$f_k := \omega_k / (2\pi), \tag{2.1.4}$$

which are measured in cycles per unit time. For example, if the time  $t$  is measured in seconds, then the normalized frequency  $f_k$  is measured in cycles per second, or Hertz; if  $t$  is in years, then  $f_k$  is in cycles per year. So physical interpretation becomes easier using the normalized frequencies. The angular frequencies are more convenient for mathematical manipulation.

In the complex case, the Cartesian CSM takes the form

$$x_t = \sum_{k=1}^p \beta_k \exp(i\omega_k t) \quad (t \in \mathbb{Z}), \tag{2.1.5}$$

where  $\omega_k \in (-\pi, \pi] \setminus \{0\}$  is the frequency of the  $k$ th sinusoid and  $\beta_k \in \mathbb{C}$  is its complex amplitude. Let  $\beta_k$  be further parameterized by real-valued parameters  $A_k \in \mathbb{R}$  and  $B_k \in \mathbb{R}$  such that

$$\begin{cases} \beta_k := A_k - iB_k, \\ A_k := \Re(\beta_k), \quad B_k := -\Im(\beta_k). \end{cases} \tag{2.1.6}$$

Then, the polar CSM can be expressed as

$$x_t = \sum_{k=1}^p C_k \exp\{i(\omega_k t + \phi_k)\} \quad (t \in \mathbb{Z}). \tag{2.1.7}$$

where  $C_k > 0$  and  $\phi_k \in (-\pi, \pi]$  are related to  $A_k$  and  $B_k$  through the transformation (2.1.3). Note that  $B_k$  is defined as  $-\Im(\beta_k)$  rather than  $\Im(\beta_k)$  in order for

the transformation (2.1.3) to be valid for both RSM and CSM. Figure 2.2(b) illustrates this relationship for the CSM. The polar parameters  $C_k$  and  $\phi_k$  can also be expressed in terms of the complex amplitude  $\beta_k$  such that

$$C_k = |\beta_k|, \quad \phi_k = \angle \beta_k.$$

Without loss of generality, the frequencies  $\omega_k$  are assumed to satisfy

$$-\pi < \omega_1 < \cdots < \omega_p < \pi.$$

The frequencies can also be redefined by the aliasing transformation

$$\omega_k \mapsto \begin{cases} \omega_k + 2\pi & \text{if } \omega_k < 0, \\ \omega_k & \text{otherwise,} \end{cases}$$

so that they take values in the interval  $(0, 2\pi)$ . In this case, it will be assumed that

$$0 < \omega_1 < \cdots < \omega_p < 2\pi.$$

As with the RSM, it is often more meaningful to consider the normalized frequencies  $f_k$  defined by (2.1.4) instead of the angular frequencies  $\omega_k = 2\pi f_k$ .

A real sinusoid can be expressed as the sum of two conjugate complex sinusoids. Indeed, for any  $\omega \in (0, \pi)$ ,  $C > 0$ , and  $\phi \in (-\pi, \pi]$ , let

$$\beta := \frac{1}{2} C \exp(i\phi).$$

Then,

$$\begin{aligned} C \cos(\omega t + \phi) &= \beta \exp(i\omega t) + \beta^* \exp(-i\omega t) \\ &= \beta \exp(i\omega t) + \beta^* \exp(i(2\pi - \omega)t). \end{aligned}$$

For this reason, the RSM can be written in the form of (2.1.5) with  $p := 2q$  and

$$\begin{cases} \omega_{p-k+1} := -\omega_k \text{ or } 2\pi - \omega_k \\ \beta_{p-k+1} := \beta_k^* := \frac{1}{2} C_k \exp(-i\phi_k) \end{cases} \quad (k = 1, \dots, q), \quad (2.1.8)$$

where  $C_k$  and  $\phi_k$  are given by (2.1.2). The RSM can also be written in the form of (2.1.7) with  $p := 2q$  and

$$\begin{cases} \omega_{p-k+1} := -\omega_k \text{ or } 2\pi - \omega_k \\ \phi_{p-k+1} := -\phi_k = -\arctan(-B_k, A_k) \\ C_{p-k+1} := C_k := \frac{1}{2} \sqrt{A_k^2 + B_k^2} \end{cases} \quad (k = 1, \dots, q), \quad (2.1.9)$$

where  $A_k$  and  $B_k$  are given by (2.1.1). With this transformation, results derived for the CSM can be translated into the corresponding results for the RSM.

It is worth pointing out that in the RSM (2.1.1)–(2.1.2) the power of the  $k$ th sinusoid is equal to  $\frac{1}{2}C_k^2$  asymptotically for large sample sizes, because

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n |C_k \cos(\omega_k t + \phi_k)|^2 = \frac{1}{2}C_k^2 = \frac{1}{2}(A_k^2 + B_k^2).$$

In the CSM (2.1.5)–(2.1.7), the power of the  $k$ th sinusoid is equal to  $C_k^2$  exactly for all sample sizes, because

$$n^{-1} \sum_{t=1}^n |C_k \exp\{i(\omega_k t + \phi_k)\}|^2 = C_k^2 = |\beta_k|^2 = A_k^2 + B_k^2.$$

This distinction should be kept in mind when comparing the results derived for the RSM with those derived for the CSM in later chapters.

While both RSM and CSM are discussed in this book, not all methods and analyses are presented for both models. A primary reason is that the original sources in the literature focus on different models. In many cases, one can infer the results for one model from the results for the other. In some cases, however, the inference is not straightforward or even not valid (for example, there is no RSM counterpart for the CSM with  $p = 1$ ).

Frequency separation is an important concept in dealing with sinusoids. If two frequencies are not sufficiently separated, the corresponding sinusoids cannot be easily distinguished from each other based on a finite data record. In the complex case, the separation of frequencies can be measured by

$$\Delta := \min_{k \neq k'} \{|\omega_k - \omega_{k'}|, 2\pi - |\omega_k - \omega_{k'}|\},$$

where the second term is necessary because of the aliasing. In the real case, the separation of frequencies is measured by

$$\Delta := \min_{k \neq k'} \{|\omega_k - \omega_{k'}|, \omega_k, \pi - \omega_k\},$$

where the last two terms are required to take the conjugate frequencies into account (a frequency  $\omega_k$  near 0 or  $\pi$  is also near its conjugate frequency  $-\omega_k$  in the CSM representation).

Once the sample size is given, an inherent limit of frequency resolution is determined. In comparison with this resolution limit, the frequencies can be described as well separated or closely spaced. The amount of frequency separation relative to the resolution limit, and hence the sample size, turns out to be a key factor that determines the technique and accuracy with which the frequencies can be estimated on the basis of a finite data record. It also contributes to the increased complexity in the mathematical analysis. In asymptotic analysis, a common and useful technique is to allow the frequency separation to approach zero at a certain rate as the sample size grows. Such frequency separation conditions

are typically expressed in terms of a power function of  $n^{-1}$  and in the order of magnitude, for example,  $\mathcal{O}(n^{-1})$  or  $\mathcal{O}(n^{-1/2})$ , rather than the exact amount. The separation conditions play a key role in the asymptotic analysis of the estimation methods discussed in later chapters.

Throughout the book, we are concerned with time series of the form

$$y_t = x_t + \epsilon_t \quad (t = 1, \dots, n), \quad (2.1.10)$$

where the  $\epsilon_t$  represent the random measurement errors or noise. Additional assumptions about  $\{\epsilon_t\}$  will be made for specific methods and analytical studies. The simplest assumption is that the  $\epsilon_t$  are independent and identically distributed, or i.i.d., random variables with mean zero and variance  $\sigma^2$ . As a convention, the noise is always assumed real when referring to the RSM and complex when referring to the CSM. Unless noted otherwise, all parameters in the RSM and CSM are regarded as deterministic but possibly unknown constants that are fixed in all realizations of the random process  $\{y_t\}$ . However, in some analytical results, the phases  $\phi_k$  of the sinusoids are allowed to be random variables that are independent of the noise  $\{\epsilon_t\}$ . In some simulations, the frequencies  $\{\omega_k\}$  are also generated randomly to mitigate the possible dependence of the results on specific frequency values.

## 2.2 Spectral Analysis of Stationary Processes

Stationarity is an important property of random processes that makes statistical analysis meaningful. Although not always true in practice, the stationarity assumption is typically applicable to sufficiently short data records taken from slowly-varying nonstationary processes.

**Definition 2.1** (Stationary Random Processes). A zero-mean real or complex random process  $\{X_t\}$  is said to be (weakly or wide-sense) stationary if the covariance function  $c(s, t) := E(X_s X_t^*)$  is finite and time-invariant, i.e.,  $c(t + u, t)$  does not depend on  $t$ . In this case,  $r(u) := c(t + u, t) = E(X_{t+u} X_t^*)$  is called the autocovariance function (ACF) of  $\{X_t\}$ .

The following proposition summarizes some useful properties of the ACF.

**Proposition 2.1** (Properties of the Autocovariance Function). *The ACF  $r(u)$  of a zero-mean stationary process  $\{X_t\}$  is nonnegative definite in the sense that*

$$\sum_{u=1}^m \sum_{v=1}^m a_u a_v^* r(u - v) \geq 0$$

for any constant sequence  $\{a_1, \dots, a_m\} \subset \mathbb{C}$ . The ACF is symmetric in the sense that  $r(-u) = r^*(u)$  and is bounded by the variance of  $X_t$ , i.e.,  $|r(u)| \leq r(0)$ .

In addition to the time-domain characterization, the ACF can also be represented in the frequency domain as a linear combination of infinitely many sinusoids with different frequencies. The subject of spectral analysis of stationary processes is largely built upon the following proposition.

**Proposition 2.2** (Spectral Representation of the Autocovariance Function [46, pp. 117–118]). *If  $r(u)$  is the ACF of a stationary process  $\{X_t\}$ , then there exists a unique nondecreasing right-continuous function  $S(\omega)$ , with  $S(-\pi) = 0$ , such that*

$$r(u) = (2\pi)^{-1} \int_{-\pi}^{\pi} \exp(i\omega u) dS(\omega).$$

The function  $S(\omega)$  is called the spectral distribution function. If  $r(u)$  is absolutely summable, i.e.,  $\sum |r(u)| < \infty$ , then there exists a unique uniformly continuous nonnegative function  $f(\omega)$ , called the spectral density function (SDF), such that

$$r(u) = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\omega) \exp(i\omega u) d\omega$$

and

$$f(\omega) = \sum_{u=-\infty}^{\infty} r(u) \exp(-iu\omega),$$

in which case,  $S(\omega)$  can be expressed as  $S(\omega) = \int_{-\pi}^{\omega} f(\lambda) d\lambda$  and therefore is differentiable with  $\dot{S}(\omega) = f(\omega)$ . If  $r(u)$  is real, then  $f(\omega)$  and  $S(\omega)$  are symmetric in the sense that  $f(-\omega) = f(\omega)$  and  $S(-\omega) = S(\pi) - S(\omega)$ . Both  $f(\omega)$  and  $S(\omega)$  can be extended as  $2\pi$ -periodic functions in  $\mathbb{R}$ .

**Remark 2.1** The SDF defined in Proposition 2.2 has the property

$$(2\pi)^{-1} \int_{-\pi}^{\pi} f(\omega) d\omega = \int_{-1/2}^{1/2} f(2\pi x) dx = r(0).$$

Therefore, it represents the distribution of the total variance  $r(0)$  with respect to the normalized frequency  $\omega/(2\pi)$  (cycles per unit time). This definition is often used in engineering textbooks such as [369]. In statistical textbooks such as [46], the SDF is often defined as  $f(\omega) := 2\pi \sum_{u=-\infty}^{\infty} r(u) \exp(-iu\omega)$ . In this case, the SDF satisfies  $\int_{-\pi}^{\pi} f(\omega) d\omega = r(0)$ , so it represents the distribution of the total variance with respect to the angular frequency  $\omega$  (radians per unit time).

**Remark 2.2** The absolute summability of the ACF is a sufficient but not necessary condition for the existence of the spectral density function. The spectral distribution function is also known as the integrated spectrum [298, p. 209].

A typical stationary process can be classified according to the type of its spectrum into one of the following three categories.

**Definition 2.2** (Discrete, Continuous, and Mixed Spectra). A stationary process with spectral distribution function  $S(\omega)$  is said to have a discrete spectrum if  $S(\omega) = \sum a_k^2 \mathcal{I}(\omega_k \leq \omega)$  for some finite sequences  $\{\omega_k\} \subset (-\pi, \pi)$  and  $\{a_k\} \subset \mathbb{R}$ , where  $\mathcal{I}(\cdot)$  is the indicator function. It is said to have a continuous spectrum if there exists a function  $f(\omega) \geq 0$  such that  $S(\omega) = \int_{-\pi}^{\omega} f(\lambda) d\lambda$ . It is said to have a mixed spectrum if  $S(\omega) = S_d(\omega) + S_c(\omega)$ , where  $S_d(\omega)$  is a discrete spectrum and  $S_c(\omega)$  is a continuous spectrum.

**Remark 2.3** A discrete spectrum of the form  $S(\omega) = \sum a_k^2 \mathcal{I}(\omega_k \leq \omega)$  is a non-decreasing step function with jumps of magnitude  $a_k^2$  at  $\omega = \omega_k$ . The spectral distribution function not having a jump at  $\omega = 0$  is the necessary and sufficient condition for the process to be ergodic in the mean, i.e., the sample mean converges in mean-square to the expected value of the process as the sample size approaches infinity [298, p. 342].

The following proposition shows that a stationary process itself can be represented as a linear combination of infinitely many sinusoids.

**Proposition 2.3** (Spectral Representation of Stationary Processes [46, p. 145]). *If  $\{X_t\}$  is a zero-mean stationary process with spectral distribution function  $S(\omega)$ , then there exists a random process  $\{Z(\omega), -\pi \leq \omega \leq \pi\}$ , with  $Z(-\pi) = 0$ , such that*

$$X_t = \int_{-\pi}^{\pi} \exp(i\omega t) dZ(\omega) \quad (2.2.1)$$

almost surely, where  $Z(\omega)$  satisfies the following conditions:

- (a)  $E\{Z(\omega)\} = 0$  for  $\omega \in [-\pi, \pi]$ ,
- (b)  $E\{|Z(\omega)|^2\} < \infty$  for  $\omega \in [-\pi, \pi]$ ,
- (c)  $E\{|Z(\omega) - Z(\lambda)|^2\} = (2\pi)^{-1}\{S(\omega) - S(\lambda)\}$  for  $-\pi \leq \omega < \lambda \leq \pi$ ,
- (d)  $E\{(Z(\omega) - Z(\lambda))(Z(\omega') - Z(\lambda'))^*\} = 0$  for  $-\pi \leq \lambda \leq \omega < \lambda' \leq \omega' \leq \pi$ .

A process satisfying (a), (b), and (d) is called an orthogonal-increment process. The representation (2.2.1) is unique in the sense that if it also holds with another orthogonal-increment process  $Z'(\omega)$ , then  $P\{Z(\omega) = Z'(\omega)\} = 1$  for all  $\omega$ .

According to Proposition 2.3, if  $\{X_t\}$  is a zero-mean stationary process of discrete spectrum with spectral distribution function

$$S(\omega) = 2\pi \sum_{k=1}^p C_k^2 \mathcal{I}(\omega_k \leq \omega), \quad (2.2.2)$$

then it can be expressed in the form of (2.1.5) with  $\beta_k := Z(\omega_k) - Z(\omega_k^-)$ , where  $Z(\omega_k^-)$  stands for the limit of  $Z(\omega)$  (in mean-square) as  $\omega$  approaches  $\omega_k$  from

the left. Because  $Z(\omega)$  satisfies (a), (c), and (d) in Proposition 2.3, we have

$$E(\beta_k) = 0, \quad E(\beta_k \beta_{k'}^*) = C_k^2 \delta_{k-k'}, \quad (2.2.3)$$

where  $\{\delta_u\}$  ( $u \in \mathbb{Z}$ ) is the Kronecker delta sequence, named after the German mathematician Leopold Kronecker (1823–1891), such that  $\delta_0 = 1$  and  $\delta_u = 0$  for all  $u \neq 0$ . Therefore, a zero-mean stationary process of discrete spectrum is nothing but a sum of complex sinusoids of the form (2.1.5) with uncorrelated zero-mean random coefficients  $\beta_k$ . Using the notation of Dirac delta  $\delta(\omega)$ , named after the British physicist Paul Dirac (1902–1984), the SDF that corresponds to  $S(\omega)$  in (2.2.2) can be formally expressed as

$$f(\omega) = 2\pi \sum_{k=1}^p C_k^2 \delta(\omega - \omega_k), \quad (2.2.4)$$

which is an impulsive function consisting of discrete impulses located at the  $\omega_k$ . An SDF of this form is also known as a line spectrum [369]. By Proposition 2.2, the corresponding ACF can be expressed as

$$r(u) = \sum_{k=1}^p C_k^2 \exp(i\omega_k u), \quad (2.2.5)$$

which is a weighted sum of complex sinusoids with frequencies  $\omega_k$ .

On the other hand, if  $\{X_t\}$  takes the form (2.1.5) with  $\beta_k = C_k \exp(i\phi_k)$  and  $\omega_k \in (-\pi, \pi)$ , and if the  $\phi_k$  are i.i.d. random variables with uniform distribution in  $(-\pi, \pi)$  and the  $C_k$  are real constants, then the condition (2.2.3) is satisfied. In this case,  $\{X_t\}$  can be expressed as (2.2.1) with

$$Z(\omega) := \sum_{k=1}^p \beta_k \mathcal{I}(\omega_k \leq \omega),$$

which implies that  $\{X_t\}$  is a zero-mean stationary process of discrete spectrum whose spectral distribution and density functions and whose ACF are given by (2.2.2), (2.2.4), and (2.2.5), respectively.

More generally, consider the random process  $\{y_t\}$  which is given by (2.1.10) with  $\{x_t\}$  taking the form (2.1.5). If the  $\phi_k$  are i.i.d. random variables with uniform distribution in  $(-\pi, \pi]$  and the noise  $\{\epsilon_t\}$  is a zero-mean stationary process of continuous spectrum with spectral distribution function  $S_\epsilon(\omega)$  and is independent of the  $\phi_k$ , then  $\{y_t\}$  is a zero-mean stationary process of mixed spectrum with spectral distribution function

$$S_y(\omega) = S_x(\omega) + S_\epsilon(\omega),$$

where  $S_x(\omega)$  takes the form (2.2.2). The corresponding SDF can be expressed as

$$f_y(\omega) = f_x(\omega) + f_\epsilon(\omega),$$

where  $f_x(\omega)$  takes the form (2.2.4) and  $f_\epsilon(\omega) := \hat{S}_\epsilon(\omega)$  is the SDF of  $\{\epsilon_t\}$ .

If the  $\phi_k$  are constants, then  $\{y_t\}$  is no longer a stationary process in the sense of Definition 2.1. But, we can still interpret  $\{y_t\}$  as a stationary process of mixed spectrum by considering the sample autocovariance function of  $\{y_1, \dots, y_n\}$  as  $n \rightarrow \infty$ . As will be explained in Chapter 4, the sample ACF of  $\{y_1, \dots, y_n\}$ , defined as  $\hat{r}_y(u) := n^{-1} \sum_{t=1}^{n-|u|} y_{t+|u|} y_t^*$  for  $|u| < n$ , converges (in a suitable sense) to a finite limit which can be expressed as the sum of the ACF given by (2.2.5) and the ACF of  $\{\epsilon_t\}$ . In other words, this limit of the sample ACF coincides with the ordinary ACF in the previous case where the phases are random. Therefore, it has a mixed spectrum of the same form as a stationary process.

In practice, the noise process  $\{\epsilon_t\}$  in (2.1.10) can be assumed to have a continuous spectrum, which is in contrast to the signal  $\{x_t\}$  in (2.1.1)–(2.1.5) that has a discrete spectrum. This is the key feature that distinguishes the signal from the noise. A special case is where  $\{\epsilon_t\}$  can be modeled as a white noise process, i.e., the ACF of  $\{\epsilon_t\}$  takes the form  $r_\epsilon(u) = \sigma^2 \delta_u$ . This will be denoted as  $\{\epsilon_t\} \sim \text{WN}(0, \sigma^2)$ . Linear processes constitute a more general model for  $\{\epsilon_t\}$ , which is useful especially in asymptotic analysis.

**Definition 2.3** (Linear Processes). A random process  $\{X_t\}$  is said to be a linear process if there exists a sequence of constants  $\{\psi_j\}$  and a white noise process  $\{\zeta_t\} \sim \text{WN}(0, \sigma^2)$  such that

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \zeta_{t-j} \quad (2.2.6)$$

converges in mean square. A linear process of the form (2.2.6) can be regarded as the output of a linear time-invariant (LTI) filter with transfer function  $\psi(z) := \sum \psi_j z^{-j}$  ( $z \in \mathbb{C}$ ) and input  $\{\zeta_t\}$ . If  $z$  and  $z^{-1}$  are interpreted as the forward-shift and backward-shift operators such that  $z\zeta_t = \zeta_{t+1}$  and  $z^{-1}\zeta_t = \zeta_{t-1}$ , then (2.2.6) can also be expressed as  $X_t = \psi(z)\zeta_t$ .

Useful properties of linear processes are given in the following proposition.

**Proposition 2.4** (Properties of Linear Processes [46, pp. 122, 154–155]). *Let  $\{\psi_j\}$  be a sequence of constants. If  $\sum |\psi_j|^2 < \infty$ , then the infinite series in (2.2.6) converges in mean square. If  $\sum |\psi_j| < \infty$ , then the infinite series in (2.2.6) converges almost surely and the filter  $\psi(z) := \sum \psi_j z^{-j}$  is said to be bounded-in-bounded-out (BIBO) stable. In both cases, the SDF of  $\{X_t\}$  takes the form*

$$f(\omega) = \sigma^2 |\Psi(\omega)|^2,$$

where  $\Psi(\omega) := \psi(\exp(i\omega)) = \sum \psi_j \exp(-ij\omega)$ , and the ACF of  $\{X_t\}$  is given by

$$r(u) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_{j+u} \psi_j^*.$$

Moreover, if the filter  $\psi(z)$  is BIBO stable, then  $f(\omega)$  is a uniformly continuous function and  $\{r(u)\}$  is an absolutely summable sequence.

The following proposition shows that linear processes encompass a large class of stationary processes with continuous spectra.

**Proposition 2.5** (Wold Decomposition [46, p. 187]). *Any purely nondeterministic zero-mean stationary process  $\{X_t\}$  can be expressed as  $X_t = \sum_{j=0}^{\infty} \psi_j \zeta_{t-j}$ , where  $\psi_0 := 1$ ,  $\sum |\psi_j|^2 < \infty$ , and  $\{\zeta_t\} \sim \text{WN}(0, \sigma^2)$  for some constant  $\sigma^2 > 0$ . Moreover, for any given  $t$ , the random variable  $\zeta_t$  is a member of the closed linear space that comprises all linear combinations of  $X_t, X_{t-1}, \dots$  and their mean-square limits.*

Proposition 2.5 is part of a general theorem [46, Theorem 5.7.2, pp. 187–188] which asserts that any stationary process can be decomposed into a purely non-deterministic component as a linear process, and a deterministic component, such as a sum of sinusoids with i.i.d. random amplitudes, that can be predicted perfectly by a linear combination of its past values.

The process  $\{\zeta_t\}$  in Proposition 2.5 is merely guaranteed to be white noise, i.e., its ACF takes the form  $\sigma^2 \delta_{ii}$ . In the linear process model (2.2.6), stronger assumptions, such as  $\{\zeta_t\}$  being a sequence of i.i.d. random variables, are often needed for some asymptotic analyses. Moreover, because Proposition 2.5 only guarantees the square-summability of  $\{\psi_j\}$ , the corresponding spectrum may not be smooth (or may not even be well-defined at some frequencies). The assumption of BIBO stability ensures that the corresponding SDF is a continuous function (which should not be confused with the concept of continuous spectra). Some analyses require the assumption of strong BIBO stability in the sense that  $\sum |j|^r |\psi_j| < \infty$  for some constant  $r > 1$ , in which case the SDF can be thought of as being smoother than a continuous function. An important example of linear processes with BIBO-stable filters is the autoregressive (AR) process that satisfies the difference equation

$$X_t + \sum_{j=1}^m \varphi_j X_{t-j} = \zeta_t, \quad \{\zeta_t\} \sim \text{WN}(0, \sigma^2), \quad (2.2.7)$$

where  $\varphi(z) := 1 + \sum_{j=1}^m \varphi_j z^{-j}$  has all its roots inside the unit circle of the complex plane  $z \in \mathbb{C}$ . In this case,  $\psi(z) = 1/\varphi(z)$ .

---

## 2.3 Gaussian Processes and White Noise

The concept of Gaussian or normal distribution is well known for real-valued random variables and vectors. For a real random vector  $\mathbf{X}$ , the Gaussian dis-

tribution is specified completely by the mean  $E(\mathbf{X})$  and the covariance matrix  $\text{Cov}(\mathbf{X}) := E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\}$ . In the complex case, the covariance between random variables  $X$  and  $Y$  is defined as

$$\text{Cov}(X, Y) := E\{(X - E(X))(Y - E(Y))^*\}.$$

Therefore, for a complex random vector  $\mathbf{X}$ , not only do we need to consider the mean  $\boldsymbol{\mu} := E(\mathbf{X})$  and the covariance matrix

$$\boldsymbol{\Sigma} := \text{Cov}(\mathbf{X}, \mathbf{X}) := E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^H\},$$

but also the *complementary* covariance matrix, which is defined as

$$\tilde{\boldsymbol{\Sigma}} := \text{Cov}(\mathbf{X}, \mathbf{X}^*) := E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\}.$$

Given these quantities, the mean and the covariance matrix of the real-valued random vector  $\mathbf{X}_r := [\Re(\mathbf{X})^T, \Im(\mathbf{X})^T]^T$  are completely specified by

$$\boldsymbol{\mu}_r := E(\mathbf{X}_r) = [\Re(\boldsymbol{\mu})^T, \Im(\boldsymbol{\mu})^T]^T \quad (2.3.1)$$

and

$$\boldsymbol{\Sigma}_r := \text{Cov}(\mathbf{X}_r) = \frac{1}{2} \begin{bmatrix} \Re(\boldsymbol{\Sigma} + \tilde{\boldsymbol{\Sigma}}) & -\Im(\boldsymbol{\Sigma} - \tilde{\boldsymbol{\Sigma}}) \\ \Im(\boldsymbol{\Sigma} + \tilde{\boldsymbol{\Sigma}}) & \Re(\boldsymbol{\Sigma} - \tilde{\boldsymbol{\Sigma}}) \end{bmatrix}. \quad (2.3.2)$$

On the other hand, if the mean  $\boldsymbol{\mu}_r$  and the covariance matrix  $\boldsymbol{\Sigma}_r$  of a real-valued random vector  $\mathbf{X}_r := [\mathbf{X}_1^T, \mathbf{X}_2^T]^T$  are given in the form of

$$\boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \quad (2.3.3)$$

then  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$ , and  $\tilde{\boldsymbol{\Sigma}}$  of the complex random vector  $\mathbf{X} := \mathbf{X}_1 + i\mathbf{X}_2$  are determined by

$$\begin{cases} \boldsymbol{\mu} = \boldsymbol{\mu}_1 + i\boldsymbol{\mu}_2, \\ \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_{22} + i(\boldsymbol{\Sigma}_{12}^T - \boldsymbol{\Sigma}_{12}), \\ \tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{22} + i(\boldsymbol{\Sigma}_{12}^T + \boldsymbol{\Sigma}_{12}). \end{cases} \quad (2.3.4)$$

Note that the matrix  $\Im(\boldsymbol{\Sigma})$  is antisymmetric, i.e.,  $\Im(\boldsymbol{\Sigma})^T = -\Im(\boldsymbol{\Sigma})$ . This implies in particular that the diagonal elements of  $\Im(\boldsymbol{\Sigma})$  are all equal to zero.

With these properties in mind, let us define complex Gaussian distributions and processes as follows.

**Definition 2.4** (Complex Gaussian Distributions and Processes). A random vector  $\mathbf{X}$  is said to have a general complex Gaussian distribution with mean  $\boldsymbol{\mu}$ , covariance matrix  $\boldsymbol{\Sigma}$ , and complementary covariance matrix  $\tilde{\boldsymbol{\Sigma}}$ , denoted by  $\mathbf{X} \sim N_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}})$ , if  $\mathbf{X}_r := [\Re(\mathbf{X})^T, \Im(\mathbf{X})^T]^T$  is Gaussian with mean  $\boldsymbol{\mu}_r$  and covariance

matrix  $\Sigma_r$ , i.e.,  $\mathbf{X}_r \sim N(\boldsymbol{\mu}_r, \Sigma_r)$ , satisfying (2.3.1)–(2.3.4). If  $\tilde{\Sigma} = \mathbf{0}$ , then  $\mathbf{X}$  is said to have a symmetric complex Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ , denoted as  $\mathbf{X} \sim N_c(\boldsymbol{\mu}, \Sigma)$ . A random process  $\{X_t\}$  is said to be general or symmetric complex Gaussian if  $\mathbf{X} := [X_{t_1}, \dots, X_{t_n}]^T$  is general or symmetric complex Gaussian, respectively, for any distinct points  $\{t_1, \dots, t_n\}$  and any  $n$ .

A very important property of real Gaussian distributions is their invariance under linear transformation — a linear transform of Gaussian variables remains Gaussian. This property is also possessed by the general and symmetric complex Gaussian variables, respectively, as stated in the following proposition.

**Proposition 2.6** (Linear Transform of Complex Gaussian Random Variables). *Let  $\mathbf{A}$  be a complex matrix and  $\mathbf{b}$  be a complex vector.*

- (a) *If  $\mathbf{X} \sim N_c(\boldsymbol{\mu}, \Sigma, \tilde{\Sigma})$ , then  $\mathbf{AX} + \mathbf{b} \sim N_c(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^H, \mathbf{A}\tilde{\Sigma}\mathbf{A}^T)$ .*
- (b) *If  $\mathbf{X} \sim N_c(\boldsymbol{\mu}, \Sigma)$ , then  $\mathbf{AX} + \mathbf{b} \sim N_c(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^H)$ .*
- (c) *If  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , then  $\mathbf{AX} + \mathbf{b} \sim N_c(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^H, \mathbf{A}\Sigma\mathbf{A}^T)$ .*

PROOF. The real and imaginary parts of  $\mathbf{AX} + \mathbf{b}$  are jointly Gaussian because they are linear transforms of the real and imaginary parts of  $\mathbf{X}$  which are jointly Gaussian. Part (a) follows from the fact that  $\text{Cov}(\mathbf{AX} + \mathbf{b}, \mathbf{AX} + \mathbf{b}) = \mathbf{A}\Sigma\mathbf{A}^H$  and  $\text{Cov}(\mathbf{AX} + \mathbf{b}, (\mathbf{AX} + \mathbf{b})^*) = \mathbf{A}\tilde{\Sigma}\mathbf{A}^T$ . Part (b) is a direct result of (a) with  $\tilde{\Sigma} = \mathbf{0}$ . Part (c) follows from the fact that the real and imaginary parts of  $\mathbf{AX} + \mathbf{b}$  are linear transforms of  $\mathbf{X}$  and  $\text{Cov}(\mathbf{AX} + \mathbf{b}, (\mathbf{AX} + \mathbf{b})^*) = \mathbf{A}\Sigma\mathbf{A}^T$ .  $\square$

According to (2.3.4),  $\tilde{\Sigma} = \mathbf{0}$  if and only if  $\Sigma_r$  is symmetric in the sense that

$$\Sigma_{11} = \Sigma_{22}, \quad \Sigma_{12} = -\Sigma_{12}^T. \quad (2.3.5)$$

Hence, a symmetric complex Gaussian random vector is characterized by its real and imaginary parts having a symmetric Gaussian distribution satisfying (2.3.5). The probability density function (PDF) of a symmetric complex Gaussian random vector  $\mathbf{X} \sim N_c(\boldsymbol{\mu}, \Sigma)$  takes the form

$$p(\mathbf{x}) = \pi^{-n} |\Sigma|^{-1} \exp\{-(\mathbf{x} - \boldsymbol{\mu})^H \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\} \quad (\mathbf{x} \in \mathbb{C}^n),$$

which should be interpreted as a function of  $\Re(\mathbf{x})$  and  $\Im(\mathbf{x})$ . Therefore, if  $\mathbf{X}$  and  $\mathbf{Y}$  are jointly symmetric complex Gaussian with  $\text{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$ , then  $\mathbf{X}$  and  $\mathbf{Y}$  are mutually independent. A univariate random variable  $X$  has a symmetric complex Gaussian distribution with mean zero and variance  $\sigma^2$ , i.e.,  $X \sim N_c(0, \sigma^2)$ , if and only if  $\Re(X)$  and  $\Im(X)$  are i.i.d.  $N(0, \frac{1}{2}\sigma^2)$ .

By definition, a symmetric complex Gaussian process  $\{X_t\}$  has zero complementary covariance function, i.e.,  $E(X_s X_t) = 0$  for all  $s$  and  $t$ . This is equivalent to the following symmetry condition for the real and imaginary parts of  $\{X_t\}$ :

$$\begin{cases} \text{Cov}\{\Re(X_s), \Re(X_t)\} = \text{Cov}\{\Im(X_s), \Im(X_t)\}, \\ \text{Cov}\{\Re(X_s), \Im(X_t)\} = -\text{Cov}\{\Im(X_s), \Re(X_t)\}. \end{cases} \quad (2.3.6)$$

Such processes arise naturally in real-world applications, including radar, sonar, and communications, as a result of modulation. For example, let  $\{Y_t\}$  be a zero-mean real Gaussian process and  $\phi$  be a real random variable, independent of  $\{Y_t\}$ , such that  $E\{\exp(i2\phi)\} = 0$ . Consider the modulated complex process

$$X_t := Y_t \exp\{i(\omega_c t + \phi)\},$$

where  $\omega_c$ , a nonzero constant, is known as the carrier frequency. It is easy to show that  $E(X_t) = 0$  and

$$\begin{aligned} E\{\Re(X_s)\Re(X_t)\} &= \frac{1}{2}E(Y_s Y_t) [\cos(\omega_c(s-t)) + E\{\cos(\omega_c(s+t) + 2\phi)\}] \\ &= \frac{1}{2}E(Y_s Y_t) \cos(\omega_c(s-t)), \\ E\{\Im(X_s)\Im(X_t)\} &= \frac{1}{2}E(Y_s Y_t) [\cos(\omega_c(s-t)) - E\{\cos(\omega_c(s+t) + 2\phi)\}] \\ &= \frac{1}{2}E(Y_s Y_t) \cos(\omega_c(s-t)), \\ E\{\Re(X_s)\Im(X_t)\} &= \frac{1}{2}E(Y_s Y_t) [\sin(\omega_c(s-t)) + E\{\sin(\omega_c(s+t) + 2\phi)\}] \\ &= \frac{1}{2}E(Y_s Y_t) \sin(\omega_c(s-t)). \end{aligned}$$

So the condition (2.3.6) is satisfied. Observe that  $\{\Re(X_t)\}$  and  $\{\Im(X_t)\}$  are jointly Gaussian, because conditioning on  $\phi$  they are jointly Gaussian with mean zero and with covariance and cross-covariance functions being independent of  $\phi$ . Therefore, by definition,  $\{X_t\}$  is a zero-mean symmetric complex Gaussian process with covariance function  $c(s, t) = E(Y_s Y_t) \exp\{i\omega_c(s-t)\}$ . In addition, if  $\{Y_t\}$  is stationary, so is  $\{X_t\}$ . Note that (2.3.6) is not satisfied if  $E\{\exp(i2\phi)\} \neq 0$ , in which case,  $\{X_t\}$  is only a general complex Gaussian process.

For convenience, we will drop the word “symmetric” when referring to symmetric complex Gaussian variables, distributions, or processes and simply call them *complex Gaussian*. We will retain the word “general” when referring to general complex Gaussian distributions.

White noise is a special type of stationary process that plays an important role in statistical theory. It is a useful model for ambient noise and measurement errors in many applications. It can also be used to simulate stationary processes with a nonwhite, or colored, spectrum through linear filtering.

**Definition 2.5** (White Noise). A zero-mean stationary process  $\{X_t\}$  with variance  $\sigma^2$  is said to be white noise, denoted by  $\{X_t\} \sim \text{WN}(0, \sigma^2)$ , if the ACF of  $\{X_t\}$  takes the form  $r(u) = \sigma^2 \delta_u$ , where  $\{\delta_u\}$  is the Kronecker delta sequence; or equivalently, if the SDF of  $\{X_t\}$  takes the form  $f(\omega) = \sigma^2$  for all  $\omega$ . A sequence  $\{X_t\}$  of i.i.d. random variables with mean zero and variance  $\sigma^2$  is a white noise process, denoted by  $\{X_t\} \sim \text{IID}(0, \sigma^2)$ . A sequence  $\{X_t\}$  is called Gaussian white noise, denoted by  $\{X_t\} \sim \text{GWN}(0, \sigma^2)$  in the real case and by  $\{X_t\} \sim \text{GWN}_c(0, \sigma^2)$  in the complex case, if the  $X_t$  are i.i.d.  $N(0, \sigma^2)$  or  $N_c(0, \sigma^2)$ , respectively.

Martingale differences with constant variance constitute another, perhaps less familiar, example of white noise.

**Definition 2.6** (Martingale Differences). A random sequence  $\{X_t\}$  is said to be a sequence of martingale differences if there exists an increasing sequence of  $\sigma$ -fields (or filtration)  $\{\mathfrak{F}_t\}$  such that (a)  $X_t$  is measurable with respect to  $\mathfrak{F}_t$  for each  $t$  and (b)  $E(|X_t|^2) < \infty$  and  $E(X_t|\mathfrak{F}_{t-1}) = 0$  almost surely for all  $t$ . A sequence  $\{X_t\}$  of martingale differences with constant variance  $\sigma^2$  is a white noise process and is denoted by  $\{X_t\} \sim \text{MD}(0, \sigma^2)$ .

If  $\{X_t\}$  is a sequence of martingale differences, then  $S_t := \sum_{j=1}^t X_j$  is called a *martingale*, characterized by the property [34, p. 458]

$$E(S_t|\mathfrak{F}_{t-1}) = S_{t-1}.$$

The notion of martingale difference stems from the fact that  $X_t = S_t - S_{t-1}$ .

The  $\sigma$ -field  $\mathfrak{F}_t$  in Definition 2.6 can be interpreted as a representation of certain historical information about the random process up to (and including) time  $t$ . The requirement that  $X_t$  be measurable with respect to  $\mathfrak{F}_t$  simply means that  $X_t$  is determined completely by the historical information up to time  $t$ . If the historical information is known only up to time  $t-1$ , then the martingale difference  $X_t$  remains unpredictable (in the minimum mean-square sense) because the best prediction,  $E(X_t|\mathfrak{F}_{t-1})$ , is equal to the mean of  $X_t$ , which is zero. Given this interpretation, it is not difficult to see that the one-step prediction errors,  $X_t := Y_t - E(Y_t|Y_{t-1}, Y_{t-2}, \dots)$  ( $t = 1, 2, \dots$ ), of a random process  $\{Y_t\}$  are martingale differences with  $\mathfrak{F}_t$  being the  $\sigma$ -field generated by  $\{Y_t, Y_{t-1}, \dots\}$ .

Martingale differences in general are uncorrelated but not necessarily independent or identically distributed or even stationary. For example, the variance can change with  $t$ . A trivial sequence of martingale differences is a sequence of i.i.d. random variables with mean zero, in which case  $\mathfrak{F}_t$  is defined as the  $\sigma$ -field generated by  $\{X_t, X_{t-1}, \dots\}$ . Therefore, the concept of martingale differences is a generalization of the concept of zero-mean i.i.d. random variables.

The assertion in Definition 2.6 that a sequence of martingale differences with constant variance is white noise can be justified as follows. By definition, if  $\{X_t\}$  is a sequence of martingale differences, then, by the iterated expectation,

$$\begin{aligned} E(X_t) &= E\{E(X_t|\mathfrak{F}_{t-1})\} = 0 \quad \text{for all } t, \\ E(X_t X_s^*) &= E\{E(X_t|\mathfrak{F}_{t-1})X_s^*\} = 0 \quad \text{for all } t > s. \end{aligned}$$

This means that  $\{X_t\}$  is an uncorrelated process with mean zero. If, in addition,  $E(|X_t|^2) = \sigma^2$  for all  $t$ , then

$$E(X_t X_s^*) = \sigma^2 \delta_{t-s},$$

which, by definition, implies that  $\{X_t\} \sim \text{WN}(0, \sigma^2)$ . Note that if  $E(|X_t|^2|\mathfrak{F}_{t-1}) = \sigma^2$  for all  $t$ , then  $E(|X_t|^2) = \sigma^2$  for all  $t$ .

## 2.4 Linear Prediction Theory

Linear prediction is a powerful technique for time series analysis. Autoregressive models are directly related to linear prediction. The following example shows that linear prediction is also useful for modeling sinusoids.

Consider a simple case where  $y_t = \cos(\omega_0 t)$ , with  $\omega_0 \in (0, \pi)$  being a constant. Using trigonometric identities, we obtain

$$\begin{aligned}\cos(\omega_0(t-1)) &= \cos(\omega_0)\cos(\omega_0 t) + \sin(\omega_0)\sin(\omega_0 t), \\ \cos(\omega_0(t-2)) &= \{2\cos^2(\omega_0) - 1\}\cos(\omega_0 t) + 2\cos(\omega_0)\sin(\omega_0)\sin(\omega_0 t).\end{aligned}$$

Combining these equations leads to

$$2\cos(\omega_0)\cos(\omega_0(t-1)) - \cos(\omega_0(t-2)) = \cos(\omega_0 t).$$

In other words, we can write

$$y_t = 2\cos(\omega_0)y_{t-1} - y_{t-2}.$$

This expression means that the current value  $y_t$  of the sinusoid can be predicted without error by a suitable linear combination of the previous values  $y_{t-1}$  and  $y_{t-2}$ . Furthermore, the coefficient  $c := 2\cos(\omega_0)$  can be used to identify the frequency  $\omega_0$  because  $\omega_0 = \arccos(c/2)$ . This observation has motivated the development of many linear-prediction-based algorithms for frequency estimation which will be discussed later in Chapter 9.

In this section, we provide a brief review of the linear prediction theory to facilitate the later discussions. First, the following proposition describes the properties of the best linear predictor for a zero-mean stationary process.

**Proposition 2.7** (Best Linear Prediction [46, p. 64] [177, p. 157]). *Let  $\{X_t\}$  be a zero-mean stationary process with ACF  $r(u)$ . Then, for any  $m \geq 1$ , the best linear predictor of  $X_t$  based on  $\{X_{t-1}, \dots, X_{t-m}\}$ , defined as the minimizer of  $E\{|X_t - Y_t|^2\}$  with respect to  $Y_t$  being any linear function of  $\{X_{t-1}, \dots, X_{t-m}\}$ , can be expressed as  $\hat{X}_t = -\sum_{j=1}^m \varphi_j X_{t-j}$ , where  $\boldsymbol{\varphi} := [\varphi_1, \dots, \varphi_m]^T$  satisfies  $\mathbf{R}\boldsymbol{\varphi} = -\mathbf{r}$ , with*

$$\mathbf{R} := \begin{bmatrix} r(0) & r^*(1) & \cdots & r^*(m-1) \\ r(1) & r(0) & \cdots & r^*(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(m-1) & r(m-2) & \cdots & r(0) \end{bmatrix}, \quad \mathbf{r} := \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(m) \end{bmatrix}.$$

*The prediction error  $X_t - \hat{X}_t$  has the orthogonality property  $E\{(X_t - \hat{X}_t)X_{t-j}^*\} = 0$  ( $j = 1, \dots, m$ ) and its variance  $\sigma^2 := E\{|X_t - \hat{X}_t|^2\}$  satisfies  $\sigma^2 = r(0) + \mathbf{r}^H \boldsymbol{\varphi}$ .*

The next proposition presents a fast algorithm to compute the inverse of a covariance matrix formed by a stationary process. With this algorithm, the coefficients of the best predictor in Proposition 2.7 can be computed efficiently.

**Proposition 2.8** (Levinson-Durbin Algorithm [46, p. 169] [177, pp. 171–176]). *Let  $\{X_t\}$  be a zero-mean stationary process with ACF  $r(u)$ . Let  $\mathbf{X} := [X_1, \dots, X_n]^T$  and  $\mathbf{\Sigma} := E(\mathbf{X}\mathbf{X}^H) = [r(s-t)]$  ( $s, t = 1, \dots, n$ ). Then, the Cholesky decomposition of  $\mathbf{\Sigma}^{-1}$  takes the form  $\mathbf{\Sigma}^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^H$ , where  $\mathbf{D} := \text{diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_{n-1}^2)$  is a diagonal matrix and  $\mathbf{U} := [u_{st}]$  ( $s, t = 1, \dots, n$ ) is an upper triangular matrix with  $u_{st} := \varphi_{t-1, t-s}^*$  for  $1 \leq s \leq t \leq n$  ( $\varphi_{t0} := 1$  for all  $t = 0, 1, \dots$ ). The  $\varphi_{tj}$  and  $\sigma_t^2$  can be computed recursively as follows: for  $t = 1, 2, \dots$  and with the initial value  $\sigma_0^2 = r(0)$ ,*

$$\begin{aligned}\varphi_{tt} &= -\frac{1}{\sigma_{t-1}^2} \left\{ r(t) + \sum_{j=1}^{t-1} \varphi_{t-1, j} r(t-j) \right\}, \\ \varphi_{tj} &= \varphi_{t-1, j} + \varphi_{tt} \varphi_{t-1, t-j}^* \quad (j = 1, \dots, t-1), \\ \sigma_t^2 &= \sigma_{t-1}^2 (1 - |\varphi_{tt}|^2).\end{aligned}$$

*The best linear predictor of  $X_{t+1}$  based on  $\{X_1, \dots, X_t\}$  can be expressed as  $\tilde{X}_{t+1} := -\sum_{j=1}^t \varphi_{tj} X_{t+1-j}$  with the prediction error variance  $\sigma_t^2 = E\{|X_{t+1} - \tilde{X}_{t+1}|^2\}$ .*

**Remark 2.4** The best linear prediction of  $X_t$  defined in Proposition 2.7 can be expressed as  $\hat{X}_t = -\sum_{j=1}^m \varphi_{mj} X_{t-j}$  with  $\sigma^2 := E\{|X_t - \hat{X}_t|^2\} = \sigma_m^2$ .

**Remark 2.5** The quantity  $\varphi_{mm}$  is called the  $m$ th reflection coefficient. It coincides with the negative of the lag- $m$  partial autocorrelation coefficient defined as the autocorrelation coefficient between the best forward prediction error of  $X_t$  based on  $\{X_{t-1}, \dots, X_{t-m+1}\}$  and the best backward prediction error of  $X_{t-m}$  based on the same set of predictors. It follows from the third equation in Proposition 2.8 that  $\sigma_m^2 > 0$  if and only if  $|\varphi_{tt}| < 1$  for all  $t = 1, \dots, m$ .

In Proposition 2.8, the  $\varphi_{tj}$  and  $\sigma_t^2$  are computed from the ACF of  $\{X_t\}$ . When  $\{X_t\}$  is an AR process, they can be computed directly from the AR parameters based on the following algorithm.

**Proposition 2.9** (Levinson-Durbin Algorithm for AR Processes [46, p. 242] [177, pp. 172–173]). *Let  $\{X_t\}$  be an AR( $m$ ) process of the form (2.2.7). Let  $\varphi_{tj}$  and  $\sigma_t^2$  be defined in Proposition 2.8. Then, for any  $t \geq m$ ,  $\varphi_{tj} = \varphi_j$  if  $1 \leq j \leq m$ ,  $\varphi_{tj} = 0$  if  $m < j \leq t$ , and  $\sigma_t^2 = \sigma^2$ ; for  $t = m-1, m-2, \dots$ , the  $\varphi_{tj}$  and  $\sigma_t^2$  can be computed recursively as follows:*

$$\begin{aligned}\varphi_{tj} &= \frac{\varphi_{t+1, j} - \varphi_{t+1, t+1} \varphi_{t+1, t+1-j}^*}{1 - |\varphi_{t+1, t+1}|^2} \quad (j = 1, \dots, t), \\ \sigma_t^2 &= \frac{\sigma_{t+1}^2}{1 - |\varphi_{t+1, t+1}|^2}.\end{aligned}$$

Moreover, let  $\Sigma$  be the covariance matrix defined in Proposition 2.8 and let  $\Sigma^{-1} := [\eta_{st}]$  ( $s, t = 1, \dots, n$ ). Then,  $\eta_{st} = 0$  for all  $|s - t| > m$ . In other words,  $\Sigma^{-1}$  is a band matrix with bandwidth  $2m + 1$ .

While the Levinson-Durbin algorithm produces the Cholesky decomposition for the inverse of a covariance matrix, the following innovations algorithm provides the Cholesky decomposition for the covariance matrix itself. It also gives an expression for the best predictor in terms of the prediction errors or innovations, hence the name of the algorithm.

**Proposition 2.10** (Innovations Algorithm [46, pp. 172, 193, 255] [177, pp. 29–30]). *Let  $\{X_t\}$  be a zero-mean stationary process with ACF  $r(u)$ . Let  $\mathbf{X} := [X_1, \dots, X_n]^T$  and  $\Sigma := E(\mathbf{X}\mathbf{X}^H) = [r(s - t)]$  ( $s, t = 1, \dots, n$ ). Then, the Cholesky decomposition of  $\Sigma$  takes the form  $\Sigma = \mathbf{L}\mathbf{D}\mathbf{L}^H$ , where  $\mathbf{D} := \text{diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_{n-1}^2)$  is the diagonal matrix defined in Proposition 2.8 and  $\mathbf{L} = \mathbf{U}^{-H} := [c_{st}]$  is a lower triangular matrix with  $c_{st} := \psi_{s-1, s-t}$  for  $1 \leq t \leq s \leq n$  ( $\psi_{t0} := 1$  for all  $t = 0, 1, \dots$ ). The  $\psi_{tj}$  and  $\sigma_t^2$  can be computed recursively as follows: for  $t = 1, 2, \dots$  and with  $\sigma_0^2 = r(0)$ ,*

$$\psi_{t, t-j} = \frac{1}{\sigma_j^2} \left\{ r(t-j) - \sum_{l=0}^{j-1} \psi_{t, t-l} \psi_{j, j-l}^* \sigma_l^2 \right\} \quad (j = 0, 1, \dots, t-1),$$

$$\sigma_t^2 = r(0) - \sum_{j=0}^{t-1} |\psi_{t, t-j}|^2 \sigma_j^2.$$

The best linear predictor of  $X_{t+1}$  based on  $\{X_1, \dots, X_t\}$  can be expressed as  $\tilde{X}_{t+1} = \sum_{j=1}^t \psi_{tj} Z_{t+1-j}$ , where  $Z_t := X_t - \tilde{X}_t$  ( $\tilde{X}_0 := 0$ ) and  $\text{Var}(Z_t) = \sigma_{t-1}^2$ . If  $\{X_t\}$  is a moving-average (MA) process of order  $m$ , i.e.,  $X_t = \sum_{j=0}^m \psi_j \zeta_{t-j}$ , where  $\psi_0 := 1$  and  $\{\zeta_t\} \sim \text{WN}(0, \sigma^2)$ , then  $\Sigma$  is a band matrix with bandwidth  $2m + 1$  such that  $\psi_{tj} = 0$  for  $t \geq j > m$ . Moreover, if the filter  $\psi(z) := \sum_{j=0}^m \psi_j z^{-j}$  is invertible, then, as  $t \rightarrow \infty$ ,  $E\{|Z_t - \zeta_t|^2\} \rightarrow 0$ ,  $\sigma_t^2 \rightarrow \sigma^2$ , and  $|\psi_{tj} - \psi_j| = \mathcal{O}(\rho^t) \rightarrow 0$  for all  $j = 1, \dots, m$  and for some constant  $\rho \in (0, 1)$ .

**Remark 2.6** With  $c_{st} := \psi_{s-1, s-t}$ , the recursions can be written as

$$c_{st} = \frac{1}{\sigma_{t-1}^2} \left\{ r(s-t) - \sum_{l=1}^{t-1} c_{sl} c_{tl}^* \sigma_{l-1}^2 \right\} \quad (t = 1, \dots, s-1),$$

$$\sigma_{s-1}^2 = r(0) - \sum_{l=1}^{s-1} |c_{sl}|^2 \sigma_{l-1}^2.$$

This is the standard Cholesky decomposition algorithm [177, pp. 29–30].

**Remark 2.7** If  $\{X_t\}$  is an MA process of order  $m$ , then it suffices to calculate  $\psi_{t, t-j}$  for  $t - m \leq j \leq t - 1$ , and hence the complexity of the Cholesky decomposition by the innovations algorithm takes the form  $\mathcal{O}(n)$  for large  $n$ .

## 2.5 Asymptotic Statistical Theory

In many estimation problems, it is often difficult to mathematically characterize and quantify the statistical behavior of an estimator for finite sample sizes. A practical solution to the problem is simulation. While simulation does provide valuable insights and is indispensable to practitioners, simulation results are usually inconclusive and cannot be extrapolated under different conditions. Asymptotic analysis serves as an important tool to fill in the gap. Although it does not directly answer questions regarding the finite-sample behavior, asymptotic analysis often yields conclusive and useful results that are valid under very general conditions and for large but finite sample sizes; in many cases it is also the only way to obtain such results.

As a performance criterion, it is desirable that an estimator calculated from a finite data record converges in some sense to the parameter of interest as the sample size approaches infinity. It is also desirable that the unknown distribution of an estimator converges to a known one so that the randomness of the estimator can be easily characterized and quantified. These concerns lead to three widely-used modes of convergence: convergence in probability, almost sure convergence, and convergence in distribution. Associated with the first two modes of convergence are two ways of evaluating the magnitude of a random variable: the boundedness in probability and the almost sure boundedness. These concepts are summarized in the following.

**Definition 2.7** (Convergence and Boundedness in Probability). A sequence of random variables  $\{X_n\}$  is said to converge to zero in probability (or weakly), denoted by  $X_n = \mathcal{O}_P(1)$  or  $X_n \xrightarrow{P} 0$ , if for any constant  $\delta > 0$ ,  $P(|X_n| > \delta) \rightarrow 0$  as  $n \rightarrow \infty$ . It is said to converge to a random variable  $X$ , denoted by  $X_n \xrightarrow{P} X$ , if  $X_n - X = \mathcal{O}_P(1)$ . It is said to be bounded in probability, denoted by  $X_n = \mathcal{O}_P(1)$ , if for any constant  $\delta > 0$  there exists a constant  $c > 0$  such that  $P(|X_n| > c) < \delta$  for large  $n$ . Moreover, for any sequence of positive constants  $\{a_n\}$ , we write  $X_n = \mathcal{O}_P(a_n)$  if  $a_n^{-1}X_n = \mathcal{O}_P(1)$ , and we write  $X_n = \mathcal{O}_P(a_n)$  if  $a_n^{-1}X_n = \mathcal{O}_P(1)$ . A statement of convergence or boundedness in probability is valid for a sequence of random vectors if it is valid componentwise.

The concept of convergence in probability can be easily understood by interpreting  $\delta$  in Definition 2.7 as a prescribed *tolerance level* for deviation of  $X_n$  from its target value. In so doing, the statement  $X_n \xrightarrow{P} 0$  simply means that no matter how small the tolerance level is, the probability of  $X_n$  exceeding that level will approach zero as  $n$  approaches infinity.

Note that the probability here is calculated on the basis of repeatedly observing  $X_n$  from different random experiments (or scenarios) for each given and

fixed  $n$ . Therefore, the fact that this probability approaches zero does not necessarily imply that for each given experiment the resulting infinite sequence  $\{X_n\}$  converges to zero in the ordinary sense. For example, in an infinite sequence  $\{X_n\}$  produced by a given experiment, one may find infinitely many instances in which the tolerance level is violated, but the probability that the tolerance level is violated at a fixed instance  $n$  when considering all possible outcomes from the repeated experiments can still approach zero as  $n \rightarrow \infty$ .

Similarly,  $X_n = \mathcal{O}_P(1)$  does not necessarily imply that the  $X_n$  are bounded in the ordinary sense for each given experiment, i.e., it does not necessarily imply that there exists a constant  $c > 0$  such that  $|X_n| \leq c$  for all  $n$  and all experiments. The boundedness in probability only means that the probability of  $X_n$  exceeding a bound  $c$  can be made arbitrarily small for all  $n$  if  $c$  is sufficiently large.

To ensure the convergence and the boundedness of a random sequence in the ordinary sense for all possible experiments, the concepts of almost sure convergence and almost sure boundedness are needed.

**Definition 2.8** (Almost Sure Convergence and Boundedness). A sequence of random variables  $\{X_n\}$  is said to converge to zero almost surely, denoted by  $X_n = \mathcal{O}(1)$  or  $X_n \xrightarrow{a.s.} 0$ , if  $P(X_n \rightarrow 0) = 1$  as  $n \rightarrow \infty$ , or equivalently, if  $P(|X_n| > \delta \text{ i.o.}) = 0$  for any constant  $\delta > 0$ , where i.o. stands for “infinitely often.” The sequence is said to converge to a random variable  $X$ , denoted by  $X_n \xrightarrow{a.s.} X$ , if  $X_n - X = \mathcal{O}(1)$ . It is said to be bounded almost surely, denoted by  $X_n = \mathcal{O}(1)$ , if there is a constant  $c > 0$  such that  $P(|X_n| > c) = 0$  for all  $n$ . For any sequence of positive constants  $\{a_n\}$ , we write  $X_n = \mathcal{O}(a_n)$  if  $a_n^{-1}X_n = \mathcal{O}(1)$ , and we write  $X_n = \mathcal{O}(a_n)$  if  $a_n^{-1}X_n = \mathcal{O}(1)$ . A statement of almost sure convergence or boundedness is valid for a sequence of random vectors if it is valid componentwise.

Convergence in distribution is another mode of convergence. It is useful when the probability distribution of an estimator is of interest, as is the case when constructing a confidence interval for the parameter being estimated.

**Definition 2.9** (Convergence in Distribution). A sequence of random vectors  $\{\mathbf{X}_n\}$  is said to converge in distribution to a random vector  $\mathbf{X}$ , denoted by  $\mathbf{X}_n \xrightarrow{D} \mathbf{X}$ , if the cumulative distribution function (CDF) of  $\mathbf{X}_n$  converges to the CDF of  $\mathbf{X}$  at every continuity point of the latter.

Unlike the other modes of convergence, the convergence in distribution does not directly address the convergence of a random sequence itself. Instead, it concerns the convergence of the CDFs. Just like two random variables from completely unrelated experiments can have the same CDF, the random variables in Definition 2.9 may come from unrelated experiments.

The three modes of convergence are related to each other by a hierarchy with the almost sure convergence being the strongest mode and the convergence in distribution the weakest mode.

**Proposition 2.11** (Hierarchy of Convergence Modes [34, p. 330]). *If  $X_n \xrightarrow{a.s.} X$ , then  $X_n \xrightarrow{P} X$ . If  $X_n \xrightarrow{P} X$ , then  $X_n \xrightarrow{D} X$ . The converses are not true in general. However, if  $X_n \xrightarrow{D} c$  for some constant  $c$ , then  $X_n \xrightarrow{P} c$ .*

In the asymptotic analysis of an estimator, it often suffices to consider the major terms in an expansion of the estimator that dominate the other terms in magnitude. This requires to combine random variables of different orders of magnitude, for which the following proposition is very useful.

**Proposition 2.12** (Arithmetics of Big O and Small O [46, p. 199]). *Let  $\{X_n\}$  and  $\{Y_n\}$  be random sequences. Let  $\{a_n\}$  and  $\{b_n\}$  be sequences of positive constants. If  $X_n = \mathcal{O}_P(a_n)$  and  $Y_n = \mathcal{O}_P(b_n)$ , then  $X_n Y_n = \mathcal{O}_P(a_n b_n)$  and  $X_n + Y_n = \mathcal{O}_P(a_n + b_n)$ . The same implication holds if  $\mathcal{O}_P$  is everywhere replaced by  $\mathcal{O}$ . If  $X_n = \mathcal{O}_P(a_n)$  and  $Y_n = \mathcal{O}_P(b_n)$ , then  $X_n Y_n = \mathcal{O}_P(a_n b_n)$ . All these assertions remain valid if  $\mathcal{O}_P$  and  $\mathcal{O}$  are everywhere replaced by  $\mathcal{o}$  and  $\mathcal{O}$ , respectively.*

The following proposition ensures that the modes of convergence are preserved after certain transformations.

**Proposition 2.13** (Convergence after Simple Transformation).

- (a) *If  $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ , then  $\mathbf{g}(\mathbf{X}_n) \xrightarrow{P} \mathbf{g}(\mathbf{X})$  for any continuous function  $\mathbf{g}(\cdot)$ . This assertion remains valid if  $\xrightarrow{P}$  is replaced by  $\xrightarrow{a.s.}$  or  $\xrightarrow{D}$ .*
- (b) *If  $\mathbf{X}_n \xrightarrow{D} \mathbf{X}$  and  $\mathbf{Y}_n \xrightarrow{P} \mathbf{c}$  for some constant vector  $\mathbf{c}$ , then  $\mathbf{Y}_n^H \mathbf{X}_n \xrightarrow{D} \mathbf{c}^H \mathbf{X}$  and  $\mathbf{X}_n + \mathbf{Y}_n \xrightarrow{D} \mathbf{X} + \mathbf{c}$ . This assertion is known as Slutsky's Theorem.*

PROOF. These results are well known in the real case [34, p. 332 and p. 334] [46, pp. 200–201 and 206–207]. In the complex case, part (a) is trivial with regard to the convergence in probability and almost sure convergence. For the convergence in distribution, observe that by definition  $\mathbf{X}_n \xrightarrow{D} \mathbf{X}$  is the same as  $\mathbf{Y}_n := [\Re(\mathbf{X}_n^T), \Im(\mathbf{X}_n^T)]^T \xrightarrow{D} \mathbf{Y} := [\Re(\mathbf{X}^T), \Im(\mathbf{X}^T)]^T$ . Let  $\mathbf{g} = \mathbf{g}_1 + i\mathbf{g}_2$ . Because  $\mathbf{g}_1$  and  $\mathbf{g}_2$  are continuous real functions, it follows that  $[\mathbf{g}_1^T(\mathbf{Y}_n), \mathbf{g}_2^T(\mathbf{Y}_n)]^T \xrightarrow{D} [\mathbf{g}_1^T(\mathbf{Y}), \mathbf{g}_2^T(\mathbf{Y})]^T$ , which is the same as  $\mathbf{g}(\mathbf{X}_n) \xrightarrow{D} \mathbf{g}(\mathbf{X})$ . Part (b) can be proved similarly.  $\square$

Taylor expansion is a useful tool in the asymptotic analysis of estimators. The following proposition ensures its validity for random variables.

**Proposition 2.14** (Taylor Expansion). *Let  $\{r_n\}$  be a sequence of positive constants such that  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ .*

- (a) *For a sequence of real random variables  $\{X_n\}$ , if  $X_n = a_n + \mathcal{O}_P(r_n)$ , where  $\{a_n\}$  is a sequence of constants, then for any real function  $g(\cdot)$  which has  $m$  continuous derivatives in a neighborhood of  $a_n$ ,*

$$g(X_n) = \sum_{j=0}^m \{g_j(a_n) / j!\} (X_n - a_n)^j + \mathcal{O}_P(r_n^m),$$

where  $g_j(\cdot)$  denotes the  $j$ th derivative of  $g(\cdot)$ .

(b) For a sequence of real random vectors  $\{\mathbf{X}_n\}$ , if  $\mathbf{X}_n = \mathbf{a}_n + \mathcal{O}_P(r_n)$ , where  $\{\mathbf{a}_n\}$  is a sequence of constant vectors, then for any real function  $g(\cdot)$  which is continuously differentiable in a neighborhood of  $\mathbf{a}_n$ ,

$$g(\mathbf{X}_n) = \{\nabla g(\mathbf{a}_n)\}^T (\mathbf{X}_n - \mathbf{a}_n) + \mathcal{O}_P(r_n),$$

where  $\nabla$  denotes the gradient operator. Both assertions remain valid if  $\mathcal{O}_P$  and  $\mathcal{O}_P$  are everywhere replaced by  $\mathcal{O}$  and  $\mathcal{O}$ , respectively.

PROOF. Consider part (a). From calculus, we have

$$g(x) = \sum_{j=0}^m \{g_j(a_n)/j!\} (x - a_n)^j + \mathcal{O}(|x - a_n|^m).$$

The assertion follows immediately from the fact that  $\mathcal{O}(|X_n - a_n|^m) = \mathcal{O}_P(r_n^m)$ . Part (b) can be proved similarly.  $\square$

Asymptotic normality is an important concept in estimation theory.

**Definition 2.10** (Asymptotic Normality). A sequence of random vectors  $\{\mathbf{X}_n\}$  with mean  $\boldsymbol{\mu}_n$  and covariance matrix  $\boldsymbol{\Sigma}_n$  is said to be asymptotically Gaussian (complex Gaussian), denoted by  $\mathbf{X}_n \stackrel{\Delta}{\sim} N(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$  ( $\mathbf{X}_n \stackrel{\Delta}{\sim} N_c(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ ), if  $\boldsymbol{\Sigma}_n^{-H/2}(\mathbf{X}_n - \boldsymbol{\mu}_n) \xrightarrow{D} \mathbf{Z}$ , where  $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$  ( $\mathbf{Z} \sim N_c(\mathbf{0}, \mathbf{I})$ ). It is said to be asymptotically general complex Gaussian with mean  $\boldsymbol{\mu}_n$ , covariance matrix  $\boldsymbol{\Sigma}_n$ , and complementary covariance matrix  $\tilde{\boldsymbol{\Sigma}}_n$ , denoted by  $\mathbf{X}_n \stackrel{\Delta}{\sim} N_c(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n, \tilde{\boldsymbol{\Sigma}}_n)$ , if  $[\Re(\mathbf{X}_n)^T, \Im(\mathbf{X}_n)^T]^T \stackrel{\Delta}{\sim} N(\boldsymbol{\mu}_{nr}, \boldsymbol{\Sigma}_{nr})$ , where  $(\boldsymbol{\mu}_{nr}, \boldsymbol{\Sigma}_{nr})$  and  $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_n, \tilde{\boldsymbol{\Sigma}}_n)$  satisfy (2.3.1) and (2.3.2).

Many estimators enjoy the asymptotic normality. The following proposition ensures that this property is preserved after differentiable transformations. It is a generalization of the result in [46, p. 211].

**Proposition 2.15** (Asymptotic Normality after Transformation). Let  $\{a_n\}$  and  $\{b_n\}$  be sequences of positive constants such  $a_n \rightarrow 0$  and  $b_n \rightarrow 0$  as  $n \rightarrow \infty$ . Assume that  $[a_n^{-1}(\mathbf{X}_n - \boldsymbol{\mu}_n)^T, b_n^{-1}(\mathbf{Y}_n - \mathbf{v}_n)^T]^T \stackrel{\Delta}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_n)$  with  $\boldsymbol{\Sigma}_n = \mathcal{O}(1)$  and  $\boldsymbol{\Sigma}_n \geq \boldsymbol{\Sigma}$  (i.e.,  $\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}$  is nonnegative definite) for all  $n$ , where  $\boldsymbol{\Sigma}$  is nonsingular covariance matrix. If  $\mathbf{g}_1(\cdot)$  and  $\mathbf{g}_2(\cdot)$  are real functions, continuously differentiable in a neighborhood of  $\boldsymbol{\mu}_n$  and  $\mathbf{v}_n$  respectively, with  $\nabla^T \mathbf{g}_1(\boldsymbol{\mu}_n) \neq \mathbf{0}$  and  $\nabla^T \mathbf{g}_2(\mathbf{v}_n) \neq \mathbf{0}$ , then

$$[a_n^{-1}(\mathbf{g}_1(\mathbf{X}_n) - \mathbf{g}_1(\boldsymbol{\mu}_n))^T, b_n^{-1}(\mathbf{g}_2(\mathbf{Y}_n) - \mathbf{g}_2(\mathbf{v}_n))^T]^T \stackrel{\Delta}{\sim} N(\mathbf{0}, \mathbf{J}_n \boldsymbol{\Sigma}_n \mathbf{J}_n^T),$$

where  $\mathbf{J}_n := \text{diag}\{\nabla^T \mathbf{g}_1(\boldsymbol{\mu}_n), \nabla^T \mathbf{g}_2(\mathbf{v}_n)\}$  is the Jacobian matrix.

PROOF. Because  $a_n^{-1}(\mathbf{X}_n - \boldsymbol{\mu}_n) \stackrel{\Delta}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{n1})$  and  $b_n^{-1}(\mathbf{Y}_n - \mathbf{v}_n) \stackrel{\Delta}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{n2})$  for some  $\boldsymbol{\Sigma}_{n1} = \mathcal{O}(1)$  and  $\boldsymbol{\Sigma}_{n2} = \mathcal{O}(1)$ , we can write

$$\mathbf{X}_n = \boldsymbol{\mu}_n + \mathcal{O}_P(a_n), \quad \mathbf{Y}_n = \mathbf{v}_n + \mathcal{O}_P(b_n).$$

It follows from Proposition 2.14(b) that

$$\begin{aligned}\mathbf{g}_1(\mathbf{X}_n) - \mathbf{g}_1(\boldsymbol{\mu}_n) &= \mathbf{J}_{n1}(\mathbf{X}_n - \boldsymbol{\mu}_n) + \mathcal{O}_P(a_n), \\ \mathbf{g}_2(\mathbf{Y}_n) - \mathbf{g}_2(\mathbf{v}_n) &= \mathbf{J}_{n2}(\mathbf{Y}_n - \mathbf{v}_n) + \mathcal{O}_P(b_n),\end{aligned}$$

where  $\mathbf{J}_{n1} := \nabla^T \mathbf{g}_1(\boldsymbol{\mu}_n)$  and  $\mathbf{J}_{n2} := \nabla^T \mathbf{g}_2(\mathbf{v}_n)$ . Therefore,

$$\begin{bmatrix} a_n^{-1}(\mathbf{g}_1(\mathbf{X}_n) - \mathbf{g}_1(\boldsymbol{\mu}_n)) \\ b_n^{-1}(\mathbf{g}_2(\mathbf{Y}_n) - \mathbf{g}_2(\mathbf{v}_n)) \end{bmatrix} = \mathbf{J}_n \begin{bmatrix} a_n^{-1}(\mathbf{X}_n - \boldsymbol{\mu}_n) \\ b_n^{-1}(\mathbf{Y}_n - \mathbf{v}_n) \end{bmatrix} + \mathcal{O}_P(1) \stackrel{\Delta}{\sim} \mathbf{N}(\mathbf{0}, \mathbf{J}_n \boldsymbol{\Sigma}_n \mathbf{J}_n^T),$$

where  $\mathbf{J}_n := \text{diag}(\mathbf{J}_{n1}, \mathbf{J}_{n2})$ . The proof is complete.  $\square$

The last proposition summarizes some useful techniques for proving different modes of convergence.

**Proposition 2.16** (Methods of Proving Convergence).

- (a) If  $E(|X_n|^r) = \mathcal{O}(a_n)$  for some constants  $a_n > 0$  and  $r > 0$ , then  $X_n = \mathcal{O}_P(a_n^{1/r})$ .
- (b) If  $E(|X_n|^r) \rightarrow 0$  for some constant  $r > 0$ , then  $X_n \xrightarrow{P} 0$ .
- (c) If  $\sum_{n=1}^{\infty} E(|X_n|^r) < \infty$  for some constant  $r > 0$ , then  $X_n \xrightarrow{a.s.} 0$ .
- (d) Let  $\{\mathbf{X}_n\}$  be a sequence of  $r$ -dimensional real random vectors. Then,  $\mathbf{X}_n \xrightarrow{D} \mathbf{X}$  if and only if  $\mathbf{a}^T \mathbf{X}_n \xrightarrow{D} \mathbf{a}^T \mathbf{X}$  for all  $\mathbf{a} \in \mathbb{R}^r$ . This way of proving the asymptotic distribution of random vectors is known as the Cramér-Wold device.
- (e) Let  $\{\mathbf{X}_n\}$  and  $\{\mathbf{Y}_{nm}\}$  be sequences of random vectors such that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\|\mathbf{X}_n - \mathbf{Y}_{nm}\| > \delta) = 0$$

for any constant  $\delta > 0$ . If  $\mathbf{Y}_{nm} \xrightarrow{D} \mathbf{Y}_m$  as  $n \rightarrow \infty$  for each fixed  $m$  and  $\mathbf{Y}_m \xrightarrow{D} \mathbf{Y}$  as  $m \rightarrow \infty$ , then  $\mathbf{X}_n \xrightarrow{D} \mathbf{Y}$  as  $n \rightarrow \infty$ .

PROOF. A simple proof of part (d) using the characteristic function can be found in [34, p. 383] and [46, p. 204]. Part (e) is well known in the real case [34, p. 332] [46, p. 207]. In the complex case, it can be easily proved by considering, as in the proof of Proposition 2.13, the real vectors formed by the real and imaginary parts of  $\mathbf{X}_n$ ,  $\mathbf{Y}_{nm}$ ,  $\mathbf{Y}_m$ , and  $\mathbf{Y}$ . Therefore, let us focus on (a)–(c). For any  $c > 0$ , Markov's inequality [34, p. 80] gives  $P(a_n^{-1/r} |X_n| > c) \leq E(|X_n|^r) / (a_n c^r) \leq b/c^r$ , where  $b := \sup\{E(|X_n|^r) / a_n\} < \infty$ . By taking  $c > (b/\varepsilon)^{1/r}$ , we obtain  $P(a_n^{-1/r} |X_n| > c) < \varepsilon$ . Part (a) is thus proved. Part (b) follows from part (a) with  $a_n \rightarrow 0$ . To prove part (c), we use Markov's inequality  $P(|X_n| > \delta) \leq \delta^{-r} E(|X_n|^r)$  and obtain

$$\sum_{n=1}^{\infty} P(|X_n| > \delta) \leq \delta^{-r} \sum_{n=1}^{\infty} E(|X_n|^r) < \infty$$

for any constant  $\delta > 0$ . This, according to the Borel-Cantelli lemma [34, p. 59], leads to  $P(|X_n| > \delta \text{ i.o.}) = 0$ . Hence, by definition,  $X_n \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$ .  $\square$



# Chapter 3

---

## Cramér-Rao Lower Bound

The Cramér-Rao inequality, also known as the information inequality, provides a lower limit, called the Cramér-Rao lower bound (CRLB), for the covariance matrix of unbiased estimators. It is named after the Swedish statistician Harald Cramér (1893–1985) and the Indian American statistician Calyampudi Radhakrishna Rao (1920–). An unbiased estimator that attains the CRLB is called a statistically efficient estimator because it has the smallest variance among all unbiased estimators. The CRLB has been studied for the estimation of sinusoidal parameters under various conditions. It is widely used as a performance benchmark for comparing the accuracy and statistical efficiency of different estimators.

In this chapter, we derive the CRLB under the assumption that the noise has a Gaussian distribution. We also derive some asymptotic expressions of the CRLB for large sample sizes with well-separated or closely spaced frequencies. Finally, we discuss the CRLB under the condition of nonGaussian white noise and its relationship with the CRLB under the Gaussian assumption. We show in particular that the CRLB is maximized by the Gaussian distribution among all noise distributions that have the same or smaller variance. In this sense, the Gaussian distribution can be regarded as the least favorable distribution for the noise in the estimation of sinusoidal parameters. We also show that the Laplace distribution is the least favorable distribution in another family of noise distributions.

---

### 3.1 Cramér-Rao Inequality

The Cramér-Rao inequality for the general problem of parameter estimation is stated in the following proposition. A proof can be found in Section 3.5.

**Proposition 3.1** (Cramér-Rao or Information Inequality). *Let  $\mathbf{Y}$  be a real or complex random vector that has a PDF  $p(\mathbf{y}|\boldsymbol{\vartheta})$  with respect to certain measure  $\nu$ , where  $\boldsymbol{\vartheta}$  is a real-valued parameter taking on values in  $\Theta \subset \mathbb{R}^r$ . Let  $\boldsymbol{\eta}(\boldsymbol{\vartheta}) \in \mathbb{R}^m$  be a real-valued differentiable function with Jacobian matrix  $\mathbf{J}(\boldsymbol{\vartheta}) := \nabla^T \boldsymbol{\eta}(\boldsymbol{\vartheta})$ , where  $\nabla := \partial/\partial \boldsymbol{\vartheta}$  denotes the gradient operator with respect to  $\boldsymbol{\vartheta}$ . Let  $\hat{\boldsymbol{\eta}} := \hat{\boldsymbol{\eta}}(\mathbf{Y})$  be an*

unbiased estimator of  $\boldsymbol{\eta}(\boldsymbol{\theta})$  on the basis of  $\mathbf{Y}$ . Assume that the following regularity conditions are satisfied.

- (a)  $p(\mathbf{y}|\boldsymbol{\theta})$  has a common support for all  $\boldsymbol{\theta} \in \Theta$ ;
- (b)  $\nabla p(\mathbf{y}|\boldsymbol{\theta})$  exists almost surely in  $\mathbf{y}$  for any given  $\boldsymbol{\theta} \in \Theta$ ;
- (c)  $\ell(\boldsymbol{\theta}|\mathbf{y}) := \log\{p(\mathbf{y}|\boldsymbol{\theta})\}$ , the log likelihood function, has the property

$$E\{\nabla \ell(\boldsymbol{\theta}|\mathbf{Y})\} = \mathbf{0}, \quad (3.1.1)$$

$$E\{\hat{\boldsymbol{\eta}}[\nabla \ell(\boldsymbol{\theta}|\mathbf{Y})]^T\} = \mathbf{J}(\boldsymbol{\theta}), \quad (3.1.2)$$

$$0 < \mathbf{I}(\boldsymbol{\theta}) := E\{[\nabla \ell(\boldsymbol{\theta}|\mathbf{Y})][\nabla \ell(\boldsymbol{\theta}|\mathbf{Y})]^T\} < \infty. \quad (3.1.3)$$

Then, for any constant vector  $\mathbf{a} \in \mathbb{R}^m$ ,

$$\text{Var}(\mathbf{a}^T \hat{\boldsymbol{\eta}}) = \mathbf{a}^T \text{Cov}(\hat{\boldsymbol{\eta}}) \mathbf{a} \geq \mathbf{a}^T \mathbf{J}(\boldsymbol{\theta}) \mathbf{I}(\boldsymbol{\theta})^{-1} \mathbf{J}(\boldsymbol{\theta})^T \mathbf{a}, \quad (3.1.4)$$

where the equality holds for some  $\mathbf{a} \neq \mathbf{0}$  and  $\boldsymbol{\theta} \in \Theta$  if and only if there exists a constant  $c \neq 0$  such that  $\mathbf{a}^T \{c(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}(\boldsymbol{\theta})) + \mathbf{J}(\boldsymbol{\theta}) \mathbf{I}(\boldsymbol{\theta})^{-1} \nabla \ell(\boldsymbol{\theta}|\mathbf{Y})\} = 0$  almost surely.

**Remark 3.1** The following observations are useful in verifying the regularity conditions in Proposition 3.1. First, because

$$E\{\nabla \ell(\boldsymbol{\theta}|\mathbf{Y})\} = \int \nabla p(\mathbf{y}|\boldsymbol{\theta}) d\nu,$$

the condition (3.1.1) is satisfied if the orders of differentiation and integration are interchangeable in the identity  $\nabla \int p(\mathbf{y}|\boldsymbol{\theta}) d\nu = \nabla 1 = \mathbf{0}$ . A sufficient condition that allows the interchange of orders is that there exists an integrable function  $g(\mathbf{y})$  such that  $\|\nabla p(\mathbf{y}|\boldsymbol{\theta})\| \leq g(\mathbf{y})$  for all  $\boldsymbol{\theta} \in \Theta$ . Similarly, because

$$E\{\hat{\boldsymbol{\eta}}[\nabla \ell(\boldsymbol{\theta}|\mathbf{Y})]^T\} = \int \hat{\boldsymbol{\eta}}(\mathbf{y})[\nabla p(\mathbf{y}|\boldsymbol{\theta})]^T d\nu,$$

the condition (3.1.2) is satisfied if the orders of differentiation and integration are interchangeable in the identity  $\nabla^T \int \hat{\boldsymbol{\eta}}(\mathbf{y}) p(\mathbf{y}|\boldsymbol{\theta}) d\nu = \nabla^T \boldsymbol{\eta}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})$ . A sufficient condition for the interchange of orders is that all elements in  $\hat{\boldsymbol{\eta}}(\mathbf{y})[\nabla p(\mathbf{y}|\boldsymbol{\theta})]^T$  are upper-bounded in absolute value by an integrable function of  $\mathbf{y}$  for all  $\boldsymbol{\theta} \in \Theta$ .

**Remark 3.2** The regularity conditions in Proposition 3.1 can be replaced, according to [162, p. 73], by the following conditions:

- (a)  $p(\mathbf{y}|\boldsymbol{\theta})$  is a continuous function of  $\boldsymbol{\theta} \in \Theta$  for almost every  $\mathbf{y}$ ;
- (b) for each  $\boldsymbol{\theta} \in \Theta$ , the function  $\sqrt{p(\mathbf{y}|\boldsymbol{\theta})}$  has a mean-square derivative at  $\boldsymbol{\theta}$ , i.e., there exists a square-integrable function  $\boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\theta})$  such that

$$\int |\sqrt{p(\mathbf{y}|\boldsymbol{\theta} + \boldsymbol{\delta})} - \sqrt{p(\mathbf{y}|\boldsymbol{\theta})} - \boldsymbol{\delta}^T \boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\theta})|^2 d\nu = o(\|\boldsymbol{\delta}\|^2), \quad \boldsymbol{\delta} \rightarrow \mathbf{0};$$

(c)  $\boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\vartheta})$  is mean-square continuous in  $\boldsymbol{\vartheta}$ , i.e.,

$$\int \|\boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\vartheta} + \boldsymbol{\delta}) - \boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\vartheta})\|^2 d\nu = o(1), \quad \boldsymbol{\delta} \rightarrow 0.$$

Under these conditions,  $\mathbf{I}(\boldsymbol{\vartheta}) = 4 \int \|\boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\vartheta})\|^2 d\nu$ . In addition, if  $p(\mathbf{y}|\boldsymbol{\vartheta})$  is differentiable with respect to  $\boldsymbol{\vartheta}$ , then  $\boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\vartheta}) = \nabla \sqrt{p(\mathbf{y}|\boldsymbol{\vartheta})} = \frac{1}{2} \sqrt{p(\mathbf{y}|\boldsymbol{\vartheta})} \nabla \ell(\mathbf{y}|\boldsymbol{\vartheta})$ .

The matrix  $\mathbf{J}(\boldsymbol{\vartheta})\mathbf{I}(\boldsymbol{\vartheta})^{-1}\mathbf{J}(\boldsymbol{\vartheta})^T$  in (3.1.4) is called the *Cramér-Rao lower bound* (CRLB) for estimating the parameter  $\boldsymbol{\eta}(\boldsymbol{\vartheta})$  on the basis of  $\mathbf{Y}$ . It will be denoted as CRLB( $\boldsymbol{\eta}(\boldsymbol{\vartheta})$ ). As a convention, (3.1.4) can also be written as

$$\text{Cov}(\hat{\boldsymbol{\eta}}) \geq \text{CRLB}(\boldsymbol{\eta}(\boldsymbol{\vartheta})) := \mathbf{J}(\boldsymbol{\vartheta})\mathbf{I}(\boldsymbol{\vartheta})^{-1}\mathbf{J}(\boldsymbol{\vartheta})^T, \quad (3.1.5)$$

which holds for any unbiased estimator  $\hat{\boldsymbol{\eta}}$  of  $\boldsymbol{\eta}(\boldsymbol{\vartheta})$  based on  $\mathbf{Y}$ .

In the special case where  $\boldsymbol{\eta}(\boldsymbol{\vartheta}) = \boldsymbol{\vartheta}$ , the Cramér-Rao inequality reduces to

$$\text{Cov}(\hat{\boldsymbol{\vartheta}}) \geq \text{CRLB}(\boldsymbol{\vartheta}) := \mathbf{I}(\boldsymbol{\vartheta})^{-1}, \quad (3.1.6)$$

where  $\hat{\boldsymbol{\vartheta}}$  is any unbiased estimator of  $\boldsymbol{\vartheta}$  based on  $\mathbf{Y}$ . In other words, for any constant vector  $\mathbf{a} \in \mathbb{R}^r$ , we have

$$\text{Var}(\mathbf{a}^T \hat{\boldsymbol{\vartheta}}) \geq \mathbf{a}^T \mathbf{I}(\boldsymbol{\vartheta})^{-1} \mathbf{a},$$

where the equality holds for some  $\mathbf{a} \neq \mathbf{0}$  and  $\boldsymbol{\vartheta} \in \Theta$  if and only if  $\mathbf{a}^T \{c(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) + \mathbf{I}(\boldsymbol{\vartheta})^{-1} \nabla \ell(\boldsymbol{\vartheta}|\mathbf{Y})\} = 0$  almost surely for some constant  $c \neq 0$ .

In the more general case where  $\boldsymbol{\eta}(\boldsymbol{\vartheta})$  has the same dimension as  $\boldsymbol{\vartheta}$  but  $\boldsymbol{\eta}(\boldsymbol{\vartheta}) \neq \boldsymbol{\vartheta}$  for some or all  $\boldsymbol{\vartheta} \in \Theta$ , one can regard  $\hat{\boldsymbol{\eta}}$  as a *biased* estimator of  $\boldsymbol{\vartheta}$  with the bias given by  $\mathbf{b}(\boldsymbol{\vartheta}) := \boldsymbol{\eta}(\boldsymbol{\vartheta}) - \boldsymbol{\vartheta}$ . It follows from (3.1.5) that

$$\begin{aligned} E\{(\hat{\boldsymbol{\eta}} - \boldsymbol{\vartheta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\vartheta})^T\} &= \text{Cov}(\hat{\boldsymbol{\eta}}) + \mathbf{b}(\boldsymbol{\vartheta})\mathbf{b}(\boldsymbol{\vartheta})^T \\ &\geq \mathbf{J}(\boldsymbol{\vartheta})\mathbf{I}(\boldsymbol{\vartheta})^{-1}\mathbf{J}(\boldsymbol{\vartheta})^T + \mathbf{b}(\boldsymbol{\vartheta})\mathbf{b}(\boldsymbol{\vartheta})^T, \end{aligned}$$

where  $\mathbf{J}(\boldsymbol{\vartheta}) := \nabla^T \boldsymbol{\eta}(\boldsymbol{\vartheta}) = \nabla^T \{\mathbf{b}(\boldsymbol{\vartheta}) + \boldsymbol{\vartheta}\} = \nabla^T \mathbf{b}(\boldsymbol{\vartheta}) + \mathbf{I}$ . This inequality generalizes (3.1.6) and is valid for any biased estimator of  $\boldsymbol{\vartheta}$  under the regularity conditions in Proposition 3.1. It implies that the mean-square error (MSE) of  $\mathbf{a}^T \hat{\boldsymbol{\eta}}$  as a biased estimator of  $\mathbf{a}^T \boldsymbol{\vartheta}$  satisfies

$$\text{MSE}(\mathbf{a}^T \hat{\boldsymbol{\eta}}) := E\{|\mathbf{a}^T \hat{\boldsymbol{\eta}} - \mathbf{a}^T \boldsymbol{\vartheta}|^2\} \geq \mathbf{a}^T \mathbf{J}(\boldsymbol{\vartheta})\mathbf{I}(\boldsymbol{\vartheta})^{-1}\mathbf{J}(\boldsymbol{\vartheta})^T \mathbf{a} + \{\mathbf{a}^T \mathbf{b}(\boldsymbol{\vartheta})\}^2.$$

Note that  $\mathbf{a}^T \mathbf{b}(\boldsymbol{\vartheta})$  is nothing but the bias  $\mathbf{a}^T \hat{\boldsymbol{\eta}}$  for estimating  $\mathbf{a}^T \boldsymbol{\vartheta}$ .

The matrix  $\mathbf{I}(\boldsymbol{\vartheta})$ , defined in (3.1.3), is known as *Fisher's information matrix* (FIM), named after the English statistician Ronald Aylmer Fisher (1890–1962). Owing to its inverse relationship with the CRLB, Fisher's information matrix can be interpreted as a measure of the intrinsic easiness in estimating  $\boldsymbol{\vartheta}$ , whereas its inverse, the CRLB, measures the intrinsic difficulty.

Proposition 3.1 can be easily generalized to the case where  $\boldsymbol{\theta}$  is a random vector. Indeed, if the assumptions in Proposition 3.1 are true when  $\boldsymbol{\theta}$  is treated as a deterministic variable, then the inequality (3.1.5) remains valid for the conditional covariance matrix  $\text{Cov}(\hat{\boldsymbol{\eta}}|\boldsymbol{\theta})$ , i.e.,

$$\text{Cov}(\hat{\boldsymbol{\eta}}|\boldsymbol{\theta}) \geq \mathbf{J}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})^{-1}\mathbf{J}(\boldsymbol{\theta})^T. \quad (3.1.7)$$

Therefore, the unconditional covariance matrix satisfies

$$\text{Cov}(\hat{\boldsymbol{\eta}}) = E\{\text{Cov}(\hat{\boldsymbol{\eta}}|\boldsymbol{\theta})\} \geq E\{\mathbf{J}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})^{-1}\mathbf{J}(\boldsymbol{\theta})^T\}, \quad (3.1.8)$$

where the expected value is taken with respect to  $\boldsymbol{\theta}$  as a random variable. In this case, we refer to the lower bound in (3.1.7) as the *conditional* CRLB and refer to the lower bound in (3.1.8) as the *unconditional* CRLB, or simply the CRLB.

Equipped with Proposition 3.1, the remainder of this chapter is devoted to the special case of estimating the sinusoidal parameters. For finite sample sizes, the CRLB can be derived easily under the Gaussian assumption as will be discussed in Section 3.2. This result is useful for numerical calculation, but it offers little insight except for some very special cases. An easier way of analyzing the CRLB is to make the assumption of large sample sizes, as will be discussed in Section 3.3. Under this assumption, the finite sample CRLB can be approximated by much simpler expressions from which interesting conclusions can be drawn. Simple expressions can also be obtained under the condition of nonGaussian white noise. These results lead to very interesting findings concerning the performance limit in nonGaussian cases, which will be discussed in Section 3.4.

## 3.2 CRLB for Sinusoids in Gaussian Noise

Let us begin with the case of finite sample sizes under the assumption that  $\{\epsilon_t\}$  is a zero-mean (real or complex) Gaussian process. Let us also assume that the covariance matrix  $\mathbf{R}_\epsilon$  of  $\boldsymbol{\epsilon} := [\epsilon_1, \dots, \epsilon_n]^T$  may depend on an unknown auxiliary parameter  $\boldsymbol{\eta}$  (for example, the variance of  $\epsilon_t$ ) which is not a function of the sinusoidal parameter  $\boldsymbol{\theta}$ . The problem is to find the CRLB for estimating the sinusoidal parameter  $\boldsymbol{\theta}$  in the presence of the auxiliary parameter  $\boldsymbol{\eta}$  from a data record  $\mathbf{y} := [y_1, \dots, y_n]^T$  that satisfies

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}, \quad (3.2.1)$$

where  $\mathbf{x} := [x_1, \dots, x_n]^T$  is given by (2.1.1), (2.1.2), (2.1.5), or (2.1.7).

In the real case where  $\mathbf{x}$  is given by (2.1.1) or (2.1.2), the following theorem can be established. See Section 3.5 for a proof.

**Theorem 3.1** (CRLB for RSM). *Let  $\mathbf{y}$  be given by (3.2.1) with  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\epsilon)$ . If  $\mathbf{x}$  takes the Cartesian form (2.1.1), then the CRLB for estimating  $\boldsymbol{\theta} := [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_q^T]^T$  with  $\boldsymbol{\theta}_k := [A_k, B_k, \omega_k]^T$  can be expressed as*

$$\text{CRLB}(\boldsymbol{\theta}) = (\mathbf{X}^T \mathbf{R}_\epsilon^{-1} \mathbf{X})^{-1}, \quad (3.2.2)$$

where  $\mathbf{X} := \partial \mathbf{x} / \partial \boldsymbol{\theta}^T = [\mathbf{X}_1, \dots, \mathbf{X}_q]$ ,  $\mathbf{X}_k := \partial \mathbf{x} / \partial \boldsymbol{\theta}_k^T = [\mathbf{x}_{1k}, \mathbf{x}_{2k}, \mathbf{x}_{3k}]$ , and

$$\begin{cases} \mathbf{x}_{1k} := [\cos(\omega_k), \dots, \cos(\omega_k n)]^T, \\ \mathbf{x}_{2k} := [\sin(\omega_k), \dots, \sin(\omega_k n)]^T, \\ \mathbf{x}_{3k} := -A_k [\sin(\omega_k), \dots, n \sin(\omega_k n)]^T \\ \quad + B_k [\cos(\omega_k), \dots, n \cos(\omega_k n)]^T. \end{cases} \quad (3.2.3)$$

If  $\mathbf{x}$  takes the polar form (2.1.2), then the CRLB for estimating  $\boldsymbol{\theta} := [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_q^T]^T$  with  $\boldsymbol{\theta}_k := [C_k, \phi_k, \omega_k]^T$  can be expressed as (3.2.2) with the  $\mathbf{x}_{jk}$  defined by

$$\begin{cases} \mathbf{x}_{1k} := [\cos(\omega_k + \phi_k), \dots, \cos(\omega_k n + \phi_k)]^T, \\ \mathbf{x}_{2k} := -C_k [\sin(\omega_k + \phi_k), \dots, \sin(\omega_k n + \phi_k)]^T, \\ \mathbf{x}_{3k} := -C_k [\sin(\omega_k + \phi_k), \dots, n \sin(\omega_k n + \phi_k)]^T. \end{cases} \quad (3.2.4)$$

Similarly, the following theorem can be obtained for the complex case where  $\mathbf{x}$  is given by (2.1.5) or (2.1.7). See Section 3.5 for a proof.

**Theorem 3.2** (CRLB for CSM). *Let  $\mathbf{y}$  be given by (3.2.1) with  $\boldsymbol{\epsilon} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{R}_\epsilon)$ . If  $\mathbf{x}$  takes the Cartesian form (2.1.5), then the CRLB for estimating  $\boldsymbol{\theta} := [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_p^T]^T$  with  $\boldsymbol{\theta}_k := [A_k, B_k, \omega_k]^T$  can be expressed as*

$$\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{2} \{\Re(\mathbf{X}^H \mathbf{R}_\epsilon^{-1} \mathbf{X})\}^{-1}, \quad (3.2.5)$$

where  $\mathbf{X} := \partial \mathbf{x} / \partial \boldsymbol{\theta}^T = [\mathbf{X}_1, \dots, \mathbf{X}_p]$ ,  $\mathbf{X}_k := \partial \mathbf{x} / \partial \boldsymbol{\theta}_k^T = [\mathbf{x}_{1k}, \mathbf{x}_{2k}, \mathbf{x}_{3k}]$ , and

$$\begin{cases} \mathbf{x}_{1k} := [\exp(i\omega_k), \dots, \exp(i\omega_k n)]^T, \\ \mathbf{x}_{2k} := -i \mathbf{x}_{1k}, \\ \mathbf{x}_{3k} := i(A_k - iB_k) [\exp(i\omega_k), \dots, n \exp(i\omega_k n)]^T. \end{cases} \quad (3.2.6)$$

If  $\mathbf{x}$  takes the polar form (2.1.7), then the CRLB in (3.2.5) remains valid for estimating  $\boldsymbol{\theta} := [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_p^T]^T$  with  $\boldsymbol{\theta}_k := [C_k, \phi_k, \omega_k]^T$ , provided the  $\mathbf{x}_{jk}$  are given by

$$\begin{cases} \mathbf{x}_{1k} := [\exp\{i(\omega_k + \phi_k)\}, \dots, \exp\{i(\omega_k n + \phi_k)\}]^T, \\ \mathbf{x}_{2k} := i C_k \mathbf{x}_{1k}, \\ \mathbf{x}_{3k} := i C_k [\exp\{i(\omega_k + \phi_k)\}, \dots, n \exp\{i(\omega_k n + \phi_k)\}]^T. \end{cases} \quad (3.2.7)$$

**Remark 3.3** The Jacobian matrix of the function that transforms the Cartesian parameter  $\{A_k, B_k, \omega_k\}$  into the polar parameter  $\{C_k, \phi_k, \omega_k\}$  takes the form

$$\mathbf{J}_k := \begin{bmatrix} A_k/C_k & B_k/C_k & 0 \\ B_k/C_k^2 & -A_k/C_k^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.2.8)$$

Let  $\Sigma_P$  and  $\Sigma_C$  denote the CRLB in (3.2.2) or (3.2.5) for the polar and Cartesian parameters, respectively. Then, we can write

$$\Sigma_P = \mathbf{J} \Sigma_C \mathbf{J}^T,$$

where  $\mathbf{J} := \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_q)$  for the RSM and  $\mathbf{J} := \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_p)$  for the CSM.

To gain some insights from these results, consider the following examples.

**Example 3.1** (Single Complex Sinusoid in Gaussian White Noise: Cartesian). Let  $\mathbf{x}$  be given by the Cartesian CSM (2.1.5) with  $p = 1$  and let  $\boldsymbol{\epsilon} \sim \mathcal{N}_c(\mathbf{0}, \sigma^2 \mathbf{I})$ . By Theorem 3.2, the CRLB for estimating  $\boldsymbol{\theta} := [A_1, B_1, \omega_1]^T$  takes the form

$$\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{2} \sigma^2 \{\Re(\mathbf{X}^H \mathbf{X})\}^{-1}, \quad (3.2.9)$$

where  $\mathbf{X} := [\mathbf{x}_{11}, \mathbf{x}_{21}, \mathbf{x}_{31}]$  is given by (3.2.6). It is easy to verify that

$$\begin{aligned} \mathbf{x}_{11}^H \mathbf{x}_{11} &= \mathbf{x}_{21}^H \mathbf{x}_{21} = n, & \mathbf{x}_{11}^H \mathbf{x}_{21} &= -i \mathbf{x}_{11}^H \mathbf{x}_{11} = -in, & \mathbf{x}_{11}^H \mathbf{x}_{31} &= i(A_1 - iB_1) \sum_{t=1}^n t, \\ \mathbf{x}_{21}^H \mathbf{x}_{31} &= i \mathbf{x}_{11}^H \mathbf{x}_{31} = -(A_1 - iB_1) \sum_{t=1}^n t, & \mathbf{x}_{31}^H \mathbf{x}_{31} &= (A_1^2 + B_1^2) \sum_{t=1}^n t^2. \end{aligned}$$

Therefore,

$$\Re(\mathbf{X}^H \mathbf{X}) = \begin{bmatrix} n & 0 & B_1 \sum t \\ 0 & n & -A_1 \sum t \\ B_1 \sum t & -A_1 \sum t & (A_1^2 + B_1^2) \sum t^2 \end{bmatrix}. \quad (3.2.10)$$

Because  $\sum t = \frac{1}{2} n(n+1)$ ,  $\sum t^2 = \frac{1}{6} n(n+1)(2n+1)$ , and  $C_1^2 = A_1^2 + B_1^2$ , we obtain

$$\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{\gamma_1} \begin{bmatrix} \frac{1}{2n} A_1^2 + \frac{2n+1}{n(n-1)} B_1^2 & -\frac{3(n+1)}{2n(n-1)} A_1 B_1 & -\frac{3}{n(n-1)} B_1 \\ & \frac{2n+1}{n(n-1)} A_1^2 + \frac{1}{2n} B_1^2 & \frac{3}{n(n-1)} A_1 \\ \text{symmetric} & & \frac{6}{n(n^2-1)} \end{bmatrix}, \quad (3.2.11)$$

where  $\gamma_1 := C_1^2 / \sigma^2$  is the signal-to-noise ratio (SNR) of the complex sinusoid.  $\diamond$

**Example 3.2** (Single Complex Sinusoid in Gaussian White Noise: Polar). Let  $\mathbf{x}$  be given by the polar CSM (2.1.7) with  $p = 1$  and let  $\boldsymbol{\epsilon} \sim N_c(\mathbf{0}, \sigma^2 \mathbf{I})$ . Then, by Theorem 3.2, the CRLB for estimating  $\boldsymbol{\theta} := [C_1, \phi_1, \omega_1]^T$  takes the form (3.2.9) with  $\mathbf{X} := [\mathbf{x}_{11}, \mathbf{x}_{21}, \mathbf{x}_{31}]$  given by (3.2.7). Observe that

$$\begin{aligned} \mathbf{x}_{11}^H \mathbf{x}_{11} &= n, & \mathbf{x}_{11}^H \mathbf{x}_{21} &= i C_1 \mathbf{x}_{11}^H \mathbf{x}_{11} = i C_1 n, & \mathbf{x}_{11}^H \mathbf{x}_{31} &= i C_1 \sum_{t=1}^n t, \\ \mathbf{x}_{21}^H \mathbf{x}_{21} &= C_1^2 n, & \mathbf{x}_{21}^H \mathbf{x}_{31} &= -i C_1 \mathbf{x}_{11}^H \mathbf{x}_{31} = C_1^2 \sum_{t=1}^n t, & \mathbf{x}_{31}^H \mathbf{x}_{31} &= C_1^2 \sum_{t=1}^n t^2. \end{aligned}$$

Therefore,

$$\Re(\mathbf{X}^H \mathbf{X}) = \begin{bmatrix} n & 0 & 0 \\ 0 & C_1^2 n & C_1^2 \sum t \\ 0 & C_1^2 \sum t & C_1^2 \sum t^2 \end{bmatrix}. \quad (3.2.12)$$

Straightforward calculation yields

$$\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{\gamma_1} \begin{bmatrix} \frac{1}{2n} C_1^2 & 0 & 0 \\ 0 & \frac{2n+1}{n(n-1)} & -\frac{3}{n(n-1)} \\ 0 & -\frac{3}{n(n-1)} & \frac{6}{n(n^2-1)} \end{bmatrix}, \quad (3.2.13)$$

where  $\gamma_1 := C_1^2 / \sigma^2$  is the SNR of the complex sinusoid.  $\diamond$

As can be seen from Examples 3.1 and 3.2, the CRLB for the frequency parameter in both models remains the same and depends solely on the SNR and the sample size. The CRLB under the polar CSM (2.1.7) takes a much simpler form than the CRLB under the Cartesian CSM (2.1.5). In particular, the amplitude is decoupled with the frequency and the phase in (3.2.13). Moreover, because (3.2.13) does not depend on the phase of the sinusoid, it remains valid when the phase is a random variable. This can be justified by first conditioning on the phase to obtain  $\text{Cov}(\hat{\boldsymbol{\theta}} | \phi_1) \geq \text{CRLB}(\boldsymbol{\theta})$  and then taking the expected value on both sides with respect to  $\phi_1$  to get  $\text{Cov}(\hat{\boldsymbol{\theta}}) = E\{\text{Cov}(\hat{\boldsymbol{\theta}} | \phi_1)\} \geq \text{CRLB}(\boldsymbol{\theta})$ . The ability to accommodate random phase is the major advantage of the polar system.

For large  $n$ , the CRLB for the amplitude or phase parameters takes the form  $\mathcal{O}(n^{-1})$ , but the CRLB for the frequency parameter takes the form  $\mathcal{O}(n^{-3})$ . This indicates that the frequency can be estimated with potentially much higher accuracy than the amplitude and the phase.

The higher rate of accuracy for frequency estimation may seem surprising to some who are accustomed to the usual  $\mathcal{O}(n^{-1})$  rate for parameter estimation.

An intuitive explanation of this phenomenon can be obtained by examining the sensitivity of the sinusoidal model to its parameters. It follows from (2.1.7) that

$$|dx_t| = |dC_k|, \quad |dx_t| = C_k |d\phi_k|, \quad |dx_t| = tC_k |d\omega_k|.$$

These expressions reveal that the error in  $x_t$  caused by an amplitude or phase offset remains constant over time but the error caused by a frequency offset grows linearly with time  $t$ . This means that the sinusoidal signal is more sensitive to the frequency offset than the amplitude or phase offset. As a result, an error in frequency estimation will be amplified much more strongly than the same amount of error in amplitude or phase estimation, sending a message for further improvement on frequency estimation.

The next example illustrates the interaction among multiple sinusoids in the CRLB which is absent in Examples 3.1 and 3.2.

**Example 3.3** (Two Complex Sinusoids in Gaussian White Noise). Let  $\mathbf{x}$  be given by the polar CSM (2.1.7) with  $p = 2$  and let  $\boldsymbol{\epsilon} \sim N_c(\mathbf{0}, \sigma^2 \mathbf{I})$ . In this case,  $\boldsymbol{\theta}_k := [C_k, \phi_k, \omega_k]^T$ ,  $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T]^T$ , and  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ , where  $\mathbf{X}_k := [\mathbf{x}_{1k}, \mathbf{x}_{2k}, \mathbf{x}_{3k}]$  is given by (3.2.7). As in Example 3.2, it is easy to show that

$$\Re(\mathbf{X}_1^H \mathbf{X}_1) = \mathbf{M}_{11}, \quad \Re(\mathbf{X}_2^H \mathbf{X}_2) = \mathbf{M}_{22},$$

where  $\mathbf{M}_{kk}$  takes the form (3.2.12) with  $C_k$  in place of  $C_1$ . Similarly,

$$\Re(\mathbf{X}_1^H \mathbf{X}_2) = \mathbf{M}_{12}, \quad \Re(\mathbf{X}_2^H \mathbf{X}_1) = \mathbf{M}_{21},$$

where

$$\mathbf{M}_{kk'} := \begin{bmatrix} \sum C_{kk'}(t) & C_{k'} \sum s_{kk'}(t) & C_{k'} \sum t s_{kk'}(t) \\ -C_k \sum s_{kk'}(t) & C_k C_{k'} \sum c_{kk'}(t) & C_k C_{k'} \sum t c_{kk'}(t) \\ -C_k \sum t s_{kk'}(t) & C_k C_{k'} \sum t c_{kk'}(t) & C_k C_{k'} \sum t^2 c_{kk'}(t) \end{bmatrix}, \quad (3.2.14)$$

with  $c_{kk'}(t) := \cos\{d_{kk'}(t)\}$ ,  $s_{kk'}(t) := \sin\{d_{kk'}(t)\}$ , and  $d_{kk'}(t) := (\omega_k - \omega_{k'})t + \phi_k - \phi_{k'}$ . According to Theorem 3.2,

$$\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{2} \sigma^2 \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}^{-1}. \quad (3.2.15)$$

Moreover, using the second matrix inversion formula in Lemma 12.4.1, the CRLB for estimating  $\boldsymbol{\theta}_1$  can be expressed as

$$\text{CRLB}(\boldsymbol{\theta}_1) = \frac{1}{2} \sigma^2 (\mathbf{M}_{11} - \mathbf{M}_{12} \mathbf{M}_{22}^{-1} \mathbf{M}_{21})^{-1} \geq \frac{1}{2} \sigma^2 \mathbf{M}_{11}^{-1}, \quad (3.2.16)$$

where the inequality is due to the fact that  $\mathbf{M}_{12} \mathbf{M}_{22}^{-1} \mathbf{M}_{21}$  is nonnegative definite and hence  $\mathbf{M}_{11} - \mathbf{M}_{12} \mathbf{M}_{22}^{-1} \mathbf{M}_{21} \leq \mathbf{M}_{11}$ . Note that the lower bound in (3.2.16) is

nothing but the CRLB in Example 3.2 for estimating  $\boldsymbol{\theta}_1$  in the absence of the second sinusoid. This means that the presence of the other sinusoid can raise the CRLB for estimating  $\boldsymbol{\theta}_1$ .

Similarly, it can be shown that (3.2.15) remains valid for estimating the Cartesian parameters  $\boldsymbol{\theta}_k := [A_k, B_k, \omega_k]^T$  except that  $\mathbf{X}_k$  is given by (3.2.6) and hence

$$\mathbf{M}_{kk'} = \Re \begin{bmatrix} \sum e_{kk'}(t) & -i \sum e_{kk'}(t) & i \beta_{k'} \sum t e_{kk'}(t) \\ i \sum e_{kk'}(t) & \sum e_{kk'}(t) & -\beta_{k'} \sum t e_{kk'}(t) \\ -i \beta_k \sum t e_{kk'}(t) & -\beta_k \sum t e_{kk'}(t) & \beta_k^* \beta_{k'} \sum t^2 e_{kk'}(t) \end{bmatrix}, \quad (3.2.17)$$

where  $e_{kk'}(t) := \exp\{i(\omega_{k'} - \omega_k)t\}$  and  $\beta_k := A_k - iB_k$ .

In general, for estimating  $p$  complex sinusoids in Gaussian white noise,

$$\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{2} \sigma^2 \mathbf{M}^{-1},$$

where  $\mathbf{M} := [\mathbf{M}_{kk'}]$  ( $k, k' = 1, \dots, p$ ), with  $\mathbf{M}_{kk'}$  given by (3.2.14) or (3.2.17).  $\diamond$

Example 3.3 shows that the CRLB for multiple sinusoids is generally larger than the corresponding CRLB for a single sinusoid. This can be attributed entirely to the interference among multiple sinusoids. Moreover, unlike the CRLB for a single complex sinusoid, the CRLB for multiple complex sinusoids depends on the frequency as well as the phase of the sinusoids. The dependency is through the functions  $d_{kk'}(t)$  ( $k \neq k'$ ) which are determined solely by the frequency difference  $\omega_k - \omega_{k'}$  and the phase difference  $\phi_k - \phi_{k'}$ .

Numerical studies in [89] and [317] demonstrate that the CRLB for two complex sinusoids becomes much higher than the corresponding CRLB for a single complex sinusoid as the frequency separation  $\Delta := |\omega_1 - \omega_2|$  falls below  $2\pi/n$ . This is an indication of increased difficulty in estimating closely spaced frequencies. Moreover, because a real sinusoid with frequency  $\omega_1 \in (0, \pi)$  can be regarded as two complex conjugate sinusoids with frequencies  $\omega_1$  and  $\omega_2 := -\omega_1$ , the increased difficulty also applies to the estimation of low-frequency real sinusoids. The same is true for real sinusoids with frequencies near  $\pi$ . Further analysis of the CRLB for closely spaced frequencies is provided later in Section 3.3.

Let us consider the computation of the CRLB under colored noise. To be more specific, let us focus on the real case for which the CRLB equals  $(\mathbf{X}^T \mathbf{R}_e^{-1} \mathbf{X})^{-1}$  by Theorem 3.1. Observe that direct calculation of the CRLB requires the inversion of the  $n$ -by- $n$  matrix  $\mathbf{R}_e$ , which can be burdensome when  $n$  is large. The burden can be reduced considerably if we use the Levinson-Durbin algorithm in Proposition 2.8 to compute  $\mathbf{X}^T \mathbf{R}_e^{-1} \mathbf{X}$  without explicitly inverting  $\mathbf{R}_e$ .

According to Proposition 2.8,  $\mathbf{R}_e^{-1} = \mathbf{U} \mathbf{D}^{-1} \mathbf{U}^H$ , where  $\mathbf{D} := \text{diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_{n-1}^2)$  and  $\mathbf{U} := [\varphi_{t-1, t-s}^*]$  ( $s, t = 1, \dots, n$ ;  $\varphi_{t-1, t-s} := 0$  for  $s > t$  and  $\varphi_{t-1, 0} := 1$ ). Observe that  $\mathbf{D}$  and  $\mathbf{U}$  can be computed recursively from  $\mathbf{R}_e$ . Now, let

$$\mathbf{z}_{jk} := [z_{jk}(1), \dots, z_{jk}(n)]^T := \mathbf{U}^H \mathbf{x}_{jk} \quad (j = 1, 2, 3).$$

Because  $\mathbf{U}^H = [\varphi_{s-1, s-t}]$  ( $s, t = 1, \dots, n$ ), we have

$$z_{jk}(t) = \sum_{u=0}^{t-1} \varphi_{t-1, u} x_{jk}(t-u) \quad (t = 1, \dots, n). \quad (3.2.18)$$

Moreover, let  $\mathbf{Z}_k := [\mathbf{z}_{1k}, \mathbf{z}_{2k}, \mathbf{z}_{3k}] = \mathbf{U}^H \mathbf{X}_k$  and  $\mathbf{Z} := [\mathbf{Z}_1, \dots, \mathbf{Z}_q] = \mathbf{U}^H \mathbf{X}$ . Then,

$$\mathbf{X}^H \mathbf{R}_\epsilon^{-1} \mathbf{X} = \mathbf{X}^H \mathbf{U} \mathbf{D}^{-1} \mathbf{U}^H \mathbf{X} = \mathbf{Z}^H \mathbf{D}^{-1} \mathbf{Z} = [\mathbf{I}_{kk'}] \quad (k, k' = 1, \dots, q), \quad (3.2.19)$$

where

$$\mathbf{I}_{kk'} := \mathbf{Z}_k^H \mathbf{D}^{-1} \mathbf{Z}_{k'} = [z_{jk}^H \mathbf{D}^{-1} z_{j'k'}] \quad (j, j' = 1, 2, 3), \quad (3.2.20)$$

$$z_{jk}^H \mathbf{D}^{-1} z_{j'k'} = \sum_{t=1}^n z_{jk}^*(t) z_{j'k'}(t) / \sigma_{t-1}^2 \quad (j, j' = 1, 2, 3). \quad (3.2.21)$$

Equations (3.2.18)–(3.2.21), together with the recursion in Proposition 2.8, constitute a fast algorithm for computing  $\mathbf{X}^H \mathbf{R}_\epsilon^{-1} \mathbf{X}$ . Note that if  $\{\epsilon_t\}$  is an AR process, then the matrices  $\mathbf{I}_{kk'}$  can be obtained directly from the AR parameters, rather than the covariance matrix  $\mathbf{R}_\epsilon$ , by using the algorithm in Proposition 2.9.

Finally, let us investigate an important special case where the signal frequencies are known but the amplitude and phase parameters are unknown. In this case, it suffices to use the CRLB given by the following corollary for the amplitude and phase estimation. The result is stated without proof, as it is a direct result of Theorem 3.1 for the RSM and of Theorem 3.2 for the CSM.

**Corollary 3.1** (CRLB for Amplitude and Phase Parameters). *Let  $\mathbf{y}$  be given by (3.2.1). If  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}_\epsilon)$ , then, the CRLB for estimating  $\boldsymbol{\theta} := [A_1, B_1, \dots, A_q, B_q]^T$  in the RSM (2.1.1) or  $\boldsymbol{\theta} := [C_1, \phi_1, \dots, C_q, \phi_q]^T$  in the RSM (2.1.2) can be expressed as (3.2.2) with  $\mathbf{X}_k := [\mathbf{x}_{1k}, \mathbf{x}_{2k}]$  given by (3.2.3) or (3.2.4), respectively. If  $\boldsymbol{\epsilon} \sim N_c(\mathbf{0}, \mathbf{R}_\epsilon)$ , then the CRLB for estimating  $\boldsymbol{\theta} := [A_1, B_1, \dots, A_p, B_p]^T$  in the CSM (2.1.5) or  $\boldsymbol{\theta} := [C_1, \phi_1, \dots, C_p, \phi_p]^T$  in the CSM (2.1.7) takes the form (3.2.5) with  $\mathbf{X}_k := [\mathbf{x}_{1k}, \mathbf{x}_{2k}]$  given by (3.2.6) or (3.2.7), respectively.*

**Remark 3.4** In the complex case, it is sometimes convenient to arrange the amplitude parameters such that

$$\boldsymbol{\theta} := [A_1, \dots, A_p, B_1, \dots, B_p]^T = [\Re(\boldsymbol{\beta})^T, -\Im(\boldsymbol{\beta})^T]^T,$$

where  $\beta_k := A_k - iB_k$ . For estimating this parameter, the expression in (3.2.5) remains true except that  $\mathbf{X} := [\mathbf{F}, -i\mathbf{F}]$ , where  $\mathbf{F} := [\mathbf{f}(\omega_1), \dots, \mathbf{f}(\omega_p)]$  and  $\mathbf{f}(\omega_k) := [\exp(i\omega_k), \dots, \exp(in\omega_k)]^T = \mathbf{x}_{1k}$ . In the special case of Gaussian white noise, it reduces to  $\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{2} \sigma^2 \boldsymbol{\Sigma}$ , where

$$\boldsymbol{\Sigma} := \begin{bmatrix} \Re(\mathbf{F}^H \mathbf{F}) & \Im(\mathbf{F}^H \mathbf{F}) \\ -\Im(\mathbf{F}^H \mathbf{F}) & \Re(\mathbf{F}^H \mathbf{F}) \end{bmatrix}^{-1} = \begin{bmatrix} \Re\{(\mathbf{F}^H \mathbf{F})^{-1}\} & \Im\{(\mathbf{F}^H \mathbf{F})^{-1}\} \\ -\Im\{(\mathbf{F}^H \mathbf{F})^{-1}\} & \Re\{(\mathbf{F}^H \mathbf{F})^{-1}\} \end{bmatrix}.$$

The second expression can be verified by simple matrix algebra.

The following example illustrates this result.

**Example 3.4** (CRLB for Amplitude Parameters: Single Complex Sinusoid in Gaussian White Noise). Consider the complex case where  $\boldsymbol{\epsilon} \sim N_c(\mathbf{0}, \sigma^2 \mathbf{I})$  and  $p = 1$ . Because  $\mathbf{f}^H(\omega_1)\mathbf{f}(\omega_1) = n$ , it follows from Remark 3.4 that the CRLB for estimating  $\boldsymbol{\theta} := [A_1, B_1]^T$  can be expressed as  $\text{CRLB}(\boldsymbol{\theta}) = \frac{1}{2}n^{-1}\sigma^2\mathbf{I}$ .  $\diamond$

It is interesting to compare this example with Example 3.1 where the signal frequency is unknown and jointly estimated with the amplitude parameters. In Example 3.1, the CRLB for estimating  $A_1$ , which is the (1,1) entry of the matrix in (3.2.11), can be different from that for estimating  $B_1$ , which is the (2,2) entry. Furthermore, it is easy to verify that the CRLB for estimating  $A_1$  is strictly less than the CRLB for estimating  $B_1$  if  $|A_1| > |B_1|$ , regardless of the sample size and the SNR. Conversely, the CRLB for estimating  $B_1$  is strictly less than the CRLB for estimating  $A_1$  if  $|B_1| > |A_1|$ . In other words, the larger of the two parameters has a smaller CRLB. This is in contrast with Example 3.4 where the signal frequency is known. In Example 3.4, the CRLB is the same for both  $A_1$  and  $B_1$  regardless of their magnitude. Moreover, in the case of  $A_1 = B_1 \neq 0$  and large  $n$ , the CRLB in Example 3.1 for estimating the amplitude parameters is approximately equal to 2.5 times the CRLB in Example 3.4; in the case of  $A_1 = 0$  and  $B_1 \neq 0$ , the CRLB in Example 3.1 for estimating  $A_1$  is approximately 4 times the CRLB in Example 3.4. These comparisons illustrate the increased difficulty in estimating the amplitude parameters when the frequencies are unknown.

---

### 3.3 Asymptotic CRLB for Sinusoids in Gaussian Noise

As demonstrated by Example 3.3, the interaction among the sinusoids makes it difficult to grasp the full implication of the CRLB for multiple sinusoids. This is true even in the simple case of a single real sinusoid. However, when the sample size is large, the CRLB can be approximated by much simpler expressions.

Let us begin by continuing the discussion in Examples 3.1 and 3.2.

**Example 3.5** (Single Complex Sinusoid in Gaussian White Noise). Consider the case discussed in Example 3.1. let  $\mathbf{K}_n := \text{diag}(n^{1/2}, n^{1/2}, n^{3/2})$ . Then, for large  $n$ , the CRLB in (3.2.11) for estimating  $\boldsymbol{\theta} = \boldsymbol{\theta}_1 := [A_1, B_1, \omega_1]^T$  can be expressed as

$$\text{CRLB}(\boldsymbol{\theta}) = \mathbf{K}_n^{-1} \{ \boldsymbol{\Gamma}(\boldsymbol{\theta}) + \mathcal{O}(n^{-1}) \} \mathbf{K}_n^{-1}, \quad (3.3.1)$$

where

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}) := \frac{1}{2} \gamma_1^{-1} \boldsymbol{\Lambda}_C(\boldsymbol{\theta}_1)$$

and

$$\mathbf{\Lambda}_C(\boldsymbol{\theta}_k) := \begin{bmatrix} A_k^2 + 4B_k^2 & -3A_kB_k & -6B_k \\ -3A_kB_k & 4A_k^2 + B_k^2 & 6A_k \\ -6B_k & 6A_k & 12 \end{bmatrix}. \quad (3.3.2)$$

Similarly, for the case discussed in Example 3.2, the CRLB in (3.2.13) for estimating  $\boldsymbol{\theta} = \boldsymbol{\theta}_1 := [C_1, \phi_1, \omega_1]^T$  can be expressed as (3.3.1) with

$$\mathbf{\Gamma}(\boldsymbol{\theta}) := \frac{1}{2}\gamma_1^{-1}\mathbf{\Lambda}_P(\boldsymbol{\theta}_1),$$

where

$$\mathbf{\Lambda}_P(\boldsymbol{\theta}_k) := \begin{bmatrix} C_k^2 & 0 & 0 \\ 0 & 4 & -6 \\ 0 & -6 & 12 \end{bmatrix}. \quad (3.3.3)$$

In both cases,  $\gamma_1 := C_1^2/\sigma^2$  is the SNR of the complex sinusoid.  $\diamond$

The matrix  $\mathbf{K}_n^{-1}\mathbf{\Gamma}(\boldsymbol{\theta})\mathbf{K}_n^{-1}$  in (3.3.1) is called the asymptotic CRLB or ACRLB. For any unbiased estimator  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$ , it follows from (3.3.1) that

$$\text{Cov}(\mathbf{K}_n\hat{\boldsymbol{\theta}}) = \mathbf{K}_n\text{Cov}(\hat{\boldsymbol{\theta}})\mathbf{K}_n \geq \mathbf{K}_n\text{CRLB}(\boldsymbol{\theta})\mathbf{K}_n = \mathbf{\Gamma}(\boldsymbol{\theta}) + \mathcal{O}(n^{-1}).$$

Hence the matrix  $\mathbf{\Gamma}(\boldsymbol{\theta})$  serves as an asymptotic lower bound for the covariance matrix of the normalized estimator  $\mathbf{K}_n\hat{\boldsymbol{\theta}}$ . The different orders of magnitude in the CRLB are made very clear through the normalizing factors in  $\mathbf{K}_n$ : for amplitude and phase estimation, the normalizing factor is equal to  $n^{1/2}$ ; for frequency estimation, it is equal to  $n^{3/2}$ . When the normalizing factors are self-evident, we may also refer to  $\mathbf{\Gamma}(\boldsymbol{\theta})$  as the ACRLB.

Next, consider the more interesting case in Example 3.3. Because two sinusoids are involved in this example, the degree of frequency separation becomes a key factor in determining the ACRLB. The frequency separation is meaningful only if it is measured with respect to the sample size  $n$ . To facilitate this analysis, let us assume that the frequencies may depend on  $n$  with the possibility that the distance between them approaches zero as  $n \rightarrow \infty$ .

**Example 3.6** (Two Complex Sinusoids in Gaussian White Noise). In Example 3.3 for the polar parameters, define  $\Delta := \min\{|\omega_1 - \omega_2|, 2\pi - |\omega_1 - \omega_2|\}$  and assume

$$\lim_{n \rightarrow \infty} n\Delta = \infty. \quad (3.3.4)$$

By Lemma 12.1.4, the matrix  $\mathbf{M}_{12}$ , defined by (3.2.14), can be expressed as

$$\mathbf{M}_{12} = \begin{bmatrix} \mathcal{O}(\Delta^{-1}) & \mathcal{O}(\Delta^{-1}) & \mathcal{O}(n\Delta^{-1}) \\ \mathcal{O}(\Delta^{-1}) & \mathcal{O}(\Delta^{-1}) & \mathcal{O}(n\Delta^{-1}) \\ \mathcal{O}(n\Delta^{-1}) & \mathcal{O}(n\Delta^{-1}) & \mathcal{O}(n^2\Delta^{-1}) \end{bmatrix} = \mathbf{K}_n\mathcal{O}(n^{-1}\Delta^{-1})\mathbf{K}_n.$$

Similarly,  $\mathbf{M}_{21} = \mathbf{K}_n \mathcal{O}(n^{-1} \Delta^{-1}) \mathbf{K}_n$ . Moreover, because  $\mathbf{M}_{kk}$  takes the form (3.2.12) with  $C_k$  in place of  $C_1$ , it follows that

$$\mathbf{M}_{kk} = \mathbf{K}_n \{\mathbf{W}_k + \mathcal{O}(n^{-1})\} \mathbf{K}_n,$$

where

$$\mathbf{W}_k := \begin{bmatrix} 1 & 0 & 0 \\ 0 & C_k^2 & \frac{1}{2} C_k^2 \\ 0 & \frac{1}{2} C_k^2 & \frac{1}{3} C_k^2 \end{bmatrix}. \quad (3.3.5)$$

By matrix algebra, we obtain

$$\mathbf{W}_k^{-1} = (1/C_k^2) \mathbf{\Lambda}_P(\boldsymbol{\theta}_k).$$

Let  $\mathbf{\Gamma}(\boldsymbol{\theta}_k) := \frac{1}{2} \sigma^2 \mathbf{W}_k^{-1} = \frac{1}{2} \gamma_k^{-1} \mathbf{\Lambda}_P(\boldsymbol{\theta}_k)$ , where  $\gamma_k := C_k^2 / \sigma^2$ . Then, the CRLB for estimating  $\boldsymbol{\theta} := [\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T]^T$  with  $\boldsymbol{\theta}_k := [C_k, \phi_k, \omega_k]^T$  can be expressed as

$$\begin{aligned} \text{CRLB}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{K}_n^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_n^{-1} \end{bmatrix} \left\{ \begin{bmatrix} \mathbf{\Gamma}(\boldsymbol{\theta}_1) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}(\boldsymbol{\theta}_2) \end{bmatrix} + \mathcal{O}(n^{-1} \Delta^{-1}) \right\} \begin{bmatrix} \mathbf{K}_n^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_n^{-1} \end{bmatrix} \\ &= \mathbf{K}^{-1} \{\mathbf{\Gamma}(\boldsymbol{\theta}) + \mathcal{O}(n^{-1} \Delta^{-1})\} \mathbf{K}^{-1}, \end{aligned} \quad (3.3.6)$$

where  $\mathbf{K} := \text{diag}(\mathbf{K}_n, \mathbf{K}_n)$  and  $\mathbf{\Gamma}(\boldsymbol{\theta}) = \text{diag}\{\mathbf{\Gamma}(\boldsymbol{\theta}_1), \mathbf{\Gamma}(\boldsymbol{\theta}_2)\}$ .

The CRLB for the Cartesian parameters  $\boldsymbol{\theta}_k := [A_k, B_k, \omega_k]^T$  also takes the form (3.3.6) except that  $\mathbf{\Gamma}(\boldsymbol{\theta}_k) = \frac{1}{2} \gamma_k^{-1} \mathbf{\Lambda}_C(\boldsymbol{\theta}_k)$ . Indeed, by Lemma 12.1.5,  $\mathbf{M}_{kk'}$  defined by (3.2.17) takes the form  $\mathbf{M}_{kk'} = \mathbf{K}_n \mathcal{O}(n^{-1} \Delta^{-1}) \mathbf{K}_n$  for  $k \neq k'$ . Moreover, it follows from (3.2.10) that  $\mathbf{M}_{kk} = \mathbf{K}_n \{\mathbf{W}_k + \mathcal{O}(n^{-1})\} \mathbf{K}_n$ , where

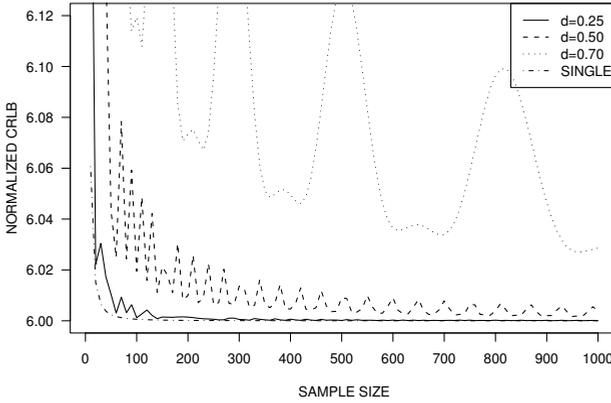
$$\mathbf{W}_k := \begin{bmatrix} 1 & 0 & \frac{1}{2} B_k \\ 0 & 1 & -\frac{1}{2} A_k \\ \frac{1}{2} B_k & -\frac{1}{2} A_k & \frac{1}{3} (A_k^2 + B_k^2) \end{bmatrix}. \quad (3.3.7)$$

Because

$$\mathbf{W}_k^{-1} = (1/C_k^2) \mathbf{\Lambda}_C(\boldsymbol{\theta}_k),$$

we obtain (3.3.6) with  $\mathbf{\Gamma}(\boldsymbol{\theta}_k) := \frac{1}{2} \sigma^2 \mathbf{W}_k^{-1} = \frac{1}{2} \gamma_k^{-1} \mathbf{\Lambda}_C(\boldsymbol{\theta}_k)$  and  $\gamma_k := C_k^2 / \sigma^2$ .  $\diamond$

This example shows that for large sample sizes the CRLB for two sinusoids is decoupled so that the CRLB obtained for single complex sinusoids in Examples 3.1 and 3.5 can be used separately to approximate the CRLB for each sinusoid. The key prerequisite for the decoupling is the frequency separation condition (3.3.4). It is required to ensure that  $\mathbf{K}_n^{-1} \mathbf{M}_{kk'} \mathbf{K}_n^{-1} \rightarrow \mathbf{0}$  as  $n \rightarrow \infty$  for  $k \neq k'$ . For finite sample sizes, the accuracy of the approximation depends on  $\Delta$ . In fact,



**Figure 3.1.** Plot of  $n^3\text{CRLB}(\omega_1)$  as a function of  $n$  in the case of two unit-amplitude complex sinusoids in Gaussian white noise ( $\omega_1 = 2\pi \times 0.1$ ,  $\omega_2 = \omega_1 + \Delta$ ,  $\phi_1 = 0$ ,  $\phi_2 = \pi/2$ ,  $\gamma_1 = \gamma_2 = 1$ ) under different frequency separation conditions of the form  $\Delta = 2\pi/n^d$ . Solid line,  $d = 0.25$ ; dashed line,  $d = 0.5$ ; dotted line,  $d = 0.7$ . The dash-dotted line depicts the normalized single-sinusoid CRLB which approaches its asymptotic value 6 as  $n$  grows.

if  $n\Delta$  does not approach infinity fast enough as  $n$  increases, the approximation can be very poor. This point is illustrated by the example shown in Figure 3.1.

In this example, the frequency separation parameter  $\Delta$  takes the form  $\Delta = 2\pi/n^d$  with  $d = 0.25, 0.5, 0.7$ , so the condition (3.3.4) is always satisfied. But, as we can see from Figure 3.1, the accuracy of the single-sinusoid CRLB as an approximation to the exact CRLB deteriorates rapidly when  $d$  gets closer to unity. This result also serves as a confirmation to the earlier comment on Example 3.5 that the CRLB increases rapidly as frequency separation decreases. Note that the oscillations of the CRLB as a function of the sample size are due entirely to the interplay between the sample size and the phases of the sinusoids.

Now, let us investigate the case of real sinusoids through two examples.

**Example 3.7** (Single Real Sinusoid in Gaussian White Noise: Cartesian). Let  $\mathbf{x}$  be given by (2.1.1) with  $q = 1$  and let  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ . By Theorem 3.1, the CRLB for estimating  $\boldsymbol{\theta} := [A_1, B_1, \omega_1]^T$  takes the form  $\text{CRLB}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$ , where  $\mathbf{X} := [\mathbf{x}_{11}, \mathbf{x}_{21}, \mathbf{x}_{31}]$  is given by (3.2.3). Let  $c_t := \cos(\omega_1 t)$  and  $s_t := \sin(\omega_1 t)$ . Then,

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \sum c_t^2 & \sum c_t s_t & \sum t(-A_1 c_t s_t + B_1 c_t^2) \\ & \sum s_t^2 & \sum t(-A_1 s_t^2 + B_1 c_t s_t) \\ \text{symmetric} & & \sum t^2(-A_1 s_t + B_1 c_t)^2 \end{bmatrix}. \quad (3.3.8)$$

This matrix is not as simple as its counterpart (3.2.10) for a single complex sinusoid. Although an explicit expression of the CRLB can be obtained by inverting the matrix, let us derive a simpler expression for large sample sizes.

Toward that end, assume that  $\omega_1$  is not too close to 0 or  $\pi$  in the sense that  $\Delta := \min\{\omega_1, \pi - \omega_1\}$  satisfies (3.3.4). Then, as  $n \rightarrow \infty$ , Lemma 12.1.5 gives

$$\left\{ \begin{array}{l} \sum_{t=1}^n t^r c_t s_t = \mathcal{O}(n^r \Delta^{-1}) \quad (r = 0, 1, 2), \\ \sum_{t=1}^n c_t^2 = \frac{1}{2}n + \mathcal{O}(\Delta^{-1}), \quad \sum_{t=1}^n s_t^2 = \frac{1}{2}n + \mathcal{O}(\Delta^{-1}), \\ \sum_{t=1}^n t c_t^2 = \frac{1}{4}n^2 + \mathcal{O}(n\Delta^{-1}), \quad \sum_{t=1}^n t s_t^2 = \frac{1}{4}n^2 + \mathcal{O}(n\Delta^{-1}), \\ \sum_{t=1}^n t^2 c_t^2 = \frac{1}{6}n^3 + \mathcal{O}(n^2 \Delta^{-1}), \quad \sum_{t=1}^n t^2 s_t^2 = \frac{1}{6}n^3 + \mathcal{O}(n^2 \Delta^{-1}). \end{array} \right. \quad (3.3.9)$$

Therefore, (3.3.8) can be written as

$$\mathbf{X}^T \mathbf{X} = \mathbf{K}_n \left\{ \frac{1}{2} \mathbf{W}_1 + \mathcal{O}(n^{-1} \Delta^{-1}) \right\} \mathbf{K}_n,$$

where  $\mathbf{W}_1$  takes the form (3.3.7) with  $k = 1$ . This result leads to

$$\text{CRLB}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} = \mathbf{K}_n^{-1} \{ \boldsymbol{\Gamma}(\boldsymbol{\theta}) + \mathcal{O}(n^{-1} \Delta^{-1}) \} \mathbf{K}_n^{-1}, \quad (3.3.10)$$

where  $\boldsymbol{\Gamma}(\boldsymbol{\theta}) := 2\sigma^2 \mathbf{W}_1^{-1} = \gamma_1^{-1} \boldsymbol{\Lambda}_C(\boldsymbol{\theta}_1)$ , with  $\boldsymbol{\Lambda}_C(\boldsymbol{\theta}_1)$  given by (3.3.2) and with  $\gamma_1 := \frac{1}{2}(A_1^2 + B_1^2)/\sigma^2$  being the SNR of the real sinusoid.  $\diamond$

**Example 3.8** (Single Real Sinusoid in Gaussian White Noise: Polar). Let  $\mathbf{x}$  be given by (2.1.2) with  $q = 1$  and let  $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ . By Theorem 3.1, the CRLB for estimating  $\boldsymbol{\theta} := [C_1, \phi_1, \omega_1]^T$  is  $\text{CRLB}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$ , where  $\mathbf{X} := [\mathbf{x}_{11}, \mathbf{x}_{21}, \mathbf{x}_{31}]$  is given by (3.2.4). Let  $c_t := \cos(\omega_1 t + \phi_1)$  and  $s_t := \sin(\omega_1 t + \phi_1)$ . Then,

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \sum c_t^2 & -C_1 \sum c_t s_t & -C_1 \sum t c_t s_t \\ & C_1^2 \sum s_t^2 & C_1^2 \sum t s_t^2 \\ \text{symmetric} & & C_1^2 \sum t^2 s_t^2 \end{bmatrix}. \quad (3.3.11)$$

Under the assumption that  $\Delta := \min\{\omega_1, \pi - \omega_1\}$  satisfies (3.3.4), the results in (3.3.9) remain valid for  $c_t = \cos(\omega_1 t + \phi_1)$  and  $s_t = \sin(\omega_1 t + \phi_1)$ . Substituting these expressions in (3.3.11) yields

$$\mathbf{X}^T \mathbf{X} = \mathbf{K}_n \left\{ \frac{1}{2} \mathbf{W}_1 + \mathcal{O}(n^{-1} \Delta^{-1}) \right\} \mathbf{K}_n,$$

where  $\mathbf{W}_1$  is defined by (3.3.5) for  $k = 1$ . Therefore, we obtain (3.3.10) with  $\boldsymbol{\Gamma}(\boldsymbol{\theta}) := 2\sigma^2 \mathbf{W}_1^{-1} = \gamma_1^{-1} \boldsymbol{\Lambda}_P(\boldsymbol{\theta}_1)$ , where  $\boldsymbol{\Lambda}_P(\boldsymbol{\theta}_1)$  is defined by (3.3.3) and  $\gamma_1 := \frac{1}{2} C_1^2 / \sigma^2$  is the SNR of the real sinusoid.  $\diamond$

The following theorem summarizes the results for the ACRLB in the general case of multiple real or complex sinusoids. Given Examples 3.5–3.8, it suffices to state the theorem without proof.

**Theorem 3.3** (Asymptotic CRLB: Gaussian White Noise). *Let  $\mathbf{y}$  be given by (3.2.1) with  $\mathbf{x}$  satisfying (2.1.1), (2.1.2), (2.1.5), or (2.1.7), and with  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$  in the real case and  $\boldsymbol{\epsilon} \sim \mathcal{N}_c(\mathbf{0}, \sigma^2 \mathbf{I})$  in the complex case. Assume that (3.3.4) is satisfied by*

$$\Delta := \begin{cases} \min_{k \neq k'} \{|\omega_k - \omega_{k'}|, \omega_k, \pi - \omega_k\} & \text{for RSM,} \\ \min_{k \neq k'} \{|\omega_k - \omega_{k'}|, 2\pi - |\omega_k - \omega_{k'}|\} & \text{for CSM.} \end{cases} \quad (3.3.12)$$

Let  $\boldsymbol{\theta}_k \in \mathbb{R}^3$  denote the parameters of the  $k$ th sinusoid and define

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}_k) := \begin{cases} \gamma_k^{-1} \boldsymbol{\Lambda}_C(\boldsymbol{\theta}_k) & \text{for Cartesian RSM (2.1.1),} \\ \gamma_k^{-1} \boldsymbol{\Lambda}_P(\boldsymbol{\theta}_k) & \text{for polar RSM (2.1.2),} \\ \frac{1}{2} \gamma_k^{-1} \boldsymbol{\Lambda}_C(\boldsymbol{\theta}_k) & \text{for Cartesian CSM (2.1.5),} \\ \frac{1}{2} \gamma_k^{-1} \boldsymbol{\Lambda}_P(\boldsymbol{\theta}_k) & \text{for polar CSM (2.1.7),} \end{cases} \quad (3.3.13)$$

where  $\boldsymbol{\Lambda}_C(\boldsymbol{\theta}_k)$  and  $\boldsymbol{\Lambda}_P(\boldsymbol{\theta}_k)$  are given by (3.3.2) and (3.3.3), respectively, and where  $\gamma_k := \frac{1}{2} C_k^2 / \sigma^2$  in the real case and  $\gamma_k := C_k^2 / \sigma^2$  in the complex case. Then, as  $n \rightarrow \infty$ , the CRLB for estimating  $\boldsymbol{\theta} := [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_r^T]^T$  ( $r := q$  for the RSM and  $r := p$  for the CSM) can be expressed as

$$\text{CRLB}(\boldsymbol{\theta}) = \mathbf{K}^{-1} \{ \boldsymbol{\Gamma}(\boldsymbol{\theta}) + \mathcal{O}(n^{-1} \Delta^{-1}) \} \mathbf{K}^{-1},$$

where  $\boldsymbol{\Gamma}(\boldsymbol{\theta}) := \text{diag}\{\boldsymbol{\Gamma}(\boldsymbol{\theta}_1), \dots, \boldsymbol{\Gamma}(\boldsymbol{\theta}_r)\}$  and  $\mathbf{K} := \text{diag}\{\mathbf{K}_n, \dots, \mathbf{K}_n\}$ .

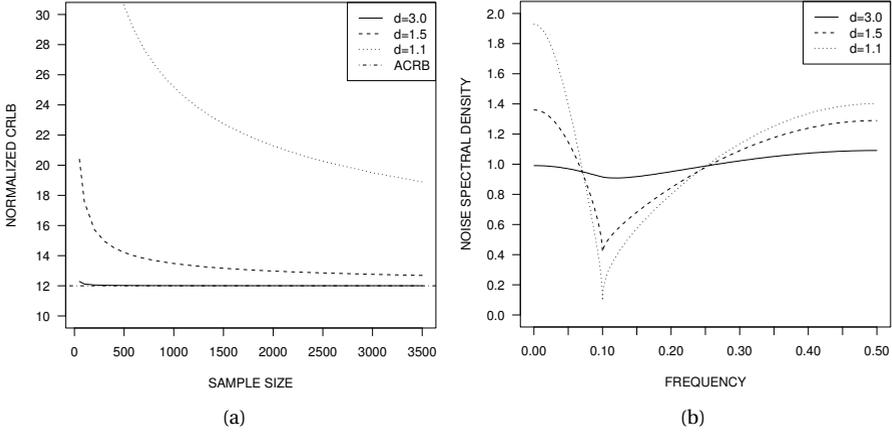
Next, we consider the ACRLB under the condition of Gaussian colored noise with continuous spectrum. In this case, the following theorem can be established as a generalization of Theorem 3.3. See Section 3.5 for a proof.

**Theorem 3.4** (Asymptotic CRLB: Gaussian Colored Noise). *Let the conditions of Theorem 3.3 be satisfied except that  $\{\epsilon_t\}$  is a stationary real or complex Gaussian process with mean zero and SDF  $f_\epsilon(\omega)$ , where  $f_\epsilon(\omega)$  is a continuous function with  $f_0 := \min_\omega f_\epsilon(\omega) > 0$ . Then, as  $n \rightarrow \infty$ , the CRLB for estimating  $\boldsymbol{\theta}$  defined in Theorem 3.3 can be expressed as*

$$\text{CRLB}(\boldsymbol{\theta}) = \mathbf{K}^{-1} \{ \boldsymbol{\Gamma}(\boldsymbol{\theta}) + \mathcal{O}(1) \} \mathbf{K}^{-1}, \quad (3.3.14)$$

where the matrix  $\boldsymbol{\Gamma}(\boldsymbol{\theta})$  is the same as in Theorem 3.3 except that  $\gamma_k$  in (3.3.13) is defined with  $f_\epsilon(\omega_k)$  in place of  $\sigma^2$ , i.e.,  $\gamma_k := \frac{1}{2} C_k^2 / f_\epsilon(\omega_k)$  in the real case and  $\gamma_k := C_k^2 / f_\epsilon(\omega_k)$  in the complex case.

Theorem 3.4 reveals that the ACRLB depends on the noise spectrum solely through its values at the signal frequencies, so the noise spectrum elsewhere has zero contribution to the asymptotic performance limit. An intuitive explanation for this interesting result is as follows: Because the sinusoids are extremely well localized in the frequency domain, any reasonable frequency estimator, let alone



**Figure 3.2.** (a) Plot of  $n^3\text{CRLB}(\omega_1)$  as a function of  $n$  in the case of a single real sinusoid in Gaussian colored noise with SDF  $f_c(\omega) = 1/|\sum_{j=0}^{\infty} \varphi_j \exp(-ij\omega)|^2$ , where  $\varphi_j := (j+1)^{-d} \cos(j\omega_1)$ ,  $\omega_1 = 2\pi \times 0.1$ , and  $\gamma_1 = 1$ . Solid line,  $d = 3$ ; dashed line,  $d = 1.5$ ; dotted line,  $d = 1.1$ . The dash-dotted line depicts the asymptotic value of the normalized CRLB. (b) Plot of  $f_c(\omega)$  as a function of the normalized frequency  $f := \omega/(2\pi)$ .

the optimal ones, must have the capability of suppressing the noise outside a small neighborhood of the signal frequencies. Furthermore, as the sample size grows, the small neighborhood must shrink toward zero in order to suppress the noise most effectively. Therefore, in the limit, only the values of the noise spectrum at the signal frequencies have impact on the estimation accuracy.

It is worth pointing out that although the continuity of the noise spectrum suffices for the validity of Theorem 3.4, the degree of smoothness of the noise spectrum plays an important role in determining the accuracy of the ACRLB as an approximation to the CRLB for finite sample sizes. To demonstrate this point, consider the example shown in Figure 3.2.

In this example, the noise is a linear process of the form

$$\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \zeta_{t-j}, \tag{3.3.15}$$

where  $\sum |\psi_j| < \infty$  and  $\{\zeta_t\} \sim \text{IID}(0, \sigma^2)$ . By Proposition 2.4,

$$f_c(\omega) := \sigma^2 |\Psi(\omega)|^2,$$

where  $\Psi(\omega) := \sum \psi_j \exp(-ij\omega)$ . Take  $\{\zeta_t\} \sim \text{GWN}(0, 1)$  and  $\Psi(\omega) = 1/\Phi(\omega)$ , where  $\Phi(\omega) := \sum_{j=0}^{\infty} \varphi_j \exp(-ij\omega)$  with  $\varphi_j := (j+1)^{-d} \cos(0.2\pi j)$ . In other words, the noise is a special Gaussian AR( $\infty$ ) process. The smoothness of the noise spectrum is controlled by the parameter  $d$  which takes values 1.1, 1.5, and 3. As

shown in Figure 3.2(b), the noise spectrum is very smooth with  $d = 3$  and becomes less smooth at the signal frequency  $\omega_1 = 2\pi \times 0.1$  as  $d$  decreases. The exact CRLB shown in Figure 3.2(a) is computed directly from  $\{\varphi_j\}$  by using the algorithm in Proposition 2.9 and an AR(6000) truncation of the AR( $\infty$ ) noise spectrum. As we can see, the ACRLB, which equals 12, is an excellent approximation to the exact CRLB when  $d = 3$ , even for small sample sizes. But, as  $d$  decreases, especially when  $d = 1.1$ , it becomes poorer even for very large sample sizes. A similar, but less dramatic, effect is also observed (not shown) when the noise is an AR(2) process with a sharp spectral peak located at the signal frequency.

As an important special case, the ACRLB for estimating the amplitude and phase parameters can be easily derived from Theorem 3.3 and Theorem 3.4. The following assertion is stated without proof.

**Corollary 3.2** (Asymptotic CRLB for the Amplitude and Phase Parameters). *Let the conditions of Theorem 3.3 or 3.4 be satisfied. For the Cartesian RSM and CSM, let  $\boldsymbol{\theta}_k := [A_k, B_k]^T$ . For the polar RSM and CSM, let  $\boldsymbol{\theta}_k := [C_k, \phi_k]^T$ . Define*

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}_k) := \begin{cases} \boldsymbol{\Gamma}_C(\boldsymbol{\theta}_k) & \text{for Cartesian RSM (2.1.1),} \\ \boldsymbol{\Gamma}_P(\boldsymbol{\theta}_k) & \text{for polar RSM (2.1.2),} \\ \frac{1}{4}\boldsymbol{\Gamma}_C(\boldsymbol{\theta}_k) & \text{for Cartesian CSM (2.1.5),} \\ \frac{1}{4}\boldsymbol{\Gamma}_P(\boldsymbol{\theta}_k) & \text{for polar CSM (2.1.7),} \end{cases} \quad (3.3.16)$$

where

$$\begin{aligned} \boldsymbol{\Gamma}_C(\boldsymbol{\theta}_k) &:= \text{diag}\{2f_\epsilon(\omega_k), 2f_\epsilon(\omega_k)\}, \\ \boldsymbol{\Gamma}_P(\boldsymbol{\theta}_k) &:= \text{diag}\{2f_\epsilon(\omega_k), 2f_\epsilon(\omega_k)/C_k^2\}. \end{aligned}$$

Then, as  $n \rightarrow \infty$ , the CRLB for estimating  $\boldsymbol{\theta} := [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_r^T]^T$  ( $r := q$  for the RSM and  $r := p$  for the CSM) can be expressed as

$$\text{CRLB}(\boldsymbol{\theta}) = n^{-1}\{\boldsymbol{\Gamma}(\boldsymbol{\theta}) + \mathcal{O}(1)\}, \quad (3.3.17)$$

where  $\boldsymbol{\Gamma}(\boldsymbol{\theta}) := \text{diag}\{\boldsymbol{\Gamma}(\boldsymbol{\theta}_1), \dots, \boldsymbol{\Gamma}(\boldsymbol{\theta}_r)\}$ .

**Remark 3.5** By comparing Corollary 3.2 with Theorem 3.3 and Theorem 3.4, we can see that the same remark we made at the end of Section 3.2 regarding the CRLB for estimating the amplitude parameters  $(A_k, B_k)$  in the complex case applies to the ACRLB for estimating  $(A_k, B_k)$  in the real case.

So far, the signal frequencies are assumed to satisfy (3.3.4). It is under this condition that the sinusoids become decoupled in  $\boldsymbol{\Gamma}(\boldsymbol{\theta})$  which takes a block-diagonal form. In the remainder of this section, let us investigate a case where the condition (3.3.4) is not satisfied.