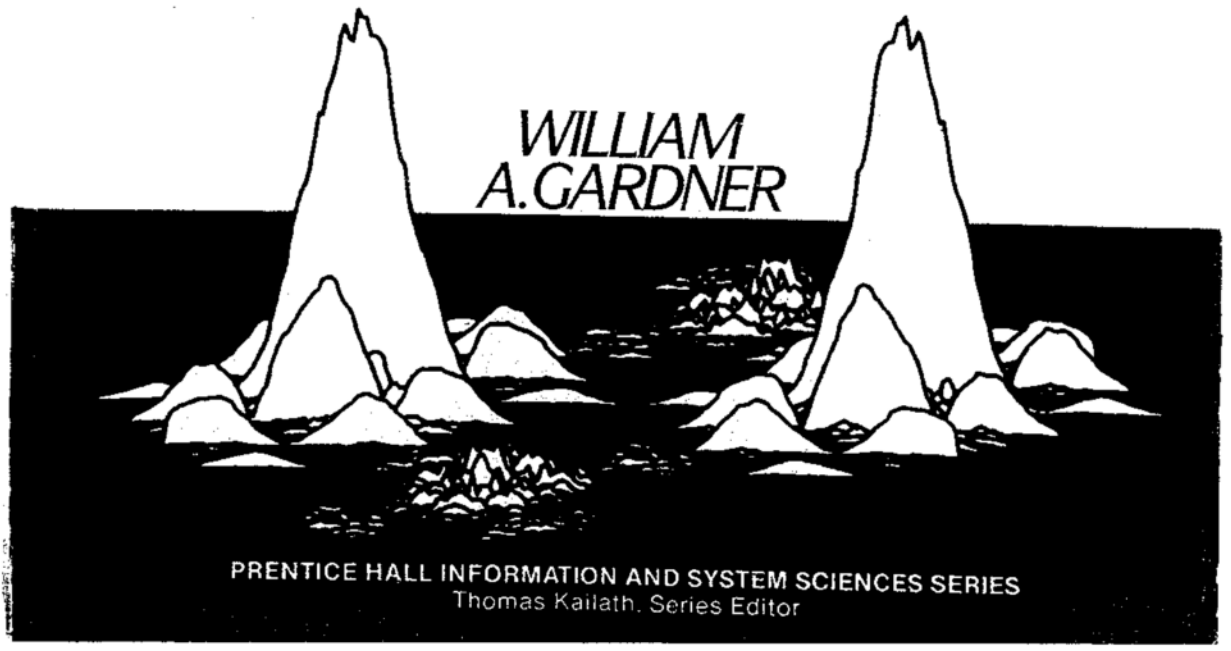
STATISTICAL SPECTRAL ANALYSIS

A NONPROBABILISTIC THEORY

WILLIAM A. GARDNER



PRENTICE HALL INFORMATION AND SYSTEM SCIENCES SERIES
Thomas Kailath, Series Editor

CONTENTS

FOREWORD

A good deal of our statistical theory, although it is mathematical in nature, originated not in mathematics but in problems of astronomy, geomagnetism and meteorology: examples of fruitful problems in these subjects have included the clustering of stars, also galaxies, on the celestial sphere, tidal analysis, the correlation of fluctuations of the Earth's magnetic field with other solar-terrestrial effects, and the determination of seasonal variations and climatic trends from weather data. All three of these fields are observational. Great figures of the past, such as C. F. Gauss (1777–1855) (who worked with both astronomical and geomagnetic data, and discovered the method of least square fitting of data, the normal error distribution, and the Fast Fourier Transform algorithm), have worked on observational data analysis and have contributed much to our body of knowledge on time series and randomness.

Much other theory has come from gambling, gunnery, and agricultural research, fields that are experimental. Measurements of the fall of shot on a firing range will reveal a pattern that can be regarded as a sample from a normal distribution in two dimensions, together with whatever bias is imposed by pointing and aiming, the wind, air temperature, atmospheric pressure and Earth rotation. The deterministic part of any one of these influences may be characterized with further precision by further firing tests. In the experimental sciences, as well as in the observational, great names associated with the foundations of statistics and probability also come to mind.

Experimental subjects are traditionally distinguished from observational ones by the property that conditions are under the control of the experimenter. The design of experiments leads the experimenter to the idea of an ensemble, or random process, an abstract probabilistic creation illustrated by the bottomless

barrel of well-mixed marbles that is introduced in elementary probability courses. A characteristic feature of the contents of such a barrel is that we know in advance how many marbles there are of each color, because it is we who put them in; thus, a sample set that is withdrawn after stirring must be compatible with the known mix.

The observational situation is quite unlike this. Our knowledge of what is in the barrel, or of what Nature has in store for us, is to be deduced from what has been observed to come out of the barrel, to date. The probability distribution, rather than being a given, is in fact to be intuited from experience. The vital stage of connecting the world of experience to the different world of conventional probability theory may be glossed over when foreknowledge of the barrel and its contents—a probabilistic model—are posited as a point of departure. Many experimental situations are like this observational one.

The theory of signal processing, as it has developed in electrical and electronics engineering, leans heavily toward the random process, defined in terms of probability distributions applicable to ensembles of sample signal waveforms. But many students who are adept at the useful mathematical techniques of the probabilistic approach and quite at home with joint probability distributions are unable to make even a rough drawing of the underlying sample waveforms. The idea that the sample waveforms are the deterministic quantities being modeled somehow seems to get lost.

When we examine the pattern of fall of shot from a gun, or the pattern of bullet holes in a target made by firing from a rifle clamped in a vise, the distribution can be characterized by its measurable centroid and second moments or other spread parameters. While such a pattern is necessarily discrete, and never much like a normal distribution, we have been taught to picture the pattern as a sample from an infinite ensemble of such patterns; from this point of view the pattern will of course be compatible with the adopted parent population, as with the marbles. In this probabilistic approach, to simplify mathematical discussion, one begins with a model, or specification of the continuous probability distribution from which each sample is supposed to be drawn. Although this probability distribution is not known, one is comforted by the assurance that it is potentially approachable by expenditure of more ammunition. But in fact it is not.

The assumption of randomness is an expression of ignorance. Progress means the identification of systematic effects which, taken as a whole, may initially give the appearance of randomness or unpredictability. Continuing to fire at the target on a rifle range will not refine the probability distribution currently in use but will reveal, to a sufficiently astute planner of experiments, that air temperature, for example, has a determinate effect which was always present but was previously accepted as stochastic. After measurement, to appropriate precision, temperature may be allowed for. Then a new probability model may be constructed to cover the effects that remain unpredictable.

Many authors have been troubled by the standard information theory approach via the random process or probability distribution because it seems to put the cart before the horse. Some sample parameters such as mean amplitudes or powers, mean durations and variances may be known, to precision of measurement,

but if we are to go beyond pure mathematical deduction and make advances in the realm of phenomena, theory should start from the data. To do otherwise risks failure to discover that which is not built into the model. Estimating the magnitude of an earthquake from seismograms, assessing a stress-test cardiogram, or the pollutant in a stormwater drain, are typical exercises where noise, systematic or random, is to be fought against. Problems on the forefront of development are often ones where the probability distributions of neither signal nor noise is known; and such distributions may be essentially unknowable because repetition is impossible. Thus, any account of measurement, data processing, and interpretation of data that is restricted to probabilistic models leaves something to be desired.

The techniques used in actual research with real data do not loom large in courses in probability. Professor Gardner's book demonstrates a consistent approach from data, those things which in fact are given, and shows that analysis need not proceed from assumed probability distributions or random processes. This is a healthy approach and one that can be recommended to any reader.

Ronald N. Bracewell
Stanford, California

PREFACE

This book grew out of an enlightening discovery I made a few years ago, as a result of a long-term attempt to strengthen the tenuous conceptual link between the abstract probabilistic theory of cyclostationary stochastic processes and empirical methods of signal processing that accommodate or exploit periodicity in random data. After a period of unsatisfactory progress toward using the concept of *ergodicity*[1] to strengthen this link, it occurred to me (perhaps wishfully) that the abstraction of the probabilistic framework of the theory might not be necessary. As a first step in pursuing this idea, I set out to clarify for myself the extent to which the probabilistic framework is needed to explain various well-known concepts and methods in the theory of stationary stochastic processes, especially spectral analysis theory. To my surprise, I discovered that all the concepts and methods of empirical spectral analysis can be explained in a more straightforward fashion in terms of a deterministic theory, that is, a theory based on time-averages of a single time-series rather than ensemble-averages of hypothetical random samples from an abstract probabilistic model. To be more specific, I found that the fundamental concepts and methods of empirical spectral analysis can be explained without use of probability calculus or the concept of probability and that probability calculus, which is indeed useful for quantification of the notion of *degree of randomness* or *variability,* can be based on time-averages of a single time-series without any use of the concept or theory of a stochastic process defined on an abstract probability space. This seemed to be of such fundamental importance for practicing engineers and scientists and so intuitively satisfying that I felt it must already be in the literature.

To put my discovery in perspective, I became a student of the history of the subject. I found that the apparent present-day complacence with the abstraction of the probabilistic theory of stochastic processes, introduced by A. N. Kolmogorov in 1941, has been the trend for about 40 years. Nevertheless, I found also that

[1] *Ergodicity* is the property of a mathematical model for an infinite set of time-series that guarantees that an ensemble average over the infinite set will equal an infinite time average over one member of the set.

many probabilists throughout this period, including Kolmogorov himself, have felt that the concept of *randomness* should be defined as directly as possible, and that from this standpoint it seems artificial to conceive of a time-series as a sample of a stochastic process. (The first notable attempt to set up the probability calculus more directly was the theory of *Collectives* introduced by Von Mises in 1919; the mathematical development of such alternative approaches is traced by P. R. Masani [Masani 1979].) In the engineering literature, I found that in the early 1960s two writers, D. G. Brennan [Brennan 1961] and E. M. Hofstetter [Hofstetter 1964], had made notable efforts to explain that much of the theory of stationary time-series need not be based on the abstract probabilistic theory of stochastic processes and then linked with empirical method only through the abstract concept of ergodicity, but rather that a probabilistic theory based directly on time-averages will suffice; however, they did not pursue the idea that a theory of empirical spectral analysis can be developed without any use of probability. Similarly, the more recent book by D. R. Brillinger on time-series [Brillinger 1975] briefly explains precisely how the probabilistic theory of stationary time-series can be based on time-averages, but it develops the theory of empirical spectral analysis entirely within the probabilistic framework. Likewise, the early engineering book by R. B. Blackman and J. W. Tukey [Blackman and Tukey 1958] on spectral analysis defines an idealized spectrum in terms of time-averages but then carries out all analysis of measurement techniques within the probabilistic framework of stochastic processes. In the face of this 40-year trend, I was perplexed to find that the one most profound and influential work in the entire history of the subject of empirical spectral analysis, Norbert Wiener's *Generalized Harmonic Analysis*, written in 1930 [Wiener 1930], was entirely devoid of probability theory; and yet I found only one book written since then for engineers or scientists that provides more than a brief mention of Wiener's deterministic theory. All other such books that I found emphasize the probabilistic theory of A. N. Kolmogorov usually to the complete exclusion of Wiener's deterministic theory. This one book was written by a close friend and colleague of Wiener's, Y. W. Lee, in 1960 [Lee 1960]. Some explanation of this apparent historical anomaly is given by P. R. Masani in his recent commentary on Wiener's *Generalized Harmonic Analysis* [Masani 1979]: "The quick appearance of the Birkhoff ergodic theorem and the Kolmogorov theory of stochastic processes after the publication of Wiener's *Generalized Harmonic Analysis* created an intellectual climate favoring stochastic analysis rather than generalized harmonic analysis." But Masani goes on to explain that the current opinion, that Wiener's 1930 memoir [Wiener 1930] marks the culmination of generalized harmonic analysis and its supercession by the more advanced theories of stochastic processes, is questionable on several counts, and he states that the "integrity and wisdom" in the attitude expressed in the early 1960s by Kolmogorov suggesting a possible return to the ideas of Von Mises ". . . should point the way toward the future. Side by side with the vigorous pursuit of the theory of stochastic processes, must coexist a more direct process-free [deterministic] inquiry of randomness of different classes of functions." In an even stronger stance, T. L. Fine in the concluding section of his book *Theories of Probability* [Fine, 1973] states "Judging from the present confused

status of probability theory, the time is at hand for those concerned about the characterization of chance and uncertainty and the design of inference and decision-making systems to reconsider their long-standing dependence on the traditional statistical and probabilistic methodology. . . . Why not ignore the complicated and hard to justify probability-statistics structure and proceed 'directly' to those, perhaps qualitative, assumptions that characterize our source of random phenomena, the means at our disposal, and our task?''

As a result of my discovery and my newly gained historical perspective, I felt compelled to write a book that would have the same goals, in principle, as many existing books on spectral analysis—to present a general theory and methodology for empirical spectral analysis—but that would present a more relevant and palatable (for many applications) deterministic theory following Wiener's original approach rather than the conventional probabilistic theory. As the book developed, I continued to wonder about the apparent fact that no one in the 50 years since Wiener's memoir had considered such a project worthy enough to pursue. However, as I continued to search the literature, I found that one writer, J. Kampé de Fériet, did make some progress along these lines in a tutorial paper [Kampé de Fériet 1954], and other authors have contributed to development of deterministic theories of related subjects in time-series analysis, such as linear prediction and extrapolation [Wold 1948], [Finch 1969], [Fine 1970]. Furthermore, as the book progressed and I observed the favorable reactions of my students and colleagues, my conviction grew to the point that I am now convinced that it is generally beneficial for students of the subject of empirical spectral analysis to study the deterministic theory *before* studying the more abstract probabilistic theory.

When I had completed most of the development for a book on a deterministic theory of empirical spectral analysis of stationary time-series, I was then able to return to the original project of presenting the results of my research work on cyclostationary time-series but within a nonprobabilistic framework. Once I started, it quickly became apparent that I was able to conceptualize intuitions, hunches, conjectures, and so forth far more clearly than before when I was laboring within the probabilistic framework. The original relatively fragmented research results on cyclostationary stochastic processes rapidly grew into a comprehensive theory of random time-series from periodic phenomena that is every bit as satisfying as the theory of random time-series from constant phenomena (stationary time-series). This theory, which brings to light the fundamental role played by spectral correlation in the study of periodic phenomena, is presented in Part II.

Part I of this book is intended to serve as both a graduate-level textbook and a technical reference. The only prerequisite is an introductory course on Fourier analysis. However, some prior exposure to probability would be helpful for Section B in Chapter 5 and Section A in Chapter 15. The body of the text in Part I presents a thorough development of fundamental concepts and results in the theory of statistical spectral analysis of empirical time-series from *constant phenomena,* and a brief overview is given at the end of Chapter 1. Various supplements that expand on topics that are in themselves important or at least

illustrative but that are not essential to the foundation and framework of the theory, are included in appendices and exercises at the ends of chapters.

Part II of this book, like Part I, is intended to serve as both textbook and reference, and the same unifying philosophical framework developed in Part I is used in Part II. However, unlike Part I, the majority of concepts and results presented in Part II are new. Because of the novelty of this material, a brief preview is given in the Introduction to Part II. The only prerequisite for Part II is Part I.

The focus in this book is on fundamental concepts, analytical techniques, and basic empirical methods. In order to maintain a smooth flow of thought in the development and presentation of concepts that steadily build on one another, various derivations and proofs are omitted from the text proper, and are put into the exercises, which include detailed hints and outlines of solution approaches. Depending on the students' background, the instructor can either assign these as homework exercises, or present them in the lectures. Because the treatment of experimental design and applications is brief and is also relegated to the exercises and concise appendices, some readers might desire supplements on these topics.

REFERENCES

BLACKMAN, R. B. and J. W. TUKEY. 1958. *The Measurement of Power Spectra*. New York: American Telephone and Telegraph Co.

BRENNAN, D. G. 1961. Probability theory in communication system engineering. Chapter 2 in *Communication System Theory*. Ed. E. J. Baghdady, New York: McGraw-Hill.

BRILLINGER, D. R. 1975. *Time Series*. New York: Holt, Rinehart and Winston.

FINCH, P. D. 1969. Linear least squares prediction in non-stochastic time-series. *Advances in Applied Prob.* 1:111–22.

FINE, T. L. 1970. Extrapolation when very little is known about the source. *Information and Control*. 16:331–359.

FINE, T. L. 1973. *Theories of Probability: An Examination of Foundations*. New York: Academic Press.

HOFSTETTER, E. M. 1964. Random processes. Chapter 3 in *The Mathematics of Physics and Chemistry*, vol. II. Ed. H. Margenau and G. M. Murphy. Princeton, N.J.: D. Van Nostrand Co.

KAMPÉ DE FÉRIET, J. 1954. Introduction to the statistical theory of turbulence, I and II. *J. Soc. Indust. Appl. Math.* 2, Nos. 1 and 3:1–9 and 143–74.

LEE, Y. W. 1960. *Statistical Theory of Communication*. New York: John Wiley & Sons.

MASANI, P. R. 1979. "Commentary on the memoir on generalized harmonic analysis." pp. 333–379 in *Norbert Wiener: Collected Works, Volume II*. Cambridge, Mass.: Massachusetts Institute of Technology.

WIENER, N. 1930. Generalized harmonic analysis. *Acta Mathematika*. 55:117–258.

WOLD, H. O. A. 1948. On prediction in stationary time-series. *Annals of Math Stat.* 19:558–67.

William A. Gardner

ACKNOWLEDGMENTS

I would like to express my gratitude to Mr. William A. Brown for his important technical and moral support in the early stages of this project, and to Professors Enders A. Robinson, Ronald N. Bracewell, and James L. Massey for their enthusiastic encouragement. I also would like to express my appreciation to Professor Thomas Kailath for bringing to my attention several early fundamental papers on nonprobabilistic statistical theory. In addition, I would like to thank Professor Herschel H. Loomis and Dr. Crawford W. Scott for their interest in applications of the theory in Part II, and the resultant financial support, and Messrs. Brian G. Agee, William A. Brown, and Chihkang Chen for their participation in applying the theory of Part II. Credit is due Messrs. Brown and Chen for their contributions to some of the technical material in Chapter 12, and also special credit is due Mr. Brown for his major contribution to Chapter 15, especially section B. Further credit is due Messrs. Chen and Brown for their substantial joint effort to produce the many excellent computer-generated graphs. It is a pleasure to express my appreciation to Mrs. Patty A. Gemulla and Mrs. Marion T. Franke for their excellent job of typing the manuscript, Dr. Sheldon N. Salinger for critically reading the manuscript, Mr. Randy S. Roberts and Messrs. Brown and Chen for their substantial proofreading efforts, and many other past and present students for their feedback and assistance. My deepest gratitude is expressed to my wife, Nancy, for her patience, understanding, and support throughout this demanding project and the years of work leading up to it.

William A. Gardner

GLOSSARIES

GLOSSARY OF NOTATIONS AND TERMINOLOGY FOR WINDOW FUNCTIONS

$a_T(t)$	General data-tapering window of unity height and approximate width T.
$A_{1/T}(f)$	Fourier transform of $a_T(t)$.
$E(f)$	Effective spectral smoothing window.
$g_{\Delta t}(t)$	General time-smoothing window of unity area and approximate width Δt.
$G_{1/\Delta t}(f)$	Fourier transform of $g_{\Delta t}(t)$.
$h_{1/\Delta f}(\tau)$	General autocorrelation-tapering window of unity height and approximate width $1/\Delta f$.
$H_{\Delta f}(f)$	General spectral smoothing window of unity area and approximate width Δf; Fourier transform of $h_{1/\Delta f}(\tau)$.
$u_T(t)$	Rectangle window of unity area and width T.
$v_T(t)$	Triangle window of unity area and base width $2T$.
$w_T(t)$	Sinc window of unity area and null-to-null width $2T$.
$z_T(t)$	Squared sinc window of unity area and null-to-null width $2T$.

GLOSSARY OF NOTATIONS AND TERMINOLOGY FOR CORRELATIONS AND SPECTRA IN PART I[1]

$r_h(\tau)$	*Finite autocorrelation of* h: (36), Chapter 2.
$\tilde{r}_h(\tau)$	*Finite autocorrelation* of discrete-time h: (86), Chapter 3.

[1] Some notation that is used only within a single chapter is not included in this glossary.

$R_{x_T}(t, \tau)$	*Time-variant correlogram* (*time-variant finite-time autocorrelation*) of segment of x of duration T: (19), Chapter 2; for tapered data, (20), Chapter 2.
$\tilde{R}_{x_T}(t, \tau)$	*Time-variant correlogram* (*time-variant finite-time autocorrelation*) of segment of discrete-time x of duration T: (67), Chapter 2, (44), Chapter 6.
$R_x(t, \tau)_T$	*Time-variant finite-average autocorrelation* of x: (21), Chapter 2.
$\hat{R}_x(\tau)$	*Limit autocorrelation* of x: (6), Chapter 1; see also (31), Chapter 2.
$\tilde{R}_x(\tau)$	*Limit autocorrelation* of discrete-time x: (79), Chapter 3.
$\mathscr{R}_x(t, \tau)$	*Probabilistic instantaneous autocorrelation* of x: (7), Chapter 8.
$S_{x_T}(t, f)$	*Time-variant periodogram* (*time-variant finite-time spectrum*) of segment of x of duration T: (1), (2), Chapter 2; for tapered data, (1), (11), Chapter 2; Fourier transform of $R_{x_T}(t, \tau)$.
$\tilde{S}_{x_T}(t, f)$	*Time-variant periodogram* (*time-variant finite-time spectrum*) of segment of discrete-time x of duration T: (69b), Chapter 3, (30), Chapter 6; Fourier series transform of $\tilde{R}_{x_T}(t, \tau)$.
$\mathscr{S}_{x_T}(t, f)$	*Expected time-variant periodogram* (*expected time-variant finite-time spectrum*) of segment of x of duration T: (3), Chapter 8.
$S_x(t, f)_T$	*Time-variant pseudospectrum* of x: (22), Chapter 2; Fourier transform of $R_x(t, \tau)_T$.
$S_{x_{1/\Delta f}}(t, f)_{\Delta t}$	*Temporally smoothed spectrum* of tapered x: (11), Chapter 3, (1), Chapter 4.
$\tilde{S}_{x_{1/\Delta f}}(t, f)_{\Delta t}$	*Temporally smoothed spectrum* of tapered discrete-time x: (29), Chapter 6.
$S_{x_{\Delta t}}(t, f)_{\Delta f}$	*Spectrally smoothed spectrum* of x: (16)–(17), Chapter 3, (2) and (21), Chapter 4.
$\tilde{S}_{x_{\Delta t}}(t, f)_{\Delta f}$	*Spectrally smoothed spectrum* of discrete-time x: (36) and (43), Chapter 6.
$S_x(t, f)_{1/\Delta f,\, \Delta t}$	*Temporally smoothed pseudospectrum* of x: (3), Chapter 4.
$S_x(t, f)_{\Delta t,\, \Delta f}$	*Spectrally smoothed pseudospectrum* of x: (4) and (22), Chapter 4.
$\hat{S}_x(f)$	*Limit spectrum* of x: (26), Chapter 3; Fourier transform of $\hat{R}_x(\tau)$.
$\tilde{S}_x(f)$	*Limit spectrum* of discrete-time x: (69a), Chapter 3.
$\mathscr{S}_x(t, f)$	*Probabilistic instantaneous spectrum* of x: (6), Chapter 8.
$X_T(t, f)$	*Time-variant finite-time complex spectrum* of segment of x of duration T: (2), (11), Chapter 2; *complex demodulate,* (44)–(45), Chapter 4.
$\overline{X}_T(t, f)$	*Normalized time-variant finite-time complex spectrum* of segment of x of duration T: (27), Chapter 5.

$\vec{X}_T(t, f)$	*Time-variant finite-time complex spectrum* of segment of discrete-time x of length $N = 1 + T/T_s$: (53), Chapter 2, (28), Chapter 6.
$x_T(t, f)$	*Local sine wave component* of x: (44), Chapter 4.
$\tilde{x}_T(t, f)$	*Local sine wave component* of discrete-time x: (26), Chapter 6; for tapered data, (31), Chapter 6.

GLOSSARY OF NOTATIONS AND TERMINOLOGY FOR CROSS CORRELATIONS AND CROSS SPECTRA IN PART I

$\hat{C}_{xy}(f)$	*Complex coherence function* of x and y: (32), Chapter 7.
$R_{xy_T}(t, \tau)$	*Time-variant cross correlogram* (*time-variant finite-time cross correlation*) of segments of x and y of duration T: (5), Chapter 7.
$R_{xy}(t, \tau)_T$	*Time-variant finite-average cross correlation* of x and y: (13), Chapter 7.
$\hat{R}_{xy}(\tau)$	*Limit cross correlation* of x and y: (18), (20), Chapter 7.
$S_{xy_T}(t, f)$	*Time-variant cross periodogram* (*time-variant finite-time cross spectrum*) of segments of x and y of duration T: (3), Chapter 7; Fourier transform of $R_{xy_T}(t, \tau)$.
$S_{xy}(t, f)_T$	*Time-variant pseudo–cross spectrum* of x and y: (12), Chapter 7; Fourier transform of $R_{xy}(t, \tau)_T$.
$S_{xy_{1/\Delta f}}(t, f)_{\Delta t}$	*Temporally smoothed cross spectrum* of x and y: (9), Chapter 7.
$S_{xy_{\Delta t}}(t, f)_{\Delta f}$	*Spectrally smoothed cross spectrum* of x and y: (8), Chapter 7.
$S_{xy}(t, f)_{1/\Delta f,\, \Delta t}$	*Temporally smoothed pseudo–cross spectrum* of x and y: (11), Chapter 7.
$S_{xy}(t, f)_{\Delta t,\, \Delta f}$	*Spectrally smoothed pseudo–cross spectrum* of x and y: (10), Chapter 7.
$\hat{S}_{xy}(f)$	*Limit cross spectrum* of x and y: (17), Chapter 7; Fourier transform of $\hat{R}_{xy}(\tau)$.

GLOSSARY OF NOTATIONS AND TERMINOLOGY FOR CYCLIC CORRELATIONS AND CYCLIC SPECTRA IN PART II

$\hat{C}_x^{\alpha}(f)$	*Spectral autocoherence* of x: (35), Chapter 10.
$\hat{C}_{xy}^{\alpha}(f)$	*Cyclic cross coherence* of x and y, (45b), Chapter 14.
$r_h^{\alpha}(\tau)$	*Finite cyclic autocorrelation* of h: (137b), Chapter 11.
$R_{x_T}^{\alpha}(t, \tau)$	*Time-variant cyclic cross correlogram* of segment of x of duration T: (12), Chapter 11.

$R_x^\alpha(t, \tau)_T$	*Time-variant finite-average cyclic autocorrelation* of x: (14), Chapter 11.
$R_{xy_T}^\alpha(t, \tau)$	*Time-variant cyclic cross correlogram* of segments of x and y of duration T: (67), Chapter 11.
$\hat{R}_x^\alpha(\tau)$	*Limit cyclic autocorrelation* of x: (25), Chapter 10; (13), (15), Chapter 11.
$\bar{R}_x^\alpha(\tau)$	*Limit cyclic autocorrelation* of discrete-time x: (109), (111), Chapter 11.
$\hat{R}_{xy}^\alpha(\tau)$	*Limit cyclic cross correlation* of x and y: (69), Chapter 11.
$\hat{R}_x(t, \tau; T_0)$	*Limit periodic autocorrelation* of x (period $= T_0$): (99), (102), Chapter 10.
$\hat{R}_x(t, \tau)$	*Limit almost periodic autocorrelation* of x with multiple periodicity: (100), (103), Chapter 10; or *limit periodic autocorrelation* of x with single periodicity: (23), (24), Chapter 10.
$S_{x_T}^\alpha(t, f)$	*Time-variant cyclic periodogram* of segment of x of duration T: (8), (11), Chapter 11.
$S_{xy_T}^\alpha(t, f)$	*Time-variant cyclic cross periodogram* of segments of x and y of duration T: (65), (66), Chapter 11.
$S_{x_{1/\Delta f}}^\alpha(t, f)_{\Delta t}$	*Temporally smoothed cyclic spectrum* of x: (1), Chapter 11, (4a), Chapter 13.
$\bar{S}_{x_{1/\Delta f}}^\alpha(t, f)_{\Delta t}$	*Temporally smoothed cyclic spectrum* of discrete-time x: (6), Chapter 13.
$S_{x_{\Delta t}}^\alpha(t, f)_{\Delta f}$	*Spectrally smoothed cyclic spectrum* of x: (4b), Chapter 13.
$\bar{S}_{x_{\Delta t}}^\alpha(t, f)_{\Delta f}$	*Spectrally smoothed cyclic spectrum* of discrete-time x: (5), Chapter 13.
$\hat{S}_x^\alpha(f)$	*Limit cyclic spectrum* of x: (30), Chapter 10, (43), Chapter 11.
$\bar{S}_x^\alpha(f)$	*Limit cyclic spectrum* of discrete-time x: (110), (112), Chapter 11.
$\hat{S}_{xy}^\alpha(f)$	*Limit cyclic cross spectrum* of x and y: (63), (68), Chapter 11.
$\hat{S}_x(t, f; T_0)$	*Limit periodic spectrum* of x (period $= T_0$); Fourier transform of $\hat{R}_x(t, \tau; T_0)$.
$\hat{S}_x(t, f)$	*Limit almost periodic spectrum* of x with multiple periodicity: (106), (107), Chapter 10; or *limit periodic spectrum* of x with single periodicity: (58), Chapter 10.

Part I

CONSTANT PHENOMENA

INTRODUCTION

The subject of Part I is the statistical spectral analysis of empirical time-series. The term *empirical* indicates that the time-series represents data from a physical phenomenon; the term *spectral analysis* denotes decomposition of the time-series into sine wave components; and the term *statistical* indicates that the squared magnitude of each measured or computed sine wave component, or the product of pairs of such components, is averaged to reduce random effects in the data that mask the spectral characteristics of the phenomenon under study. The purpose of Part I is to present a comprehensive *deterministic* theory of statistical spectral analysis and thereby to show that contrary to popular belief, the theoretical foundations of this subject need not be based on *probabilistic* concepts. The motivation for Part I is that for many applications the conceptual gap between practice and the deterministic theory presented herein is narrower and thus easier to bridge than is the conceptual gap between practice and the more abstract probabilistic theory. Nevertheless, probabilistic concepts are not ignored. A means for obtaining probabilistic interpretations of the deterministic theory is developed in terms of fraction-of-time distributions, and ensemble averages are occasionally discussed.

A few words about the terminology used are in order. Although the terms *statistical* and *probabilistic* are used by many as if they were synonymous, their meanings are quite distinct. According to the *Oxford English Dictionary, statistical* means nothing more than "consisting of or founded on collections of numerical facts". Therefore, an average of a collection of spectra is a *statistical spectrum*.

And this has nothing to do with probability. Thus, there is nothing contradictory in the notion of a deterministic or nonprobabilistic theory of statistical spectral analysis. (An interesting discussion of variations in usage of the term *statistical* is given in *Comparative Statistical Inference* by V. Barnett [Barnett 1973]). The term *deterministic* is used here as it is commonly used, as a synonym for nonprobabilistic. Nevertheless, the reader should be forewarned that the elements of the nonprobabilistic theory presented herein are defined by infinite limits of time averages and are therefore no more deterministic in practice than are the elements of the probabilistic theory. (In mathematics, the deterministic and probabilistic theories referred to herein are sometimes called the *functional* and *stochastic* theories, respectively.) The term *random* is often taken as an implication of an underlying probabilistic model. But in this book, the term is used in its broader sense to denote nothing more than the vague notion of erratic unpredictable behavior.

1

INTRODUCTION TO SPECTRAL ANALYSIS

This introductory chapter sets the stage for the in-depth study of spectral analysis taken up in the following chapters by explaining objectives and motives, answering some basic questions about the nature and uses of spectral analysis, and establishing a historical perspective on the subject.

A. OBJECTIVES AND MOTIVES

A premise of this book is that the way engineers and scientists are commonly taught to think about empirical statistical spectral analysis of time-series data is fundamentally inappropriate for many applications. The subject is not really as abstruse as it appears to be from the conventional point of view. The problem is that the subject has been imbedded in the abstract probabilistic framework of stochastic processes, and this abstraction impedes conceptualization of the fundamental principles of empirical statistical spectral analysis. Hence, the probabilistic theory of statistical spectral analysis should be taught to engineers and scientists only after they have learned the fundamental deterministic principles—both qualitative and quantitative. For example, one should first learn 1) when and why sine wave analysis of time-series is appropriate, 2) how and why temporal and spectral resolution interact, 3) why statistical (averaged) spectra are of interest, and 4) what the various methods for measuring and computing statistical spectra are and how they are related. One should also learn how simultaneously to control the spectral and temporal resolution and the degree of randomness (reliability) of a statistical spectrum. All this can be accomplished in a nonsuperficial way without reference to the probabilistic theory of stochastic processes.

The concept of a deterministic theory of statistical spectral analysis is not new. Much deterministic theory was developed prior to and after the infusion, beginning in the 1930s, of probabilistic concepts into the field of time-series analysis. The most fundamental concept underlying present-day theory of statistical spectral analysis is the concept of an *ideal spectrum,* and the primary objective of statistical spectral analysis is to estimate the ideal spectrum using a finite amount of data. The first theory to introduce the concept of an ideal spectrum is Norbert Wiener's theory of *generalized harmonic analysis* [Wiener 1930], and this theory is deterministic. Later, Joseph Kampé de Fériet presented a deterministic theory of statistical spectral analysis that ties Wiener's theory more closely to the empirical reality of finite-length time-series [Kampé de Fériet 1954]. But the very great majority of treatments in the ensuing 30 years consider only a probabilistic theory of statistical spectral analysis, although a few authors do briefly mention the dual deterministic theory (e.g., [Koopmans 1974; Brillinger 1976]).

The primary objective of Part I of this book is to adopt the deterministic viewpoint of Wiener and Kampé de Fériet and show that a *comprehensive* deterministic theory of statistical spectral analysis, which for many applications relates more directly to empirical reality than does its more popular probabilistic counterpart, can be developed. A secondary objective of Part I is to adopt the empirical viewpoint of Donald G. Brennan [Brennan 1961] and Edward M. Hofstetter [Hofstetter 1964], from which they develop an objective probabilistic theory of stationary random processes based on fraction-of-time distributions and show that probability theory can be applied to the deterministic theory of statistical spectral analysis without introducing a more abstract mathematical model of empirical reality based on the axiomatic or subjective probabilistic theory of stochastic processes. This can be interpreted as an exploitation of Herman O. A. Wold's isomorphism between an empirical time-series and a probabilistic model of a stationary stochastic process. This isomorphism is responsible for the duality between probabilistic (ensemble-average) and deterministic (time-average) theories of time-series [Wold 1948] [Gardner 1985].

There are two motives for Part I of this book. The first is to stimulate a reassessment of the way engineers and scientists are often taught to think about statistical spectral analysis by showing that probability theory need not play a primary role. The second motive is to pave the way for introducing a new theory and methodology for statistical spectral analysis of random data from periodically time-variant phenomena, which is presented in Part II. The fact that this new theory and methodology, which unifies various emerging—as well as long-established—time-series analysis concepts and techniques, is most transparent when built on the foundation of the deterministic theory developed in Part I is additional testimony that probability theory need not play a primary role in statistical spectral analysis.

The book, although concise, is tutorial and is intended to be comprehensible by graduate students and professionals in engineering, science, mathematics, and statistics. The accomplishments of the book should be appreciated most by

those who have studied statistical spectral analysis in terms of the popular probabilistic theory and have struggled to bridge the conceptual gaps between this abstract theory and empirical reality.

B. ORIENTATION

1. What Is Spectral Analysis?

Spectral analysis of functions is used for solving a wide variety of practical problems encountered by engineers and scientists in nearly every field of engineering and science. The functions of primary interest in most fields are temporal or spatial waveforms or discrete data. The most basic purpose of spectral analysis is to represent a function by a sum of weighted sinusoidal functions called *spectral components;* that is, the purpose is to decompose (analyze) a function into these spectral components. The weighting function in the decomposition is a density of spectral components. This *spectral density* is also called a *spectrum.*[1] The reason for representing a function by its spectrum is that the spectrum can be an efficient, convenient, and often revealing description of the function.

As an example of the use of spectral representation of temporal waveforms in the field of signal processing, consider the *signal extraction* problem of extracting an information-bearing signal from corrupted (noisy) measurements. In many situations, the spectrum of the signal differs substantially from the spectrum of the noise. For example, the noise might have more high-frequency content; hence, the technique of *spectral filtering* can be used to attenuate the noise while leaving the signal intact. Another example is the *data-compression* problem of using *coding* to compress the amount of data used to represent information for the purpose of efficient storage or transmission. In many situations, the information contained in a complex temporal waveform (e.g., a speech segment) can be coded more efficiently in terms of the spectrum.

There are two types of spectral representations. The more elementary of the two shall be referred to as simply the *spectrum,* and the other shall be referred to as the *statistical spectrum.* The term *statistical* indicates that averaging or smoothing is used to reduce random effects in the data that mask the spectral characteristics of the phenomenon under study. For time-functions, the spectrum is obtained from an *invertible* transformation from a time-domain description of a function, $x(t)$, to a frequency-domain description, or more generally to a joint time- and frequency-domain description. The (complex) spectrum of a segment of data of length T centered at time t and evaluated at frequency f is

$$X_T(t,f) \triangleq \int_{t-T/2}^{t+T/2} x(u)e^{-i2\pi fu}\,du, \tag{1}$$

for which $i = \sqrt{-1}$. Because of the invertibility of this transformation, a

[1] The term *spectrum,* which derives from the Latin for image, was originally introduced by Sir Isaac Newton (see [Robinson 1982]).

function can be recovered from its spectrum,

$$x(u) = \int_{-\infty}^{\infty} X_T(t, f) e^{i2\pi f u}\, df, \qquad u \in [t - T/2, t + T/2]. \tag{2}$$

In contrast to this, a statistical spectrum involves an averaging or smoothing operation that is not invertible. For example, the statistical spectrum

$$S_{x_T}(t, f)_{\Delta t} \triangleq \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} S_{x_T}(v, f)\, dv, \tag{3}$$

for which $S_{x_T}(t, f)$ is the normalized squared magnitude spectrum

$$S_{x_T}(t, f) \triangleq \frac{1}{T} |\, X_T(t, f)|^2, \tag{4}$$

is obtained from a temporal smoothing operation. Thus, a statistical spectrum is a summary description of a function from which the function $x(t)$ cannot be recovered. Therefore, although the spectrum is useful for both signal extraction and data compression, the statistical spectrum is not *directly* useful for either. It is, however, quite useful *indirectly* for analysis, design, and adaptation of schemes for *signal extraction* and *data compression*. It is also useful for *forecasting* or *prediction* and more directly for other signal-processing tasks such as 1) the *modeling* and *system-identification* problems of determining the characteristics of a system from measurements on it, such as its response to excitation, and 2) *decision* problems, such as the *signal-detection problem* of detecting the presence of a signal buried in noise. As a matter of fact, the problem of detecting hidden periodicities in random data motivated the earliest work in the development of spectral analysis, as discussed in Section D.

Statistical spectral analysis has diverse applications in areas such as mechanical vibrations, acoustics, speech, communications, radar, sonar, ultrasonics, optics, astronomy, meteorology, oceanography, geophysics, economics, biomedicine, and many other areas. To be more specific, let us briefly consider a few applications. Spectral analysis is used to characterize various signal sources. For example, the spectral purity of a sine wave source (oscillator) is determined by measuring the amounts of harmonics from distortion due, for example, to nonlinear effects in the oscillator and also by measuring the spectral content close in to the fundamental frequency of the oscillator, which is due to random phase noise. Also, the study of modulation and coding of sine wave carrier signals and pulse-train signals for communications, telemetry, radar, and sonar employs spectral analysis as a fundamental tool, as do surveillance systems that must detect and identify modulated and coded signals in a noisy environment. Spectral analysis of the response of electrical networks and components such as amplifiers to both sine wave and random-noise excitation is used to measure various properties such as nonlinear distortion, rejection of unwanted components, such as power-supply components and common-mode components at the inputs of differential amplifiers, and the characteristics of filters, such as center frequencies, bandwidths, pass-band ripple, and stop-band rejection. Similarly, spectral analysis is used to study the magnitude and phase characteristics of the transfer functions as well as nonlinear distortion of various electrical, mechanical, and other systems,

including loudspeakers, communication channels and modems (modulator-demodulators), and magnetic tape recorders in which variations in tape motion introduce signal distortions. In the monitoring and diagnosis of rotating machinery, spectral analysis is used to characterize random vibration patterns that result from wear and damage that cause imbalances. Also, structural analysis of physical systems such as aircraft and other vehicles employs spectral analysis of vibrational response to random excitation to identify natural modes of vibration (resonances). In the study of natural phenomena such as weather and the behavior of wildlife and fisheries populations, the problem of identifying cause-effect relationships is attacked using techniques of spectral analysis. Various physical theories are developed with the assistance of spectral analysis, for example, in studies of atmospheric turbulence and undersea acoustical propagation. In various fields of endeavor involving large, complex systems such as economics, spectral analysis is used in fitting models to time-series for several purposes, such as simulation and forecasting. As might be surmised from this sampling of applications, the techniques of spectral analysis permeate nearly every field of science and engineering.

Spectral analysis applies to both continuous-time functions, called *waveforms,* and discrete-time functions, called *sampled data.* Other terms are commonly used also; for example, the terms *data* and *time-series* are each used for both continuous-time and discrete-time functions. Since the great majority of data sources are continuous-time phenomena, continuous-time data are focused on in this book, because an important objective is to maintain a close tie between theory and empirical reality. Furthermore, since optical technology has emerged as a new frontier in signal processing and optical quantities vary continuously in time and space, this focus on continuous time data is well suited to upcoming technological developments. Nevertheless, since some of the most economical implementations of spectrum analyzers and many of the newly emerging parametric methods of spectral analysis operate with discrete time and discrete frequency and since some data are available only in discrete form, discrete-time and discrete-frequency methods also are described.

2. Why Analyze Waveforms Into Sine Wave Components?[2]

The primary reason why sine waves are especially appropriate components with which to analyze waveforms is our preoccupation with *linear time-invariant* (LTI) *transformations,* which we often call *filters.* A secondary reason why statistical (time-averaged) analysis into sine wave components is especially appropriate is our preoccupation with *time-invariant phenomena* (data sources). To be specific, a transformation of a waveform $x(t)$ into another waveform, say $y(t)$, is an LTI transformation if and only if there exists a weighting function $h(t)$ (here assumed

[2] Readers in need of a brief remedial review of the prerequisite topic of linear time-invariant transformations and the Fourier transform should consult Appendix I.

to be absolutely integrable in the generalized sense, which accommodates Dirac deltas) such that $y(t)$ is the *convolution* (denoted by $\otimes$) of x with h:

$$y(t) = x(t) \otimes h(t) = \int_{-\infty}^{\infty} h(t - u)x(u)\, du \tag{5a}$$

$$= \int_{-\infty}^{\infty} h(v)x(t - v)\, dv. \tag{5b}$$

The *time-invariance property* of a transformation is, more precisely, a *translation-invariance property* that guarantees that a translation, by w, of $x(t)$ to $x(t + w)$ has no effect on $y(t)$ other than a corresponding translation to $y(t + w)$ (exercise 1). A phenomenon is said to be *time-invariant* only if it is *persistent* in the sense that it is appropriate to conceive of a mathematical model of $x(t)$ for which the following limit time-average exists for each value of τ and is not identically zero,[3]

$$\hat{R}_x(\tau) \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x\left(t - \frac{\tau}{2}\right) dt. \tag{6}$$

This function is called the *limit autocorrelation function*[4] for $x(t)$. For $\tau = 0$, (6) is simply the time-averaged value of the instantaneous power.[5]

Sine wave analysis is especially appropriate for studying a convolution because the *principal components* (*eigenfunctions*) of the convolution operator are the complex sine wave functions, $e^{i2\pi ft}$ for all real values of f. This follows from the facts that (1) the convolution operation produces a *continuous linear combination* of time-translates, that is, $y(t)$ is a weighted sum (over v) of $x(t - v)$, and (2) the complex sine wave is the only bounded function whose form is invariant (except for a scale factor) to time-translation, that is, a bounded function $x(t)$ satisfies

$$x(t - v) = cx(t) \tag{7}$$

for all t if and only if

$$x(t) = Xe^{i2\pi ft} \tag{8}$$

for some real values of X and f (exercise 3). As a consequence, the form of a bounded function $x(t)$ is invariant to convolution if and only if $x(t) = Xe^{i2\pi ft}$, in

[3] In Part II, it is explained that periodic and almost periodic phenomena as well as constant (time-invariant) phenomena satisfy (6). For $x(t)$ to be from a constant phenomenon, it must satisfy not only (6) but also $\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t + \tau/2)x(t - \tau/2)e^{-i2\pi\alpha t}\, dt \equiv 0$ for all $\alpha \neq 0$.

[4] In some treatments of time-series analysis (see [Jenkins and Watts 1968]), the function (6) modified by subtraction of the mean

$$\hat{m}_x \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, dt$$

from $x(t)$, is called the *autocovariance function,* and when normalized by $\hat{R}_x(0)$ it is called the *autocorrelation function.*

[5] If $x(t)$ is the voltage (in *volts*) across a one-*ohm* resistance, then $x^2(t)$ is the power dissipation (in *watts*).

which case (5) yields

$$y(t) = H(f)x(t), \tag{9}$$

for which

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\, dt. \tag{10}$$

This fact can be exploited in the study of convolution by decomposing a waveform $x(t)$ into a continuous linear combination of sine waves,[6]

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{i2\pi ft}\, df, \tag{11}$$

with weighting function

$$X(f) \triangleq \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\, dt, \tag{12}$$

because then substitution of (11) into (5) yields

$$y(t) = \int_{-\infty}^{\infty} Y(f)e^{i2\pi ft}\, df, \tag{13}$$

for which

$$Y(f) = H(f)X(f). \tag{14}$$

Thus, any particular sine wave component in $y(t)$, say

$$y_f(t) \triangleq Y(f)e^{i2\pi ft}, \tag{15}$$

can be determined solely from the corresponding sine wave component in $x(t)$, since (14) and (15) yield

$$y_f(t) = H(f)x_f(t). \tag{16}$$

The scale factor $H(f)$ is the *eigenvalue* associated with the eigenfunction $e^{i2\pi ft}$ of the convolution operator. Transformations (11) and (12) are the *Fourier transform* and its *inverse,* abbreviated by

$$X(\cdot) = F\{x(\cdot)\}$$
$$x(\cdot) = F^{-1}\{X(\cdot)\}.$$

Statistical (time-averaged) analysis of waveforms into sine wave components is especially appropriate for time-invariant phenomena because an ideal statistical spectrum, in which all random effects have been averaged out, exists if and only if the limit autocorrelation (6) exists. Specifically, it is shown in Chapter 3 that the ideal statistical spectrum obtained from (3) by smoothing over *all* time,

$$\lim_{\Delta t \to \infty} S_{x_T}(t, f)_{\Delta t},$$

exists if and only if the limit autocorrelation $\hat{R}_x(\tau)$ exists. Moreover, this ideal

[6] If $x(t)$ is absolutely integrable, then (11) and (12) are the usual *Fourier transform* pair, but if $x(t)$ is a persistent waveform (which does not die out as $|t| \to \infty$) from a time-invariant phenomenon, then (11) and (12) must be replaced with the *generalized* (integrated) *Fourier transform* [Wiener 1930], in which case (14) becomes the Stieltjes integral $Y(f) = \int_{-\infty}^{f} H(\nu)\, dX(\nu)$ [Gardner 1985].

statistical spectrum can be characterized in terms of the Fourier transform of $\hat{R}_x(\tau)$, denoted by

$$\hat{S}_x(f) \triangleq \int_{-\infty}^{\infty} \hat{R}_x(\tau) e^{-i2\pi f\tau}\, d\tau. \tag{17}$$

Specifically,

$$\lim_{\Delta t\to\infty} S_{x_T}(t, f)_{\Delta t} = \int_{-\infty}^{\infty} \hat{S}_x(f - \nu) z_{1/T}(\nu)\, d\nu = \hat{S}_x(f) \otimes z_{1/T}(f), \tag{18}$$

for which $z_{1/T}(f)$ is the unit-area sinc-squared function with width parameter $1/T$,

$$z_{1/T}(f) = \frac{1}{T}\left[\frac{\sin(\pi f T)}{\pi f}\right]^2. \tag{19}$$

As the time-interval of spectral analysis is made large, we obtain (in the limit)

$$\lim_{T\to\infty} \lim_{\Delta t\to\infty} S_{x_T}(t, f)_{\Delta t} = \hat{S}_x(f), \tag{20}$$

because the limit of $z_{1/T}(f)$ is the *Dirac delta*

$$\lim_{T\to\infty} z_{1/T}(f) = \delta(f), \tag{21}$$

and convolution of a function with the Dirac delta as in (18) leaves the function unaltered (exercise 2). The ideal statistical spectrum $\hat{S}_x(f)$ defined by (20) is called the *limit spectrum.*

Before leaving this topic of justifying the focus on sine wave components for time-series analysis, it is instructive (especially for the reader with a background in stochastic processes) to consider how the justification must be modified if we are interested in probabilistic (ensemble-averaged) statistical spectra rather than deterministic (time-averaged) statistical spectra. Let us therefore consider an ensemble of *random samples* of waveforms $\{x(t, s)\}$, indexed by s; for convenience in the ensuing heuristic argument, let us assume that the ensemble is a continuous ordered set for which the ensemble index, s, can be any real number. For each member $x(t, s)$ of the ensemble, we can obtain an analysis into principal components (sine wave components). A characteristic property of a set of principal components is that they are mutually *uncorrelated*[7] in the sense that

$$\langle x_f, x_\nu \rangle_t \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_f(t, s) x_\nu^*(t, s)\, dt = 0, \qquad f \neq \nu, \tag{22}$$

where * denotes complex conjugation (exercise 5). But in the probabilistic theory,

[7] For a persistent waveform (which does not die out as $|t| \to \infty$) from a time-invariant phenomenon, the property of sine wave components being mutually uncorrelated is deeper than suggested by (22). In particular, the envelopes (from (1)), $X_T(t, f)$ and $X_T(t, \nu)$, of the *local sine wave components* (cf. Chapter 4, Section E) become uncorrelated in the limit $T \to \infty$ for all $f \neq \nu$ as explained in Chapter 7, Section C.

it is required that the principal components be uncorrelated over the ensemble[8]

$$\langle x_f, x_\nu \rangle_s \triangleq \lim_{S\to\infty} \frac{1}{S} \int_{-S/2}^{S/2} x_f(t, s) x_\nu^*(t, s)\, ds = 0, \qquad f \neq \nu \tag{23}$$

as well as uncorrelated over time in order to obtain the desired simplicity in the study of time-series subjected to LTI transformations. If we proceed formally by substitution of the principal component,

$$x_f(t, s) \triangleq X(f, s) e^{i2\pi ft} = \int_{-\infty}^{\infty} x(u, s) e^{-i2\pi fu}\, du\, e^{i2\pi ft}, \tag{24}$$

into (23), we obtain[9] (after reversing the order of the limit operation and the two integration operations)

$$|\langle x_f, x_\nu \rangle_s| = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{R}_x(t, v) e^{-i2\pi(ft - \nu v)}\, dt\, dv \right|, \tag{25}$$

for which the function $\mathcal{R}_x$ is the *probabilistic autocorrelation* defined by

$$\mathcal{R}_x(t, v) \triangleq \lim_{S\to\infty} \frac{1}{S} \int_{-S/2}^{S/2} x(t, s) x(v, s)\, ds. \tag{26}$$

It can be shown (exercise 6) that (23) vanishes for all $f \neq \nu$, as desired, if and only if

$$\mathcal{R}_x(t, v) = \mathcal{R}_x(t + w, v + w) \tag{27}$$

for all translations w, in which case $\mathcal{R}_x$ depends on only the difference of its two arguments,

$$\mathcal{R}_x(t, v) = \mathcal{R}_x(t - v). \tag{28}$$

Consequently principal-component methods of study of an LTI transformation of an ensemble of waveforms are applicable if and only if the correlation of the ensemble is translation invariant. Such an ensemble of random samples of waveforms is commonly said to have arisen from a *wide-sense stationary stochastic process*.[10] But we must ask if ensembles with translation-invariant correlations are of interest in practice. As a matter of fact, they are for precisely the same reason that translation-invariant linear transformations are of practical interest. The reason is a preoccupation with *time-invariance*. That is, the ensemble of waveforms generated by some phenomenon will exhibit a translation-invariant correlation if and only if the data-generating mechanism of the phenomenon exhibits appropriate time-invariance. Such time-invariance typically results from a *stable* system being in a *steady-state* mode of operation—a *statistical equilibrium*.

[8] The limit averaging operation in (23) can be interpreted (via the *law of large numbers*) as the probabilistic *expectation operation*.

[9] To make the formal manipulation used to obtain (25) rigorous, $X(f, s)$ must be replaced with the envelope of the local sine wave component, which is obtained from (1) with $x(u)$ replaced by $x(u, s)$; then the limit, $T \to \infty$, must be taken. An in-depth treatment of this topic of spectral correlation is introduced in Chapter 7, Section C, and is the major focus of Part II.

[10] The term *stochastic* comes from the Greek *to aim* (*guess*) *at*.

The ultimate in time-invariance of a data-generating mechanism is characterized by a *translation-invariant ensemble,* which is an ensemble $\{x(t, s)\}$ for which the identity

$$x(t + w, s) = x(t, s') \tag{29}$$

holds for all s and all real w; that is, each translation by, for instance, w of each ensemble member, such as $x(t, s)$, yields another ensemble member, for example, $x(t, s')$. This time-invariance property (29) is more than sufficient for the desired time-invariance property (27). An ensemble that exhibits property (29) shall be said to have arisen from a *strict-sense stationary stochastic process.* For many applications, a natural way in which a translation-invariant ensemble would arise as a mathematical model is if the ensemble actually generated by the physical phenomenon is artificially supplemented with all translated versions of the members of the actual ensemble. In many situations, the most intuitively pleasing actual ensemble consists of one and only one waveform, $x(t)$, which shall be called the *ensemble generator.* In this case, the supplemented ensemble is defined by

$$x(t, s) = x(t + s). \tag{30}$$

The way in which a probabilistic model can, in principle, be derived from this ensemble is explained in Chapter 5, Section B. This most intuitively pleasing translation-invariant ensemble shall be said to have arisen from an *ergodic*[11] stationary stochastic process. *Ergodicity* is the property that guarantees equality between time-averages, such as (22), and ensemble-averages, such as (23). The ergodic relation (30) is known as *Herman O. A. Wold's isomorphism* between an individual time-series and a stationary stochastic process [Wold 1948].

In summary, statistical sine wave analysis—spectral analysis as we shall call it—is especially appropriate in principle if we are interested in studying linear *time-invariant* transformations of data and data from *time-invariant* phenomena. Nevertheless, in practice, statistical spectral analysis can be used to advantage for slowly time-variant linear transformations and for data from slowly time-variant phenomena (as explained in Chapter 8) and in other special cases, such as periodic time-variation (as explained in Part II) and the study of the departure of transformations from linearity (as explained in Chapter 7).

C. ORIGINS OF SPECTRAL ANALYSIS

The Fourier theory of sine wave analysis of functions has its origins in two fields of investigation into the nature of the physical world: acoustical/optical wave phenomena and astronomical and geophysical periodicities.[12] These two fields

[11] The term *ergodic* comes from the Greek for *work path,* which—in the originating field of statistical mechanics—relates to the path, in one dimension, described by $x(\cdot, s)$, of an energetic particle in a gas.

[12] The historical survey given here has been synthesized from various other more brief historical sketches found in the literature as well as from inspection of many (but not all) of the references cited here.

have furnished the primary stimuli from the natural sciences to the classical study—which extends into the first half of the twentieth century—of spectral analysis. The motions of the planets, the tides, and irregular recurrences of weather, with their hidden periodicities and disturbed harmonics, form a counterpart of the vibrating string in acoustics and the phenomena of light in optics. Although the concept of sine wave analysis has very early origins, the first bona fide uses of sine wave analysis apparently did not occur until the eighteenth century, with the work of Leonhard Euler (1707–1783) and Joseph Louis Lagrange (1736–1813) in astronomy [Lagrange 1772].[13]

The concept of *statistical* spectral analysis germinated in early studies of light, beginning with Isaac Newton's prism experiment in 1664 which led to the notion that white light is simply an additive combination of homogeneous monochromatic vibrations. The developing wave optics ideas, together with developing ideas from meteorology and astronomy, led Sir Arthur Schuster (1851–1934), around the turn of the nineteenth century, to the invention of the *periodogram* for application to the problem of detection of hidden periodicities in random data [Schuster 1894, 1897, 1898, 1900, 1904, 1906, 1911]. The periodogram, denoted by $S_{x_T}(f)$ (originally defined for discrete-time data), is simply the squared magnitude of the Fourier transform of a finite segment of data, x_T, normalized by the length, T, of the data segment (graphed versus the frequency variable, f):

$$S_{x_T}(f) \triangleq \frac{1}{T}|X_T(f)|^2 \tag{31}$$

$$X_T(f) \triangleq \int_{-T/2}^{T/2} x_T(t)e^{-i2\pi ft}\,dt, \tag{32}$$

where $x_T(t)$ is taken to be zero for $|t| > T/2$. If a substantial peak occurred in the periodogram, it was believed that an underlying periodicity of the frequency at which the peak occurred had been detected. As a matter of fact, this idea preceded Schuster in the work of George Gabriel Stokes (1819–1903) [Stokes 1879]; and a related approach to periodicity detection developed for meteorology by Christoph Hendrik Diederik Buys-Ballot (1817–1890) preceded Stokes [Buys-Ballot 1847]. The first general development of the periodogram is attributed to Evgency Evgenievich Slutsky (1880–1948) [Slutsky 1929, 1934].

Another approach to detection of periodicities that was being used in meteorology in the early part of the twentieth century was based on the *correlogram* [Clayton 1917; Alter 1927; Taylor 1920, 1938], whose earliest known use [Hooker 1901] was motivated by the studies in economics of John Henry Poynting (1852–1914) [Poynting 1884]. The correlogram, denoted by $R_{x_T}(\tau)$ (originally defined for discrete-time data), is simply the time-average of products of time-shifted versions of a finite segment of data (graphed versus the time-difference variable, τ),

$$R_{x_T}(\tau) \triangleq \frac{1}{T}\int_{-\infty}^{\infty} x_T\left(t + \frac{\tau}{2}\right) x_T\left(t - \frac{\tau}{2}\right)dt. \tag{33}$$

[13] See [Wiener 1938; Davis 1941; Robinson 1982] for the early history of spectral analysis, and [Chapman and Bartels 1940, Chapter XVI] for an account of early methods.

But since $x_T(t \pm \tau/2)$ is zero for $t \pm \tau/2$ outside $[-T/2, T/2]$, we obtain

$$R_{x_T}(\tau) = \frac{1}{T}\int_{-(T-|\tau|)/2}^{(T-|\tau|)/2} x_T\left(t + \frac{\tau}{2}\right) x_T\left(t - \frac{\tau}{2}\right)dt. \tag{34}$$

If an oscillation with τ occurred in the correlogram, it was believed that an underlying periodicity had been detected.[14]

The discovery of the *periodogram-correlogram relation* (e.g., [Stumpff 1927; Wiener 1930]) revealed that these two methods for periodicity detection were, in essence, the same. The *relation*, which is a direct consequence of the *convolution theorem* (Appendix 1-1) is that $S_{x_T}(\cdot)$ and $R_{x_T}(\cdot)$ are a Fourier transform pair (exercise 10):

$$S_{x_T}(\cdot) = F\{R_{x_T}(\cdot)\}. \tag{35}$$

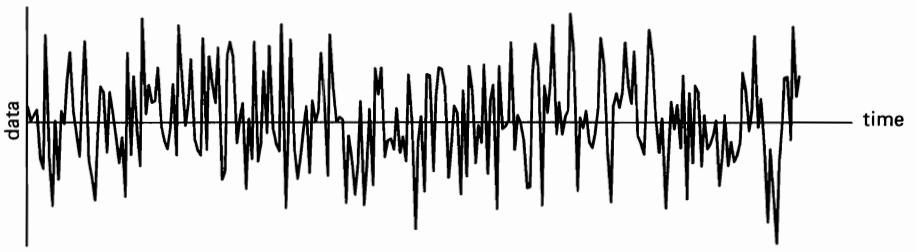
This relation was apparently understood and used by some before the turn of the century, as evidenced by the spectroscopy work of Albert Abraham Michelson (1852–1931), who in 1891 used a mechanical harmonic analyzer to compute the Fourier transform of a type of correlogram obtained from an interferometer for the purpose of examining the fine structure of the spectral lines of lightwaves.

A *completely random time-series* is defined to be one for which the discrete-time correlogram is asymptotically ($T \to \infty$) zero for all nonzero time-shifts, $\tau \neq 0$, indicating there is no correlation in the time-series. A segment of a simulated completely random time-series is shown in Figure 1-1(a), and its periodogram and correlogram are shown in Figures 1-1(b) and 1-1(c). This concept arose (originally for discrete-time data) around the turn of the century [Goutereau 1906], and a systematic theory of such completely random time-series was developed in the second decade by George Udny Yule (1871–1951) [Yule 1926]. Yule apparently first discovered the fact that an LTI transformation (a convolution) can introduce correlation into a completely random time series. It is suggested by the periodogram-correlogram relation that a completely random time series has a flat periodogram (asymptotically). By analogy with the idea of white light containing equal amounts of all spectral components (in the optical band), a completely random time series came to be called *white noise*. As a consequence of the discoveries of the correlation-inducing effect of an LTI transformation, and the periodogram-correlogram relation, it was discovered that a completely random time series, subjected to a narrow-band LTI transformation, can exhibit a periodogram with sharp dominant peaks, when in fact there is no underlying periodicity in the data. This is illustrated in Figure 1-2. This revelation, together with several decades of experience with the erratic and unreliable behavior of periodograms, first established as an *inherent* property by Slutsky [Slutsky 1927], led during the mid–twentieth century to the development of various averaging or smoothing (statistical) methods for modifying the periodogram to improve its utility. A smoothed version of the periodogram in Figure 1-1(b) is shown in Figure 1-1(d). Such averaging techniques were apparently first proposed by Albert Einstein (1879–1955) [Einstein 1914], Norbert Wiener (1894–1964) [Wiener 1930] and later by Percy John Daniell (1889–1946) [Daniell 1946], Maurice Stevenson

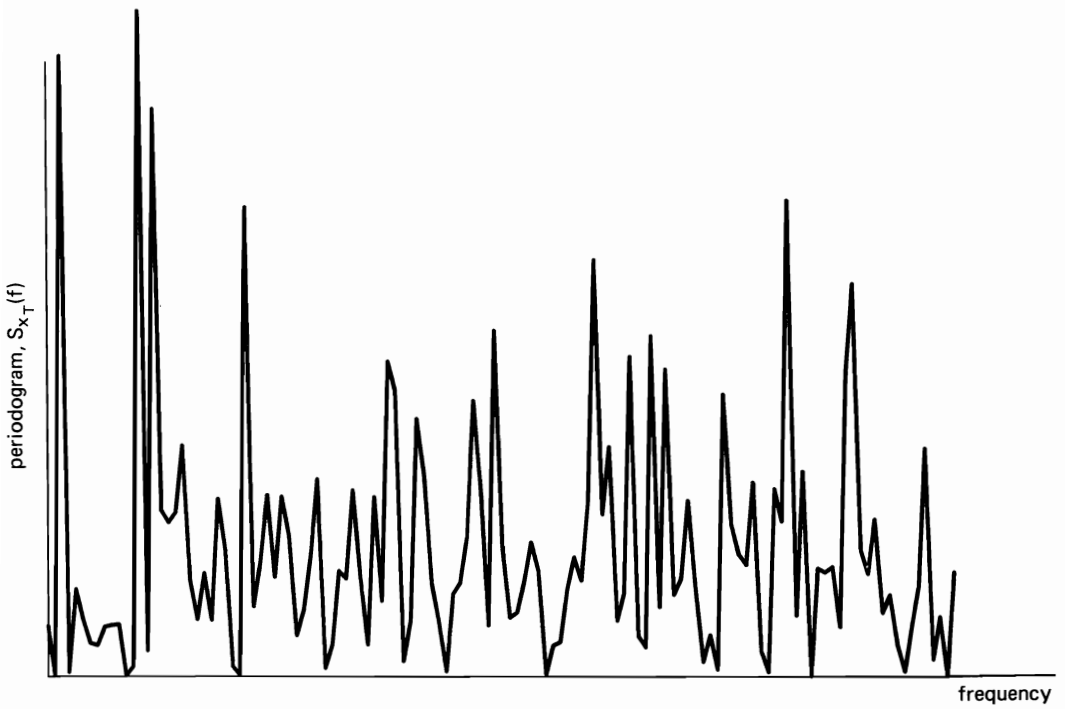
[14] The early history of correlation studies is reported in [Davis 1941].

Bartlett (1910–) [Bartlett 1948, 1950], John Wilder Tukey (1915–) [Tukey 1949], Richard Wesley Hamming (1915–), and Ralph Beebe Blackman (1904–) [Blackman and Tukey 1958]. In addition, these circumstances surrounding the periodogram led to the alternative *time-series-modeling approach to spectral analysis,* which includes various methods such as the *autoregressive-modeling method* introduced by Yule [Yule 1927] and developed by Herman O. A. Wold (1908–) [Wold 1938] and others.

Apparently independent of and prior to the introduction (by others) of empirical averaging techniques to obtain less random measurements of spectral

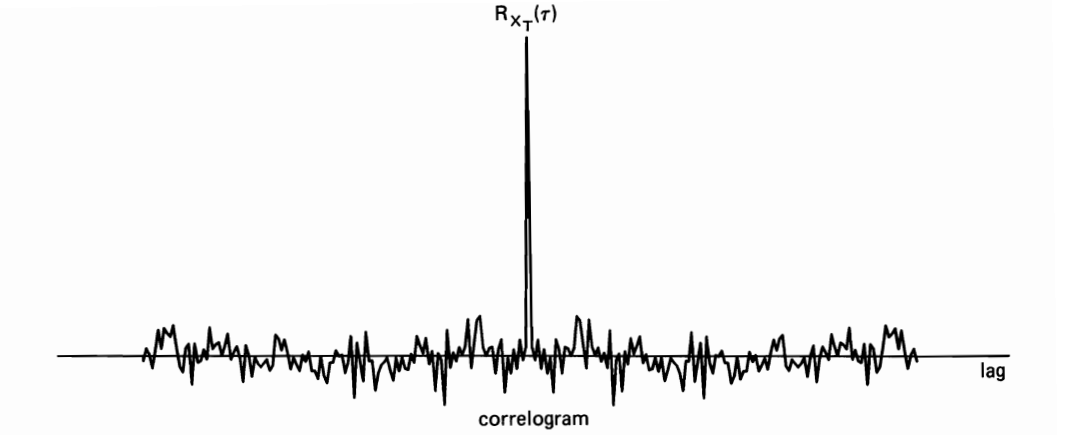


(a)

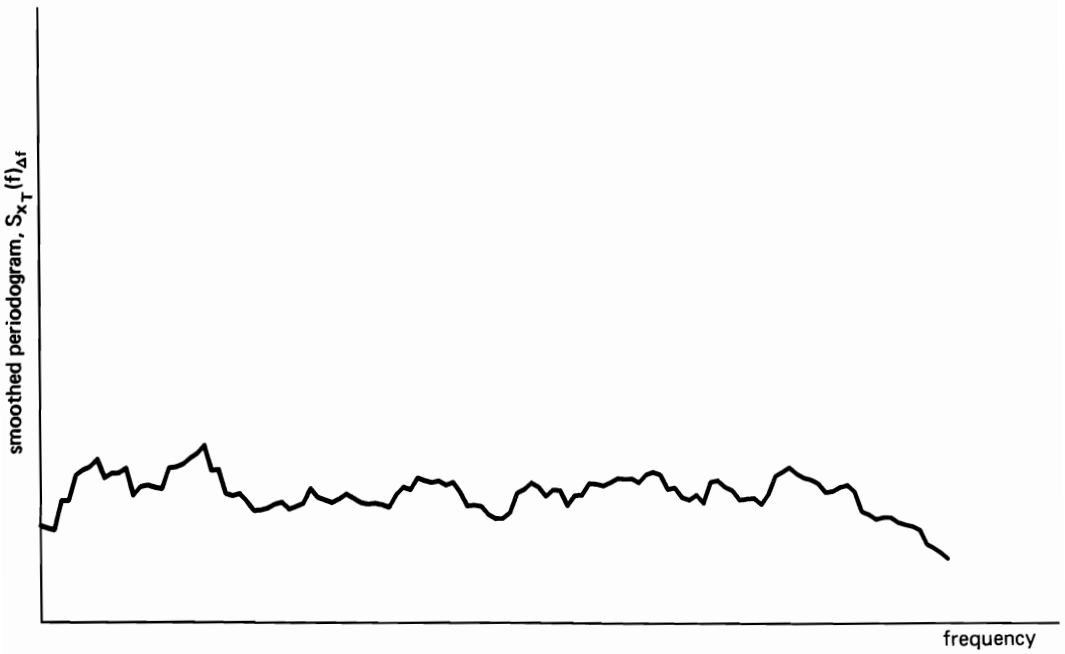


(b)

Figure 1-1 (a) Completely random data (white noise), $T = 256T_s$. (b) Periodogram of white noise, $T = 256T_s$.



(c)



(d)

Figure 1-1 (continued) (c) Correlogram of white noise, $T = 256T_s$. (d) Smoothed periodogram of white noise, $T = 256T_s$, $\Delta f = 21/256T_s$.

content of random time-series, Wiener developed his theory of *generalized harmonic analysis* [Wiener 1930], in which he introduced a completely nonrandom measure of spectral content. Wiener's spectrum can be characterized as a limiting form of an averaged periodogram. In terms of this limiting form of periodogram and the corresponding limiting form of correlogram, Wiener developed what might be called a calculus of averages for LTI transformations of time-series. Although

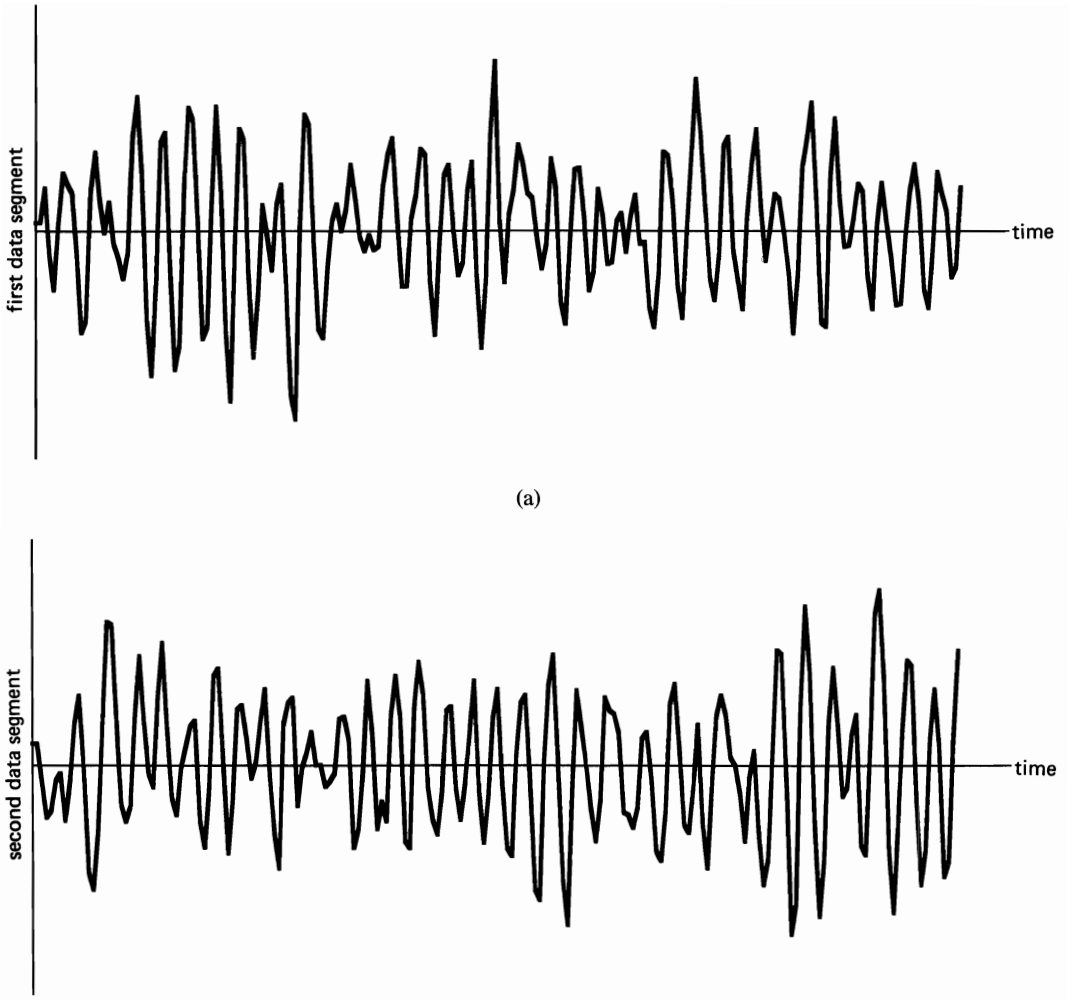
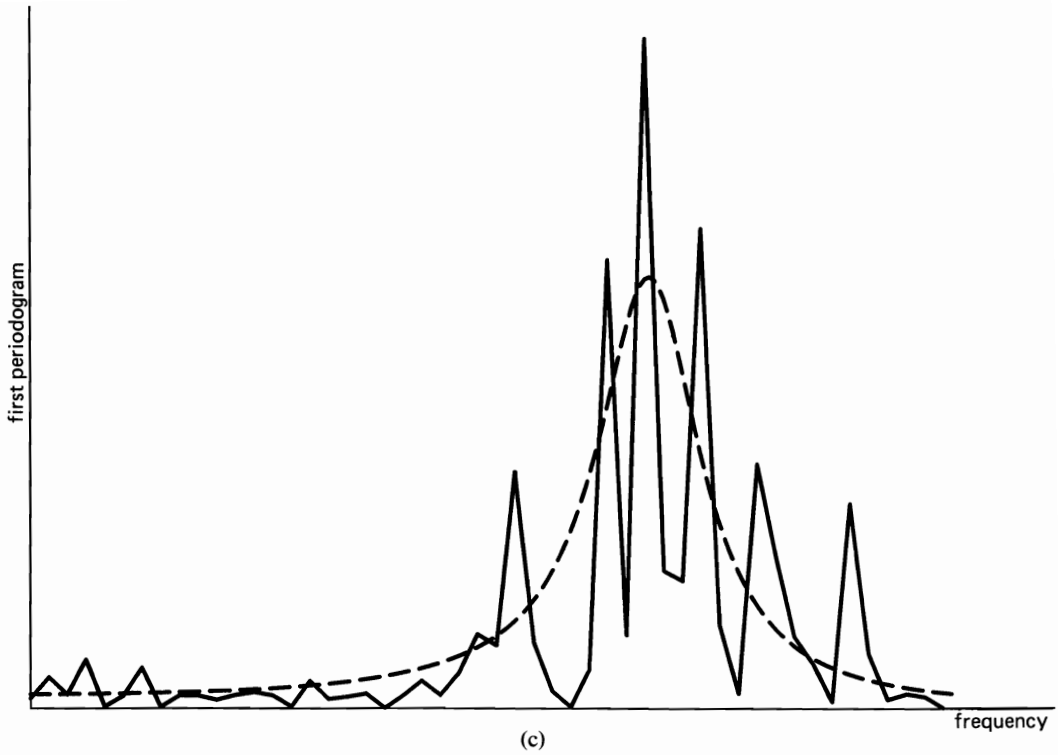


Figure 1-2 (a), (b) Two segments of narrow-band data, $T = 256T_s$.

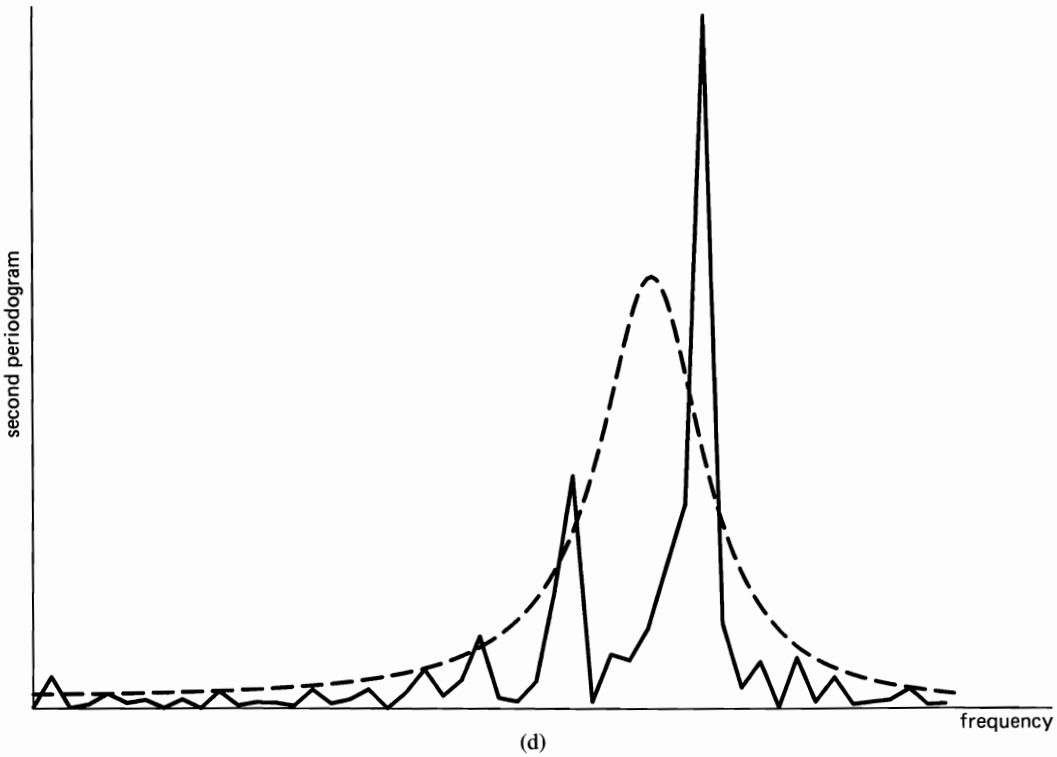
it is not well known,[15] Wiener's limit spectrum and its characterization as the Fourier transform of a limit correlogram had been previously presented (in rather terse form) by Einstein [Einstein 1914].

The autonomous development of statistical mechanics, with Josiah Willard Gibbs' (1839–1903) concept of an *ensemble average,* and the study of Brownian motion, by Maryan von Smoluchowski [von Smoluchski 1914], Einstein [Einstein 1906], and Wiener [Wiener 1923], together with the mathematical development

[15] This little-known fact was brought to the author's attention by Professor Thomas Kailath, who learned of it from Akiva Moisevich Yaglom.

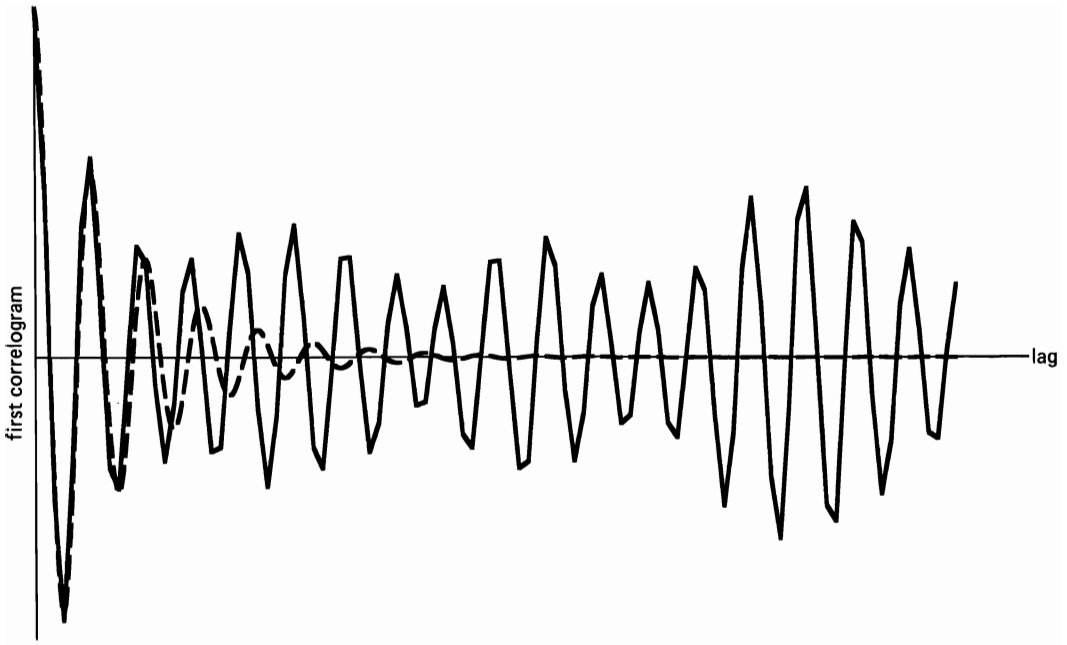


(c)

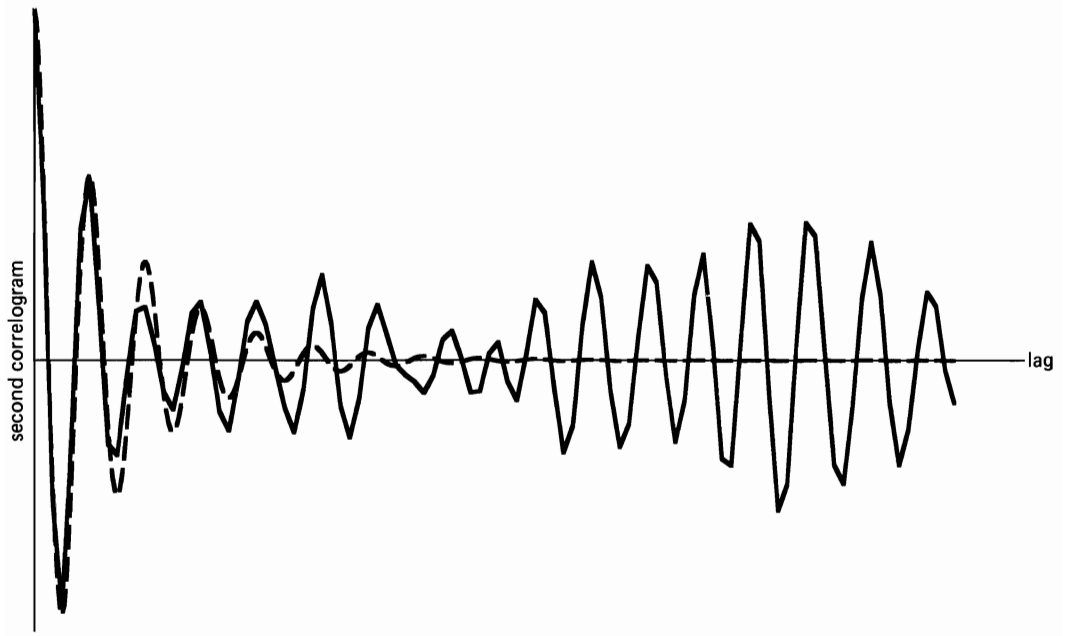


(d)

Figure 1-2 (continued) (c), (d) Periodograms of the two data segments shown in (a) and (b). (Broken curve is the limit spectrum.)



(e)



(f)

Figure 1-2 (continued) (e), (f) Correlograms of the two data segments shown in (a) and (b). (Broken curve is the limit autocorrelation.)

of probability theory based on the measure and integration theory of Henri León Lebesgue (1875–1941) around the turn of the century, led ultimately to the probabilistic theory of stochastic processes. This theory includes a probabilistic counterpart to Wiener's theory of generalized harmonic analysis, in which infinite time-averages are replaced with infinite ensemble averages. It greatly enhanced the conceptualization and mathematical modeling of erratic-data sources and the design and analysis of statistical data-processing techniques such as spectral analysis. The theory (for discrete-time processes) originated in the work of Aleksandr Jakovlevich Khinchin (1894–1959) during the early 1930s [Khinchin 1934] and was further developed in the early stages by Wold [Wold 1938], Andrei Nikolaevich Kolmogorov (1903–) [Kolmogorov 1941a,b], and Harald Cramér (1893–) [Cramér 1940, 1942].[16] Major contributions to the early development of the probabilistic theory and methodology of statistical spectral analysis were made by Ulf Grenander and Murray Rosenblatt [Grenander and Rosenblatt 1953, 1984], Emanuel Parzen [1957a, b], and Blackman and Tukey [Blackman and Tukey 1958].

The probabilistic theory of stochastic processes is currently the popular approach to time-series analysis. However, from time to time, the alternative deterministic approach, which is taken in this book, is promoted for its closer ties with empirical reality for many applications; see [Kampé de Fériet 1954; Brennan 1961; Bass 1962; Hofstetter 1964; Finch 1969; Brillinger 1975, Sec. 2.11; Masani 1979].

D. SPECTRAL ANALYSIS AND PERIODICITY

The problem of studying hidden periodicity in random data motivated the earliest work in spectral analysis and provided much of the impetus for developing spectral analysis concepts and methods during the first few decades following Schuster's pioneering work. However, the fact that most of the phenomena being studied did not exhibit periodicity but rather gave rise to data consisting of what came to be called *disturbed harmonics* (which result from subjecting a completely random time-series to a narrow-band-pass filter) resulted in a major shift in focus away from hidden periodicity and toward the time-invariance discussed in Section B. It is curious that some branch of work in the field did not retain a substantial focus on phenomena that do indeed give rise to random data with hidden periodicity and thereby did not continue the initial development of statistical theory and method for such time-series. Although existing theory and method are usually adequate for additive periodic components in random data, there is no generally appropriate theory and method for other types of hidden periodicity,

[16] The most extensive bibliography on time-series and random processes, ranging from the earliest period of contribution (mid–nineteenth century) to the recent past (1960) is the international team project bibliography edited by Wold [Wold 1965]. Starting with 1960, a running bibliography, including abstracts, is available in the *Journal of Abstracts: Statistical Theory and Method.*

such as multiplicative periodicity, that arise either from natural rhythms or from transformations intentionally designed to be periodic, as in various techniques of sampling, modulating, multiplexing, and coding employed in signal-processing systems. The lack of development of theory and method for spectral analysis of such time-series was recognized explicitly by Blackman and Tukey [Blackman and Tukey 1958, p. vi], who in effect condoned it by arguing pragmatically that no phenomenon is precisely periodic and that existing theory and method appear to be adequate. In contrast to this point of view, it is shown in Part II of this book that some phenomena can, to great advantage, be modeled as precisely periodic; also, not only is existing theory and method for spectral analysis generally inadequate in such cases, but an adequate generalization in terms of spectral correlation can be developed. This more general theory and methodology of statistical spectral analysis presented in Part II includes the theory and methodology presented in Part I as the special case for which periodicity degenerates into constancy (time-invariance). A brief introduction to the spectral correlation theory of random data from periodic phenomena is given in Chapter 7 of Part I.

E. SUMMARY

Section A explains that the objective of Part I of this book is to show that a comprehensive deterministic theory of statistical spectral analysis, which for many applications relates more directly to empirical reality than does its more popular probabilistic counterpart, can be developed—the motivation being to stimulate a reassessment of the way engineers and scientists are often taught to think about statistical spectral analysis by showing that probability theory need not play a primary role. In Section B it is explained that the most basic purpose of spectral analysis is to represent a function by a sum of weighted sinusoidal functions called *spectral components* and that procedures for statistical spectral analysis average the strengths of such components to reduce random effects. It is further explained that sine wave components, in comparison with other possible types of components, are especially appropriate for analyzing data from time-invariant phenomena, because sine waves are the principal components of time-invariant linear transformations and because an ideal sine wave spectrum exists if and only if the data source is time-invariant (in an appropriate sense). The conceptual link between this point of view and that of the probabilistic framework of ergodic stationary stochastic processes on which statistical spectral analysis is typically based is then explained in terms of Wold's isomorphism. In Section C, a historical sketch of the origins of spectral analysis is presented, and finally in Section D the need for a generalization of the theory of spectral analysis of random data, from constant phenomena to periodic phenomena, is commented upon.

Appendix 1-1 is a brief review of prerequisite material on linear time-invariant transformations and the Fourier transform.

F. OVERVIEW OF PART I

This first chapter is concluded with a brief overview of the remainder of Part I. In Chapter 2, the basic elements of empirical spectral analysis are introduced. The time-variant periodogram for nonstatistical spectral analysis is defined and characterized as the Fourier transform of the time-variant correlogram, and its temporal and spectral resolution properties are derived. The effects of linear time-invariant filtering and periodic time sampling are described. Then in Chapter 3, the fundamentals of statistical spectral analysis are introduced. The equivalence between statistical spectra obtained from temporal smoothing and statistical spectra obtained from spectral smoothing is established, and the relationship between these statistical spectra and the abstract limit spectrum is derived. The limit spectrum is characterized as the Fourier transform of the limit autocorrelation, and the effects of linear time-invariant filtering and periodic time-sampling on the limit spectrum are described. Various continuous-time and discrete-time models for time-series are introduced, and their limit spectra are calculated. Chapter 4 presents a wide variety of analog (continuous-time) methods for empirical statistical spectral analysis, and it is shown that all these methods are either exactly or approximately equivalent when a substantial amount of smoothing is done. The spectral leakage phenomenon is explained, and the concept of an effective spectral smoothing window is introduced. Then a general representation for the wide variety of statistical spectra obtained from these methods is introduced and shown to provide a means for a unified study of statistical spectral analysis. In Chapter 5, it is explained that the notion of the degree of randomness or variability of a statistical spectrum can be quantified in terms of time-averages by exploiting the concept of fraction-of-time probability. This approach is then used mathematically to characterize the temporal bias and temporal variability of statistical spectra. These characterizations form the basis for an in-depth discussion of design trade-offs involving the resolution, leakage, and reliability properties of a statistical spectrum. The general representation introduced in Chapter 4 is used here to obtain a unified treatment for the wide variety of spectral analysis methods described in Chapter 4. Chapter 6 complements Chapter 4 by presenting a variety of digital (discrete-time) methods for statistical spectral analysis. Chapter 7 generalizes the concept of spectral analysis of a single real-valued time-series to that of cross-spectral analysis of two or more complex-valued time-series. It is established that the cross spectrum, which is a measure of spectral correlation, plays a fundamental role in characterizing the degree to which two or more time-series are related by a linear time-invariant transformation. Methods for measurement of statistical cross spectra that are generalizations of the methods described in earlier chapters are presented, and the temporal bias and temporal variability of statistical cross spectra are mathematically characterized in a unified way based on a general representation. In Chapter 8, the application of statistical spectral analysis to time-variant phenomena is studied. Fundamental limitations on temporal and spectral resolution are discussed, and the roles of ensemble averaging and probabilistic models are described. Finally, in Chapter 9, an introduction to the theory of autoregressive modeling of time-series is

presented and used as the basis for describing a variety of autoregressive parametric methods of statistical spectral analysis. The chapter concludes with an extensive experimental study and comparison of various parametric and nonparametric methods of statistical spectral analysis.

EXERCISES

1. Substitute $x(t + w)$ in place of $x(t)$ in the convolution (5), and use a change of variable of integration to verify that this substitution produces $y(t + w)$ in place of $y(t)$.
2. The *impulse function,* denoted by $\delta(t)$, (also called the *Dirac delta*) is formally defined to be an idealized pulse with infinitesimal width, infinite height, and unity area; thus,

$$\delta(t) = 0, \qquad t \neq 0$$

$$\int_{-\epsilon}^{\epsilon} \delta(t)\, dt = 1, \qquad \epsilon > 0. \tag{36}$$

Consequently, the impulse function exhibits the *sampling property*

$$\delta(t - t_0)x(t) = x(t_0)\delta(t - t_0) \tag{37}$$

for every function $x(t)$ that is continuous at $t = t_0$.

(a) Verify that the Fourier transform of the impulse function

$$x(t) = \delta(t - t_0) \tag{38a}$$

is

$$X(f) = e^{-i2\pi f t_0}. \tag{38b}$$

(b) Show that $h(t) \otimes \delta(t) = h(t)$ for any continuous function $h(t)$. Then use the convolution (5) to verify that the response of an LTI transformation to an impulse excitation at $t = 0$ is $y(t) = h(t)$. Thus, the *weighting function* $h(t)$ of an LTI transformation is identical to its *impulse response.*

(c) Verify that the Fourier transform of a periodic function with Fourier series representation

$$x(t) = \Delta f \sum_{m=-\infty}^{\infty} X_m e^{i2\pi m \Delta f t} \tag{39a}$$

is

$$X(f) = \Delta f \sum_{m=-\infty}^{\infty} X_m \delta(f - m\Delta f). \tag{39b}$$

3. **(a)** To gain some insight into the fact that the only bounded waveform whose form is invariant to translation is a complex sine wave, subtract $x(t)$ from both sides of (7), divide by v, and take the limit as $v \to 0$, to obtain the following condition for (7),

$$\frac{dx}{dt} = \alpha x(t)$$

$$\alpha \triangleq \lim_{v \to 0} \frac{1 - c}{v}$$

(notice that c depends on v). Then solve this differential equation by integration to obtain

$$x(t) = \beta e^{\alpha t}$$

for arbitrary α and β. This function $x(t)$ is bounded if and only if α is imaginary

or zero. Substitute this $x(t)$ into (7) and solve for c. Then substitute this solution into the definition of α to verify consistency.

(b) As another approach, Fourier transform both sides of (7) to obtain

$$X(f)e^{i2\pi fv} = cX(f),$$

and argue that this equation is valid for all values of f if and only if $X(f)$ is nonzero for at most one value of f, say γ, to conclude that

$$X(f) = \beta\delta(f - \gamma)$$

for arbitrary β and γ. Then perform an inverse Fourier transformation of both sides of this equation to obtain the desired result.

(c) As a third approach, consider an integer t and iterate (7) for $v = 1, 2, 3, \ldots, t$, letting c_1 denote the value of c for $v = 1$:

$$x(t) = \frac{1}{c_1}x(t - 1) = \frac{1}{c_1}\left[\frac{1}{c_1}x(t - 2)\right] = \cdots = \left(\frac{1}{c_1}\right)^t x(0).$$

Show that this result can be put into the form of (8).

4. Prove the convolution theorem; that is, if

$$z(t) = x(t) \otimes y(t) = \int_{-\infty}^{\infty} x(t - v)y(v)\, dv, \tag{40a}$$

then

$$Z(f) \triangleq \int_{-\infty}^{\infty} z(t)e^{-i2\pi ft}\, dt = X(f)Y(f). \tag{40b}$$

Hint: One method is to Fourier transform both sides of (40a) and then use the change of variables $t - v = u$. Another method is to perform an inverse transformation of both sides of (40b) and then use the transform pair (38).

5. Substitute the sine waves (24) with frequencies f and ν into (22), and verify that the correlation is zero for $f \neq \nu$.

6. Show that the double Fourier transform

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{R}_x(t, v)e^{-i2\pi(ft - \nu v)}\, dt\, dv$$

vanishes for all $f \neq \nu$ if and only if

$$\mathcal{R}_x(t, v) = \mathcal{R}_x(t - v).$$

Hint: Let the above double integral be denoted by $\mathcal{S}_x(f, \nu)$; then $\mathcal{R}_x(t, v)$ is given by the inverse double Fourier transform

$$\mathcal{R}_x(t, v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{S}_x(f, \nu)e^{i2\pi(ft - \nu v)}\, df\, dv.$$

Now, in order for $\mathcal{S}_x(f, \nu)$ to vanish for all $f \neq \nu$ without $\mathcal{R}_x(t, v)$ being identically zero (or otherwise pathological), it is required that

$$\begin{aligned}\mathcal{S}_x(f, \nu) &= S_x(f)\delta(f - \nu) \\ &= S_x(\nu)\delta(f - \nu)\end{aligned}$$

for some function $S_x(f)$.

7. The Fourier transform representation (11)–(12) can be formally verified by substituting (12) into (11) and using the transform

$$\int_{-\infty}^{\infty} e^{-i2\pi f(t - s)}\, df = \delta(t - s).$$

Do this by assuming that the required interchange of integrals is justified.

8. To illustrate that negative frequencies $f < 0$ for a real waveform $x(t)$ are simply mathematical artifacts with no physical significance, show that they can be dispensed with, while preserving the Fourier transform relation. Specifically, define the *one-sided Fourier transform*, denoted by

$$F_+\{x(\cdot)\} = X_+(\cdot),$$

by

$$X_+(f) \triangleq \begin{cases} 2X(f), & f > 0 \\ 0, & f < 0. \end{cases}$$

Then show that $x(t)$ can be recovered from $X_+(f)$ by the *inverse one-sided Fourier transform,* denoted by

$$x(\cdot) = F_+^{-1}\{X_+(\cdot)\},$$

and defined by

$$x(t) = \text{Re}\left\{\int_0^\infty X_+(f)e^{i2\pi ft}\, df\right\},$$

where Re{·} denotes the real part of the complex quantity in the braces.
Hint: First prove that for a real waveform $x(t)$, $X(f)$ exhibits the Hermitian symmetry

$$X(-f) = X^*(f).$$

Then use the result of exercise 7:

$$F^{-1}\{F[x(\cdot)]\} = x(\cdot).$$

9. Verify *Parseval's relation* for Fourier transforms:

$$\int_{-\infty}^{\infty} x(t)[y(t)]^*\, dt = \int_{-\infty}^{\infty} X(f)[Y(f)]^*\, df. \tag{41}$$

Hint: Substitute the Fourier transform integrals for $X(f)$ and $Y(f)$ into the right member of (41) and then use the transform (from exercise 2)

$$\int_{-\infty}^{\infty} e^{-i2\pi ft} df = \delta(t). \tag{42}$$

10. Verify that a double convolution is given by the double integral

$$x(t) \otimes y(t) \otimes z(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t - u)y(u - v)z(v)\, du\, dv. \tag{43}$$

11. Show that the periodogram-correlogram relation is simply an application of the convolution theorem (exercise 4). *Hint:* Use the change of variables $t = t' + \tau/2$ in (33).

12. Verify the Fourier transform pairs given in Table 1-1 on page 26. The waveforms in this table are defined by (12)–(15) in Chapter II. *Hint:* For $v_T(t)$, use the convolution theorem together with $v_T(t) = u_T(t) \otimes u_T(t)$. For $w_T(t)$, start with $\frac{1}{T}u_{1/T}(f)$, and evaluate the inverse transform.

13. Verify the transform pairs

$$F\{\cos(2\pi f_0 t)\} = \frac{1}{2}\delta(f - f_0) + \frac{1}{2}\delta(f + f_0)$$

$$F\{\sin(2\pi f_0 t)\} = \frac{1}{2i}\delta(f - f_0) - \frac{1}{2i}\delta(f + f_0).$$

Hint: Use Euler's identity, $e^{i\theta} = \cos\theta + i\sin\theta$.

TABLE 1-1 Windows and Transforms

Time-function	Fourier transform
Rectangle $= u_T(t)$	$\frac{1}{T}w_{1/T}(f) = \text{Sinc}$
Triangle $= v_T(t)$	$\frac{1}{T}z_{1/T}(f) = \text{Sinc}^2$
$\frac{1}{T}\,\text{Sinc} = w_T(t)$	$\frac{1}{T}u_{1/T}(f) = \text{Rectangle}$
$\frac{1}{T}\,\text{Sinc}^2 = z_T(t)$	$\frac{1}{T}v_{1/T}(f) = \text{Triangle}$

14. Use the results of exercises 12 and 13 to determine the response of an LTI transformation with impulse-response function $h(t)$ to an excitation $x(t)$ for the following cases.
(a) $x(t) = \cos(2\pi f_0 t)$, $h(t) = u_T(t)$
(b) $x(t) = \sin(2\pi f_0 t)$, $h(t) = v_T(t)$
(c) $x(t) = u_T(t)$, $h(t) = u_T(t)$

15. Use Parseval's relation (exercise 9) to evaluate the integrals

$$\int_{-\infty}^{\infty}\left(\frac{\sin \pi t/T}{\pi t}\right)^2 dt$$

and

$$\int_{-\infty}^{\infty}\left(\frac{\sin \pi t/T}{\pi t}\right)^3 dt.$$

16. The time-frequency dual of the convolution theorem (exercise 4) establishes that the Fourier transform of a product of time functions is the convolution of their Fourier transforms, that is, if

$$z(t) = x(t)y(t)$$

then

$$Z(f) = X(f) \otimes Y(f).$$

Use this theorem and the result of exercise 13 to determine the Fourier transforms of the waveforms $z(t) = x(t)\cos(2\pi f_0 t)$ and $z(t) = x(t)\sin(2\pi f_0 t)$.

APPENDIX 1-1
Linear Time-Invariant Transformations and the Fourier Transform: A Review

Let us begin with a problem that illustrates the utility of sine wave analysis. We consider the problem of determining the current flow through a series connection of a voltage source, resistor, capacitor, and inductor, as depicted in

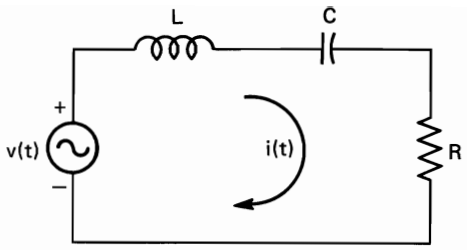


Figure 1-1-1 Electrical circuit.

Figure 1-1-1. By equating the sum of voltages around the circuit to zero, we obtain the integro-differential equation

$$L\frac{di(t)}{dt} + \frac{1}{C}\int i(t)\,dt + Ri(t) - v(t) = 0. \tag{1}$$

We first consider a sine wave excitation

$$v(t) = A\cos(2\pi ft + \theta), \tag{2}$$

and we assume that the response current is a sine wave of the same frequency, f,

$$i(t) = B\cos(2\pi ft + \phi). \tag{3}$$

Substitution of expressions (2) and (3) into (1), the equation relating $i(t)$ to $v(t)$, yields (after some algebraic and trigonometric manipulation)

$$\begin{aligned} B &= A|H(f)| \\ \phi &= \theta + \arg\{H(f)\}, \end{aligned} \tag{4}$$

where

$$H(f) = \frac{i2\pi fC}{1-(2\pi f)^2LC + i2\pi fRC} = |H(f)|e^{i\arg\{H(f)\}}. \tag{5}$$

In (5), i denotes the imaginary number, $i = \sqrt{-1}$, and $|\cdot|$ and $\arg\{\cdot\}$ denote the magnitude and angle, respectively, of a complex number. Thus, (3) and (4) yield the solution

$$i(t) = |H(f)|A\cos(2\pi ft + \theta + \arg\{H(f)\}). \tag{6}$$

Hence, the generally complicated problem of solving an integro-differential equation reduces, in this case of sine wave excitation, to relatively simple algebraic and trigonometric manipulation. The simplicity of the solution is most evident for a *complex sine wave* excitation

$$v(t) = Ae^{i(2\pi ft+\theta)} = A\cos(2\pi ft + \theta) + iA\sin(2\pi ft + \theta). \tag{7}$$

Following the same procedure, we obtain

$$i(t) = H(f)v(t). \tag{8}$$

The solution (6) for a real sine wave excitation is simply the real part of the complex solution (8).

Now, let us inquire if this same simplicity of solution is possible for excitations other than sine waves. To show that it is, in essence, still possible, the *Fourier*

transforms

$$V(\cdot) = F\{v(\cdot)\}, \qquad I(\cdot) = F\{i(\cdot)\}$$

of the waveforms v and i are introduced:

$$V(f) = \int_{-\infty}^{\infty} v(t)e^{-i2\pi ft}\, dt$$

$$I(f) = \int_{-\infty}^{\infty} i(t)e^{-i2\pi ft}\, dt. \tag{9}$$

By Fourier transforming both sides of the integro-differential equation (1) and using the properties summarized by

$$F\left\{L\frac{di}{dt} + \frac{1}{C}\int i\, dt\right\} = L(i2\pi f)F\{i\} + \frac{1}{C}\left(\frac{1}{i2\pi f}\right)F\{i\},$$

we obtain the solution for I in terms of V,

$$I(f) = H(f)V(f), \tag{10}$$

which looks much like the solution (8) for the complex sine wave excitation. However, now we must inverse Fourier transform this result,

$$i(\cdot) = F^{-1}\{I(\cdot)\}, \qquad i(t) = \int_{-\infty}^{\infty} I(f)e^{i2\pi ft}\, df, \tag{11}$$

to obtain

$$i(t) = \int_{-\infty}^{\infty} H(f)V(f)e^{i2\pi ft}\, df. \tag{12}$$

The reason we obtain essentially the same simplicity of solution to the potentially complicated equation (1) by using the Fourier transform is that this transform decomposes the waveforms i and v into continuous sums (integrals) of weighted sine waves (11). Thus, we are using sine wave analysis of waveforms. Let us now consider in more general terms the precise situation for which it is especially useful to analyze (decompose) a waveform into sine wave components.

Sine waves are especially appropriate components with which to analyze waveforms when we are studying *linear time-invariant* (LTI) transformations of waveforms (such as the transformation relating $v(t)$ to $i(t)$ in the preceding circuit problem). An LTI transformation can be characterized by the *convolution operation* that transforms a waveform, say $x(t)$, into another waveform, say $y(t)$, according to the formula

$$\boxed{\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(t-u)x(u)\, du = \int_{-\infty}^{\infty} h(v)x(t-v)\, dv \\ &= h(t) \otimes x(t), \end{aligned}} \tag{13}$$

for some weighting function $h(t)$. The *time-invariance* property is, more precisely, a *translation invariance* property that guarantees that a translation by w of $x(t)$ to $x(t + w)$ has no effect on $y(t)$ other than a corresponding translation to $y(t + w)$ (exercise 1). The linearity property guarantees that the transformation of a linear combination of component waveforms, say $x(t) = a_1x_1(t) + a_2x_2(t)$, is simply

the same linear combination of the transformations of the component waveforms, $y(t) = a_1y_1(t) + a_2y_2(t)$, where $y_1(t) = h(t) \otimes x_1(t)$ is the transformation of $x_1(t)$ and $y_2(t) = h(t) \otimes x_2(t)$ is the transformation of $x_2(t)$ and a_1 and a_2 are arbitrary real or complex numbers. To understand better the nature of the convolution operation that characterizes all LTI transformations, we consider a discrete-time approximation to the integral in (13),

$$y(t) \cong \sum_{n=-\infty}^{\infty} h(n\Delta t)x(t - n\Delta t)\Delta t = \cdots + h(-\Delta t)x(t + \Delta t)\Delta t + h(0)x(t)\Delta t$$
$$+ h(\Delta t)x(t - \Delta t)\Delta t + h(2\Delta t)x(t - 2\Delta t)\Delta t + \cdots, \tag{14}$$

in which Δt is a fixed increment of the variable t. We see from (14) that convolution is approximately a *linear combination* of time-translates, $x(t - n\Delta t)$, of a waveform $x(t)$; that is, it is approximately a sum of weighted versions of the translates $x(t - n\Delta t)$. In fact, (13) reveals directly that convolution is precisely a *continuous linear combination* of time-translates. In order to discover the consequences of this, we observe (exercise 3) that the only type of bounded waveform whose form is invariant to translation—in the sense that

$$x(t + w) = cx(t) \tag{15}$$

for all t and w and for some scalar, c, whose value can depend on w but not on t—is the complex sine wave

$$x(t) = e^{i2\pi ft} \tag{16}$$

for arbitrary real f. The invariance of (15) holds for (16) with

$$c = e^{i2\pi fw}. \tag{17}$$

As a consequence of this unique translation-invariance property, *the complex sinewave is the only bounded waveform whose form is invariant to LTI transformation*. To illustrate, we substitute (16) into (13) to obtain

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(v)e^{i2\pi f(t-v)}\, dv \\ &= \int_{-\infty}^{\infty} h(v)e^{-i2\pi fv}\, dv\, e^{i2\pi ft} \\ &= H(f)e^{i2\pi ft} = H(f)x(t), \end{aligned}$$

where

$$H(f) = \int_{-\infty}^{\infty} h(v)e^{i2\pi fv}\, dv. \tag{18}$$

Thus, with $x(t)$ given by (16), we obtain

$$y(t) = H(f)x(t). \tag{19}$$

In common terminology, the weighting function $h(\cdot)$ is the *impulse-response function* of the LTI transformation (exercise 2), and $H(\cdot)$ is the *transfer function*. Moreover, (18) reveals that these two functions are related by the *Fourier transformation*

$$H(\cdot) = F\{h(\cdot)\}. \tag{20}$$

As a result of this invariance property of complex sine waves with respect to LTI transformations, the study of LTI transformations is greatly simplified by *analysis* (decomposition) of the waveforms subjected to the transformation into sine wave components. To explain, we first observe that a real waveform with finite *energy,*[17] denoted by $\mathscr{E}_x$,

$$0 < \mathscr{E}_x \triangleq \int_{-\infty}^{\infty} x^2(t)\, dt < \infty\,, \tag{21}$$

can be exactly represented[18] on any finite time-interval, say $[-T/2, T/2]$, by a denumerable linear combination of sine waves, namely, the *Fourier series,*

$$\begin{aligned} x(t) &= \sum_{m=-\infty}^{\infty} X_m e^{i2\pi m\Delta f t}\, \Delta f \\ &= \cdots [X_{-1}\Delta f] e^{-i2\pi\Delta f t} + [X_0\Delta f] + [X_1\Delta f] e^{i2\pi\Delta f t} \\ &\qquad + [X_2\Delta f] e^{i2\pi 2\Delta f t} + \cdots, \qquad t \in [-T/2, T/2], \end{aligned} \tag{22a}$$

for which $\Delta f = 1/T$. The amount of the sine wave component $e^{i2\pi m\Delta f t}$ in the representation is the *Fourier coefficient* $X_m\Delta f$ for which

$$X_m \triangleq \int_{-1/2\Delta f}^{1/2\Delta f} x(t) e^{-i2\pi m\Delta f t}\, dt. \tag{22b}$$

Since X_m and X_{-m} are a complex conjugate pair (for real $x(t)$), then

$$X_{-m} e^{-i2\pi m\Delta f t} + X_m e^{i2\pi m\Delta f t} = 2|X_m|\cos(2\pi m\Delta f t + \arg\{X_m\}). \tag{23}$$

To illustrate, we consider as an example the specific waveform

$$x(t) = \begin{cases} e^{-at}, & t \geqslant 0 \\ 0, & t < 0. \end{cases} \tag{24}$$

Substitution of (24) into (22b) yields

$$\begin{aligned} X_m &= \frac{1}{a + i2\pi m\Delta f}(1 - e^{-aT/2} e^{-i\pi m\Delta f T}) \\ &\cong \frac{1}{a + i2\pi m\Delta f} \qquad \text{for } T \gg \frac{1}{a}. \end{aligned} \tag{25}$$

The magnitude of the mth sine wave coefficient is therefore

$$|X_m| \cong \frac{1}{[a^2 + (2\pi m/T)^2]^{1/2}}. \tag{26}$$

A graph of these magnitudes versus the discrete frequency variable $f = m\Delta f = m/T$ is shown in Figure 1-1-2 for increasing values of T. The envelope of the coefficient magnitudes, which is described by the function

$$|X(f)| = \frac{1}{[a^2 + (2\pi f)^2]^{1/2}}, \tag{27}$$

[17] If $x(t)$ is a voltage (measured in *volts*) developed across a resistance of 1 *ohm*, then $\mathscr{E}_x$ is the energy in *joules* dissipated in the resistance.

[18] More precisely, the energy in the error between $x(t)$ and a finite term approximation (truncated version of the infinite series (22a)) approaches zero as the number of terms approaches infinity.

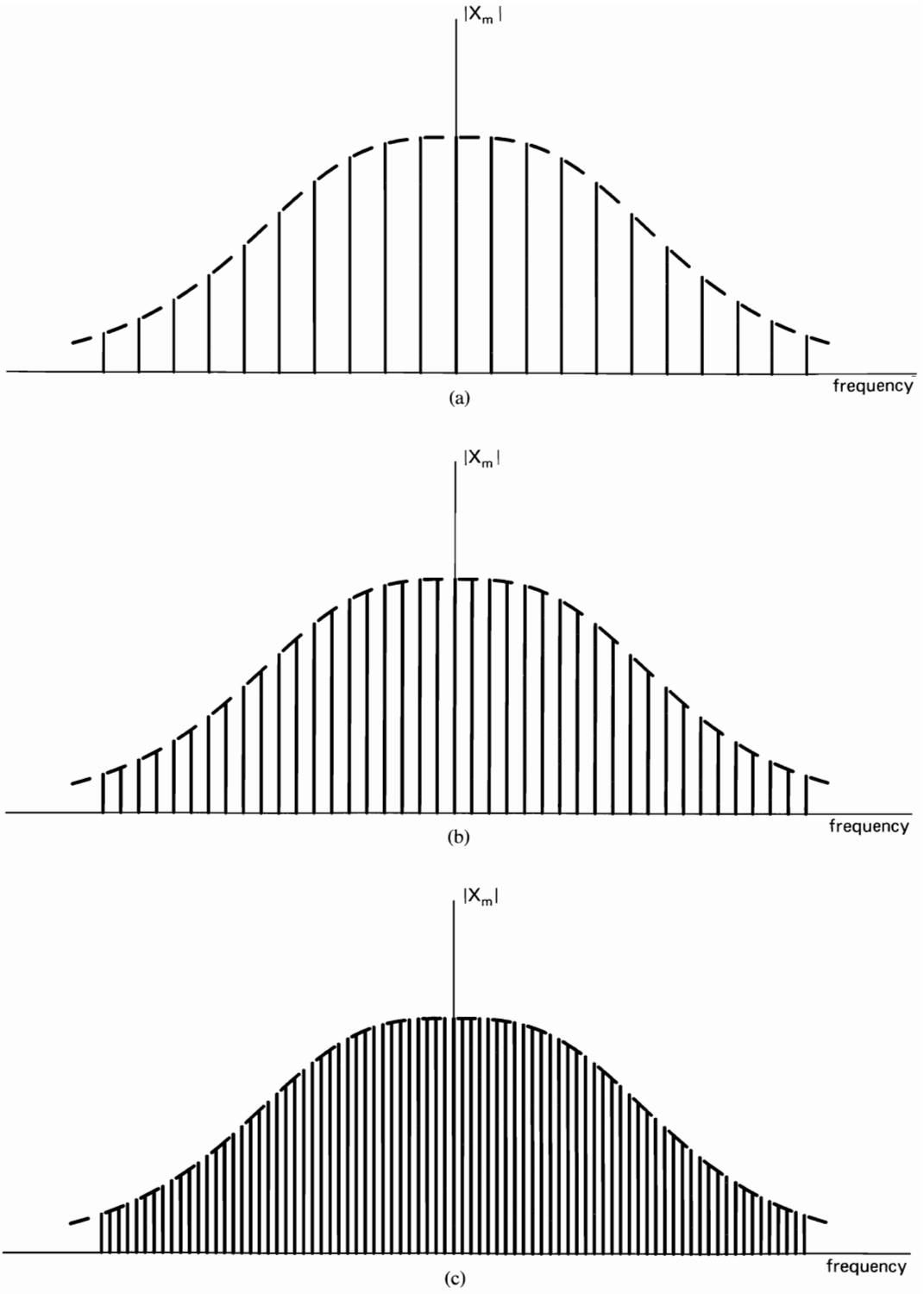


Figure 1-1-2 Magnitudes of sine wave components (26) of decaying exponential waveform (24) on an interval of increasing length T: (a) $T = T_0$, (b) $T = 2T_0$, (c) $T = 4T_0$.

remains fixed as T increases, but the individual sine wave coefficients become increasingly more dense.

By letting $T \to \infty$ ($\Delta f \to 0$) in (22a), we heuristically see that $x(t)$ can be exactly represented for all time t by a nondenumerable linear combination (i.e., a continuous weighted sum) of sine waves,

$$\boxed{x(t) = \int_{-\infty}^{\infty} X(f)e^{i2\pi ft}\, df,} \tag{28a}$$

for which $X(f)$ is heuristically obtained from

$$X(f) = \lim_{\Delta f \to 0} X_m$$

with $f = m\Delta f$, which yields the definition

$$\boxed{X(f) \triangleq \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\, dt.} \tag{28b}$$

Equations (28a) and (28b) are the *inverse Fourier transform* and *Fourier transform,* respectively:

$$x(\cdot) = F^{-1}\{X(\cdot)\}, \qquad X(\cdot) = F\{x(\cdot)\}. \tag{29}$$

The decomposition (28a) of $x(t)$ into *sine wave components* indexed by f,

$$\lim_{\Delta f \to 0} X_m e^{i2\pi m\Delta ft} = X(f)e^{i2\pi ft} \triangleq x_f(t), \tag{30}$$

when substituted into the LTI transformation (13), yields the desired result:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} X(f)e^{i2\pi f(t-u)}\, df\, du \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u)e^{-i2\pi fu}\, du\, X(f)e^{i2\pi ft}\, df \\ &= \int_{-\infty}^{\infty} H(f)X(f)e^{i2\pi ft}\, df. \end{aligned} \tag{31}$$

That is,

$$y = F^{-1}\{Y\}$$

$$\boxed{Y(f) = H(f)X(f).} \tag{32}$$

In conclusion, *the effect of an LTI transformation on a waveform is simply to scale its sine wave components.* The scaling function is the transfer function, and therefore

$$y_f(t) \triangleq Y(f)e^{i2\pi ft} = H(f)X(f)e^{i2\pi ft} \triangleq H(f)x_f(t). \tag{33}$$

There is no interaction among these sine wave components when x is transformed into y. This is due in part to the fact that *the sinewave components are uncorrelated with each other,* in the sense that

$$\langle x_{f_1}, x_{f_2} \rangle_t \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_{f_1}(t)x_{f_2}^*(t)\, dt = 0, \qquad f_1 \neq f_2 \tag{34}$$

and

$$\langle y_{f_1}, y_{f_2} \rangle_t = 0, \qquad f_1 \neq f_2 \tag{35}$$

(the asterisk denotes complex conjugation). Property (34) can be verified (exercise 5) by substitution of (30) into (34). Because of these properties, (33)–(35), of sine wave components with respect to LTI transformations, the components $y_f(t)$ are called the *principal components* of $y(t)$ with respect to the LTI transformation of $x(t)$. The values of $H(f)$ are called the *principal values,* or the *eigenvalues,* and the functions $e^{i2\pi ft}$ are called the *eigenfunctions* of the LTI transformation. (The prefix *eigen* means *characteristic.*)

It follows from (28a) that $X(f)$ is the *density of sine wave components* contained in $x(t)$. Moreover, the Fourier transformation (28b) and its inverse (28a) can be expressed explicitly in terms of sine wave components as

$$x(t) = \int_{-\infty}^{\infty} x_f(t)\, df \tag{28a}'$$

$$x_f(t) = \int_{-\infty}^{\infty} x(u) e^{i2\pi f(t-u)}\, du. \tag{28b}'$$

Since (28b)′ is a convolution, we see that $x_f(t)$ is the response to $x(t)$ of an LTI transformation with impulse-response function

$$g(t) = e^{i2\pi ft} \tag{36}$$

and corresponding transfer function (exercise 2)

$$G(\nu) = \delta(f - \nu), \tag{37}$$

which is an *ideal filter* (with infinite gain) that passes only the single sine wave of frequency f and transforms this infinitesimal sine wave component into a finite sine wave component.

As a final item in this brief review it is pointed out that inspection of (13) and (32) reveals that if a function, say y, is given by the convolution of two other functionsd, say x and h, then the Fourier transform of y, Y, is given by the *product* of Fourier transforms of x and H: $Y = XH$. This result is known as the *convolution theorem.* Additional review material is incorporated in exercises 2, 4, 7, 8, 9, 10, 12, 13, 14, and 16.

2

NONSTATISTICAL SPECTRAL ANALYSIS

This chapter introduces the basic elements of empirical spectral analysis, namely, the *time-variant periodogram* and *time-variant correlogram*, and establishes the fact that these two functions are a Fourier transform pair. The temporal and spectral resolution capability of the time-variant periodogram is determined, and a fundamental *time-frequency uncertainty principle* is established. This principle is illustrated by application to instantaneous frequency measurement in Appendix 2-1. The relationships between the time-variant periodograms of the excitation and response and between the time-variant correlograms of the excitation and response of a linear time-invariant transformation are derived. These relationships are illustrated by application to instantaneous frequency demodulation in Appendix 2-1. They are also employed to derive a time-variant local-average power spectral density function. Finally, the spectral-aliasing phenomenon associated with periodic time-sampling is explained in terms of the time-variant complex spectrum.

Throughout this chapter and the rest of the book, it is assumed that the mathematical model for each time-series of interest, say $x(t)$, unless otherwise specified is sufficiently well behaved to be lag-product integrable as well as Fourier transformable on every finite interval; that is, the integral

$$\int_a^b x\left(t + \frac{\tau}{2}\right)x\left(t - \frac{\tau}{2}\right)dt$$

exists for every finite a and b. Then all of the finite-interval correlations and spectra defined in this chapter exist. It is also assumed that the limit

$$\lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right)x\left(t - \frac{\tau}{2}\right) dt$$

exists for every finite τ and unless otherwise specified is continuous at $\tau = 0$ (and therefore is continuous at every τ [Gardner 1985].) Then all limit correlations defined in this and following chapters exist.

Motivating Example:

We consider the situation in which a physical system possesses several modes of resonance. This could be an electrical circuit, a mechanical system, or some other type of physical system. Let us assume that the system is repeatedly (aperiodically) subjected to impulse excitations and the system response (e.g., voltage or displacement) is recorded. In order to determine the natural frequencies of resonance and their associated damping factors, or bandwidths, it is desired to Fourier transform the recorded data in an attempt to estimate the transfer function of the system, which exhibits a peak for each resonant mode of the system. If the impulse excitations occur irregularly and the corresponding responses overlap each other in time, then it is not clear how long a segment of the recorded composite response should be Fourier analyzed. To complicate matters further, suppose that the physical system is changing with time, so that the resonant frequencies and bandwidths are changing. If it is desired to track these changes by allowing the time interval over which the recorded data is Fourier transformed to slide along with time, then how does this affect our choice of the segment length that is to be analyzed at each time instant? In this chapter we shall obtain answers to these and related questions by determining the temporal and spectral resolution capabilities of the time-variant periodogram obtained by Fourier transforming a sliding segment of data.

A. TEMPORAL AND SPECTRAL RESOLUTION

The *time-variant finite-time spectrum*[1] of $x(t)$, also called the *time-variant periodogram* of $x(t)$, is defined by

$$S_{x_T}(t,f) \triangleq \frac{1}{T}|X_T(t,f)|^2 \tag{1}$$

$$X_T(t,f) \triangleq \int_{t-T/2}^{t+T/2} x(u)e^{-i2\pi fu}du \tag{2a}$$

and is the normalized squared magnitude of the Fourier transform of a data-segment of length T centered at time t, as depicted in Figure 2-1. As revealed by the inverse Fourier transformation,

$$x(u) = \int_{-\infty}^{\infty} X_T(t,f)e^{i2\pi fu}\,df, \qquad u \in [t - T/2, t + T/2], \tag{2b}$$

the Fourier transform $X_T(t, f)$ is the density of complex sine wave components $\{e^{i2\pi fu} : -\infty < f < \infty\}$ contained in $x(u)$ for $u \in [t - T/2, t + T/2]$. This transform is called the *time-variant finite-time Fourier transform*, or *time-variant finite-time complex spectrum*. In order to reveal the resolving power of this spectral measurement on $x(t)$, it can be reexpressed in the two alternative forms

[1] This function was originally introduced by Harold Thayer Davis (1892–) [Davis 1941] for tracking nearly periodic components with slowly evolving amplitude and phase.

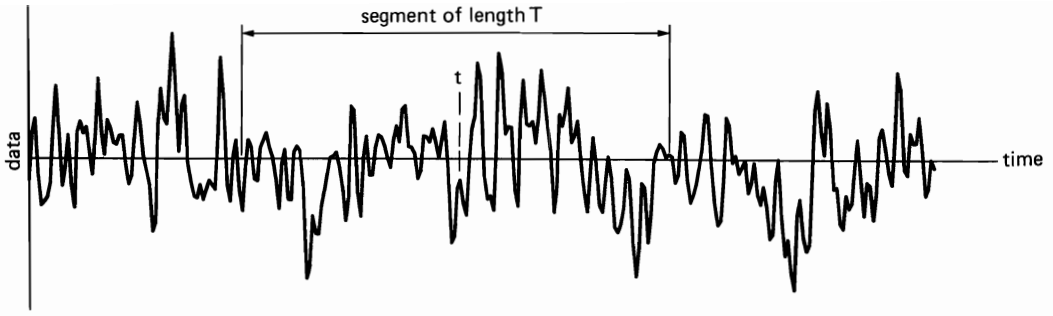


Figure 2-1 Data segment for time-variant spectral analysis.

(exercise 1)

$$S_{x_T}(t,f) = \frac{1}{T}\left|[x(t)e^{-i2\pi ft}] \otimes a_T(t)\right|^2 \tag{3}$$

$$= \frac{1}{T}\left|[X_T(t,f)e^{i2\pi ft}] \otimes A_{1/T}(f)\right|^2, \tag{4}$$

in which a_T is a rectangle function of width T (to be generalized following the discussion in this section)

$$a_T(t) = \begin{cases} 1, & |t| \leq T/2 \\ 0, & |t| > T/2, \end{cases} \tag{5}$$

and $A_{1/T}$ is a sinc function with width parameter $1/T$ (also to be generalized following this discussion):

$$A_{1/T}(f) = \frac{\sin(\pi fT)}{\pi f}. \tag{6}$$

In fact, a_T and $A_{1/T}$ are a Fourier transform pair

$$A_{1/T}(\cdot) = F\{a_T(\cdot)\}. \tag{7a}$$

That is,

$$A_{1/T}(f) = \int_{-\infty}^{\infty} a_T(t)e^{-i2\pi ft}\,dt. \tag{7b}$$

Since convolution of any function, say y, with any *pulselike function* typically removes all fine structure (wiggles) in y within intervals of the length of the pulselike function (or less),[2] as illustrated in Figure 2-2, then (3) reveals that $S_{x_T}(t,f)$ typically will not have fine structure in t within intervals of length T or less, and (4) reveals that $S_{x_T}(t,f)$ typically will not have fine structure in f within intervals of length $1/T$ or less. This is further illustrated in exercises 2 and 3.

[2] There are exceptions to this general rule, but they are typically pathological, in the sense that the function y cannot represent random data but instead must exhibit a special type of structure; see exercise 2 with $T = 10\Delta$.

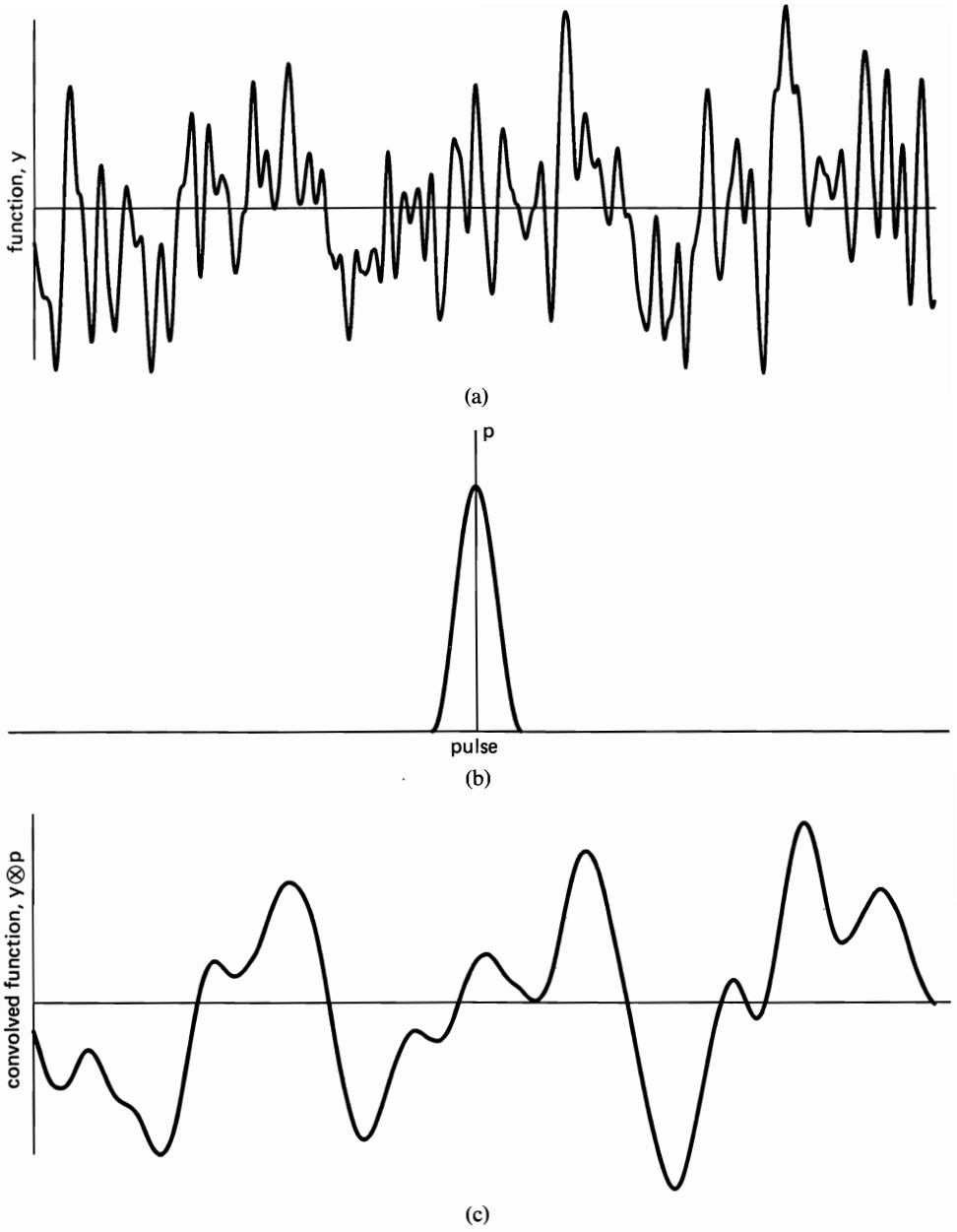


Figure 2-2 (a) A function y for illustration of the smoothing effect of convolution with a pulselike function. (b) A pulselike function p. (c) Smoothed version of function y obtained by convolving with pulselike function p.

Thus, the time-variant finite-time spectrum $S_{x_T}(t, f)$ has *temporal resolution width*, denoted by Δt°, of

$$\Delta t^{\circ} = T \tag{8}$$

and has *spectral resolution width*, denoted by Δf°, on the order of[3]

$$\Delta f^{\circ} \cong \frac{1}{T}. \tag{9}$$

The product of temporal and spectral resolution widths is therefore on the order of unity,

$$\Delta t^{\circ} \Delta f^{\circ} \cong 1, \tag{10}$$

regardless of the length T of the analysis interval $[t - T/2, t + T/2]$.

Another approach to establishing that $X_T(t, f)$—and, therefore, $S_{x_T}(t, f)$—has spectral resolution width on the order of $1/T$ is to prove that $X_T(t, f)$ can be exactly reproduced only from its frequency samples at $f = n/T$ for all integers n by interpolation. Thus there cannot be significant fine structure between points separated by $1/T$ or less. This is explained in exercise 15.

The practical significance of the preceding results ((8)–(10)) on resolution can be explained as follows. If there are spectral features in the data $x(t)$ that are as narrow as Δf^* (e.g., spectral peaks due to resonance phenomena), for example, then these cannot be accurately resolved by the periodogram unless $\Delta f^{\circ} < \Delta f^*$. Thus, the data segment length T analyzed must satisfy $T > 1/\Delta f^*$. If the spectral features are changing with time and significant changes occur in time intervals as small as Δt^*, for instance, then these time-variations cannot be accurately tracked by the time-variant periodogram unless $\Delta t^{\circ} < \Delta t^*$. Thus, the data-segment length T analyzed for each time instant t must satisfy $T < \Delta t^*$. Furthermore, (10) reveals that both tasks of spectral resolution and temporal resolution (tracking) can be performed accurately only if $\Delta t^* \Delta f^* > 1$. In fact, it can be reasoned that it makes no sense even to conceive of a spectral feature of width Δf^* changing substantially in a time-interval of length $\Delta t^* < 1/\Delta f^*$, regardless of the physical phenomenon (see exercise 23).

B. DATA TAPERING

The pulselike function a_T in (3) plays the role of a *temporal aperture*, or *window*, through which the data is seen, as revealed by reexpressing (2a) as

$$X_T(t, f) = \int_{-\infty}^{\infty} a_T(v) x(t - v) e^{-i2\pi f(t-v)}\, dv \tag{11a}$$

or

$$X_T(t, f) = [x(t) e^{-i2\pi ft}] \otimes a_T(t). \tag{11b}$$

That is, the complex spectrum $X_T(t, f)$ depends on those values of $x(t - v)$ that occur within the interval of v determined by $a_T(v)$—those values of x *seen through* a_T. Similarly, the function $A_{1/T}$ plays the role of a *spectral window* as revealed by (4) (which is analogous to (11b) except for the square). Furthermore, these temporal and spectral windows are a Fourier transform pair (7). Moreover,

[3] The particular value here depends on the particular definition of width adopted. For example, if the width is defined to be the distance between the first zero-crossings to the left and right of the center of the pulse (6), then $\Delta f^{\circ} = 2/T$.

since the Fourier transform of any pulselike function is itself pulselike and the width of this Fourier transform is on the order of the reciprocal of the width of the original pulselike function[4] (exercises 14 and 15), then relation (10) holds regardless of the particular shape of the temporal window a_T. However, the particular shape of window, especially the spectral window $A_{1/T}$, can be of considerable importance, as discussed in the following chapters, and for this reason the definition of the time-variant finite-time spectrum is generalized to allow for an arbitrary pulselike function for the temporal aperture a_T in (3). Four apertures of particular theoretical interest are defined as follows:

Rectangle:

$$u_T(t) \triangleq \begin{cases} \dfrac{1}{T}, & |t| \leq T/2 \\ 0, & |t| > T/2 \end{cases} \tag{12}$$

Triangle:

$$v_T(t) \triangleq \begin{cases} \dfrac{1}{T}\left(1 - \dfrac{|t|}{T}\right), & |t| \leq T \\ 0, & |t| > T \end{cases} \tag{13}$$

Sinc:

$$w_T(t) \triangleq \frac{\sin(\pi t/T)}{\pi t} \triangleq \frac{1}{T}\operatorname{sinc}\left(\frac{t}{T}\right) \tag{14}$$

Squared sinc:

$$z_T(t) \triangleq T\left|\frac{\sin(\pi t/T)}{\pi t}\right|^2 = \frac{1}{T}\operatorname{sinc}^2\left(\frac{t}{T}\right). \tag{15}$$

Each of these windows, which are depicted in Figure 2-3, is defined to have unity area. An aperture with unity area preserves the level of the function it is convolved with, whereas an aperture with unity height preserves the level of the function it multiplies.

Apertures other than the rectangle have a *tapering* effect on the data they multiply, since data occurring away from the aperture center are attenuated relative to the data at the aperture center. Consequently, temporal windows other than the rectangle are called *data-tapering windows*. The symbol a_T will be used from now on to denote an arbitrary data-tapering window of approximate width T and unity height at the origin. For example, a_T can represent any of the four windows Tu_T, Tv_T, Tw_T, or Tz_T. It should be noted that if a_T is not an even function, $a_T(-t) \not\equiv a_T(t)$, then in expressions (3) and (11), $a_T(-t)$ rather than $a_T(t)$ is the window that multiplies the data. This particular definition of a_T enables $X_T(t, f)$ to be expressed as a convolution with $a_T(t)$ rather than with $a_T(-t)$. It should also be noted that with a_T representing apertures other than

[4] The pulselike functions for which this reciprocal relationship holds are sometimes called *simple functions*, and functions for which the product of widths greatly exceeds unity are called *complex*, or *sophisticated, functions* (cf. [Vakman 1968]).

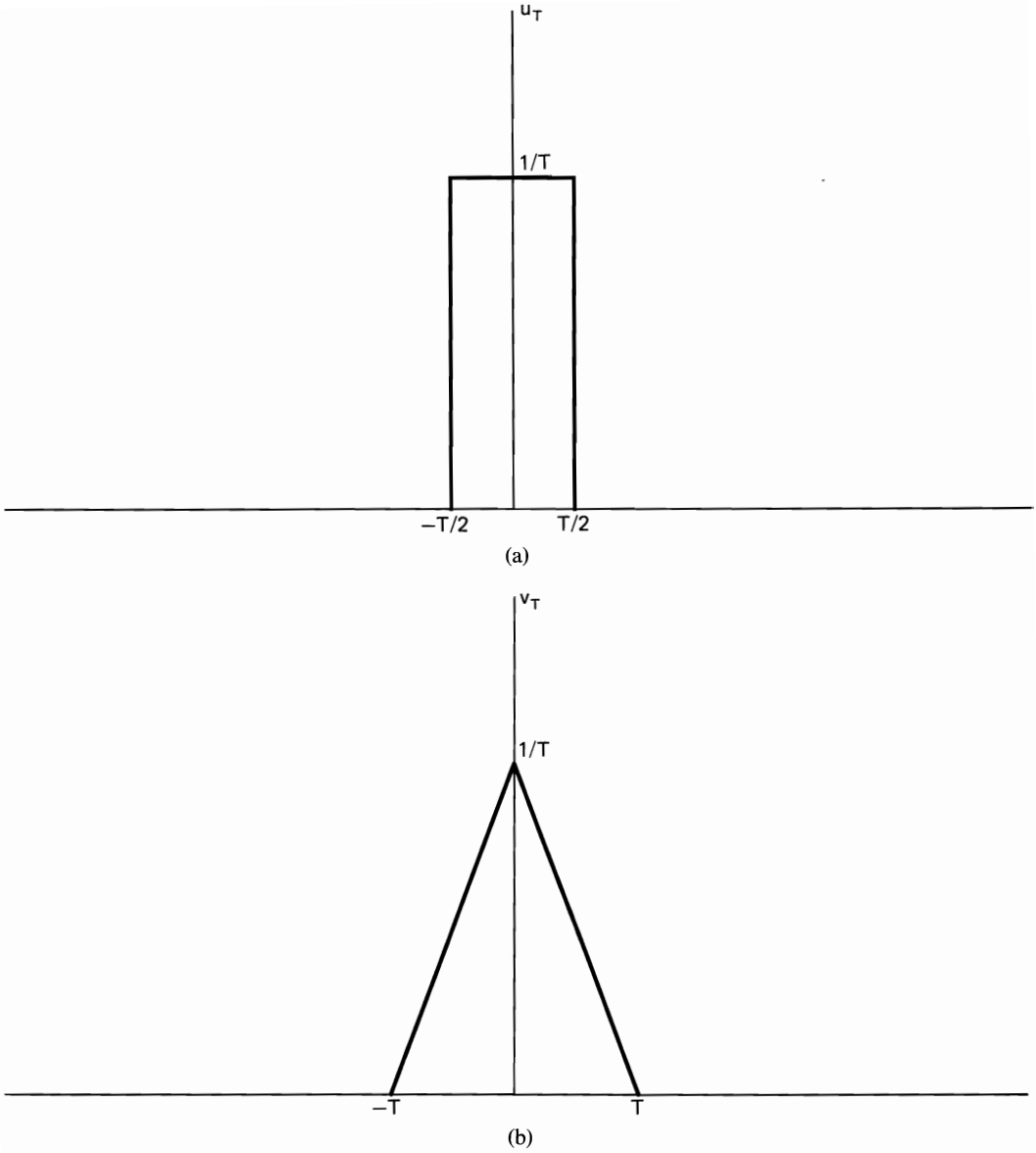
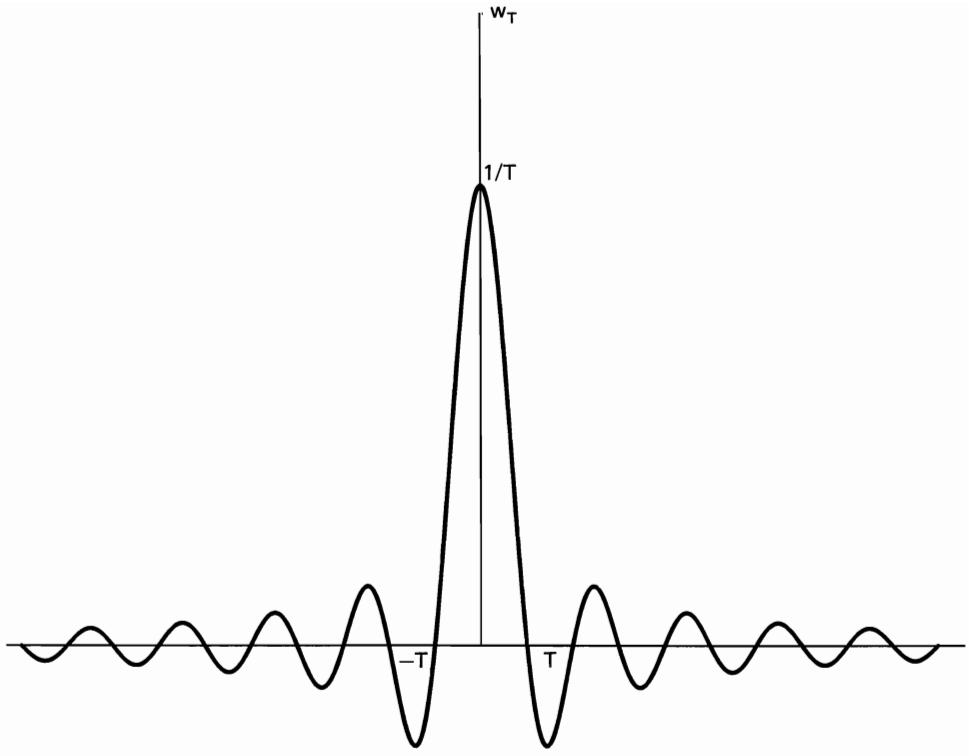
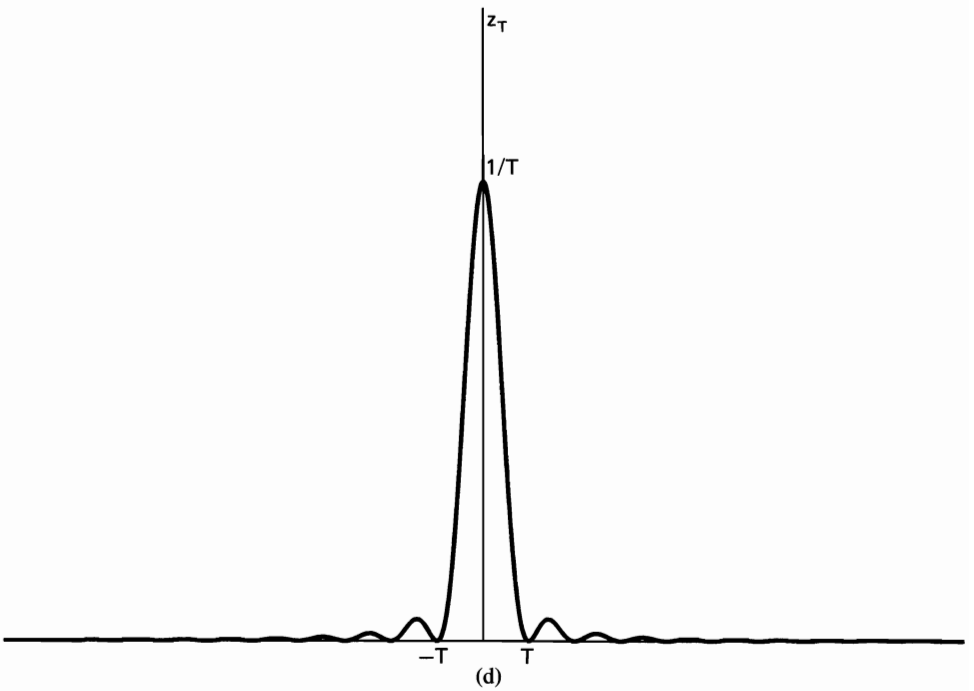


Figure 2-3 (a) Rectangle window u_T. (b) Triangle window v_T.

the rectangle, expression (4) must be modified. For example, (4) is valid, with $X_T(t, f)$ in $S_{x_T}(t, f)$ in the left member of (4) defined by (11) for any data-tapering window satisfying $a_T(t) = 0$ for $|t| > T/2$, only if $X_T(t, f)$ in the right member of (4) is defined by (2a) (which is (11) with $a_T = Tu_T$); otherwise, (4) must be further modified. For example, (4) is valid with $X_T(t, f)$ in $S_{x_T}(t, f)$ defined by (11) for any positive a_T if for the right member of (4), a_T is replaced in (7) and (11) with its square root. In any case, the interpretation of (4) or its modified



(c)



(d)

Figure 2-3 (continued) (c) Sinc window $w_T(t) = \dfrac{\sin \pi t/T}{\pi t}$. (d) Sinc-squared window $z_T(t) = T\left[\dfrac{\sin \pi t/T}{\pi t}\right]^2$.

versions is the same. It reveals that the spectral resolution width of $S_{x_T}(t, f)$ is on the order of $1/T$ for any pulselike aperture a_T of width T.

C. TIME-FREQUENCY UNCERTAINTY PRINCIPLE

When the amount of data to be analyzed is fixed at T units of time, then the spectral resolution width is fixed to be on the order of $1/T$. However, the precise value of spectral resolution width Δf° depends on the particular data-tapering window via (7) (as well as on the particular definition of width of the pulselike function $A_{1/T}$). It is therefore of interest to determine the particular data-tapering window a_T of a given width that yields the finest possible spectral resolution. By adopting the square root of the second central moment (standard deviation) of the square of a function as a particular measure of its width, it can be shown [Franks 1969] that the product of widths of temporal and spectral apertures is minimized by the Gaussian aperture,

$$a_T(t) = \exp\left[\frac{-(t/T)^2}{2}\right], \tag{16a}$$

whose Fourier transform also is Gaussian:

$$A_{1/T}(f) = T\sqrt{2\pi}\exp\left[\frac{-(2\pi f T)^2}{2}\right]. \tag{16b}$$

The minimized resolution product is

$$\min\{\Delta t^{\circ}\Delta f^{\circ}\} = \frac{1}{2\pi}, \tag{17}$$

for which Δt° and Δf° are defined to be the second central moments of the (unsquared but positive) apertures, (16). Apertures that are more convenient for implementation such as (13) (and other measures of width) yield resolution products that are closer to unity than (17).

The general relation (10) and the specific bound (17) are referred to as Dennis Gabor's *time-frequency uncertainty principle* [Gabor 1946], after Werner Heisenberg's related *principle of indeterminacy* for wave mechanics, which was formulated in 1927 (see [Vakman 1968; Robinson 1982]).

D. PERIODOGRAM-CORRELOGRAM RELATION

As a slight generalization of the periodogram-correlogram relation described in Chapter 1, Section C, we have the following relation, which can be obtained by application of the convolution theorem for Fourier transforms:

$$S_{x_T}(t, \cdot) = F\{R_{x_T}(t, \cdot)\}, \tag{18}$$

for which the function $R_{x_T}(t, \tau)$ is defined by

$$R_{x_T}(t, \tau) \triangleq \frac{1}{T}\int_{t-(T-|\tau|)/2}^{t+(T-|\tau|)/2} x(v + \tau/2)x(v - \tau/2)\, dv\, [2Tu_{2T}(\tau)], \tag{19}$$

and is called the *time-variant correlogram* and also the *time-variant finite-time autocorrelation*. This function can be obtained from the more conventional static correlogram, described in Chapter 1, Section C, by simply letting the time location

of the data segment of length T evolve with time t so that $[-T/2, T/2]$ is replaced by $[t - T/2, t + T/2]$. The limits of integration in (19) reflect the fact that the interval of overlap of data segments of length T, centered at $t + \tau/2$ and $t - \tau/2$, is $[t + T/2 - |\tau|/2, t - T/2 + |\tau|/2]$ (see Figure 2-4). The unity-height rectangle-window factor in (19) reflects the fact that this interval of overlap vanishes when $|\tau| > T$, and therefore the correlogram vanishes. This identity (18) can be further generalized (exercise 5) to incorporate data tapering by replacing definition (1) of $S_{x_T}(t, f)$ with the more general form (3), in which a_T is any data-tapering window, and by replacing definition (19) of $R_{x_T}(t, \tau)$ with the more general definition

$$R_{x_T}(t, \tau) \triangleq \frac{1}{T} \int_{-\infty}^{\infty} a_T(v + \tau/2)x(t - v - \tau/2)a_T(v - \tau/2)x(t - v + \tau/2)\, dv. \tag{20}$$

The subscript T in $R_{x_T}(t, \tau)$ denotes an approximate width of this function of τ, similar to the width T of the triangle window v_T with base $2T$. (The exact width beyond which $R_{x_T}(t, \tau) = 0$ is $2T$ if $a_T(\tau) = 0$ for $|\tau| > T/2$.) The generalized definitions (1), (11a), and (20) reduce (exercise 5) to the specific definitions (1), (2a), and (19), respectively, when a_T is the rectangle aperture (see Figure 2-4).

E. FINITE-AVERAGE AUTOCORRELATION AND PSEUDOSPECTRUM

In the development of the statistical theory presented in subsequent chapters, another definition of an autocorrelation function plays a fundamental role. By contrast with definition (19), which yields the correlation of a finite data-segment $x(t - v)[Tu_T(v)]$, the alternative definition

$$R_x(t, \tau)_T \triangleq \frac{1}{T} \int_{t-T/2}^{t+T/2} x(v + \tau/2)x(v - \tau/2)\, dv\, [2Tu_{2T}(\tau)] \tag{21}$$

yields the correlation of an unlimited data-segment, but integration of the *lag product* $x(v + \tau/2)x(v - \tau/2)$ is carried out over only a finite interval (compare with (20)), and only a finite set of *lag values* $|\tau| \leq T$ are considered. The function $R_x(t, \tau)_T$ is called the *time-variant finite-average autocorrelation*.

Motivated by the periodogram-correlogram relation (18), an alternative time-variant spectrum is *defined* by

$$S_x(t, \cdot)_T \triangleq F\{R_x(t, \cdot)_T\}. \tag{22}$$

The function $S_x(t, f)_T$ is called the *time-variant pseudospectrum*. The term *pseudo* is used because, unlike the periodogram $S_{x_T}(t, f)$, the function $S_x(t, f)_T$ can take on negative values. Moreover, it is not obtainable from a Fourier transform of a segment of data in contrast to (1).

It should be noted that the distinction between the definitions of the two types of spectra defined by (1) (or (3)) and (22), and between the two types of autocorrelations defined by (19) (or (20)) and (21), is denoted by the two different locations of the subscript T.

The primary reason that the finite-average autocorrelation and its Fourier

transform, the pseudospectrum, play a fundamental role in the theory of statistical (time-smoothed) spectral analysis is because a time-smoothed spectrum is closely approximated by a frequency-smoothed pseudospectrum. This follows from the periodogram-correlogram relation and the following identity between time-smoothed autocorrelations (for untapered data):

$$[R_{x_T}(t, \tau) \otimes u_{\Delta t}(t)]2\Delta t\, u_{2\Delta t}(\tau) = [R_x(t, \tau)_{\Delta t} \otimes u_{T-|\tau|}(t)]\, Tv_T(\tau) \tag{23}$$

(for which both convolutions are in the time variable t), which can be used to derive the approximation

$$R_{x_T}(t, \tau) \otimes u_{\Delta t}(t) \cong R_x(t, \tau)_{\Delta t}[Tv_T(\tau)], \qquad T \ll \Delta t.$$

This approximation reveals that a time-smoothed correlogram can be reinterpreted in terms of a tapered (by v_T) finite-average autocorrelation. Fourier transformation (in τ) of both sides of this approximation yields the desired result, which is explained in Chapter 3, Section B.

In preparation for discussions in subsequent chapters of several such equivalences between time-smoothing and frequency-smoothing, the underlying approximations involving autocorrelations are described here.[5] First, it is noted that (23) follows (exercise 6) from the more fundamental *window identity*

$$u_T\left(t - \frac{\tau}{2}\right)u_T\left(t + \frac{\tau}{2}\right) \equiv v_T(\tau)u_{T-|\tau|}(t) = \begin{cases} \dfrac{1}{T^2}, & |t| \leqslant \dfrac{T - |\tau|}{2} \\ 0, & |t| > \dfrac{T - |\tau|}{2}, \end{cases} \tag{24}$$

which is depicted in Figure 2-4. Now, identity (23) can be used to derive the approximation

$$R_{x_T}(t, \tau) \otimes u_{\Delta t}(t) \cong R_x(t, \tau)_{\Delta t}[Tv_T(\tau)], \qquad T \ll \Delta t. \tag{25}$$

This approximation can be made as accurate as desired for any given t and T and all τ by choosing Δt sufficiently large, provided only that the limit autocorrelation $\hat{R}_x(\tau)$ ((6), Chapter 1) exists (exercise 9). Analogous to identity (23), it can be shown that

$$[R_x(t, \tau)_T \otimes u_{\Delta t}(t)]2\Delta t\, u_{2\Delta t}(\tau) = [R_x(t, \tau)_{\Delta t} \otimes u_T(t)]\, 2Tu_{2T}(\tau), \tag{26}$$

and this identity can be used to derive the approximation

$$R_x(t, \tau)_T \otimes u_{\Delta t}(t) \cong R_x(t, \tau)_{\Delta t}\, [2Tu_{2T}(\tau)], \qquad T \ll \Delta t. \tag{27}$$

Approximation (27) can be made as accurate as desired for any given t and T and all τ by choosing Δt sufficiently large, provided only that $\hat{R}_x(\tau)$ exists (exercise 8). As a generalization of approximation (25) which is for untapered data, it can be shown by using definition (20) that for tapered data

$$R_{x_T}(t, \tau) \otimes u_{\Delta t}(t) \cong R_x(t, \tau)_{\Delta t}\frac{1}{T} r_a(\tau), \qquad T \ll \Delta t, \tag{28}$$

for which the function

$$r_a(\tau) \triangleq a_T(\tau) \otimes a_T(-\tau) = \int_{-\infty}^{\infty} a_T(u + \tau)a_T(u)\, du \tag{29}$$

[5] The reading of the material in this paragraph can be postponed until needed in subsequent chapters.

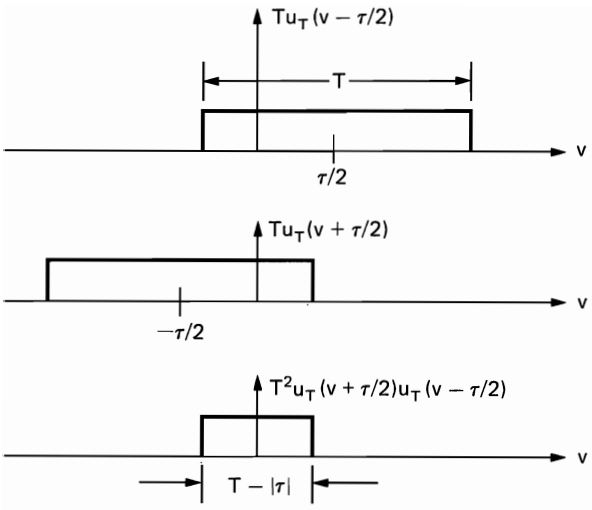


Figure 2-4 Product of shifted windows.

is called the *finite autocorrelation* of a_T. Approximation (28) can be made as accurate as desired for any given t and T, all τ, and any given bounded aperture whose support[6] is contained within a finite interval by choosing Δt sufficiently large, provided only that $\hat{R}_x(\tau)$ exists (exercise 10). Finally, it can be shown that the two types of autocorrelations in (19) and (21) are approximately equal over a limited range of τ. For convenience in the sequel, let T in (19) and (21) be replaced by Δt; then this approximation can be expressed as

$$R_x(t, \tau)_{\Delta t} \cong R_{x_{\Delta t}}(t, \tau), \qquad |\tau| \leq T \ll \Delta t. \tag{30}$$

This approximation can be made as accurate as desired for any given t and T, by choosing Δt sufficiently large, provided only that $\hat{R}_x(\tau)$ exists. The accuracy of approximations (25), (27), (28), and (30) can be quantified in terms of rms error, that is, the square root of the average over all time t of the squared error of approximation. This approach is set up in Chapter 5.

Related to the preceding approximations is the fact that the limit autocorrelation ((6), Chapter 1) can be obtained (exercise 7) from the limit of either the correlogram (for untapered data) or the finite-average autocorrelation

$$\hat{R}_x(\tau) = \lim_{T\to\infty} R_{x_T}(t, \tau) = \lim_{T\to\infty} R_x(t, \tau)_T, \tag{31}$$

and this limit is independent of t. Furthermore, for tapered data it can be shown[7] that

$$\lim_{T\to\infty} R_{x_T}(t, \tau) = \gamma \hat{R}_x(\tau), \tag{32}$$

for which the scale factor

$$\gamma \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} a_T^2(t)\, dt \tag{33}$$

[6] The term *support* denotes the domain over which a function is nonzero.

[7] (31) is established in [Wiener 1930] (see also exercise 7 and [Kampé de Fériet 1954]), and (32) is established in exercise 11.

is typically on the order of unity, since a_T has width parameter T and unity height parameter (see (5)).

F. PERIODOGRAM AND CORRELOGRAM RELATIONS FOR FILTERS

As explained in Chapter 1, Section B, a fundamental motive for spectral analysis is a preoccupation with data filtering. It is therefore important to determine the effect of filtering on the correlogram and periodogram. Let us consider the filtered data

$$y(t) = h(t) \otimes x(t). \tag{34}$$

It can be shown that the finite-average autocorrelations of x and y are related by the approximation

$$R_y(t, \tau)_T \cong R_x(t, \tau)_T \otimes r_h(\tau), \qquad |\tau| \leqslant T - 2\Delta\tau^*, \qquad T \gg \Delta\tau^*, \tag{35}$$

$$r_h(\tau) = h(\tau) \otimes h(-\tau), \tag{36}$$

for which $\Delta\tau^*$ is the *memory length* of the filter, namely, the width of $h(\tau)$. (The condition $|\tau| \leqslant T - 2\Delta\tau^*$ avoids edge effects.) Approximation (35) can be made as accurate as desired for any given t and τ and any bounded h with finite support by choosing T sufficiently large, provided only that $\hat{R}_x(\tau)$ exists (exercise 12). Similarly, it can be shown that the correlograms of x and y are related by the approximation

$$R_{y_T}(t, \tau) \cong R_{x_T}(t, \tau) \otimes r_h(\tau), \qquad T - |\tau| \gg \Delta\tau^*. \tag{37}$$

Both approximations (35) and (37) become exact equalities in the limit $T \to \infty$:

$$\hat{R}_y(\tau) = \hat{R}_x(\tau) \otimes r_h(\tau). \tag{38}$$

Approximations (35) and (37) can be used to derive the following approximations between pseudospectra and between periodograms (in which the symbol T has been replaced with the symbol Δt):

$$S_y(t, f)_{\Delta t} \otimes w_{\Delta f}(f) \cong [|H(f)|^2 S_x(t, f)_{\Delta t}] \otimes w_{\Delta f}(f), \qquad \Delta t \Delta f > 1, \Delta t \gg \Delta\tau^*, \tag{39}$$

$$S_{y_{\Delta t}}(t, f) \otimes w_{\Delta f}(f) \cong [|H(f)|^2 S_{x_{\Delta t}}(t, f)] \otimes w_{\Delta f}(f), \qquad \Delta t \Delta f > 1, \Delta t \gg \Delta\tau^*. \tag{40}$$

These approximations become exact in the limit $\Delta t \to \infty$ for any $\Delta f > 0$ (exercise 13). In fact, as shown in Chapter 3, Section C, by also letting $\Delta f \to 0$, the following relation is obtained:

$$\hat{S}_y(f) = \hat{S}_x(f)\, |H(f)|^2, \tag{41}$$

where $\hat{S}_x(f)$ is the limit spectrum introduced in Chapter 1,

$$\hat{S}_x(f) \triangleq \lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{x_{\Delta t}}(t, f) \otimes w_{\Delta f}(f). \tag{42}$$

Moreover, as also shown in Chapter 3, Section C, (42) yields the identity

$$\hat{S}_x(\cdot) = F\{\hat{R}_x(\cdot)\}, \tag{43}$$

which is called the *limit spectrum*. (Thus, (41) can be obtained directly from (38) using the convolution theorem and identity (43).) It is pointed out that if $\Delta f < 1/\Delta\tau^*$, then the operations of smoothing and multiplication by $|H(f)|^2$ in (39) and (40) can be interchanged. Also, if there is no anomalous behavior in x at the edges of the data segment (i.e., within $\Delta\tau^*$ of the two time points $t \pm \Delta t/2$), then the condition $\Delta t >> \Delta\tau^*$ can result in (39) and (40) being accurate approximations even with the frequency-smoothing operation (convolution with $w_{\Delta f}$) deleted. Since the condition $\Delta t >> \Delta\tau^*$ guarantees that the time-variant spectrum and pseudospectrum are nearly time-invariant throughout the entire response time of the filter, then the relations (39) and (40) are said to be *quasi-static* approximations.

Before proceeding it should be clarified why the condition $\Delta t >> \Delta\tau^*$ is needed in order for (40) to be a close approximation. To illustrate the necessity of this condition, consider the situation where $\Delta t = \Delta\tau^*$. In this case, the response time, or memory length, of the filter equals the length Δt of the input data segment being considered. Therefore, this input data segment will produce an output data segment that is $\Delta t + \Delta\tau^* = 2\Delta t$ in length. Consequently, the output segment of length Δt contains only half of the effects due to the input segment. Therefore, there cannot be a one-to-one relationship between input and output data segments of length $\Delta t = \Delta\tau^*$. As an extreme example, if the impulse response were $h(\tau) = \delta(\tau - \Delta\tau^*)$ and if $x(t)$ were nonzero only within the interval $t \in [-\Delta t/2, \Delta t/2]$, then $y(t) = 0$ for $t \in [-\Delta t/2, \Delta t/2]$ for $\Delta t \leqslant \Delta\tau^*$. Thus, $S_{y_{\Delta t}}(0, f) = 0$, regardless of $S_{x_{\Delta t}}(0, f)$. Hence, the approximation $S_{y_{\Delta t}}(t, f) \cong |H(f)|^2 S_{x_{\Delta t}}(t, f)$, in which $|H(f)| = 1$ in this example, would be extremely poor for $t = 0$ and, in fact, very poor for all $t \in [-\Delta t/2, \Delta t/2]$.

Let us consider a practical application of (40). If $h(\tau)$ is the impulse response of the multiply-resonant system discussed in the motivating example at the beginning of this chapter, then the width of $h(\tau)$ will be equal to several (say three to five) time constants of the most narrow-band resonance. If Δt does not greatly exceed this system memory length, then we cannot satisfy the conditions in (40). On the other hand, if Δt is sufficiently large to satisfy these conditions and if Δf is smaller than the width of the narrowest resonant peak in the transfer function magnitude $|H(f)|$, then by interchanging the order of the operations of smoothing with $w_{\Delta f}(f)$ and multiplication by $|H(f)|^2$, we obtain the close approximation

$$|H(f)|^2 \cong \frac{S_{y_{\Delta t}}(t, f) \otimes w_{\Delta f}(f)}{S_{x_{\Delta t}}(t, f) \otimes w_{\Delta f}(f)}.$$

This can be used as a basis for determining the resonant frequencies and bandwidths of the system. This idea is discussed further in the next chapter.

Relations (38) and (41) are called the *limit-autocorrelation* and *limit-spectrum relations for filters*, respectively. These relations reveal that the limit autocorrelation and limit spectrum are each *self-determinate characteristics under an LTI transformation;* that is, the only characteristic of x that determines $\hat{R}_y$ (or $\hat{S}_y$) is $\hat{R}_x$ (or $\hat{S}_x$).

G. LOCAL AVERAGE POWER SPECTRAL DENSITY

Relation (40) can be used to derive an interpretation of the time-variant periodogram $S_{x_T}(t, f)$ as the *time-variant* (or *local*) *spectral density of average power in* $x(t)$. Specifically, let $H = F\{h\}$ be the transfer function of an *ideal band-pass filter* with *center frequency* f (and image $-f$) and *bandwidth* Δf^*,

$$H(\nu) = \begin{cases} 1, & ||\nu| - f| \leq \Delta f^*/2 \\ 0, & ||\nu| - f| > \Delta f^*/2. \end{cases} \tag{44}$$

Then, in view of the discussion in Chapter 1, Section B2, we see that the filter output y represents the spectral content of the filter input x only in the band of width Δf^* centered at f (and the image band centered at $-f$). The *instantaneous power* $y^2(t)$ in this spectral band, averaged over the time-interval $[t - \Delta t^*/2, t + \Delta t^*/2]$, is therefore

$$P_{\Delta t^* \Delta f^*}(t, f) \triangleq \frac{1}{\Delta t^*} \int_{t-\Delta t^*/2}^{t+\Delta t^*/2} y^2(u)\, du \tag{45}$$

$$= R_{y_{\Delta t^*}}(t, 0) \tag{46}$$

$$= \int_{-\infty}^{\infty} S_{y_{\Delta t^*}}(t, \nu)\, d\nu. \tag{47}$$

Equation (46) is simply definition (19), and (47) is relation (18) in inverse form,

$$R_{y_T}(t, \cdot) = F^{-1}\{S_{y_T}(t, \cdot)\}, \tag{48}$$

with $T = \Delta t^*$ and evaluated at $\tau = 0$. Now, (47) can be identified with the left member of (40), with $\Delta f \to \infty$ and $\Delta t = \Delta t^*$. Therefore, substitution of (44) into the right member of (40) yields (using $S_{x_{\Delta t^*}}(t, -\nu) = S_{x_{\Delta t^*}}(t, \nu)$)

$$P_{\Delta t^* \Delta f^*}(t, f) \cong 2 \int_{f-\Delta f^*/2}^{f+\Delta f^*/2} S_{x_{\Delta t^*}}(t, \nu)\, d\nu, \qquad \Delta t^* \Delta f^* \gg 1, \tag{49}$$

in which the approximation

$$\Delta\tau^* \cong \frac{1}{\Delta f^*} \tag{50}$$

has been used. (The factor 2 in (49) is due to the image band $[-f - \Delta f^*/2, -f + \Delta f^*/2]$.) It follows from (49) that the power of $x(t)$ in the spectral band $[f - \Delta f^*/2, f + \Delta f^*/2]$, averaged over the time interval $[t - \Delta t^*/2, t + \Delta t^*/2]$, is obtained, to a close approximation, by integrating the positive function $S_{x_{\Delta t^*}}(t, \cdot)$ over this spectral band. Since (49) is a close approximation[8]

[8] Since the ideal h corresponding to (44) does not have finite support, then the straightforward method of proof analogous to that developed in exercises 12 and 13, that (40) and therefore (49) are close approximations, does not apply. However, it does apply for bandpass filters that have finite length impulse-response functions.

to (45) if

$$\Delta t^* \Delta f^* >> 1, \tag{51}$$

then the periodogram $S_{x_T}(t, f)$ (with $T = \Delta t^*$) can appropriately be called the *time-variant average power spectral density* of $x(t)$ if condition (51) holds. This condition is called the *time-frequency uncertainty condition for local-average power spectral density*. Furthermore, results in Chapter 5, Section C, reveal that condition (51) is necessary (for a broad class of data) for (49) to yield a reliable measurement of power spectral density for a constant phenomenon, in which case $P_{\Delta t^* \Delta f^*}(t, f)$ is essentially independent of t.

In summary, if it is desired to measure accurately the average power in $x(t)$ due only to spectral components in a band of width Δf^*, then the time-average must be performed over an interval Δt^*, greatly exceeding the reciprocal bandwidth $1/\Delta f^*$. And if the periodogram is to be used for this measurement as in (49), then Δt^* is simply the length of the data-segment that must be Fourier transformed. From another viewpoint, we see that for a given data-segment length Δt^*, the bandwidth Δf^* over which the periodogram must be integrated as in (49) must greatly exceed the periodogram's spectral resolution width $1/\Delta t^*$. Nevertheless, the raw (unsmoothed) periodogram can be a useful (if crude) approximate measure of time-variant average-power spectral density, $P_{\Delta t^* \Delta f^*}(t, f)$ with $\Delta f^* = 1/\Delta t^*$, in some applications involving time-variant phenomena, such as speech (see Chapter 8).

H. TIME SAMPLING AND ALIASING

Since spectral analysis is often accomplished with digital implementations, which operate in discrete time, it is important to understand the effects of time sampling on the time-variant finite-time complex spectrum,

$$X_T(t, f) \triangleq \int_{t-T/2}^{t+T/2} x(u) e^{-i2\pi f u}\, du. \tag{52}$$

If $x(u)$ is sampled every T_s units of time, then the corresponding discrete-time counterpart of (52) is

$$\tilde{X}_T(t, f) \triangleq \sum_{n=(t-T/2)/T_s}^{(t+T/2)/T_s} x(nT_s) e^{-i2\pi f n T_s}, \tag{53}$$

for which it is assumed (for simplicity of expression) that t/T_s and $T/2T_s$ are integers. $\tilde{X}_T(t, f)$ is called the *time-variant finite-time Fourier-series transform*, or the *time-variant finite-time complex spectrum for time-sampled data*. The total number of time-samples transformed in (53) is $N = T/T_s + 1$, which is an odd number. Equation (53) can be reexpressed as the integral

$$\tilde{X}_T(t, f) = \int_{t-T/2}^{t+T/2} \sum_{n=-\infty}^{\infty} \delta(u - nT_s) x(u) e^{-i2\pi f u}\, du, \tag{54}$$

and this integral can be reduced as follows:

$$\tilde{X}_T(t,f) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(u - nT_s) T u_T(t-u) x(u) e^{-i2\pi f u}\, du \tag{55}$$

$$= \int_{-\infty}^{\infty} \frac{1}{T_s} \sum_{m=-\infty}^{\infty} \delta\left(\nu - \frac{m}{T_s}\right) \int_{-\infty}^{\infty} T u_T(t-u) x(u) e^{-i2\pi(f-\nu)u}\, du\, d\nu \tag{56}$$

$$= \frac{1}{T_s} \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \delta\left(\nu - \frac{m}{T_s}\right) \int_{t-T/2}^{t+T/2} x(u) e^{-i2\pi(f-\nu)u}\, du\, d\nu \tag{57}$$

$$= \frac{1}{T_s} \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \delta\left(\nu - \frac{m}{T_s}\right) X_T(t, f-\nu)\, d\nu \tag{58}$$

$$= \frac{1}{T_s} \sum_{m=-\infty}^{\infty} X_T\left(t, f - \frac{m}{T_s}\right). \tag{59}$$

Equation (56) is the convolution theorem together with the Fourier transform pair (exercise 16)

$$F\left\{\sum_{n=-\infty}^{\infty} \delta(t - nT_s)\right\} = \frac{1}{T_s} \sum_{m=-\infty}^{\infty} \delta\left(f - \frac{m}{T_s}\right). \tag{60}$$

Summarizing (54)–(59), we have

$$\tilde{X}_T(t,f) = \frac{1}{T_s} \sum_{m=-\infty}^{\infty} X_T\left(t, f - \frac{m}{T_s}\right). \tag{61}$$

It follows from (61) that the complex spectrum of the time-sampled data is a sum of translates of the complex spectrum of the continuous time data, as illustrated in Figure 2-5. Thus, at a particular frequency f, $\tilde{X}_T(t,f)$ is the superposition of all the values $\{X_T(t, f - m/T_s) : m = 0, \pm 1, \pm 2, \pm 3, \ldots\}$. This is called the *aliasing phenomenon*. If $X_T(t, f)$ is not negligible for $|f| > 1/2T_s$, then

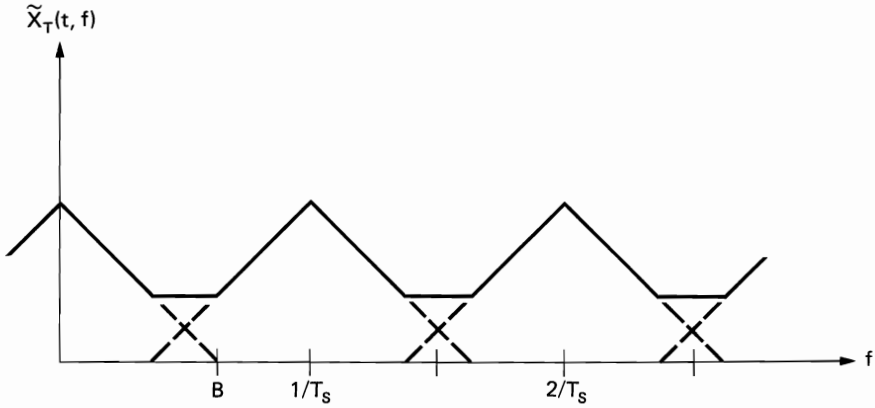


Figure 2-5 Illustration of aliasing phenomenon for a triangular-shaped complex spectrum $X_T(t, f)$ with bandwidth $B > 1/2T_s$ (t fixed).

$\tilde{X}_T(t, f)$ can differ substantially from $X_T(t, f)$, and $X_T(t, f)$ cannot be recovered from $\tilde{X}_T(t, f)$. Equivalently, in order to recover $X_T(t, f)$ from $\tilde{X}_T(t, f)$, when $X_T(t, f)$ is negligible for $|f| > B$, the sampling rate must exceed $2B$,

$$\frac{1}{T_s} > 2B, \tag{62}$$

which is referred to as the *Nyquist rate* (in honor of Harry Nyquist's pioneering work [Nyquist 1928]).

Although it is not apparent from the aliasing formula (61), it can be shown that the *discrete-time periodogram*

$$\tilde{S}_{x_T}(t, f) \triangleq \frac{1}{N} |\tilde{X}_T(t, f)|^2 \tag{63}$$

satisfies the *periodogram-correlogram relation*

$$\tilde{S}_{x_T}(t, f) = \text{FST}\{\tilde{R}_{x_T}(t, \tau)\}, \tag{64}$$

where FST denotes the *Fourier-series transform*

$$\text{FST}\{\tilde{R}_{x_T}(t, \tau)\} \triangleq \sum_{q=-\infty}^{\infty} \tilde{R}_{x_T}(t, qT_s) e^{-i2\pi f q T_s} \tag{65}$$

$$= \sum_{q=-T/T_s}^{T/T_s} \tilde{R}_{x_T}(t, qT_s) e^{-i2\pi f q T_s}, \tag{66}$$

and $\tilde{R}_{x_T}(t, \tau)$ is the *discrete-time correlogram* defined by

$$\tilde{R}_{x_T}(t, \tau) \triangleq \frac{1}{N} \sum_{n=(t-T/2)/T_s}^{(t+T/2-|\tau|)/T_s} x(nT_s + |\tau|) x(nT_s) [2T u_{2T}(\tau)],$$

$$\tau = 0, \pm T_s, \pm 2T_s, \ldots, \pm T, \tag{67}$$

which is precisely the discrete-time counterpart of the continuous-time correlogram (19) (with the change of variables $v = u - |\tau|/2$). The relation (64) can be shown to be a direct consequence of the convolution theorem for the FST (exercise 19).

I. SUMMARY

In Section A, the *time-variant periodogram*, which is the squared magnitude of the time-variant finite-time complex spectrum normalized by the data-segment length T, is introduced as an appropriate measure of local spectral content of a waveform; it is established that the *temporal resolution width* of the time-variant periodogram is T, and the *spectral resolution width* is on the order of $1/T$. In Section B, the technique of *data tapering* is introduced as a means for controlling the shape of the *spectral smoothing window* in the periodogram, and several basic tapering apertures or windows are introduced. Then Section C explains that regardless of the particular tapering aperture used, the product of temporal and spectral resolution widths is always on the order of unity, because the

corresponding temporal and spectral windows are a Fourier transform pair. In Section D, the *time-variant correlogram* is introduced as a measure of local autocorrelation of a waveform, and it is established that the time-variant periodogram is the Fourier transform of the time-variant correlogram. Then in Section E, an alternative measure of local autocorrelation termed the *finite-average autocorrelation* is introduced, and its Fourier transform, the *pseudospectrum*, is claimed to be a useful alternative to the periodogram when it is appropriately averaged to obtain a statistical spectrum. Several exact and approximate relationships among time-averaged correlograms and time-averaged finite-average autocorrelations are established for their use in the next chapter, where time-averaged measures of spectral content are studied. It is also explained that in the limit as the parameter T approaches infinity both the correlogram and finite-average autocorrelation approach the ideal *limit autocorrelation*. In Section F, an approximate convolution relation between the correlograms (and finite-average autocorrelations) at the input and output of a filter is derived and then used to derive an approximate product relation between the corresponding periodograms (and pseudospectra). It is explained that these approximate relations become exact in the limit as the parameters T in (35) and (37) and Δt in (39) and (40) approach infinity. These are referred to as the *limit-autocorrelation relation* and *limit-spectrum relation for filters*, (38) and (41). In Section G, the approximate periodogram relation for filters is used to establish that the time-variant periodogram can be interpreted as a measure of *local-average power spectral density* only if the temporal and spectral resolutions are limited in order to satisfy the time-frequency uncertainty condition (51). Finally in Section H, the discrete-time counterpart of the continuous-time complex spectrum is introduced, and the spectral aliasing phenomenon associated with time-sampling is described. Then the discrete-time counterparts of the time-variant periodogram and time-variant correlogram are introduced, and it is established that these are a Fourier-series transform pair.

In Appendix 2-1, the concept of instantaneous frequency for a sine wave with a time-variant argument is introduced and used to illustrate the resolution limitations of the time-variant periodogram.

For the sake of emphasis, two basic and fundamental results on the relationships between the overall widths and the resolution widths of Fourier transform pairs that are developed in this chapter and the exercises are repeated here at the conclusion of this summary. If a time-function has overall width (duration) on the order of T, then the spectral resolution width Δf^* of its transform must be on the order of $1/T$. Furthermore, if the time-function is pulselike, then its temporal resolution width Δt^* is on the order of its overall width T. Similarly, if a frequency function has overall width (bandwidth) on the order of B, then the temporal resolution width Δt^* of its inverse transform must be on the order of $1/B$, and if the frequency function is pulselike (low-pass or band-pass) then its spectral resolution width Δf^* is on the order of its overall width B. These simple order-of-magnitude rules are a key to understanding the principles of spectral analysis.

EXERCISES

1. **(a)** Derive the time-convolution formula (3) from the definition of the time-variant spectrum (1) and the rectangle window (5).

   (b) Derive the frequency-convolution formula (4) from the definitions (1) and (6). *Hint*: Use the change of variable $u = t - v$ in (2a), and use the identity $[x(t - v)Tu_T(v)]Tu_T(v) = x(t - v)$ for $|v| \leq T/2$ to show that

$$|X_T(t, f)| = \left| \int_{-\infty}^{\infty} [x(t - v)Tu_T(v)][Tu_T(v)e^{i2\pi fv}]\, dv \right|.$$

   Then use Parseval's relation (exercise 9, Chapter 1).

2. As a simple illustration of the effect on resolution of convolution with a pulselike function, draw graphs of the convolution $y(t) = x(t) \otimes a(t)$ for

$$x(t) = \delta(t + \Delta/2) - \delta(t - \Delta/2)$$
$$a(t) = u_T(t)$$

   for $T = 10\Delta$, $T = \Delta$, and $T = \Delta/10$. The case $T = 10\Delta$, in which $x(t)$ is a narrow doublet, is an exception to the rule that the resolution width of $y(t)$ will typically be on the order of T.

3. To illustrate the spectral resolution capability of the periodogram, consider the sum of two sine waves

$$x(t) = \cos(2\pi f_1 t) + a\cos(2\pi f_2 t - \phi),$$

   and show that

$$\begin{aligned} S_{x_T}(t, f) = T \Bigg| & \frac{1}{2}\operatorname{sinc}[(f - f_1)T]e^{-i2\pi(f-f_1)t} \\ & + \frac{1}{2}\operatorname{sinc}[(f + f_1)T]e^{-i2\pi(f+f_1)t} \\ & + \frac{a}{2}\operatorname{sinc}[(f - f_2)T]e^{-i[2\pi(f-f_2)t+\phi]} \\ & + \frac{a}{2}\operatorname{sinc}[(f + f_2)T]e^{-i[2\pi(f+f_2)t-\phi]} \Bigg|^2 \end{aligned} \tag{68}$$

$$\begin{aligned} \cong & \frac{T}{4}\operatorname{sinc}^2[(f - f_1)T] + \frac{T}{4}\operatorname{sinc}^2[(f + f_1)T] \\ & + \frac{a^2T}{4}\operatorname{sinc}^2[(f - f_2)T] + \frac{a^2T}{4}\operatorname{sinc}^2[(f + f_2)T] \\ & + \frac{aT}{4}\operatorname{sinc}[(f - f_1)T]\operatorname{sinc}[(f - f_2)T]\cos[2\pi(f_1 - f_2)t + \phi] \\ & + \frac{aT}{4}\operatorname{sinc}[(f + f_1)T]\operatorname{sinc}[(f + f_2)T]\cos[2\pi(f_1 - f_2)t + \phi] \end{aligned} \tag{69}$$

$$\begin{aligned} \cong & \frac{T}{4}\operatorname{sinc}^2[(f - f_1)T] + \frac{T}{4}\operatorname{sinc}^2[(f + f_1)T] \\ & + \frac{a^2T}{4}\operatorname{sinc}^2[(f - f_2)T] + \frac{a^2T}{4}\operatorname{sinc}^2[(f + f_2)T]. \end{aligned} \tag{70}$$

   Hint: Verify that the several cross-product terms in (68) are negligible for $2f_1T \gg 1$

and $2f_2T >> 1$. Then verify that the two cross-products in (69) are negligible for $|f_1 - f_2|T >> 1$ (for $f \neq 0$). Graphs of (68) for several values of $|f_1 - f_2|T$ shown in Figure 2-6 reveal that the two spectral components in $x(t)$ are not resolved by $S_{x_T}(t, f)$ unless $|f_1 - f_2| > 1/T$.

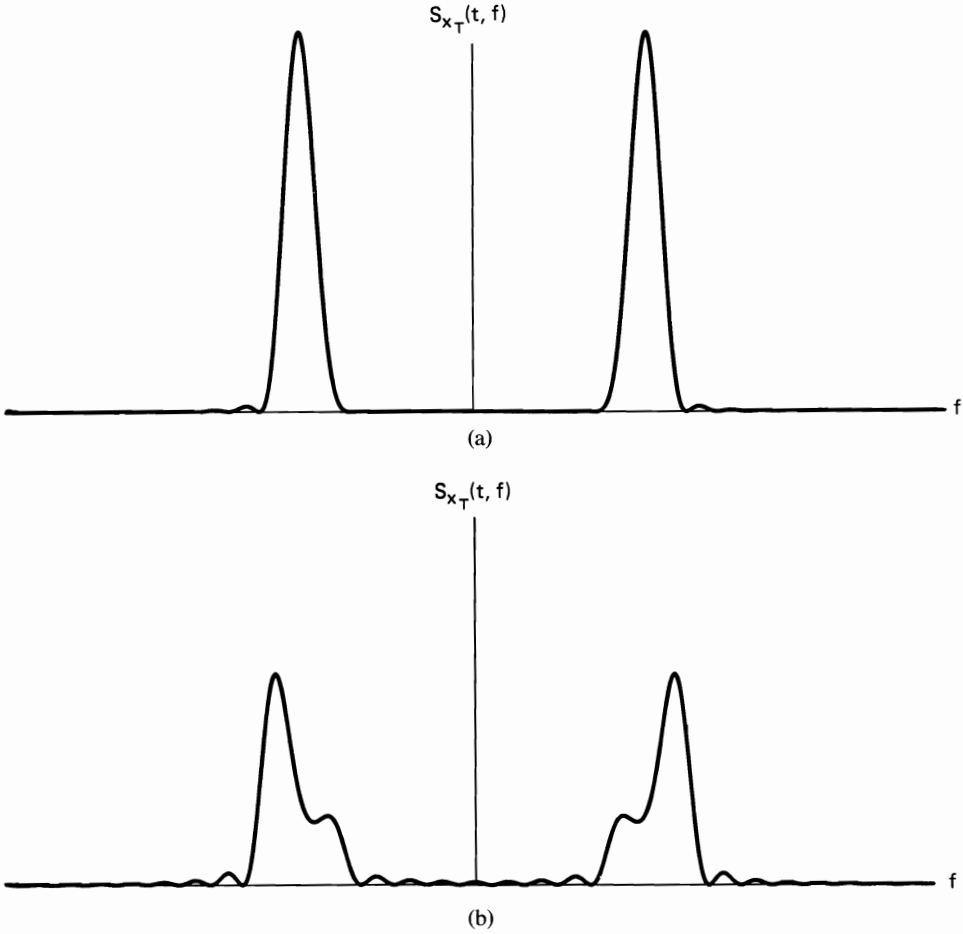


Figure 2-6 Spectra of two sine waves: (a) $f_2 - f_1 = 1/T$, $a = \sqrt{2}$. (b) $f_2 - f_1 = 3/2T$, $a = \sqrt{2}$.

4. **(a)** Prove that a spectral window A has unity height, $A(0) = 1$, if and only if the corresponding temporal window a has unity area

$$\int_{-\infty}^{\infty} a(t)\, dt = 1$$

and vice versa. Verify that the transform pairs in Table 1-1 satisfy this property.

(b) Prove that areas multiply under convolution; that is, if $a(t)$ has area α and $b(t)$ has area β, then $a(t) \otimes b(t)$ has area $\alpha\beta$. Verify that this property is satisfied by $v_T(t) = u_T(t) \otimes u_T(t)$.

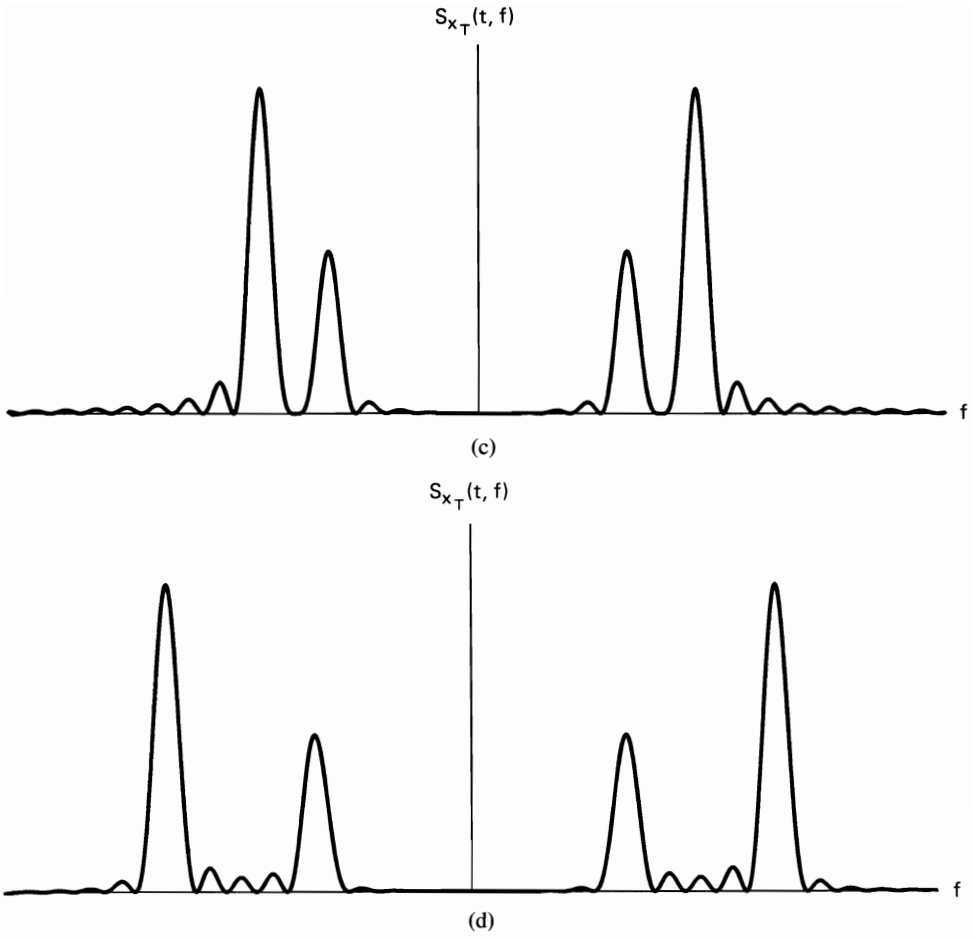


Figure 2-6 (continued) (c) $f_2 - f_1 = 2/T$, $a = \sqrt{2}$. (d) $f_2 - f_1 = 10/T$, $a = \sqrt{2}$.

(c) Evaluate the aperture parameter γ in (33) for unity-height rectangle and sinc windows. *Hint*: Use Parseval's relation for the sinc window.

(d) Use the convolution theorem to prove that

$$w_T(t) \otimes w_T(t) = w_T(t). \tag{71}$$

5. (a) Use definitions (1), (11a), and (20) to verify the *time-variant periodogram-correlogram relation for tapered data* (18). *Hint*: Express (20) as a convolution in terms of the time-series $y(w) \triangleq a_T(w)x(t - w)$ and $z(w) \triangleq a_T(-w)x(t + w)$, by using a change of variables, and then apply the convolution theorem.

(b) Show that (20) reduces to (19) for a rectangle aperture, $a_T = Tu_T$.

6. (a) Draw a graph of left and right members of the window identity (24) as a surface above the (t, τ) plane to verify this identity. *Hint*: The height of the window $u_{T-|\tau|}$ is proportional to $1/(T-|\tau|)$.

(b) Use identity (24) to verify identity (23).

(c) Verify identity (26).

7. **(a)** Prove that for $t = 0$ the two limits in (31) both equal the limit autocorrelation as defined by (6) in Chapter 1. *Hint*: Use the fact that

$$\lim_{T\to\infty} A(T)B(T) = \lim_{T\to\infty} A(T) \lim_{T\to\infty} B(T) \tag{72}$$

if the two limits on the right exist to show that

$$\lim_{T\to\infty} R_x(0, \tau)_T = \hat{R}_x(\tau)$$

by letting $B(T) = 2Tu_{2T}(\tau)$. Then show that

$$R_{x_T}(0, \tau) = R_x(0, \tau)_{T-|\tau|}\left(1 - \frac{|\tau|}{T}\right) \tag{73}$$

and again use the above fact for limits to conclude that the right side converges to $\hat{R}_x(\tau)$ (e.g., let $T' = (T - |\tau|)$. (The proof for $t \neq 0$ is not as straightforward; see [Kampé de Fériet 1954].)

8. To verify approximation (27), show that

$$R_x(t, \tau)_{\Delta t} \otimes u_T(t) \cong R_x(t, \tau)_{\Delta t}, \qquad |\tau| \leq T << \Delta t, \tag{74}$$

and use this in (26). *Hint*: To verify that (74) is a close approximation for sufficiently large Δt, proceed as follows. By virtue of (31), we know that for each t and T, all v for which $|v| \leq T/2$, and any $\epsilon > 0$, there exists a Δt such that

$$|R_x(t + v, \tau)_{\Delta t} - \hat{R}_x(\tau)| < \epsilon, \qquad |\tau| \leq T.$$

Use this to show that

$$|R_x(t, \tau)_{\Delta t} \otimes u_T(t) - \hat{R}_x(\tau)| < \epsilon \tag{75}$$

and

$$|R_x(t, \tau)_{\Delta t} - \hat{R}_x(\tau)| < \epsilon. \tag{76}$$

Then use (75)–(76) to show that

$$|R_x(t, \tau)_{\Delta t} \otimes u_T(t) - R_x(t, \tau)_{\Delta t}| < 2\epsilon, \qquad |\tau| \leq T. \tag{77}$$

Thus, for each t and T and any arbitrary small $\epsilon > 0$, there exists a Δt sufficiently large that (77) holds.

9. To verify approximation (25), show that

$$R_x(t, \tau)_{\Delta t} \otimes u_{T-|\tau|}(t) \cong R_x(t, \tau)_{\Delta t}, \qquad |\tau| \leq T << \Delta t, \tag{78}$$

and use this in (23). *Hint*: To verify that (78) is a close approximation for sufficiently large Δt, proceed by analogy with exercise 8.

10. To verify approximation (28), use definition (20) to show that for tapered data in $R_{x_T}(t, \tau)$,

$$[R_{x_T}(t, \tau) \otimes u_{\Delta t}(t)]2\Delta t u_{2\Delta t}(\tau) = \frac{1}{T}\int_{-\infty}^{\infty} a_T\left(v + \frac{\tau}{2}\right)a_T\left(v - \frac{\tau}{2}\right)R_x(t - v, \tau)_{\Delta t}\, dv, \tag{79}$$

where $R_x(t - v, \tau)_{\Delta t}$ by definition includes no data tapering, and then use the approximation (see exercise 8)

$$R_x(t + v, \tau)_{\Delta t} \cong R_x(t, \tau)_{\Delta t}, \qquad |v| << \Delta t. \tag{80}$$

Assume there exists a T^* such that $a_T(t) = 0$ for $|t| > T^*/2$, where $T^* \not> T$.

11. To verify (32)–(33) for the limit autocorrelation, use (31) and the result of exercise 10 to show that

$$\lim_{T\to\infty} R_{x_T}(t, \tau) = \lim_{T\to\infty} \lim_{\Delta t\to\infty} R_x(t, \tau)_{\Delta t} \frac{1}{T} r_a(\tau) = \gamma \lim_{\Delta t\to\infty} R_x(t, \tau)_{\Delta t} = \gamma \hat{R}_x(\tau).$$

12. **(a)** Verify the limit-autocorrelation relation for filters (38) directly by substitution of (5) in Chapter 1 into (6) in Chapter 1 (with x replaced by y). *Hint*: Use (5b) in Chapter 1 and interchange the order of the two integrations from (5b) in Chapter 1 with the limit time-average operation from (6) in Chapter 1.

(b) Verify the finite-average autocorrelation relation (35). *Hint*: Proceed as follows. In order to relate the finite-average autocorrelation for $y(t)$ to the finite-average autocorrelation for $x(t)$, substitute (34) into definition (21) (with x replaced by y) to obtain

$$\begin{aligned} R_y(t, \tau)_T &= \frac{1}{T}\int_{t-T/2}^{t+T/2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(v_1)h(v_2)x\left(u + \frac{\tau}{2} - v_1\right)x\left(u - \frac{\tau}{2} - v_2\right) dv_1\, dv_2\, du[2Tu_{2T}(\tau)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(v_1)h(v_2)\frac{1}{T}\int_{t-T/2}^{t+T/2} x\left(u + \frac{\tau}{2} - v_1\right)x\left(u - \frac{\tau}{2} - v_2\right) du\,[2Tu_{2T}(\tau)]\, dv_1\, dv_2. \end{aligned} \tag{81}$$

Assume that the amount of integration, T, used to obtain the correlation $R_y(t, \tau)_T$ is substantially larger than the filter memory length $\Delta\tau^*$:

$$T >> \Delta\tau^*. \tag{82}$$

Since the factor $h(v_1)h(v_2)$ in the integrand in (81) is negligibly small for

$$|v_1 - v_2| > 2\Delta\tau^*, \tag{83}$$

then the window $u_{2T}(\tau)$ in (81) can be replaced by the slightly shifted window $u_{2T}(\tau - v_1 + v_2)$ to obtain the close approximation

$$\begin{aligned} R_y(t, \tau)_T \cong \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(v_1)h(v_2)\left\{\frac{1}{T}\int_{t-T/2}^{t+T/2} x\left(u + \frac{\tau}{2} - \frac{v_1 + v_2}{2} - \frac{v_1 - v_2}{2}\right)\right. \\ \left. \times\, x\left(u - \frac{\tau}{2} - \frac{v_1 + v_2}{2} + \frac{v_1 - v_2}{2}\right) du\,[2Tu_{2T}(\tau - v_1 + v_2)]\right\} dv_1\, dv_2, \\ |\tau| \leqslant T - 2\Delta\tau^*. \end{aligned} \tag{84}$$

Show that this approximation is exact if $h(v) = 0$ for $|v| > \Delta\tau^*$. The quantity in braces in (84) is a time-variant finite-average autocorrelation

$$\begin{aligned} \{\cdot\} &= \frac{1}{T}\int_{t_0-T/2}^{t_0+T/2} x\left(v + \frac{\tau_0}{2}\right)x\left(v - \frac{\tau_0}{2}\right) dv\,[2Tu_{2T}(\tau_0)] = R_x(t_0, \tau_0)_T, \\ t_0 &\triangleq t - \frac{v_1 + v_2}{2} \\ \tau_0 &\triangleq \tau - (v_1 - v_2). \end{aligned} \tag{85}$$

Therefore, (84) can be reexpressed as

$$\begin{aligned} R_y(t, \tau)_T \cong \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(v_1)h(v_2)R_x\left(t - \frac{v_1 + v_2}{2}, \tau - v_1 + v_2\right)_T dv_1\, dv_2, \\ |\tau| \leqslant T - 2\Delta\tau^*. \end{aligned} \tag{86}$$

Since the temporal resolution of $R_x(t, \tau)_T$ is T, then for

$$\frac{|v_1 + v_2|}{2} \leqslant \Delta\tau^* \tag{87}$$

condition (82) suggests that

$$R_x\left(t - \frac{v_1 + v_2}{2}, \tau - v_1 + v_2\right)_T \cong R_x(t, \tau - v_1 + v_2)_T \tag{88}$$

is a close approximation. Show that the accuracy of approximation (88) for fixed τ, subject to (87) with fixed $\Delta\tau^*$, can be made as high as desired by choosing T sufficiently large (see exercise 8). Since $h(v_1)h(v_2)$ is zero for

$$\frac{|v_1 + v_2|}{2} > \Delta\tau^* \tag{89}$$

(assuming $h(v) = 0$ for $|v| > \Delta\tau^*$), then (88) yields the close approximation

$$h(v_1)h(v_2)R_x\left(t - \frac{v_1 + v_2}{2}, \tau - v_1 + v_2\right)_T \cong h(v_1)h(v_2)R_x(t, \tau - v_1 + v_2)_T \tag{90}$$

for all values of v_1 and v_2 for which these quantities are nonzero. Substitution of (90) into (86) yields the close approximation (for sufficiently large T)

$$R_y(t, \tau)_T \cong \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(v_1)h(v_2)R_x(t, \tau - v_1 + v_2)_T \, dv_1 \, dv_2,$$
$$|\tau| \leq T - 2\Delta\tau^*, \; T \gg \Delta\tau^*. \tag{91}$$

Now, show that (35) is equivalent to (91).

13. Verify the pseudospectrum relation (39). *Hint*: Proceed as follows. The condition

$$|\tau| \leq \Delta t - 2\Delta\tau^* \tag{92}$$

can be satisfied by the two conditions

$$|\tau| \leq (1 - \epsilon)\Delta t \tag{93}$$
$$\Delta t \geq \frac{2\Delta\tau^*}{\epsilon}$$

for arbitrarily small positive ϵ. Use this equivalent condition to reexpress (35) (with T replaced by Δt) as

$$R_y(t, \tau)_{\Delta t}u_T(\tau) \cong [R_x(t, \tau)_{\Delta t} \otimes r_h(\tau)]u_T(\tau), \qquad \begin{cases} \Delta t \geq T/(1 - \epsilon) \\ \Delta t \gg \Delta\tau^* \\ \Delta t \geq 2\Delta\tau^*/\epsilon \end{cases} \tag{94}$$

for sufficiently large Δt and $0 < \epsilon < 1$. Then verify that for $\epsilon \ll 1$ (94) yields

$$R_y(t, \tau)_{\Delta t}u_T(\tau) \cong [R_x(t, \tau)_{\Delta t} \otimes r_h(\tau)]u_T(\tau), \qquad \Delta t > T, \; \Delta t \gg \Delta\tau^*. \tag{95}$$

Finally, Fourier transform (95) to obtain (39) (using $\Delta f = 1/T$).

14. Let $h(\tau)$ have approximate (or exact) duration $\Delta\tau^*$, centered at $\tau = 0$. Then use the approximate (or exact) equation

$$H(f) = \int_{-\infty}^{\infty} h(\tau)e^{-i2\pi f\tau} \, d\tau$$
$$\cong \int_{-\Delta\tau^*/2}^{\Delta\tau^*/2} h(\tau)e^{-i2\pi f\tau} \, d\tau$$

to show that

$$H(f) \cong H(f) \otimes w_{1/\Delta\tau^*}(f) \tag{96}$$

approximately (or exactly). Thus, the transform of any centered function with approximate duration $\Delta\tau^*$ will have a resolution width, denoted by Δf^*, that is typically no smaller than $1/\Delta\tau^*$: $\Delta f^* \geq 1/\Delta\tau^*$. (If $h(\tau)$ is not centered, the same result can be obtained for $|H(f)|$.) In fact, it is shown in the next exercise that Δf^* and $1/\Delta\tau^*$ are on the same order of magnitude: $\Delta f^* \cong 1/\Delta\tau^*$. If $h(\tau)$ happens to be pulselike,

then $\Delta\tau^*$ is not only the duration (or overall width), it is also the resolution width of $h(\tau)$. Thus, for a pulselike function, the resolution widths of the function and its transform are approximately reciprocals. Show that these two general rules regarding the resolution width Δf^* do indeed hold for the following examples of $h(\tau)$:

(a) Rectangle pulse

(b) Triangle pulse

(c) Symmetrical exponential pulse, $e^{-|t|/\tau^*}$

(d) The product of (a), or (b), or (c) with a sum of sine waves with arbitrary frequencies.

To illustrate further the reciprocal relationship between the width of a time function and the resolution width of its Fourier transform, show that if

$$A(f) \triangleq F\{a(t)\}$$

then

$$F\{a(ct)\} = \frac{1}{|c|}A\left(\frac{f}{c}\right)$$

for any constant c. Thus, for example, if $c = \frac{1}{2}$ so that the width of a is doubled, then the resolution width of A is halved. This relationship is further elucidated in the next exercise.

15. **(a)** Consider a transient (not persistent) finite-energy waveform $x(t)$,

$$\int_{-\infty}^{\infty} x^2(t)\,dt < \infty. \tag{97}$$

Use the approach illustrated in section H to show that

$$\bar{X}(f) = \frac{1}{T_s}\sum_{m=-\infty}^{\infty} X\left(f - \frac{m}{T_s}\right), \tag{98}$$

where $X(f)$ is the Fourier transform

$$X(f) \triangleq \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\,dt, \tag{99}$$

and $\bar{X}(f)$ is the FST

$$\bar{X}(f) \triangleq \sum_{n=-\infty}^{\infty} x(nT_s)e^{-i2\pi nT_s f}. \tag{100}$$

This reveals that $X(f)$ (and therefore $x(t)$) can be recovered from $\bar{X}(f)$ (and therefore from $\{x(nT_s)\}$) provided that

$$X(f) = 0, \qquad |f| > \frac{1}{2T_s}. \tag{101}$$

(b) To obtain $x(t)$ directly from $\{x(nT_s)\}$ when (101) is satisfied, substitute (100) into

$$X(f) = \bar{X}(f)u_{1/T_s}(f) \tag{102}$$

(which follows from (98) and (101)) and then evaluate the inverse Fourier transform of (102). The result is

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\frac{\sin[\pi(t - nT_s)/T_s]}{\pi(t - nT_s)/T_s}. \tag{103}$$

This result is referred to as the *sampling theorem* for transient waveforms. A completely analogous theorem applies to persistent waveforms (see [Gardner 1985]). The sampling theorem proves that if a Fourier transform $X(f)$ has absolute width $\Delta f^* = 1/T_s$, then its inverse transform $x(t)$ can be perfectly reconstructed by interpolating its time samples separated by $\Delta t^* = T_s = 1/\Delta f^*$, with a sinc-interpolating pulse as indicated by (103). Thus, the resolution width of $x(t)$ is on the order of Δt^*. This is the time-frequency dual of the result illustrated in the previous exercise.

16. Verify the Fourier transform pair (60). *Hint*: Use (39) in Chapter 1 together with the formula for Fourier coefficients,

$$X_m \triangleq \int_{-1/2\Delta f}^{1/2\Delta f} x(t) e^{-i2\pi m \Delta f t}\, dt,$$

where $1/\Delta f = T_s$ is the period.

17. Assume that at a particular time instant $t = t_0$, the time-variant complex spectrum of $x(t)$ is given by a triangle with base extending from $f = -B$ to $f = +B$. Draw graphs of the complex spectrum $\tilde{X}_T(t_0, f)$ of the time-sampled data $x(nT_s)$ for sampling rates of $1/T_s = B$, $3B/2$, and $2B$.

18. Show that the Fourier series transform of the discrete-time rectangle window

$$a(nT_s) = \begin{cases} 1, & |n| \leq \dfrac{N-1}{2} \\ 0, & |n| > \dfrac{N-1}{2}, \quad N \text{ odd} \end{cases} \tag{104}$$

is

$$\tilde{A}(f) = \frac{\sin(\pi f N T_s)}{\sin(\pi f T_s)}. \tag{105}$$

Hint: Use the identities

$$\sum_{n=0}^{N-1} e^{-i\theta n} = \frac{1 - e^{-i\theta N}}{1 - e^{-i\theta}} \tag{106}$$

and

$$\frac{1}{2i}(e^{i\theta} - e^{-i\theta}) = \sin(\theta).$$

19. Prove the convolution theorem for the FST; that is, if (using $T_s = 1$)

$$z(t) = x(t) \otimes y(t) \triangleq \sum_{v=-\infty}^{\infty} x(t - v) y(v) \tag{107}$$

then

$$\tilde{Z}(f) \triangleq \sum_{t=-\infty}^{\infty} z(t) e^{-i2\pi f t} = \tilde{X}(f)\tilde{Y}(f). \tag{108}$$

Hint: Take the FST of both sides of (107) and then make the change of variables $u = t - v$.

20. Verify the periodogram-correlogram relation (64) for discrete time. *Hint*: Use the convolution theorem for the FST (see exercise 19).

21. Verify that the inverse FST is given by

$$x(t) = \int_{-1/2}^{1/2} \tilde{X}(f) e^{i2\pi f t}\, df, \tag{109}$$

where the FST is defined by (using $T_s = 1$)

$$\tilde{X}(f) \triangleq \sum_{t=-\infty}^{\infty} x(t)e^{-i2\pi ft}. \tag{110}$$

Hint: Substitute (110) into (109) and use the identity

$$\int_{-1/2}^{1/2} e^{i2\pi ft}\, df = \begin{cases} 1, & t = 0 \\ 0, & t = \text{integer} \neq 0. \end{cases} \tag{111}$$

22. **(a)** Consider the sum of two sine waves as in exercise 3, and assume that the two frequencies are close together: $|f_1 - f_2| \ll (f_1 + f_2)/2$. Show that this time-series can be reexpressed as a sine wave with frequency $f_0 = (f_1 + f_2)/2$ and with slowly varying amplitude and phase,

$$x(t) = a(t)\cos[2\pi f_0 t + \theta(t)].$$

Determine explicit formulas for $a(t)$ and $\theta(t)$. *Hint*: Express f_1 and f_2 in the form $f_0 \pm f_*$, and use the trigonometric identity

$$\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B).$$

(b) Show that when the sum of two sine waves, as in part (a), goes through a nonlinearity, say a square-law device, the resultant time-series, $y(t) = x^2(t)$, contains additive sine wave components at both the sum and difference frequencies, $f_1 - f_2$ and $f_1 + f_2$. These are called *beat frequencies*. *Hint*: Use the following trigonometric identity after expanding the square $x^2(t)$,

$$\cos(A)\cos(B) = \frac{1}{2}\cos(A + B) + \frac{1}{2}\cos(A - B).$$

23. Consider the frequency-modulated sine wave

$$x(t) = \sin[2\pi \int_{-\infty}^{t} f(u)\, du]$$

whose instantaneous frequency is $f(t)$ (from (1)–(2) in Appendix 2-1). We want to show that the bandwidth of the resultant spectral peak exhibited by $x(t)$ can be no narrower than the bandwidth of $f(t)$, say B. It then follows that the width, say Δf^*, of the spectral peak satisfies $\Delta f^* \geq B \cong 1/\Delta t^*$, where Δt^* is the temporal resolution width of $f(t)$. To accomplish this we consider the extreme example in which the entire frequency content of $f(t)$ is at the frequency b,

$$f(t) = a \sin(2\pi bt),$$

where $b \gg a$ so that $f(t)$ fluctuates rapidly relative to the largest value a of the instantaneous frequency $f(t)$ of $x(t)$. Use this condition and the small-angle approximation $\sin\theta \cong \theta$ to show that

$$x(t) \cong 2\pi \int_{-\infty}^{t} f(u)\, du.$$

This same result holds for arbitrary $f(t)$ provided that its lowest frequency b is not too small compared with its peak value a. Use this result to show that the bandwidth of $x(t)$ (the width Δf^* of its spectral peak) is approximately equal to the bandwidth B of $f(t)$, $\Delta f^* \cong B$. Therefore, $\Delta f^* \cong 1/\Delta t^*$. It can be shown that when the above assumption that b is not too small is violated, the bandwidth of $x(t)$ can only be larger than the value B. Thus, in general, $\Delta f^* \geq B \cong 1/\Delta t^*$.

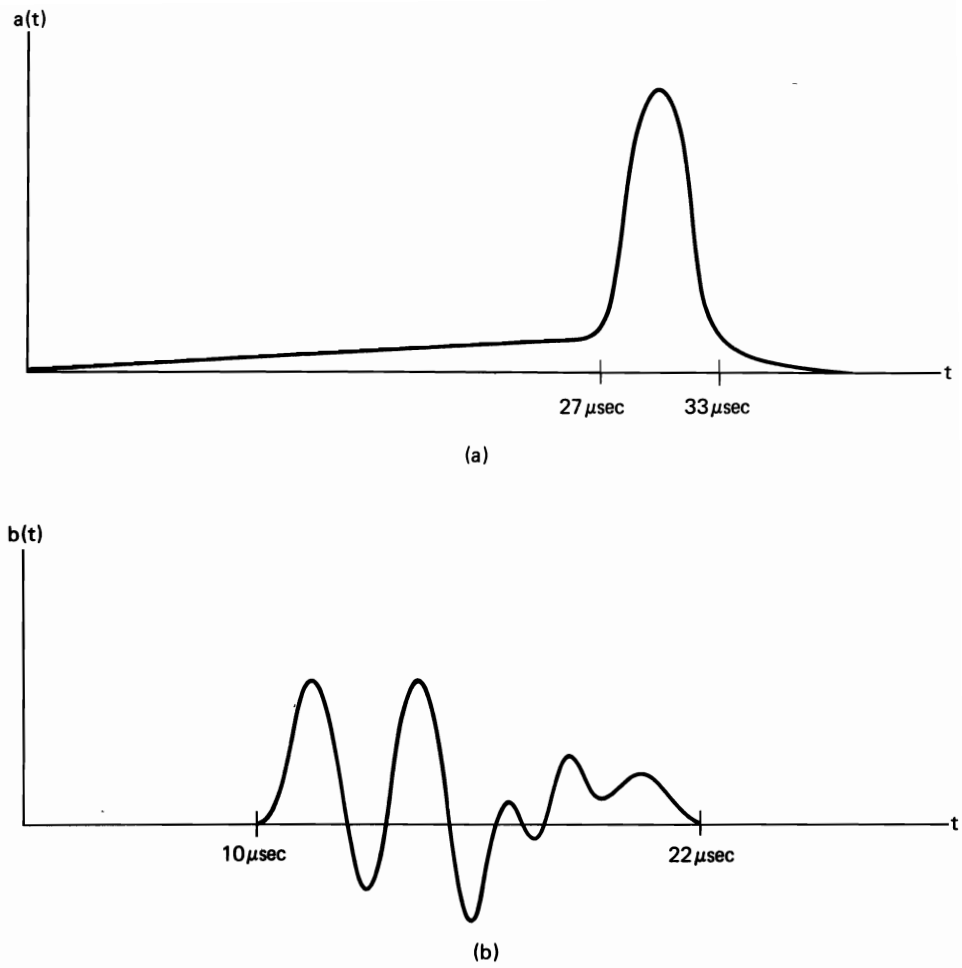


Figure 2-7 Two functions for application of the general rules of thumb regarding resolution width and overall width.

24. **(a)** Consider the pulse $a(t)$ shown in Figure 2-7(a). What is the approximate bandwidth of the Fourier transform $A(f)$?

(b) Consider the waveform $b(t)$ shown in Figure 2-7(b). Give a useful approximate lower limit on the resolution width of the Fourier transform $B(f)$. Then give an order-of-magnitude estimate of the bandwidth of $B(f)$.

(c) It is desired to determine the locations and widths of spectral peaks exhibited in some measured data that consist solely of sine waves with slowly varying amplitudes and phases. The minimum separation between peaks is known to be 10 Hz, and the narrowest peak is known to be 1 Hz. Approximately how long a time segment must be Fourier transformed in order to determine both the location and width of each spectral peak to within an error of no more than 100% of the peak width?

25. It is desired to estimate the transfer function of a linear time-invariant transformation that is continuously excited by random noise from a constant phenomenon, and it is decided to accomplish this by taking the ratio of the output periodogram to the input periodogram,

$$\frac{S_{y_T}(t - t_0, f)}{S_{x_T}(t, f)}.$$

If the impulse response $h(\tau)$ is known to be centered at $\tau = 10$ μs, and to have approximate width 1 μs, then what value for t_0 should be used and which of the values of T would yield the best results: $T = 1/10$ μs, 1 μs, or 10 μs? Explain your answer for T in words with reference to appropriate general results in the chapter. If it were decided to frequency-smooth the periodograms before taking their ratio, what value of width Δf for the smoothing window would you use, $\Delta f < 1/10$ MHz, $\Delta f > 1$ MHz, or $1/10$ MHz $< \Delta f < 1$ MHz? Explain your answer for Δf in words with reference to appropriate general results in the chapter.

26. **(a)** It is desired to estimate the amount of average power in $x(t)$ over the time interval 0 to 1s that is due to spectral components in the band $f_0 - \Delta f/2$ to $f_0 + \Delta f/2$ by integrating the periodogram $S_{x_T}(t, f)$ over this band. What values for t and T should be used, and which condition on Δf will yield the most accurate estimate, $\Delta f < 1$ Hz, $\Delta f > 1$ Hz, or $\Delta f > 10$ Hz? Explain your answers for Δf in words with reference to appropriate general results in the chapter.

(b) If it is desired to carry out the power measurement in (a) digitally and if it is known that the average power is negligible in bands above 100 Hz, then what is the lowest rate that should be used for time-sampling $x(t)$ (for digitization)? Explain your answer in words with reference to appropriate general results in the chapter.

APPENDIX 2-1

Instantaneous Frequency

Instantaneous Frequency Measurement

Consider the sine function with an argument that varies with time:

$$x(t) = \sin[\phi(t)]. \tag{1}$$

The *instantaneous frequency*[9] of this waveform is defined to be the derivative of the argument divided by 2π and is denoted by $f(t)$:

$$f(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt}. \tag{2}$$

For example, for the sine wave

$$x(t) = \sin(\omega_0 t + \theta),$$

[9] This definition of instantaneous frequency was apparently introduced by B. Van der Pol [Van der Pol 1946].

the instantaneous frequency is a constant,

$$f(t) = \frac{\omega_0}{2\pi}$$

If $f(t)$ is not a constant, then $x(t)$ in (1) is called a *frequency-modulated* (FM) *sine wave*.

Let us determine the conditions under which this mathematical definition of instantaneous frequency (2) corresponds to a quantity that can be measured in practice using spectral analysis; that is, we want to determine the conditions on $f(t)$ under which a time-variant finite-time spectrum $S_{x_T}(t, f)$ will track the time variations of $f(t)$ to within an accuracy of, say, Δf. Thus, we want $S_{x_T}(t, f)$ to be nonnegligible at time t within only the spectral band of width Δf, centered at $f(t)$ (and its image $-f(t)$). This will occur if and only if $|f(u)|$ remains in the spectral band $[|f(t)| - \Delta f/2, |f(t)| + \Delta f/2]$ during the time-interval $[t - \Delta t/2, t + \Delta t/2]$, where Δf and Δt are the spectral and temporal resolutions of $S_{x_T}(t, f)$, namely, $\Delta f \cong 1/T$ and $\Delta t \cong T$. Thus, if $f(t)$ changes by an amount δf during a small time interval δt, then it is sufficient if $|\delta f/\delta t|$ does not exceed $\Delta f/\Delta t \cong (\Delta f)^2$, which can be interpreted as a bound on the derivative of $f(t)$,

$$\left|\frac{df}{dt}\right| < (\Delta f)^2. \tag{3}$$

Thus, *the instantaneous frequency can be tracked by the time-variant periodogram within an accuracy of* Δf *if its rate does not exceed* $(\Delta f)^2$. This leads to a related condition involving only frequency parameters, as follows. The *peak frequency deviation* of the FM sine wave is defined by

$$f_{\max} \triangleq \max_t\{|f(t)|\},$$

and the frequency of the highest-frequency component of $f(t)$ is denoted by $\nu_{\max}$, which is the *bandwidth* of $f(t)$. Since the derivative of a sine wave with amplitude f and frequency ν,

$$(f)\sin(2\pi\nu t),$$

is upper bounded in magnitude by $|2\pi f\nu|$, then a sufficient condition for frequency tracking is

$$2\pi f_{\max}\nu_{\max} \leq (\Delta f)^2. \tag{4}$$

This condition is also necessary for (3) if $f(t)$ is simply a sine wave.

An example of an FM sinewave is the *linear FM signal*, for which

$$f(t) = \alpha t.$$

This signal

$$x(t) = \sin[\pi\alpha t^2]$$

is called a *chirp signal*, and α is the *chirp rate*. It follows from the preceding discussion that the instantaneous frequency of $x(t)$ can be tracked if (and in this case only if) the chirp rate does not exceed the square of the desired resolution,

$$\alpha < (\Delta f)^2.$$

In this case, it can be shown that the time-variant spectrum of the chirp signal is closely approximated by

$$S_{x_T}(t,f) \cong \frac{T}{4}\{\text{sinc}^2([f - \alpha t]T) + \text{sinc}^2([f + \alpha t]T)\}$$

for $\alpha t >> 2/T$.

Instantaneous Frequency Demodulation

As an application of the time-variant periodogram relation for filters, (40) in Chapter 2, consider the problem of measuring the instantaneous frequency of the FM sine wave

$$x(t) = \sin[\phi(t)]. \tag{5}$$

It is assumed that the instantaneous frequency rate is sufficiently low (3) that the time-variant periodogram with $T = \Delta t$ is simply

$$S_{x_{\Delta t}}(t,f) \cong \frac{\Delta t}{4}\{\text{sinc}^2([f - f(t)]\Delta t) + \text{sinc}^2([f + f(t)]\Delta t)\}, \tag{6}$$

where

$$f(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt}, \tag{7}$$

as explained in the preceding section. Consider two filters with transfer functions $H(f)$ and $G(f)$ that are approximately linear in magnitude over the spectral band for which $S_{x_T}(t, f)$ is nonnegligible, say

$$\begin{aligned} |H(f)| &\cong c(|f| - f_0) + d, \qquad ||f| - f_0| \leq B \\ |G(f)| &\cong c(f_0 - |f|) + d, \qquad ||f| - f_0| \leq B \end{aligned} \tag{8}$$

for some constants c and d, and assume that Δt greatly exceeds the memory lengths of these filters, which are generally much less than $1/B$ (so that (40) in Chapter 2 applies to the following). Now, consider the difference between the time-averaged powers of the responses of these two filters to $x(t)$:

$$\nu(t) \triangleq \frac{1}{\Delta t}\int_{t-\Delta t/2}^{t+\Delta t/2} y^2(u)\,du - \frac{1}{\Delta t}\int_{t-\Delta t/2}^{t+\Delta t/2} z^2(u)\,du \tag{9}$$

where

$$\begin{aligned} y(t) &= x(t) \otimes h(t) \\ z(t) &= x(t) \otimes g(t). \end{aligned} \tag{10}$$

The function $\nu(t)$ can be reexpressed as

$$\begin{aligned} \nu(t) &= R_{y_{\Delta t}}(t, 0) - R_{z_{\Delta t}}(t, 0) \\ &= \int_{-\infty}^{\infty} S_{y_{\Delta t}}(t,f)\,df - \int_{-\infty}^{\infty} S_{z_{\Delta t}}(t,f)\,df. \end{aligned} \tag{11}$$

Substitution of (40) in Chapter 2, for both y and H and z and G, into (11) yields

(using $\Delta\tau^* \ll 1/B$)

$$\nu(t) \cong \int_{-\infty}^{\infty} (|H(f)|^2 - |G(f)|^2) S_{x_{\Delta t}}(t, f)\, df \tag{12}$$

$$\cong a \int_{-\infty}^{\infty} [|f| - f_0]\, S_{x_{\Delta t}}(t, f)\, df, \qquad \text{for } ||f(t)| - f_0| \leq B \quad \text{and} \quad \frac{1}{\Delta t} < B, \tag{13}$$

where a is just a scale factor. Substitution of (6) into (13) and use of the fact that $S_{x_{\Delta t}}(t, f)$ is negligible for $||f| - f(t)| > 1/\Delta t$ yields

$$\nu(t) \cong b[|f(t)| - f_0], \qquad \text{for } \left|\frac{df}{dt}\right| < \left(\frac{1}{\Delta t}\right)^2 < B^2 \quad \text{and} \quad ||f(t)| - f_0| \leq B, \tag{14}$$

where b is just a scale factor. Thus, the difference (9) of time-averaged powers of the filtered signals yields the instantaneous frequency (14). Since this result requires balancing the characteristics of the two filters, as indicated by (8), this method of measurement of instantaneous frequency is called *balanced frequency discrimination*.[10] Filters that exhibit the linear characteristics (8) are called *frequency-to-amplitude converters*. Alternative methods for rapid instantaneous frequency measurement that can accommodate faster frequency deviation are available. Some of these are closely related to some of the time-series-modeling approaches to spectral analysis described in Chapter 9.

[10] This balanced scheme will cancel out even-order nonlinear terms, such as $(f \pm f_0)^2$, in (8) that are common to both $|H(f)|$ and $|G(f)|$.

3

STATISTICAL SPECTRAL ANALYSIS

This chapter introduces the fundamentals of statistical spectral analysis: the equivalence between statistical spectra obtained from *temporal smoothing* and statistical spectra obtained from *spectral smoothing,* and the relationship between these *statistical spectra* and the abstract *limit spectrum.* The motivation for smoothing—to average out undesired random effects that mask spectral features of interest—is developed by consideration of the problem of measuring the parameters of a resonance phenomenon. It is established that the *limit autocorrelation* and *limit spectrum* are a Fourier transform pair and that each is a self-determinate characteristic under a linear time-invariant transformation (filtering operation). The utility of the limit spectrum for characterizing spectral features in stationary time-series is illustrated with several examples of *modulated waveforms.* Periodically time-sampled waveforms are considered, and a formula for the limit spectrum of the discrete-time sampled data, in terms of the limit spectrum of the waveform, is derived and used to describe further the spectral aliasing phenomenon. The *moving average* and *autoregressive models* of discrete-time data are introduced, and their limit spectra are derived. In Appendix 3-1, *bandpass time-series* are considered and a general representation in terms of *lowpass time-series* is derived, and the relationships between the limit autocorrelations and limit spectra of the bandpass and lowpass time-series also are derived. In Appendix 3-2, the role of spectral analysis in the detection of random signals is explained.

A. MOTIVATING EXAMPLE

In order to understand why a statistical (average) spectrum can be preferable to a nonstatistical spectrum, we must focus our attention not on the data itself but rather on the source of the data—the mechanism that generates the data. Generally speaking, data is nothing more than a partial representation of some phenomenon—a numerical representation of some aspects of a phenomenon. The fundamental reason for interest in a statistical (e.g., time-averaged) spectrum of some given data is a belief that interesting aspects of the phenomenon being investigated have spectral influences on the data that are masked by uninteresting (for the purpose at hand) random effects and an additional belief (or, at least, hope) that these spectral influences can be revealed by averaging out the random effects. This second belief (or hope) should be based on the knowledge (or, at least, suspicion) that the spectral influences of the interesting aspects of the phenomenon are time-invariant, so that the corresponding invariant spectral features (such as peaks or valleys) will be revealed rather than destroyed by time-averaging.

This idea is illustrated with the following example. Consider the problem of determining the resonance frequency and damping ratio of a single-degree-of-freedom mechanical system (see exercise 10) that is subject to a continuous random vibrational force excitation x. The system displacement response y can be modeled as an LTI transformation of the excitation, with the transfer function magnitude $|H|$ shown in Figure 3-1, which reveals the resonance frequency f_0

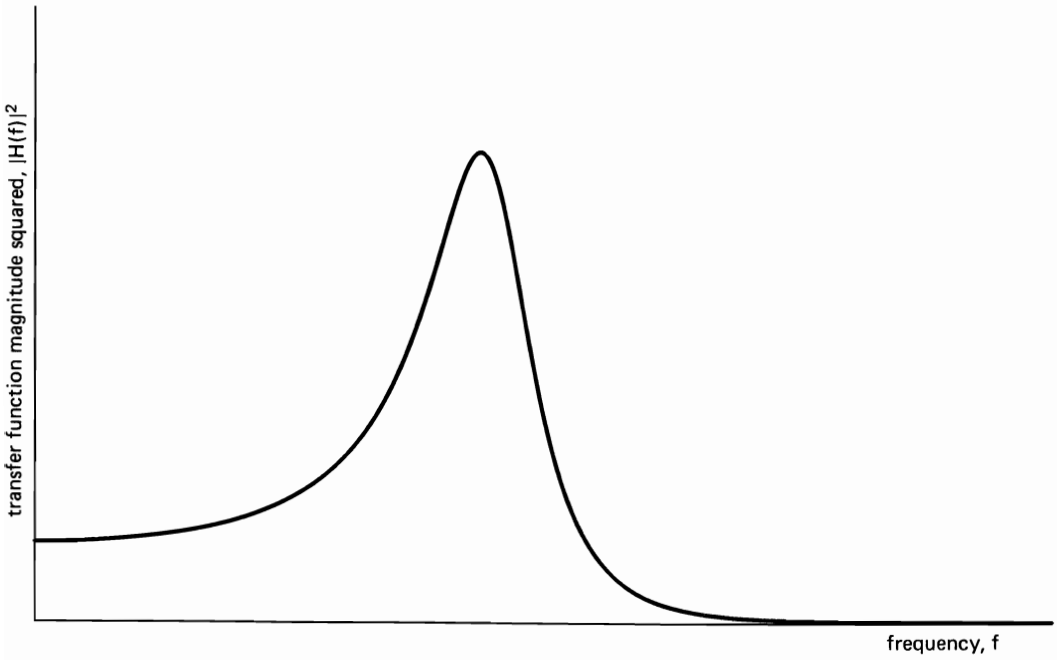


Figure 3-1 Magnitude-squared transfer function of resonant system.

and the bandwidth B (which can be related to the damping ratio). The vibrational response of the system is random by virtue of the randomness of the excitation. Consequently, the spectrum of the response data does not exhibit the desired single smooth peak shown in Figure 3-1. Rather, it is an erratic function with numerous sharp peaks and valleys, as revealed by the simulation shown in Figure 3-2(a). Moreover, as the time-interval of analysis is made longer by increasing T, the spectrum only becomes more erratic (at least locally), as revealed by the simulation shown in Figure 3-2(b). However, if the random excitation arises from a system in statistical equilibrium, the underlying time-invariance in the excitation, as well as in the resonant system, suggests that time-averaging the response spectrum will reduce the random effects while leaving the desired spectral features intact. In fact, it is shown in the next section that for $\Delta t/T >> 1$, the time-smoothed spectrum,

$$S_{y_T}(t,f)_{\Delta t} \triangleq S_{y_T}(t,f) \otimes u_{\Delta t}(t), \tag{1}$$

is closely approximated by the frequency-smoothed spectrum

$$S_{y_{\Delta t}}(t,f) \otimes z_{1/T}(f), \tag{2}$$

and for sufficiently large Δt and T the particular form of the spectral-smoothing window $z_{1/T}$ is irrelevant. Consequently, approximation (40) in Chapter 2 can be used to obtain

$$S_{y_T}(t,f)_{\Delta t} \cong |H(f)|^2 S_{x_T}(t,f)_{\Delta t}, \tag{3}$$

for which it has been assumed that[1]

$$\frac{1}{T} \leqslant \Delta f^*, \tag{4}$$

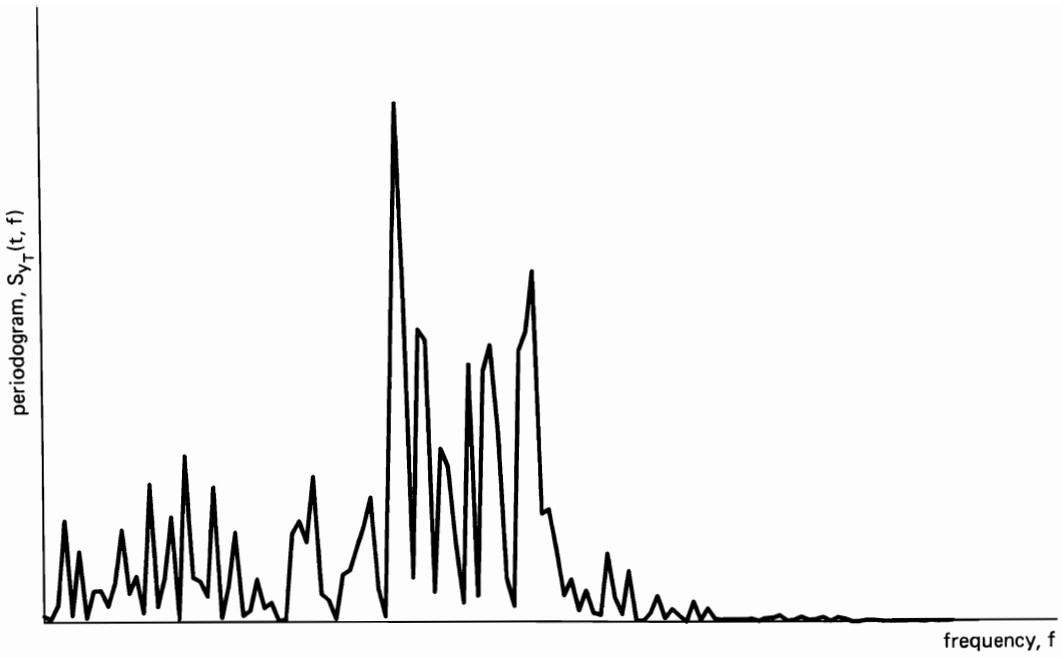
where Δf^* is the resolution width of the function $|H|^2$ (Δf^* is on the order of $1/\Delta\tau^*$, where $\Delta\tau^*$ is the system memory length—the width of h). If the system excitation is completely random so that it exhibits no spectral features, then for $\Delta t/T >> 1$, $S_{x_T}(t,f)_{\Delta t}$ will closely approximate a constant (over the support for which $|H|$ is nonnegligible), say N_0. Therefore, (3) yields the desired result:

$$S_{y_T}(t,f)_{\Delta t} \cong N_0|H(f)|^2, \qquad \Delta t \Delta f^* \geqslant \Delta t/T >> 1, \tag{5}$$

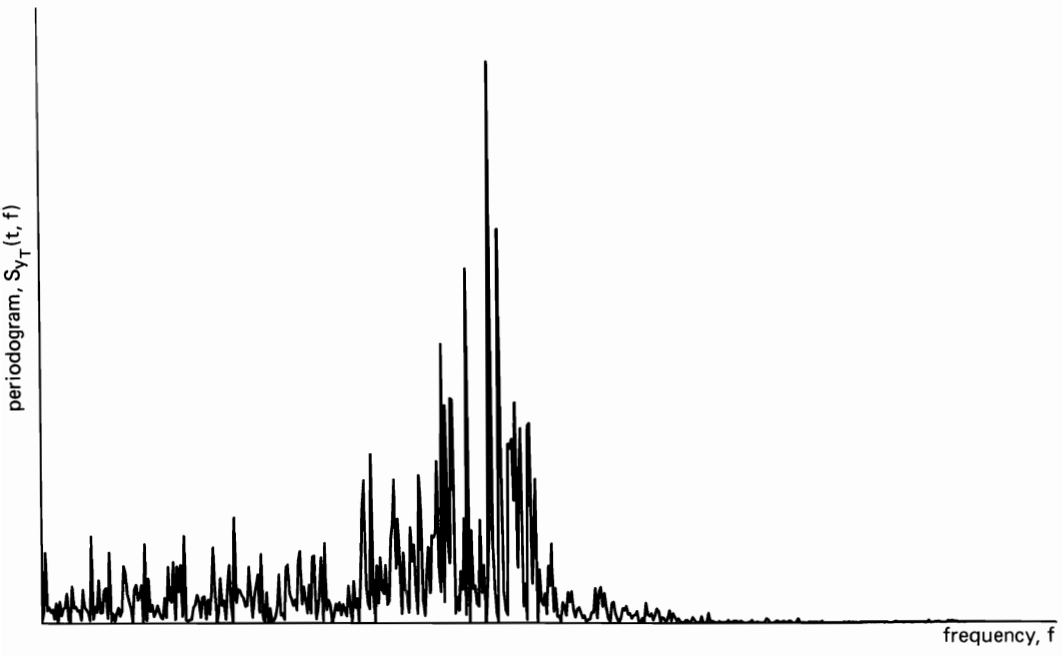
from which the resonance frequency and damping ratio can be determined. This is illustrated with the simulations shown in Figure 3-2c, d.

In addition to illustrating the use of a statistical spectrum obtained from time-smoothing a periodogram (1), this example introduces the idea that an *equivalent* statistical spectrum can also be obtained from frequency-smoothing a periodogram (2). This equivalence is established in the following section. However, before proceeding it should be clarified that in practice when automated spectrum analyzers are used to study visually the spectral features of a phenomenon, it is common practice to use very little smoothing (and in some cases no smoothing) in spite of the erratic behavior of the displayed spectrum due to random effects. But it should be remembered that human visual perception incorporates spatial

[1] Condition (4) guarantees that the order of multiplication with $|H|^2$ and convolution with $w_{1/T}$ in (40), Chapter 2, can be interchanged to obtain a close approximation.

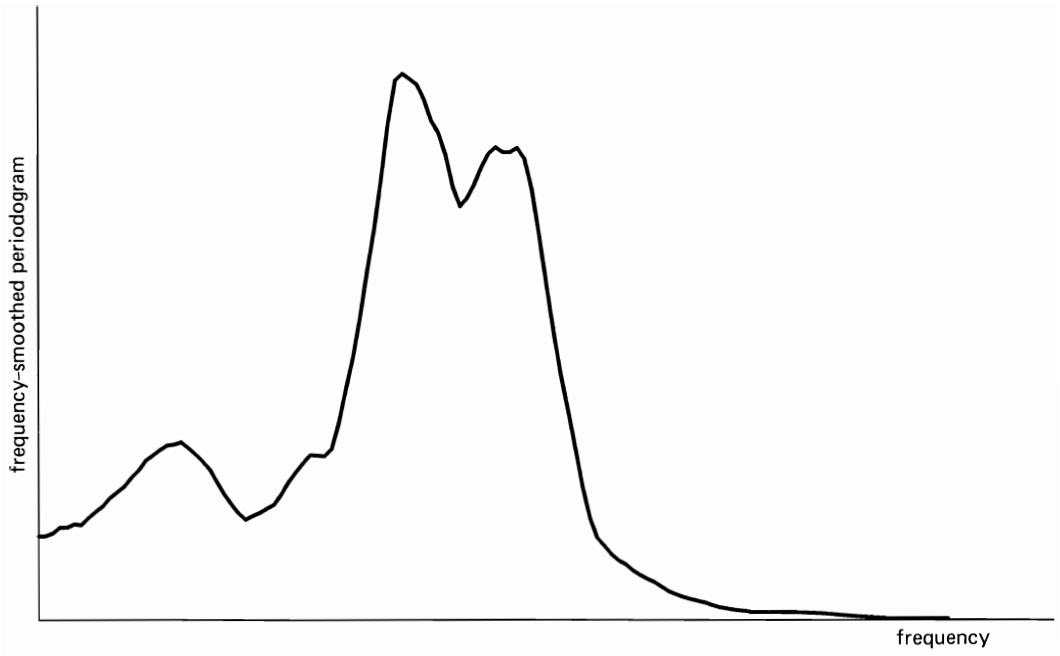


(a)

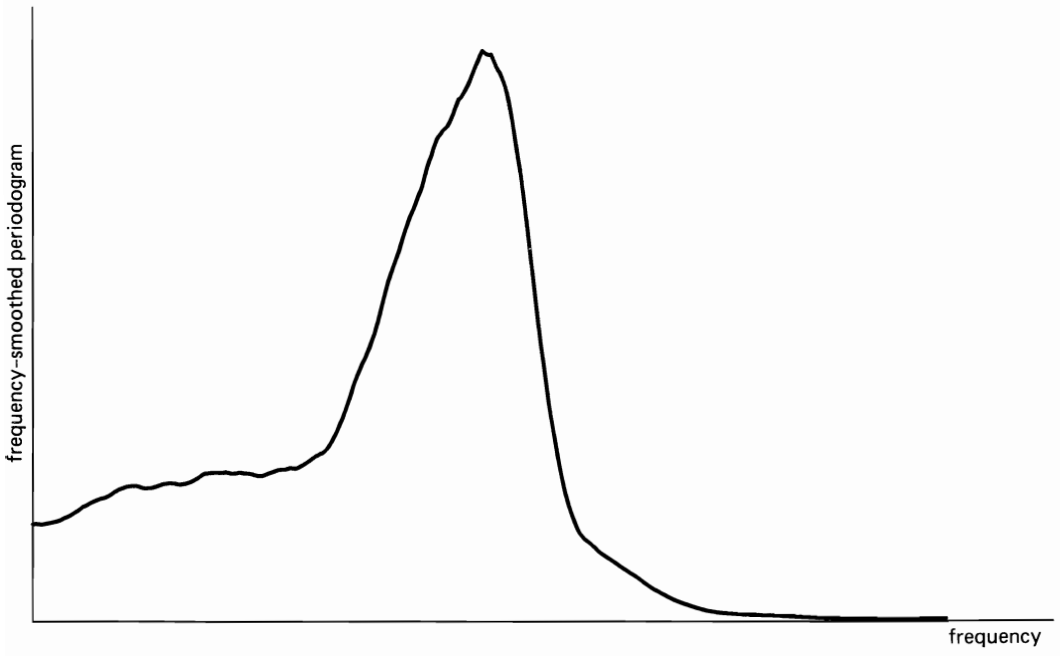


(b)

Figure 3-2 Nonstatistical spectra of response of resonant system to completely random excitation. (Length of time interval of analysis is T): (a) $T = T_0$, (b) $T = 4T_0$.



(c)



(d)

Figure 3-2 (continued) Statistical spectra obtained by frequency-smoothing the nonstatistical spectra shown in (a) and (b): (c) from (a) with $\Delta t/T = 8$, (d) from (b) with $\Delta t/T = 32$.

integration and temporal memory so that we in effect perceive a smoothed spectrum even when the analyzer uses no smoothing. This is apparent from Figure 3-2 (a) and (b), in which we can perceive the smoothed spectra that are shown in Figure 3-2 (c) and (d).

B. TEMPORAL- AND SPECTRAL-SMOOTHING EQUIVALENCE

In this section, it is shown that *time-smoothing,* with a temporal window of width $\Delta t >> T_1 = 1/\Delta f$, a time-variant spectrum with resolutions of T_1 in time and $1/T_1 = \Delta f$ in frequency, yields a statistical spectrum

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} \triangleq S_{x_{T_1}}(t,f) \otimes u_{\Delta t}(t), \tag{6}$$

that closely approximates the statistical spectrum

$$S_{x_{\Delta t}}(t,f)_{\Delta f} \triangleq S_{x_{T_2}}(t,f) \otimes z_{\Delta f}(f) \tag{7}$$

obtained by *frequency-smoothing,* with a spectral window of width on the order of $\Delta f >> 1/T_2 = 1/\Delta t$, a time-variant spectrum with resolutions of $T_2 = \Delta t$ in time and $1/T_2$ in frequency. And it is shown that this approximation,

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} \cong S_{x_{\Delta t}}(t,f)_{\Delta f}, \qquad \Delta t \Delta f >> 1, \tag{8}$$

can be made as accurate as desired for any given t and Δf and for all f by choosing Δt sufficiently large, provided only that the limit autocorrelation function

$$\hat{R}_x(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x\left(t - \frac{\tau}{2}\right) dt \tag{9}$$

exists. In fact, equivalence (8) will be obtained as a special case of a more general equivalence that incorporates an arbitrary bounded data-tapering aperture a_T with finite support T_*,

$$a_T(t) = 0, \qquad |t| > \frac{T_*}{2} \geqslant \frac{T}{2}. \tag{10}$$

Before studying the following derivation of this equivalence, the reader should interpret the signal-processing operations involved in obtaining the time-averaged periodogram, (6), and the frequency-smoothed periodogram, (7), in terms of the pictorial illustration shown in Figure 3-3.

The derivation consists of the following five steps (in which $T = 1/\Delta f$ represents T_1 in (6)):

$$S_{x_T}(t,f)_{\Delta t} \triangleq \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} S_{x_T}(t-u,f)\, du \tag{11}$$

$$= \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} \int_{-T_*}^{T_*} R_{x_T}(t-u,\tau) e^{-i2\pi f\tau}\, d\tau\, du \tag{12}$$

$$= \int_{-T_*}^{T_*} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} R_{x_T}(t-u,\tau)\, du\, e^{-i2\pi f\tau}\, d\tau \tag{13}$$

$$\cong \int_{-T_*}^{T_*} R_x(t,\tau)_{\Delta t} \frac{1}{T} r_a(\tau)\, e^{-i2\pi f\tau}\, d\tau, \qquad \Delta t >> T \tag{14}$$

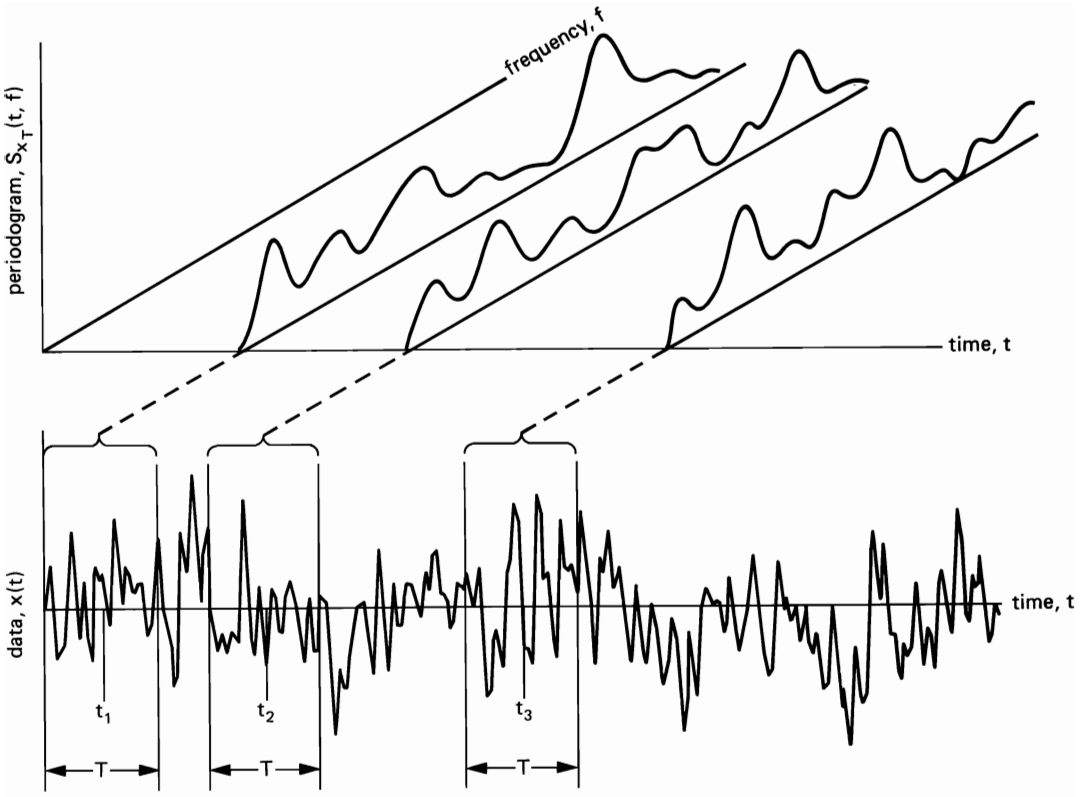


Figure 3-3 Pictorial illustration of a time-variant periodogram.

$$\cong \int_{-T_*}^{T_*} R_{x_{\Delta t}}(t, \tau) \frac{1}{T} r_a(\tau)\, e^{-i2\pi f\tau}\, d\tau, \qquad \Delta t \gg T \tag{15}$$

$$= \int_{-\infty}^{\infty} S_{x_{\Delta t}}(t, f - \nu) \frac{1}{T} |A_{1/T}(\nu)|^2\, d\nu \tag{16}$$

$$\triangleq S_{x_{\Delta t}}(t, f)_{\Delta f}. \tag{17}$$

Step 1, (12), is the time-variant periodogram-correlogram relation for tapered data, (18) in Chapter 2, together with the property that follows from (10),

$$r_a(\tau) = 0, \qquad |\tau| > T_*, \tag{18}$$

of the finite autocorrelation r_a defined by (29) in Chapter 2. Step 2, (13), is simply a reversal of the order of two finite-interval integrations. Step 3, (14), is approximation (28) in Chapter 2. Approximation (14) can be made as accurate as desired for any given t and T and for all f by choosing Δt sufficiently large, since T_* is finite, provided only that $\hat{R}_x$ exists (exercise 1). Step 4, (15), is approximation (30) in Chapter 2. Approximation (15) can be made as accurate as desired for any given t and T and for all f by choosing Δt sufficiently large, since T_* is finite, provided only that $\hat{R}_x$ exists (exercise 1). It follows from

exercise 10 in Chapter 2 that although $R_{x_T}(t, \tau)$ in (12) includes a data-tapering window, $R_{x_{\Delta t}}(t, \tau)$ in (15) does not. Step 5, (16), is the convolution theorem (which applies because (18) allows replacement of finite limits of integration in (15) with infinite limits), together with the definition

$$A_{1/T}(\cdot) \triangleq F\{a_T(\cdot)\}. \tag{19}$$

Equivalence (11)–(17) reveals that *the particular data-tapering window a_T in a time-smoothed spectrum determines an effective spectral smoothing window* $(1/T)|A_{1/T}(f)|^2$ (which is the periodogram of the data-tapering window).[2] This equivalence also reveals that effectively the same statistical spectrum can be obtained by carrying out spectral smoothing as in (16) rather than data-tapering and time-smoothing, as in (3) in Chapter 2 and (11) in this chapter. Other related equivalences, some involving the pseudospectrum, are described in Chapter 4, Section A. Approximations (14) and (15) can be quantified in terms of rms error, that is, the square root of the average over all time t of the squared error of approximation. This approach is set up in Chapter 5.

C. THE LIMIT SPECTRUM

It follows from the preceding discussion (following (17)) that in the limit $\Delta t \to \infty$, approximations (14) and (15) become exact (exercise 1), so that

$$\begin{aligned} \lim_{\Delta t \to \infty} S_{x_{1/\Delta f}}(t, f)_{\Delta t} &= \lim_{\Delta t \to \infty} S_{x_{\Delta t}}(t, f)_{\Delta f} \\ &= \lim_{\Delta t \to \infty} \int_{-T_*}^{T_*} R_x(t, \tau)_{\Delta t} \frac{1}{T} r_a(\tau)\, e^{-i2\pi f\tau}\, d\tau. \end{aligned} \tag{20}$$

Moreover, it follows from (31) in Chapter 2 that

$$\lim_{\Delta t \to \infty} \int_{-T_*}^{T_*} R_x(t, \tau)_{\Delta t} \frac{1}{T} r_a(\tau)\, e^{-i2\pi f\tau}\, d\tau = \int_{-T_*}^{T_*} \hat{R}_x(\tau) \frac{1}{T} r_a(\tau)\, e^{-i2\pi f\tau}\, d\tau, \tag{21}$$

where the interchange of limit and integration operations is justified because T_* is finite (exercise 1). Application of the convolution theorem to (21), together with (20), yields

$$\lim_{\Delta t \to \infty} S_{x_{1/\Delta f}}(t, f)_{\Delta t} = \lim_{\Delta t \to \infty} S_{x_{\Delta t}}(t, f)_{\Delta f} = \hat{S}_x(f) \otimes [\Delta f |A_{\Delta f}(f)|^2] \tag{22}$$

for which $\hat{S}_x$ is given by the Fourier transform

$$\hat{S}_x(\cdot) = F\{\hat{R}_x(\cdot)\}, \tag{23}$$

which is assumed to exist.[3] Equation (22) reveals that both of the statistical spectra, $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ and $S_{x_{\Delta t}}(t, f)_{\Delta f}$, approach the frequency-smoothed limit

[2] In all other treatments of this subject known to the author, this fundamental equivalence (11)–(17) is obtained only in a probabilistic sense, which involves unnecessary abstraction.

[3] This is equivalent to assuming that $\hat{R}_x(\tau)$, with all (if any exist) finite additive sine wave components subtracted out, is absolutely integrable.

spectrum as Δt is increased. Consequently, the (assumed) existence of the limit autocorrelation guarantees that both the time-smoothed and frequency-smoothed statistical spectra converge as $\Delta t \Delta f \to \infty$ for any fixed $\Delta f > 0$. Moreover, since

$$\lim_{\Delta f \to 0} \Delta f |A_{\Delta f}(f)|^2 = \gamma \delta(f), \tag{24}$$

in which

$$\gamma \triangleq \lim_{\Delta f \to 0} \Delta f \int_{-\infty}^{\infty} |A_{\Delta f}(f)|^2 \, df = \lim_{T \to \infty} \frac{1}{T} \int_{-\infty}^{\infty} a_T^2(t) \, dt, \tag{25}$$

then (22) yields[4] the *limit spectrum*

$$\lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{x_{1/\Delta f}}(t, f)_{\Delta t} = \gamma \hat{S}_x(f) \tag{26a}$$

$$\lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{x_{\Delta t}}(t, f)_{\Delta f} = \gamma \hat{S}_x(f), \tag{26b}$$

and for typical unity-height tapering windows, γ is on the order of unity (e.g., $\gamma = 1$ for $a_T(t) = Tu_T(t)$).

Identities (26a) and (26b) reveal that the limit spectrum, which is also given by (23), is an idealization ($\Delta t \to \infty$, $\Delta f \to 0$) of the two fundamental types of statistical spectra that are measured in practice by time-smoothing or frequency-smoothing a periodogram, and that the particular form of the data-tapering window, or the corresponding spectral smoothing window, is irrelevant[5] for sufficiently large Δt and $1/\Delta f$. It is of paramount importance that the order of the two limit operations in (26) cannot be reversed, because the sequence of periodograms

$$S_{x_T}(t, f), \qquad T \to \infty \tag{27}$$

does not converge.[6] This is proved in Chapter 5, Section C, where it is established that $S_{x_T}(t, \cdot)$ becomes an increasingly erratic function as T is increased. The important practical implications of this for smoothing also are described in Chapter 5.

The fact that the limit (26b) of the frequency-smoothed periodogram is identical to the Fourier transform of the limit autocorrelation (23) was first presented (in rather terse form) by Albert Einstein [Einstein 1914], and then independently presented and rigorously proved by Norbert Wiener, and it is

[4] The limit $\Delta f \to 0$ of (22) exists in the usual pointwise sense, provided that $\hat{R}_x$ is absolutely integrable. However, if $\hat{R}_x$ contains additive periodic components (due to additive periodic components in $x(t)$), then this limit exists only in a more general sense that accommodates Dirac deltas in $\hat{S}_x$ (see [Pfaffelhuber, 1975]).

[5] In practice Δt and $1/\Delta f$ often cannot be made as large as desired, and the particular data-tapering window or corresponding spectral-smoothing window therefore becomes important, as explained in subsequent chapters.

[6] The erroneous definition of the limit spectrum as the limit $T \to \infty$ of the periodogram (27) is a common appearance in the early literature—for example, [Carson 1931; Rice 1944; James et al. 1947; Lawson and Uhlenbeck 1949; Page 1952; Lanning and Battin 1956; Blackman and Tukey 1958; Kharkevich 1960; Papoulis 1962; Schroeder and Atal 1962]—and it continues to appear in some of the current literature, such as [Priestly 1981, Sec. 4.6; Urkowitz 1983]. An early proof that (27) does not converge for random time-series is given in [Grenander 1951].

therefore called the *Wiener relation*[7] [Wiener 1930, Sec. 3]. The fact that the limit (26a) of the time-smoothed periodogram is also identical to the Fourier transform of the limit autocorrelation—although intimately related to practical methods of statistical spectral analysis as explained in Chapter 4—is apparently not a commonly known result. In contrast to its previously unrecognized role,[8] identity (26a) is at the core of the nonprobabilistic theory presented here. Furthermore, it is (26a) rather than (26b) that leads more naturally to a probabilistic interpretation of the limit spectrum and to probabilistic methods of analysis of statistical spectra, as explained in Chapter 5.

Perhaps the single most important result based on the limit spectrum is the *limit-spectrum relation*,

$$\hat{S}_y(f) = |H(f)|^2 \hat{S}_x(f), \tag{28}$$

for filters

$$y(t) = h(t) \otimes x(t) \tag{29}$$

$$H(\cdot) = F\{h(\cdot)\}, \tag{30}$$

which follows directly from the convolution theorem, the *Wiener relation*, and the *limit-autocorrelation relation*,

$$\hat{R}_y(\tau) = r_h(\tau) \otimes \hat{R}_x(\tau) \tag{31}$$

$$r_h(\tau) = h(\tau) \otimes h(-\tau), \tag{32}$$

derived in Chapter 2, Section F.

An argument completely paralleling that given in Section G of Chapter 2 but based on the exact limit-spectrum relation for filters (28) instead of approximation (40) in Chapter 2 yields the interpretation of the limit spectrum $\hat{S}_x$ as the *spectral density of time-averaged power* in $x(t)$. That is, (49) in Chapter 2 becomes exact

[7] More precisely, Wiener worked with the smoothing interval $\Delta f = (-\infty, f)$ in (26b), which yields the *integrated limit spectrum*, whose derivative is the limit spectrum $\hat{S}_x$. Related work by A. J. Khinchin [Khinchin 1934] for random processes yielded a probabilistic counterpart of the Wiener relation, which is known as the *Wiener-Khinchin relation*. Also, Kampé de Fériet [Kampé de Fériet 1954] provided an alternative method of derivation of the *Wiener relation* based on Paul Lévy's *continuity theorem* in probability theory. Since the difference of smoothing intervals $(-\infty, f_1)$ and $(-\infty, f_2)$ used by Wiener yields a rectangle spectral-smoothing window that corresponds to a sinc data-tapering window a_T, which has infinite support, then (10) is violated and the straightforward method used here to derive (26b) does not apply; see exercise 2. In view of Einstein's prior presentation of this relation, it might be more appropriate to call it the *Einstein-Wiener relation*.

[8] It is unlikely that (26a) did not play an important conceptual role during the 1930s and 1940s, when the focus was shifting from deterministic theory to probabilistic theory, since it is (26a) rather than (26b) that leads most naturally (via H.O.A. Wold's isomorphism [Wold 1948]) to a probabilistic interpretation of the limit spectrum, as explained in Chapter 5; but the author has not been successful in finding documentation of the role of (26a) in the literature. Although the time-smoothed spectrum in (26a) is shown in Chapter 4 to be identical to the statistical spectrum obtained by the *wave-analysis* and *demodulation* methods, which do appear in the early literature, the data-tapering window a_T that corresponds to the filters typically referred to for these methods does not have finite support (10), and therefore the straightforward method used here to derive (26a) does not apply.

in the limit $\Delta t \to \infty$,

$$\lim_{\Delta t \to \infty} P_{\Delta t \Delta f}(t, f) = 2 \int_{f - \Delta f/2}^{f + \Delta f/2} \hat{S}_x(\nu)\, d\nu. \tag{33}$$

Specifically, use of the definition of time-averaged power in the spectral band $[f - \Delta f/2, f + \Delta f/2]$ (and its image $[-f - \Delta f/2, -f + \Delta f/2]$),

$$\hat{P}_{\Delta f}(f) \triangleq \lim_{\Delta t \to \infty} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} [x(t) \otimes h(t)]^2\, dt = \hat{R}_y(0) = \int_{-\infty}^{\infty} \hat{S}_y(\nu)\, d\nu, \tag{34}$$

together with (28) and

$$H(\nu) = \begin{cases} 1, & \|\nu| - f| \leq \Delta f/2 \\ 0, & \|\nu| - f| > \Delta f/2, \end{cases} \tag{35}$$

yields

$$\hat{P}_{\Delta f}(f) = 2 \int_{f - \Delta f/2}^{f + \Delta f/2} \hat{S}_x(\nu)\, d\nu, \tag{36}$$

from which we obtain the desired result

$$\lim_{\Delta f \to 0} \frac{1}{2\Delta f} \hat{P}_{\Delta f}(f) = \hat{S}_x(f), \tag{37}$$

revealing the interpretation of $\hat{S}_x$ as a *power spectral density* (PSD).

D. EXAMPLES OF SPECTRAL DENSITY

1. White Noise

A time-series is called *white noise* if its PSD is a constant, say N_0,

$$\hat{S}_x(f) = N_0. \tag{38}$$

In this case, inverse Fourier transformation yields

$$\hat{R}_x(\tau) = N_0 \delta(\tau), \tag{39}$$

revealing that there is no correlation in $x(t)$ for nonzero lag, $\tau \neq 0$. Strictly speaking, white noise is not within the scope of the theory developed here because $\hat{R}_x(\tau)$ does not exist at $\tau = 0$. But white noise can be used to great advantage with the theory in a formal way.

An interesting curiosity is the idealized linear frequency-modulated sine wave

$$x(t) = \sin(2\pi t^2) \tag{40}$$

with instantaneous frequency

$$f(t) \triangleq \frac{d}{dt}(t^2) = 2t. \tag{41}$$

It can be shown that, similar to white noise, the autocorrelation of $x(t)$ vanishes

for all $\tau \neq 0$,

$$\hat{R}_x(\tau) = \begin{cases} \frac{1}{2}, & \tau = 0 \\ 0, & \tau \neq 0 \end{cases} \tag{42}$$

[Wiener 1930]. However, unlike white noise, $\hat{R}_x(0)$ exists and[9]

$$\hat{S}_x(f) \equiv 0. \tag{43}$$

This type of pathological behavior in a time-series model can be avoided by considering only models for which $\hat{R}_x(\tau)$ is continuous at $\tau = 0$ and therefore continuous for all τ (see [Gardner 1985]). Hence, it is assumed that all limit autocorrelations are continuous, as stated at the outset in Chapter 2.

2. Sine Wave with Additive Noise

Consider the sum of time-series

$$x(t) = s(t) + n(t) \tag{44}$$

$$s(t) = a\cos(\omega_0 t + \phi_0). \tag{45}$$

It can be shown (exercise 4) that

$$\hat{R}_x(\tau) = \hat{R}_s(\tau) + \hat{R}_n(\tau) \tag{46}$$

if and only if $\hat{R}_n(\tau)$ exists and

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t)\cos(\omega_0 t + \phi)\, dt = 0 \tag{47}$$

for all ϕ, which holds if and only if $n(t)$ contains no finite additive sine wave component with frequency $f_0 = \omega_0/2\pi$. Furthermore, for $\omega_0 \neq 0$,

$$\hat{R}_s(\tau) = \frac{a^2}{2}\cos(\omega_0 \tau) \tag{48}$$

and therefore (exercise 4)

$$\hat{S}_x(f) = \frac{a^2}{4}[\delta(f - f_0) + \delta(f + f_0)] + \hat{S}_n(f) \tag{49}$$

provided that (47) holds. The Dirac deltas in the limit spectrum (49) are called *spectral lines* (see Figure 3-4).

3. Sine Wave with Multiplicative Noise (Amplitude Modulation)

Consider the product of time-series

$$x(t) = s(t)n(t) \tag{50}$$

$$s(t) = \cos(\omega_0 t + \phi_0). \tag{51}$$

[9] It should be clarified that the inverse Fourier transform does not exist for functions with discontinuities of the type (42). Consequently, although $\hat{S}_x = F\{\hat{R}_x\}$ must be valid, $\hat{R}_x = F^{-1}\{\hat{S}_x\}$ is not valid for (42). Hence the power $\hat{R}_x(0) = \frac{1}{2}$ of the waveform (40) cannot be obtained by integrating $\hat{S}_x$ over all frequencies. Here $\hat{S}_x(f) \equiv 0$ because the finite power in $x(t)$ is spread uniformly over an infinite band of frequencies.

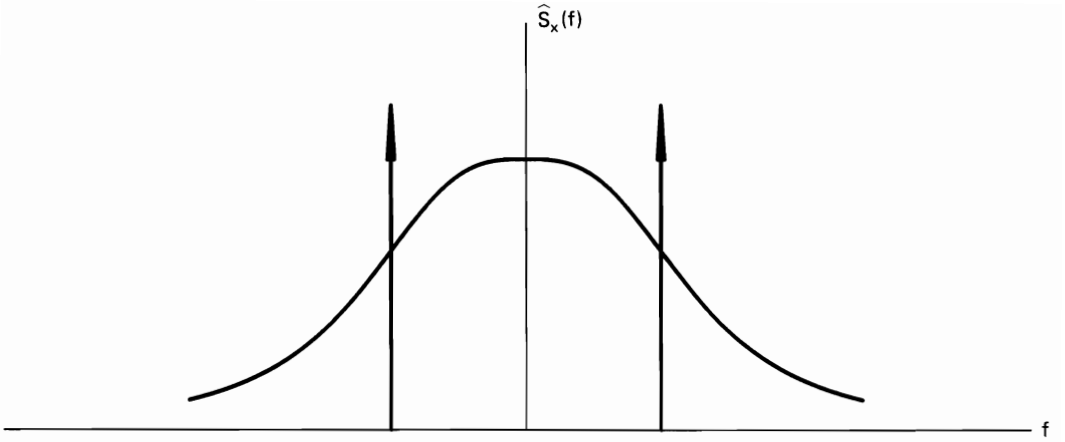


Figure 3-4 A limit spectral density for a sine wave with additive noise.

It can be shown (exercise 5) that

$$\hat{R}_x(\tau) = \hat{R}_s(\tau)\hat{R}_n(\tau) \tag{52}$$

if and only if $\hat{R}_n(\tau)$ exists and the lag products $n(t + \tau/2)n(t - \tau/2)$ contain no finite additive sine wave of frequency $2f_0$, that is,

$$\lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} n\left(t + \frac{\tau}{2}\right)n\left(t - \frac{\tau}{2}\right)\cos(2\omega_0 t + 2\phi_0)\, dt = 0 \tag{53}$$

for all τ, which holds only if $n(t)$ contains no additive, multiplicative, or other more subtle forms of periodicity with frequencies appropriately related to f_0. When condition (53) is violated, $\hat{R}_x$ can contain an additional term[10] that is commonly unrecognized in probabilistic treatments due to inappropriate phase-randomization [Gardner 1987a]. It follows from (48) and (52) that (exercise 5)

$$\hat{S}_x(f) = \tfrac{1}{4}\left[\hat{S}_n(f - f_0) + \hat{S}_n(f + f_0)\right] \tag{54}$$

provided that (53) holds.

4. Pulse-Amplitude Modulation

Consider the pulse-train time-series

$$x(t) = \sum_{j=-\infty}^{\infty} a(jT_0)p(t - jT_0) \tag{55}$$

for which $p(t)$ is a pulse with width that is typically on the order of T_0. It can be shown (exercise 6) that

$$\hat{R}_x(\tau) = \frac{1}{T_0}\sum_{k=-\infty}^{\infty} \tilde{R}_a(kT_0)r_p(\tau - kT_0) \tag{56}$$

for which r_p is the finite autocorrelation of p and $\tilde{R}_a$ is the *discrete-time limit*

[10] This additional term is studied in Part II.

autocorrelation

$$\tilde{R}_a(kT_0) \triangleq \lim_{N\to\infty} \frac{1}{2N+1} \sum_{j=-N}^{N} a(jT_0 + kT_0)a(jT_0) \tag{57}$$

(which is studied in the next section). Fourier transformation of (56) yields

$$\hat{S}_x(f) = \frac{1}{T_0}|P(f)|^2 \sum_{k=-\infty}^{\infty} \tilde{R}_a(kT_0)e^{-i2\pi kT_0 f}. \tag{58}$$

If there exists a continuous time-series $a(t)$ such that (57) is equivalent to

$$\tilde{R}_a(kT_0) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} a\left(t + \frac{kT_0}{2}\right)a\left(t - \frac{kT_0}{2}\right) dt \equiv \hat{R}_a(kT_0), \tag{59}$$

then and only then can (58) be reexpressed as (exercise 6)

$$\hat{S}_x(f) = \left(\frac{1}{T_0}\right)^2 |P(f)|^2 \sum_{m=-\infty}^{\infty} \hat{S}_a\left(f - \frac{m}{T_0}\right). \tag{60}$$

Equivalence between (57) and (59) is violated if $a(t)$ contains additive, multiplicative, or other more subtle forms of periodicity with frequencies appropriately related to $1/T_0$ (which result in finite additive sine waves in the lag products). When (59) is violated, $\hat{S}_x$ in (60) contains additional terms[11] that are commonly unrecognized in probabilistic treatments [Gardner 1987a].

5. Sine Wave with Amplitude and Phase Modulation

Consider the time-series

$$x(t) = a(t)\cos[\omega_0 t + \phi(t)], \tag{61}$$

which can be reexpressed as

$$x(t) = c(t)\cos(\omega_0 t) - s(t)\sin(\omega_0 t), \tag{62}$$

where

$$\begin{aligned} c(t) &= a(t)\cos[\phi(t)] \\ s(t) &= a(t)\sin[\phi(t)]. \end{aligned} \tag{63}$$

It can be shown (exercise 7) that

$$\hat{R}_x(\tau) = \tfrac{1}{2}[\hat{R}_c(\tau) + \hat{R}_s(\tau)]\cos(\omega_0\tau) + \tfrac{1}{2}[\hat{R}_{cs}(\tau) - \hat{R}_{cs}(-\tau)]\sin(\omega_0\tau) \tag{64}$$

if and only if $\hat{R}_c$, $\hat{R}_s$, and $\hat{R}_{cs}$, where

$$\hat{R}_{cs}(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} c\left(t + \frac{\tau}{2}\right)s\left(t - \frac{\tau}{2}\right) dt, \tag{65}$$

exist and

$$\lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} y\left(t + \frac{\tau}{2}\right)z\left(t - \frac{\tau}{2}\right)\cos(2\omega_0 t + \theta)\, dt = 0 \tag{66}$$

for all τ for (1) $y = z = c$ and $\theta = 0$, (2) $y = z = s$ and $\theta = 0$, and (3) $y = c$, $z = s$, and $\theta = \pi/2$, which holds if and only if c and s are free from certain forms of periodicity associated with the frequency f_0 (namely, finite additive sine

[11] These additional terms are studied in Part II.

waves in the lag products). When (66) is violated, $\hat{R}_x$ contains additional terms[12] that are commonly unrecognized in probabilistic treatments [Gardner 1987a]. For example, when $c(t)$ or $s(t)$ contain certain forms of periodicity (which can arise in common types of modulation), the fact that (64) is invalid reveals that *Rice's representation* for band-pass time-series [Appendix 3-1] exhibits properties that differ from those that are well known [Gardner 1987b]. It follows from (64) (exercise 7) that when (66) holds,

$$\hat{S}_x(f) = \tfrac{1}{4}[\hat{S}_c(f - f_0) + \hat{S}_c(f + f_0) + \hat{S}_s(f - f_0) + \hat{S}_s(f + f_0)] + \tfrac{1}{2}\,\mathrm{Im}\{\hat{S}_{cs}(f - f_0) - \hat{S}_{cs}(f + f_0)\}, \tag{67}$$

for which $\mathrm{Im}\{\cdot\}$ denotes *imaginary part*, and

$$\hat{S}_{cs}(\cdot) \triangleq F\{\hat{R}_{cs}(\cdot)\}. \tag{68}$$

The functions $\hat{R}_{cs}$ and $\hat{S}_{cs}$ are called the *limit cross correlation* and *limit cross spectrum*, respectively, and are studied in Chapter 7.

E. TIME-SAMPLING AND ALIASING

An idealized statistical spectrum for time-sampled data can be obtained by analogy with the procedure described in Section C—that is, by time-averaging the periodogram, and then letting the spectral resolution width, $\Delta f = 1/T$, become infinitesimal as follows:

$$\bar{S}_x(f) \triangleq \lim_{T\to\infty} \lim_{K\to\infty} \frac{1}{K} \sum_{n=-(K-1)/2}^{(K-1)/2} \bar{S}_{x_T}(t + nT_s, f), \tag{69a}$$

where $\bar{S}_{x_T}(t, f)$ is the time-variant periodogram for discrete-time data,

$$\bar{S}_{x_T}(t, f) \triangleq \frac{1}{N} |\bar{X}_T(t, f))|^2, \tag{69b}$$

and $\bar{X}_T(t, f)$ is the time-variant finite-time complex spectrum for discrete-time data, defined by (53) in Chapter 2. In (69) $N = 1 + T/T_s$ is the total number of time-samples transformed to obtain $\bar{X}_T(t, f)$, and $K = \Delta t/T$ is the number of periodograms averaged. Substitution of (61) in Chapter 2 for $\bar{X}_T(t, f)$ into (69) yields

$$\bar{S}_x(f) = \frac{1}{T_s} \lim_{T\to\infty} \lim_{K\to\infty} \frac{N-1}{N} \frac{1}{K} \sum_{n=-(K-1)/2}^{(K-1)/2} \left[\sum_{m=-\infty}^{\infty} \frac{1}{T} \left| X_T\left(t + nT_s, f - \frac{m}{T_s}\right) \right|^2 + \sum_{\substack{m,p=-\infty \\ m\neq p}}^{\infty} \frac{1}{T} X_T\left(t + nT_s, f - \frac{m}{T_s}\right) X_T^*\left(t + nT_s, f - \frac{p}{T_s}\right) \right]. \tag{70}$$

The limit of the second term (double sum with $m \neq p$) is an idealized spectral correlation and, as explained in Part II, is nonzero only for an underlying phenomenon that exhibits periodic time variation (such that the lag products contain finite additive sine waves). Thus, for a constant phenomenon only the first term

[12] These additional terms are studied in Part II.

remains, and this term is recognized as a sum of translated limit spectra

$$\bar{S}_x(f) = \frac{1}{T_s} \sum_{m=-\infty}^{\infty} \hat{S}_x\left(f - \frac{m}{T_s}\right). \tag{71}$$

This follows from the fact that for a constant phenomenon, the limit discrete-time (hopped) average and the limit continuous-time average of a periodogram yield the same result (see Part II). It can be seen from (71) that, analogous to the situation for the complex spectrum $X_T(t, f)$ described in Section H in Chapter 2, the limit spectrum of the continuous-time data $\hat{S}_x(f)$ can be recovered from the limit spectrum of the time-sampled data $\bar{S}_x(f)$ if the sampling rate exceeds twice the bandwidth B, where

$$\hat{S}_x(f) = 0, \qquad |f| \geqslant B. \tag{72}$$

Otherwise aliasing effects prevent recovery. (As shown in Part II, Chapter 11, another type of aliasing error that can result in additional terms in (71)—due to the second term in (70)—occurs for periodic phenomena [Gardner 1987a].)

Also analogous to results obtained in Section C, the limit spectrum for time-sampled data can also be obtained by frequency-smoothing the periodogram

$$\bar{S}_x(f) \triangleq \lim_{\Delta f \to 0} \lim_{T \to \infty} \frac{F_s}{\Delta f} \sum_{m=-(\Delta f/F_s - 1)/2}^{(\Delta f/F_s - 1)/2} \bar{S}_{x_T}(t, f + mF_s), \tag{73}$$

where $F_s = 1/NT_s$ is the frequency-sampling increment and $N = 1 + T/T_s$. Although a continuous-frequency average yields the same result (for a constant phenomenon), the discrete-frequency average used here is more appropriate when digital implementations are to be used for spectral analysis, as described in Chapter 6. In the inner limit, the frequency-smoothing window has fixed width Δf, and therefore the number $M = \Delta f/F_s = N\Delta f T_s$ of frequency-samples averaged in (73) approaches infinity as $N \to \infty$ and therefore as $T \to \infty$.

It follows from (71) that the limit spectrum for time-sampled data is given in terms of the limit autocorrelation for time-sampled data by the Fourier-series transform

$$\bar{S}_x(f) = \sum_{q=-\infty}^{\infty} \bar{R}_x(qT_s) e^{-i2\pi q T_s f}. \tag{74}$$

This can be verified as follows (exercise 11):

$$\sum_{q=-\infty}^{\infty} \bar{R}_x(qT_s) e^{-i2\pi q T_s f} = \int_{-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \delta(\tau - qT_s) \hat{R}_x(\tau) e^{-i2\pi f \tau} \, d\tau \tag{75}$$

$$= \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{T_s} \delta\left(\nu - \frac{m}{T_s}\right) \hat{S}_x(f - \nu) \, d\nu \tag{76}$$

$$= \frac{1}{T_s} \sum_{m=-\infty}^{\infty} \hat{S}_x\left(f - \frac{m}{T_s}\right) = \bar{S}_x(f). \tag{77}$$

Equation (75) follows from the fact that

$$\bar{R}_x(qT_s) = \hat{R}_x(qT_s) \tag{78}$$

for a constant phenomenon, where $\tilde{R}_x(\tau)$ is the discrete-time limit autocorrelation

$$\tilde{R}_x(\tau) \triangleq \lim_{K\to\infty} \frac{1}{K} \sum_{n=-(K-1)/2}^{(K-1)/2} x(nT_s + \tau)x(nT_s), \tag{79}$$

and τ is an integer multiple of T_s. As explained in Part II, if the phenomenon exhibits periodic time variation (such that the lag products contain finite additive sine waves) with frequencies appropriately related to the sampling increment T_s, then the right member of (78) contains additional terms corresponding to the second term in (70).

F. TIME-SERIES MODELS

Time-series models play a fundamental role in the class of spectrum estimation techniques called *parametric methods*, which are described in Chapter 9. Three particularly important discrete-time models are the moving average model, the autoregressive model, and their combination, the autoregressive moving average model. The limit spectra for these three models are derived in this section. Let us begin with the counterpart for discrete time,

$$\tilde{S}_y(f) = |\tilde{H}(f)|^2 \tilde{S}_x(f), \tag{80}$$

of the limit-spectrum relation (28) for filters

$$y(n) = h(n) \otimes x(n). \tag{81}$$

For convenience, the sampling increment is taken to be unity, $T_s = 1$. Analogous to the definition (5) in Chapter 1 of continuous-time convolution, the discrete-time convolution (81) is defined by

$$y(n) = x(n) \otimes h(n) \triangleq \sum_{m=-\infty}^{\infty} h(n - m)x(m) \tag{82a}$$

$$= \sum_{k=-\infty}^{\infty} h(k)x(n - k). \tag{82b}$$

The transfer function $\tilde{H}(f)$ of the filter is given by the Fourier-series transform of the discrete-impulse response $h(n)$,

$$\tilde{H}(f) = \sum_{n=-\infty}^{\infty} h(n)e^{-i2\pi nf}. \tag{83}$$

The relation (80) can be derived by Fourier-series transformation of the limit-autocorrelation relation

$$\tilde{R}_y(q) = \tilde{R}_x(q) \otimes \tilde{r}_h(q), \tag{84}$$

and application of the convolution theorem for the Fourier-series transform, for example,

$$\sum_{q=-\infty}^{\infty} \tilde{r}_h(q)e^{-i2\pi qf} = |\tilde{H}(f)|^2, \tag{85}$$

where $\bar{r}_h$ is the discrete-time finite autocorrelation sequence

$$\bar{r}_h(q) = h(q) \otimes h(-q). \tag{86}$$

The relation (84) can be derived by substitution of (82b) into (79) (with x replaced by y) (exercise 12).

1. The Moving Average Model

Consider the time-series model

$$y(n) = b_0x(n) + b_1x(n-1) + \cdots + b_Lx(n-L), \tag{87}$$

in which the excitation $x(n)$ is purely random (white noise),

$$\tilde{R}_x(q) = N_0\delta_q, \tag{88}$$

where δ_q is the discrete impulse (Kronecker delta) defined by

$$\delta_q \triangleq \begin{cases} 1, & q = 0 \\ 0, & q \neq 0. \end{cases} \tag{89}$$

It is a simple matter (exercise 13) to show that the transfer function for the linear transformation described by (87) is

$$\tilde{H}(f) = b_0 + b_1e^{-i2\pi f} + b_2(e^{-i2\pi f})^2 + \cdots + b_L(e^{-i2\pi f})^L. \tag{90}$$

Also, it easily follows from (74) and (88) that the limit spectrum of the excitation is

$$\tilde{S}_x(f) = N_0. \tag{91}$$

Thus, the limit-spectrum relation (80), together with (90) and (91), yields the result

$$\tilde{S}_y(f) = N_0 \left| \sum_{q=0}^{L} b_q(e^{-i2\pi f})^q \right|^2. \tag{92}$$

Since the current value of the response $y(n)$ is simply a moving average (sliding average) of the current and L past values of the excitation, with averaging coefficients $b_0, b_1, b_2, \ldots, b_L$, then (92) is called the limit spectrum of a *moving average (MA) model of order L*. This terminology was originally introduced in [Wold 1938].

2. The Autoregressive Model

Consider the time-series model

$$y(n) = -a_1y(n-1) - a_2y(n-2) - \cdots - a_My(n-M) + x(n), \tag{93}$$

in which the excitation is white noise (88). It is easily shown (exercise 13) that the transfer function for the linear time-invariant transformation described by (93) is

$$\tilde{H}(f) = \frac{1}{1 + a_1e^{-i2\pi f} + a_2(e^{-i2\pi f})^2 + \cdots + a_M(e^{-i2\pi f})^M}. \tag{94}$$

Thus, (80), (91), and (94) yield the result

$$\tilde{S}_y(f) = \frac{N_0}{\left|1 + \sum_{p=1}^{M} a_p(e^{-i2\pi f})^p\right|^2}. \tag{95}$$

Since the response $y(n)$ regresses on itself in the sense that its current value is a linear combination of its past M values plus the current excitation, then (95) is called the limit-spectrum of an *autoregressive* (AR) *model of order M*. This terminology was originally introduced in [Wold 1938].

3. The ARMA Model

Consider a model that is a combination of the MA and AR models (87) and (93),

$$\begin{aligned} y(n) = &-a_1y(n-1) - a_2y(n-2) - \cdots - a_My(n-M) \\ &+ b_0x(n) + b_1x(n-1) + b_2x(n-2) + \cdots + b_Lx(n-L), \end{aligned} \tag{96}$$

in which the excitation $x(n)$ is white noise, (88). It is easily shown (exercise 13) that the transfer function for the linear time-invariant transformation described by (96) is

$$\tilde{H}(f) = \frac{\sum_{q=0}^{L} b_q(e^{-i2\pi f})^q}{1 + \sum_{p=1}^{M} a_p(e^{-i2\pi f})^p}, \tag{97}$$

from which the formula for the limit spectrum of an ARMA model of order (M, L) follows directly.

As explained in Chapter 9, MA and AR models arise naturally from certain approaches to measurement of statistical spectra. Because of the particular importance of the AR model, it is studied in detail in Chapter 9. The continuous-time counterparts of these discrete-time models are studied in exercise 9.

G. STATISTICAL INFERENCE

In the spectral analysis problem considered in Section A, the spectral features of interest as described by $|H(f)|$ can be measured only approximately with a *finite* amount, Δt, of data, as indicated by approximation (5). But as shown in subsequent sections, $|H(f)|$ can be determined exactly in the abstract limit as $\Delta t \to \infty$, as indicated by (28) and (38). This reveals that exact description of the spectral characteristics of a phenomenon requires an abstract mathematical model for the data, namely, the limit spectrum. We have thus arrived at the point of view of *statistical inference*, which is that an abstract mathematical model is the desired result that can be only approximately discovered (inferred) with the use of a finite amount of data. From the point of view of statistical inference, the object of statistical spectral analysis is *spectrum estimation*, by

which is meant estimation of the limit spectrum.[13] Succinctly stated, the classical spectrum estimation design problem is: given a finite amount[14] Δt of data, determine the *best* value of spectral resolution Δf to obtain the *best* estimate of $\hat{S}_x$. This involves a trade-off between maximizing spectral resolution, which corresponds to minimizing Δf, and minimizing the *degree of randomness* or *variability* (described in Chapter 5), which in turn corresponds to maximizing Δf in order to maximize the product $\Delta t \Delta f$.

The statistical-inference or spectrum-estimation interpretation given here to spectral analysis is unconventional in that it does not rely on probabilistic concepts. However, it can be put into a probabilistic framework by reinterpreting infinite time averages as ensemble averages (expectations) via H.O.A. Wold's isomorphism (defined in Chapter 1, Section B). This is done in Chapter 5, where the notion of degree of randomness is quantified in terms of a coefficient of variation that is shown to be inversely proportional to the resolution product $\Delta t \Delta f$.

As a matter of fact, the classical spectrum estimation design problem is more involved than suggested by the preceding succinct statement, because the shape as well as the width Δf of the effective spectral window should be optimized in order to minimize the undesirable *spectral leakage effect*. This effect is described in the next chapter, and the design problem that simultaneously takes into account resolution, leakage, and degree of randomness is explained in Chapters 5 and 6.

Before proceeding, a few words about the notion of degree of randomness will be helpful to tide us over until the subject is taken up in Chapter 5. It has been shown in this chapter that randomly fluctuating (in both t and f) statistical spectra, such as $S_{x_{\Delta t}}(t, f)_{\Delta f}$ and $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$, converge in the limit ($\Delta t \to \infty$, $\Delta f \to 0$) to the nonrandom limit spectrum $\hat{S}_x(f)$ if the limit autocorrelation $\hat{R}_x(\tau)$ exists, which is necessary for a constant phenomenon. The degree of randomness or variability of a statistical spectrum can be interpreted as the degree to which the statistical spectrum varies from one point in time to another. If the underlying phenomenon is indeed constant, as hypothesized in Part I, then fluctuation with time of the statistical spectrum must be attributed to random effects. It is shown in Chapter 5 that the time-averaged squared difference between statistical spectra measured at two different times separated by an amount T_0, for example, is approximately inversely proportional to the resolution product $\Delta t \Delta f$ (for sufficiently small Δf and sufficiently large $\Delta t \Delta f$) for all $T_0 > \Delta t$. Also, the time-averaged squared difference between the statistical spectrum and the nonrandom limit

[13] In the literature, the terms *spectrum analysis* and *spectral estimation* are often used in place of the terms *spectral analysis* and *spectrum estimation*, which are used in this book. The latter two terms are more appropriate since we are not concerned with analysis of a spectrum but rather with analysis of data into spectral components, and we are not concerned with estimation using spectral methods but rather with estimation of a spectrum. Nevertheless, because of the long-standing tradition of referring to spectral analysis instruments as *spectrum analyzers*, this term is used in this book in place of the term *spectral analyzers*.

[14] The actual amount of data needed to average a periodogram of length $T = 1/\Delta f$ over an interval of length Δt is $\Delta t + 1/\Delta f$, but this is closely approximated by Δt for $\Delta t \Delta f >> 1$.

spectrum behaves in the same way. Thus, this temporal mean-square measure of the degree of randomness of a statistical spectrum reveals that the degree of randomness is made low (or the *reliability* is made high) by making the resolution product $\Delta t \Delta f$ large.

H. SUMMARY

In Section A, the problem of measuring the parameters of a resonance phenomenon from the randomly resonant response to random excitation is considered in order to motivate consideration of averaging methods for reducing random effects. It is explained that from the point of view adopted here, we focus attention on the phenomenon that gives rise to random data rather than on the data itself, and we apply averaging methods to the nonstatistical spectrum (periodogram) of the data to obtain a statistical spectrum in which the random effects in the data that mask the spectral influences from the phenomenon are reduced. In Section B, a profound fundamental result establishing an equivalence between time-smoothed and frequency-smoothed periodograms is developed. This equivalence reveals that the periodogram of the data-tapering window in a temporally smoothed periodogram of the tapered data is an effective spectral smoothing window in an equivalent spectrally smoothed periodogram of the untapered data. Then in Section C, the idealized limiting form of the statistical spectrum with $\Delta t \to \infty$ and $\Delta f \to 0$ (in this order) is shown to be simply the Fourier transform of the limit autocorrelation. This characterization of the *limit spectrum*, called the *Wiener relation*, is used to derive the limit-spectrum relation for filters (28), which in turn is used to establish the interpretation of the limit spectrum as a spectral density of time-averaged power.

In Section C, several signal and noise models are introduced, and their limit spectra are calculated. Then in Section D, the definition of the limit spectrum is adapted to discrete-time data by simply replacing the Fourier transform with the Fourier-series transform introduced in Section H of Chapter 2. A spectral aliasing formula relating the limit spectra of a waveform and its time-samples is derived. In Section F, three basic time-series models for discrete-time data are introduced. These are the MA, AR, and ARMA models. Formulas for the limit spectra for these models are derived in terms of the parameters of the models.

Finally in Section G, it is pointed out that the arguments presented in the beginning of this chapter have led us to the point of view of statistical inference, which is that an abstract mathematical model—the limit spectrum in this case—is the desired result that can be only approximately discovered (inferred) with the use of a finite amount of data. Thus statistical spectral analysis is typically called *spectrum estimation*. This section ends with a brief discussion of the dependence of the degree of randomness or variability of a statistical spectrum on the resolution product $\Delta t \Delta f$.

In Appendix 3-1, *Rice's representation* is derived. This provides a means for representing band-pass waveforms in terms of low-pass waveforms. Then the limit spectra for the low-pass representors are characterized in terms of the

limit spectrum of the band-pass waveform, and vice versa. In Appendix 3-2, the problem of detecting the presence of a random signal in additive random noise is considered, and the central role played by the periodogram and the limit spectrum is revealed.

EXERCISES

1. **(a)** In order to accomplish the verification of (20)–(22) for the limit spectrum, (13) is reexpressed as

$$S_{x_T}(t, f)_{\Delta t} = \int_{-T_*}^{T_*} [R_{x_T}(t, \tau) \otimes u_{\Delta t}(t)] e^{-i2\pi f\tau}\, d\tau. \tag{98}$$

Use the result of exercise 10, Chapter 2, and an argument parallel to that in exercise 8, Chapter 2, to argue that for any arbitrarily small positive ϵ_1, there exists a Δt_1 (which depends on t) such that

$$R_{x_T}(t, \tau) \otimes u_{\Delta t}(t) = \frac{1}{T} R_{x_{\Delta t}}(t, \tau) r_a(\tau) + \epsilon(t, \tau) \tag{99a}$$

for all $\Delta t \geq \Delta t_1$, where

$$|\epsilon(t, \tau)| \leq \epsilon_1, \qquad |\tau| \leq T_*. \tag{99b}$$

Then use (98)–(99) and (16) to show (using $T = 1/\Delta f$) that

$$|S_{x_{1/\Delta f}}(t, f)_{\Delta t} - S_{x_{\Delta t}}(t, f) \otimes [\Delta f |A_{\Delta f}(f)|^2]| \leq 2T_* \epsilon_1 \tag{100}$$

for all f and all $\Delta t \geq \Delta t_1$, which reveals that this difference between a temporally smoothed spectrum and a spectrally smoothed spectrum can be made as small as desired for each t by choosing Δt sufficiently large.

(b) To proceed with the verification of (20)–(22) consider (16), which is repeated here:

$$S_{x_{\Delta t}}(t, f) \otimes [\Delta f |A_{\Delta f}(f)|^2] = \frac{1}{T} \int_{-T_*}^{T_*} R_{x_{\Delta t}}(t, \tau) r_a(\tau) e^{-i2\pi f\tau}\, d\tau. \tag{101}$$

Use an argument parallel to that in exercise 8, Chapter 2, to argue that for any arbitrarily small positive ϵ_2, there exists a Δt_2 (which depends on t) such that

$$R_{x_{\Delta t}}(t, \tau) = \hat{R}_x(\tau) + \epsilon(t, \tau) \tag{102a}$$

for all $\Delta t \geq \Delta t_2$, where

$$|\epsilon(t, \tau)| \leq \epsilon_2, \qquad |\tau| \leq T_*. \tag{102b}$$

Use (101)–(102) to show that

$$[S_{x_{\Delta t}}(t, f) \otimes [\Delta f |A_{\Delta f}(f)|^2] - \hat{S}_x(f) \otimes [\Delta f |A_{\Delta f}(f)|^2]| \leq 2\epsilon_2 \Delta f T_* r_a(0) \tag{103}$$

for all f and all $\Delta t \geq \Delta t_2$, which reveals that this difference between a spectrally smoothed statistical spectrum and a spectrally smoothed limit spectrum can be made as small as desired for each t by choosing Δt sufficiently large. Hence, substitution of (17) into (103) yields

$$\lim_{\Delta t \to \infty} S_{x_{\Delta t}}(t, f)_{\Delta f} = \hat{S}_x(f) \otimes [\Delta f |A_{\Delta f}(f)|^2] \tag{104}$$

(since we can let $\epsilon_2 \to 0$ by letting $\Delta t \to \infty$), which establishes part of (22).

(c) To complete the verification of (20)–(22), use the results of (a) and (b) to prove that for any arbitrarily small ϵ, there exists a Δt_3 (which depends on t) such that

$$|S_{x_{1/\Delta f}}(t,f)_{\Delta t} - \hat{S}_x(f) \otimes [\Delta f |A_{\Delta f}(f)|^2]| < \epsilon \tag{105}$$

for all f and all $\Delta t \geq \Delta t_3$. *Hint:* Choose any ϵ_1 and ϵ_2 such that $\epsilon = 2T_*\epsilon_1 + 2\epsilon_2 \Delta f T_* r_a(0)$, choose $\Delta t_3 = \max\{\Delta t_1, \Delta t_2\}$, and use the fact that if $|a - b| < \delta_1$ and $|b - c| < \delta_2$, then $|a - c| < \delta_1 + \delta_2$. It follows from (105) that

$$\lim_{\Delta t \to \infty} S_{x_{1/\Delta f}}(t,f)_{\Delta t} = \hat{S}_x(f) \otimes [\Delta f |A_{\Delta f}(f)|^2], \tag{106}$$

which establishes the other part of (22).

2. Norbert Wiener [Wiener 1930, p. 134] has shown that

$$\text{l.i.m.}_{\Delta t \to \infty} \int_{-\infty}^{\infty} R_{x_{\Delta t}}(t,\tau) w_{1/\Delta f}(\tau) e^{-i2\pi f\tau}\, d\tau = \text{l.i.m.}_{T_* \to \infty} \int_{-T_*}^{T_*} \hat{R}_x(\tau) w_{1/\Delta f}(\tau) e^{-i2\pi f\tau}\, d\tau, \tag{107}$$

where $w_{1/\Delta f}$ is the sinc window and l.i.m. denotes *limit in mean square*, which means that the integrals (over f) of the squares of the differences between the sequence of functions (of f) and the limit function converge to zero. Use (107) and the identity

$$\int_{-\infty}^{\infty} \hat{R}_x(\tau) w_{1/\Delta f}(\tau) e^{-i2\pi f\tau}\, d\tau \equiv \text{l.i.m.}_{T_* \to \infty} \int_{-T_*}^{T_*} \hat{R}_x(\tau) w_{1/\Delta f}(\tau) e^{-i2\pi f\tau}\, d\tau$$

to verify that

$$\text{l.i.m.}_{\Delta t \to \infty} S_{x_{\Delta t}}(t,f)_{\Delta f} = \hat{S}_x(f) \otimes u_{\Delta f}(f), \tag{108}$$

where

$$S_{x_{\Delta t}}(t,f)_{\Delta f} \triangleq S_{x_{\Delta t}}(t,f) \otimes u_{\Delta f}(f).$$

Use result (108) together with

$$\lim_{\Delta f \to 0} u_{\Delta f}(f) = \delta(f) \tag{109}$$

to prove that

$$\lim_{\Delta f \to 0} \text{l.i.m.}_{\Delta t \to \infty} S_{x_{\Delta t}}(t,f)_{\Delta f} = \hat{S}_x(f). \tag{110}$$

(The difference between this result, (108), and (104) is that the proof of (104) requires a correlation-tapering window with finite support, unlike $w_{1/\Delta f}$ in (107)–(108). Evidently, when the support is not finite only the weaker type of convergence, (108) (rather than (104)), can be established.)

3. Use (34) to derive (36) for the time-averaged power in a spectral band.

4. (a) Use (44), (45), and (47) to verify (46) for the limit autocorrelation of a sine wave in noise.
   (b) Verify (48).
   (c) Derive the spectral-line formula (49) from (46) and (48).

5. (a) Use (50), (51), and (53) to verify formula (52) for the limit autocorrelation of an amplitude-modulated sine wave. *Hint:* Use the trigonometric identity

$$\cos(\omega_0 t_1)\cos(\omega_0 t_2) = \tfrac{1}{2}\cos[\omega_0(t_1 - t_2)] + \tfrac{1}{2}\cos[\omega_0(t_1 + t_2)]. \tag{111}$$

   (b) Derive the limit spectrum (54) from (48) and (52).
   (c) What happens to the average power in $n(t)$ when it is multiplied by $\cos(\omega_0 t + \phi)$?

6. **(a)** Derive formula (56) for the limit autocorrelation of the pulse-amplitude modulated pulse train. *Hint:* For the sequence of values of T that approach infinity in the definition

$$\hat{R}_x(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x\left(t - \frac{\tau}{2}\right) dt$$

choose

$$T = (2M + 1)T_0$$

with $M = 1, 2, 3 \ldots$, and then decompose the integral over the interval $[-T/2, T/2]$ into the sum of integrals over the $2M + 1$ intervals $[(k - \frac{1}{2})T_0, (k + \frac{1}{2})T_0]$, $k = -M, -M + 1, \ldots, M - 1, M$, to obtain

$$\hat{R}_x(\tau) = \lim_{M\to\infty} \frac{1}{T_0} \frac{1}{2M + 1} \sum_{k=-M}^{M} \int_{-T_0/2}^{T_0/2} x\left(t + kT_0 + \frac{\tau}{2}\right) x\left(t + kT_0 - \frac{\tau}{2}\right) dt. \tag{112}$$

Then substitute

$$x\left(t + kT_0 + \frac{\tau}{2}\right) x\left(t + kT_0 - \frac{\tau}{2}\right)$$
$$= \sum_{j=-\infty}^{\infty} \sum_{j'=-\infty}^{\infty} a(jT_0)a(j'T_0)p\left[t + \frac{\tau}{2} - (j - k)T_0\right]p\left[t - \frac{\tau}{2} - (j' - k)T_0\right]$$

into (112), and introduce the change of variables

$$j - k = r \quad \text{and} \quad j' - k = s,$$

to obtain

$$\hat{R}_x(\tau) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \lim_{M\to\infty} \frac{1}{T_0} \frac{1}{2M + 1} \sum_{j'=-M-s}^{M-s} a[(j' + r - s)T_0]a(j'T_0)$$
$$\times \int_{-T_0/2}^{T_0/2} p\left(t + \frac{\tau}{2} - rT_0\right)p\left(t - \frac{\tau}{2} - sT_0\right) dt. \tag{113}$$

The limit $M \to \infty$ can now be evaluated using definition (57). Finally, the remaining double sum in r and s and integral over $[-T_0/2, T_0/2]$ can be converted to a single sum in q and an integral over $(-\infty, \infty)$ by introducing the change of variables $r - s = q$ and $t' = t - rT_0$.

(b) Verify the *Poisson sum formula*

$$\sum_{p=-\infty}^{\infty} g(pT)e^{-i2\pi pTf} = \frac{1}{T} \sum_{q=-\infty}^{\infty} G\left(f - \frac{q}{T}\right) \tag{114}$$

for any Fourier transformable function g and any sampling increment T. Do this by recognizing the periodicity of this function of f and representing the right member by its Fourier series (recall (22a) and (22b) in Appendix 1-1.)

(c) Use the Poisson sum formula and (59) to derive the limit spectrum (60) from the limit autocorrelation (56).

7. **(a)** Use (66) to verify formula (64) for the limit autocorrelation of an amplitude-and-phase-modulated sine wave. *Hint:* Use the trigonometric identity in exercise 5 together with the identities

$$\sin(\omega_0 t_1)\sin(\omega_0 t_2) = \tfrac{1}{2}\cos[\omega_0(t_1 - t_2)] - \tfrac{1}{2}\cos[\omega_0(t_1 + t_2)]$$
$$\cos(\omega_0 t_1)\sin(\omega_0 t_2) = \tfrac{1}{2}\sin[\omega_0(t_1 - t_2)] + \tfrac{1}{2}\sin[\omega_0(t_1 + t_2)]. \tag{115}$$

(b) Let $\phi(t) = \phi$, a constant, and show that

$$\hat{R}_c(\tau) = \cos^2(\phi)\hat{R}_a(\tau)$$

$$\hat{R}_s(\tau) = \sin^2(\phi)\hat{R}_a(\tau)$$

$$\hat{R}_{cs}(\tau) = \cos(\phi)\sin(\phi)\hat{R}_a(\tau).$$

Use this result to verify that (64) reduces to

$$\hat{R}_x(\tau) = \tfrac{1}{2}\hat{R}_a(\tau)\cos(\omega_0\tau)$$

when there is no phase modulation.

(c) Derive the limit spectrum (67) from the limit autocorrelation (64).

8. *First-order model:* Consider the time-series model

$$x(t) + (1/\alpha)\frac{dx(t)}{dt} = z(t), \tag{116}$$

where $z(t)$ is white noise.

(a) Verify that the impulse-response function of the linear time-invariant transformation of $z(t)$ to $x(t)$ is

$$h(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geqslant 0 \\ 0, & t < 0. \end{cases} \tag{117}$$

Hint: $dh(t)/dt = \alpha\delta(t) - \alpha^2 e^{-\alpha t}$ for $t \geqslant 0$.

(b) Verify that the finite autocorrelation for h is

$$r_h(\tau) = \frac{\alpha}{2}e^{-\alpha|\tau|}. \tag{118}$$

(c) Use relation (31) to determine the limit autocorrelation for $x(t)$.

(d) Use relation (28) to show that the limit spectrum for $x(t)$ is

$$\hat{S}_x(f) = \frac{\alpha^2}{\alpha^2 + (2\pi f)^2}, \tag{119}$$

which is called the *Lorenzian spectrum.*

9. *Pole-zero models:*

(a) Consider the time-series model for $x(t)$

$$x(t) + a_1\frac{dx(t)}{dt} + a_2\frac{d^2x(t)}{dt^2} + \cdots + a_M\frac{d^Mx(t)}{dt^M} = z(t), \tag{120}$$

for which $z(t)$ is white noise. It can be shown that $x(t)$ is an LTI transformation of $z(t)$. Determine the limit spectrum for $x(t)$. (120) is called an *all-pole model* because the transfer function (and $\hat{S}_x(f)$) is the reciprocal of a polynomial. This is the continuous-time counterpart of the discrete-time AR model (Section F). *Hint:* Determine the transfer function for this LTI transformation by the method illustrated in Appendix 1-1 (namely, let $z(t) = e^{i2\pi ft}$ and $x(t) = H(f)e^{i2\pi ft}$), and then use the limit spectrum relation (28).

(b) Consider the time-series model for $x(t)$,

$$x(t) = z(t) + b_1\frac{dz(t)}{dt} + b_2\frac{d^2z(t)}{dt^2} + \cdots + b_L\frac{d^Lz(t)}{dt^L}, \tag{121}$$

for which $z(t)$ is white noise. It can be shown that $x(t)$ is an LTI transformation of $z(t)$. Determine the limit spectrum for $x(t)$. (121) is called an *all-zero model*

because the transfer function (and $\hat{S}_x(f)$) is a polynomial. This is the continuous-time counterpart of the MA model.

(c) Show that the time-series $x(t)$ modeled by

$$x(t) + a_1\frac{dx(t)}{dt} + \cdots + a_M\frac{d^M x(t)}{dt^M} = z(t) + b_1\frac{dz(t)}{dt} + \cdots + b_L\frac{d^L z(t)}{dt^L}, \tag{122}$$

for which $z(t)$ is white noise, has a limit spectrum that is a rational function of f. (122) is called a *pole-zero model*. This is the continuous-time counterpart of the ARMA model.

(d) Determine an explicit pole-zero model (values of a's and b's in (122)) for a time-series with the limit spectrum

$$\hat{S}_x(f) = \frac{4 + (2\pi f)^2}{9 + (2\pi f)^2}.$$

Hint: $|\alpha + i2\pi f|^2 = \alpha^2 + (2\pi f)^2$ for real α.

(e) Determine the limit autocorrelation for the time-series model obtained in (d). *Hint:* Express $\hat{S}_x$ as a sum of two fractions, and use the result of exercise 8 and the fact that multiplication by $(2\pi f)^2$ corresponds to $-d^2/d\tau^2$.

10. *Resonance:* To illustrate an all-pole system of common interest, consider the single-degree-of-freedom mechanical system consisting of a mass, spring, and dashpot, as shown in Figure 3-5. Assume that the mass is subject to a force excitation $x(t)$, producing a displacement response $y(t)$. From Newton's laws, the differential equation of motion describing the response of this system is given by

$$m\frac{d^2y(t)}{dt^2} + c\frac{dy(t)}{dt} + ky(t) = x(t). \tag{123}$$

(a) Show that the transfer function of the system is given by

$$H(f) = \frac{1/k}{1 - (f/f_n)^2 + i2\xi f/f_n}, \tag{124a}$$

where

$$\xi \triangleq \frac{c}{2\sqrt{km}} \tag{124b}$$

$$f_n \triangleq \frac{1}{2\pi}\sqrt{\frac{k}{m}}. \tag{124c}$$

The parameter ξ in (124b) is called the *damping ratio* of the system and describes the system damping as a fractional portion of critical damping c_c. If the mass is displaced from its neutral position and released, c_c is that value of damping for which the mass will just return with the greatest speed to its neutral position without oscillation; for the system in Figure 3-5, $c_c = 2\sqrt{km}$. The parameter f_n in (124c) is called the *undamped natural frequency* of the system. If the system had no damping and the mass were displaced from its neutral position and released, the system would perpetually oscillate at the frequency f_n. The magnitude of the transfer function is given by

$$|H(f)| = \frac{1/k}{\sqrt{[1 - (f/f_n)^2]^2 + (2\xi f/f_n)^2}}. \tag{125}$$

Graphs of both this magnitude and the phase of $H(f)$ for various values of ξ are

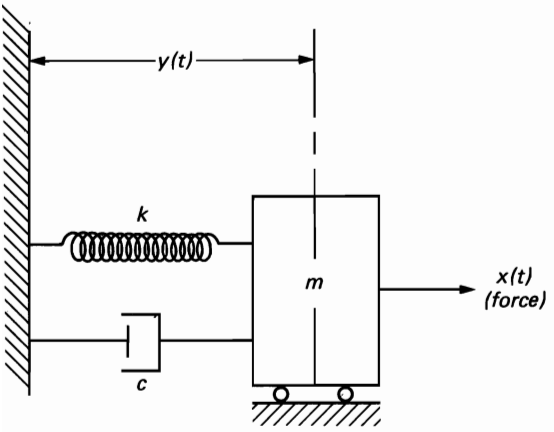


Figure 3-5 A single-degree-of-freedom mechanical system.

shown in Figure 3-6. Observe that there is a peak at some frequency less than f_n for all cases where $\xi \leq 1/\sqrt{2}$. The frequency at which this peak occurs is called the *resonance frequency* of the system, and it approaches f_n as ξ decreases.

(b) Show (by minimizing the denominator of $|H(f)|$) that the resonance frequency, denoted by f_0, is given by

$$f_0 = f_n\sqrt{1 - 2\xi^2}, \qquad \xi^2 \leq \tfrac{1}{2}, \tag{126}$$

and that the peak value, which occurs at the resonance frequency, is given by

$$|H(f_0)| = \frac{1/k}{2\xi\sqrt{1 - \xi^2}}, \qquad \xi^2 \leq \tfrac{1}{2}. \tag{127}$$

(c) Some physical systems have very small values of damping such that $\xi << 1$. For example, mechanical structures often have damping ratios of $\xi < 0.05$. Hence it is common in practice to find physical systems with transfer functions that display very sharp peaks. Such systems in effect appear to be narrow band-pass filters, and their bandwidth is commonly measured in terms of the *half-power point bandwidth*, defined by

$$B = f_2 - f_1, \quad \text{where} \quad |H(f_1)|^2 = |H(f_2)|^2 = \tfrac{1}{2}|H(f_0)|^2. \tag{128}$$

Show that for the case where the damping ratio is relatively small, the bandwidth is given by

$$B \cong 2\xi f_0. \tag{129}$$

(d) Verify that the impulse-response function of this system (for $\xi < 1$) is given by

$$h(\tau) = Ae^{-2\pi f_n \xi \tau}\sin(2\pi f_d \tau), \qquad \tau > 0, \tag{130a}$$

where

$$f_d = f_n\sqrt{1 - \xi^2} = \textit{damped natural frequency} \tag{130b}$$

$$A = \frac{2\pi f_n^2}{kf_d} = \frac{2\pi f_n}{k\sqrt{1 - \xi^2}}. \tag{130c}$$

It follows from (b) and (d) that the autocorrelation and spectral density functions

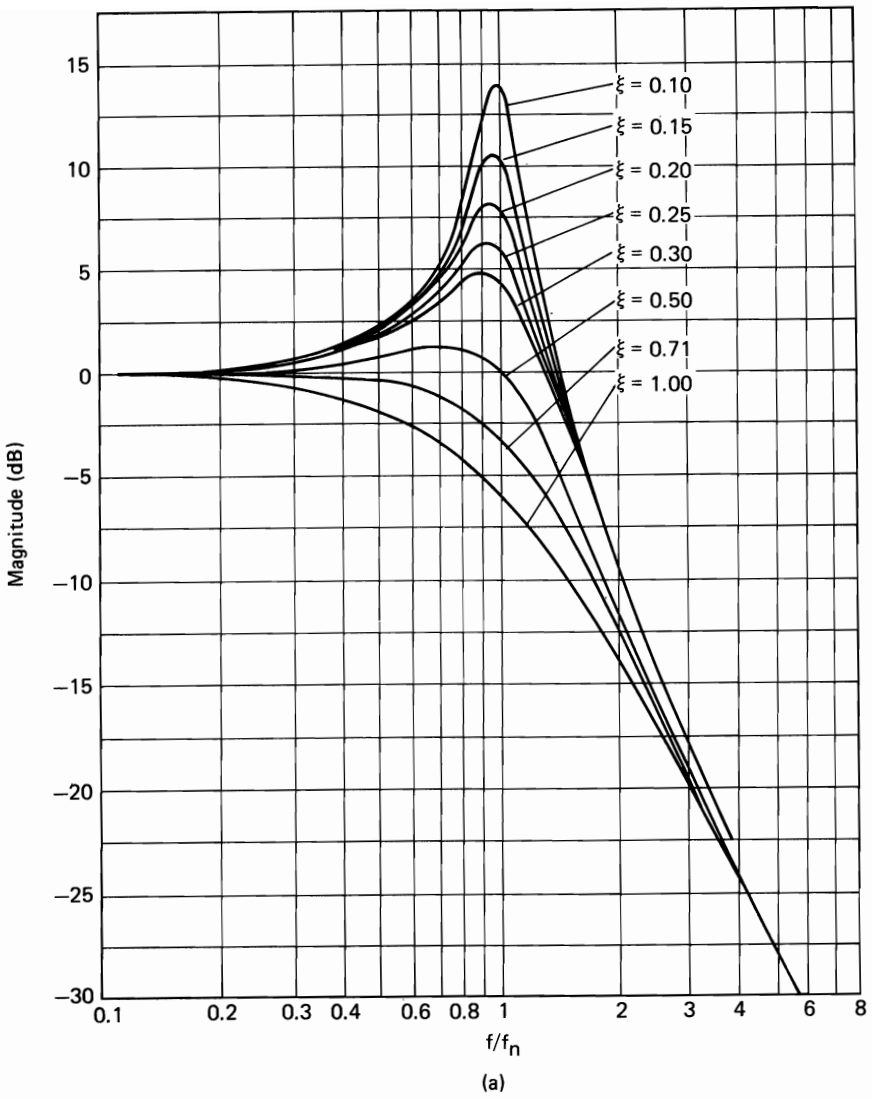


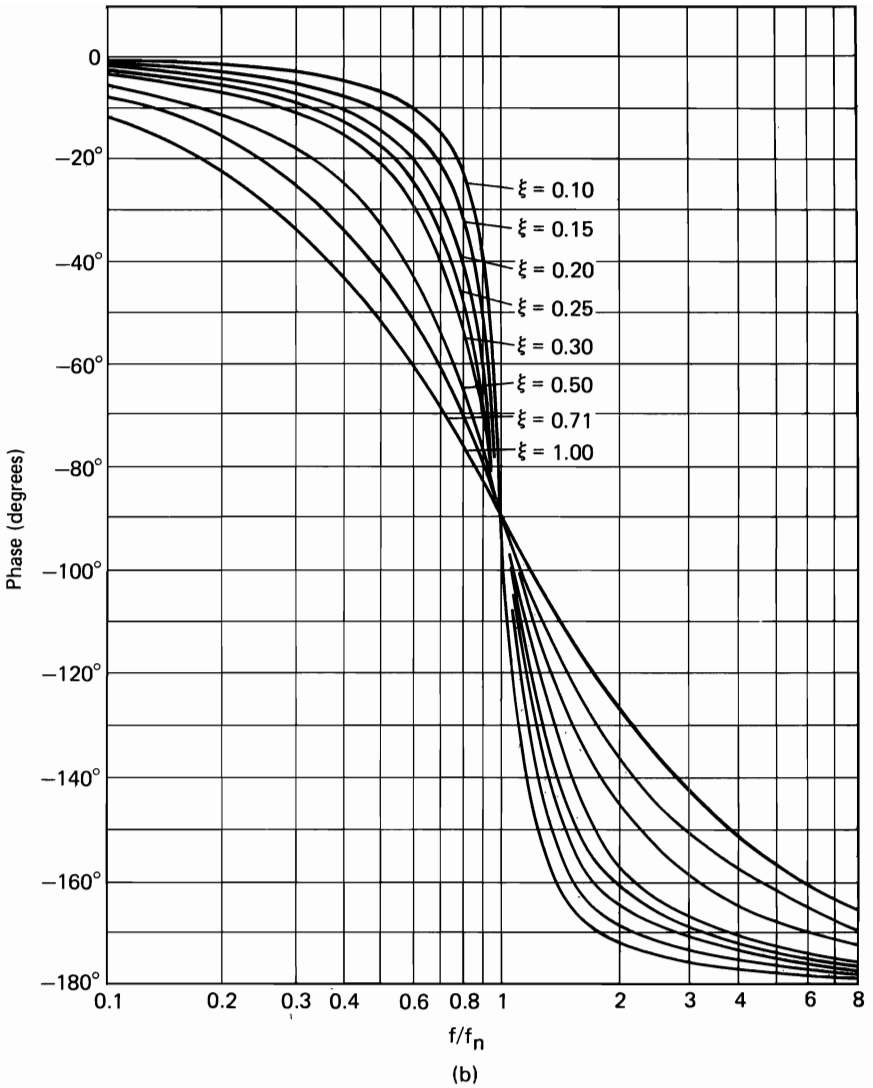
Figure 3-6 Transfer function (124) of the single-degree-of-freedom system shown in Figure 3-5. (a) Magnitude.

of the system response $y(t)$ are given by

$$\hat{R}_y(\tau) = A^2 \int_0^\infty \int_0^\infty e^{-2\pi f_n \xi(\alpha+\beta)} \sin(2\pi f_d \alpha)\sin(2\pi f_d \beta)\hat{R}_x(\tau + \alpha - \beta)d\alpha\, d\beta, \tag{131}$$

$$\hat{S}_y(f) = \frac{\hat{S}_x(f)/k^2}{[1 - (f/f_n)^2]^2 + (2\xi f/f_n)^2}. \tag{132}$$

(e) To illustrate a specific case, assume that the input $x(t)$ is white noise, and show



(b)

Figure 3-6 (continued) (b) Phase.

that

$$\hat{R}_y(\tau) = \frac{N_0 \pi f_n e^{-2\pi f_n \xi |\tau|}}{2\xi k^2} \left[\cos(2\pi f_d \tau) + \frac{\xi}{\sqrt{1-\xi^2}} \sin(2\pi f_d |\tau|) \right] \tag{133}$$

$$\cong \frac{N_0 \pi f_n e^{-2\pi f_n \xi |\tau|}}{2\xi k^2} \cos(2\pi f_n \tau), \qquad \text{for } \xi << 1, \tag{134}$$

$$\hat{S}_y(f) = \frac{N_0/k^2}{[1 - (f/f_n)^2]^2 + (2\xi f/f_n)^2}. \tag{135}$$

For the common situation where $\xi << 1$, (133)–(135) provide reasonably accurate

approximations of the system response as long as $\hat{S}_x(f) \cong N_0$ over the frequency range $(1 - 6\xi)f_n < f < (1 + 6\xi)f_n$ and $\hat{S}_x(f)$ decays outside this range.

(f) As an electrical analog of the mechanical system shown in Figure 3-5, consider the resistive-inductive-capacitive circuit shown in Figure 3-7. From Kirchhoff's laws the differential equation describing the voltage across the capacitor is given by

$$LC\frac{d^2y(t)}{dt^2} + RC\frac{dy(t)}{dt} + y(t) = x(t). \tag{136}$$

Determine the parameters k, f_n, f_0, f_d, ξ, and B in terms of R, L, and C. Then relate the quality factor

$$Q \triangleq \frac{1}{R}\sqrt{\frac{L}{C}} \tag{137}$$

to the damping ratio (124b).

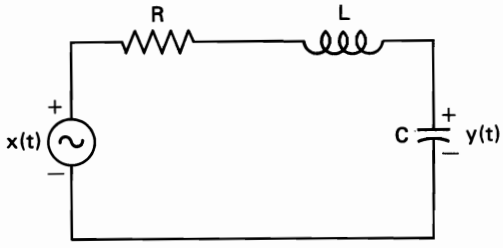


Figure 3-7 Electrical analog of mechanical system shown in Figure 3-5.

(g) An especially graphic example of the utility of statistical spectral analysis in connection with resonant systems is given by the problem of designing the suspension of vehicles such as automobiles, motorcycles, and aircraft (takeoff and landing gear) to minimize the discomfort and destructive effects, such as fatigue, of random vibrations induced by rough road surface. A simplified model of a suspension system for a single wheel is a single-degree-of-freedom system consisting of the mass of the vehicle, a spring, and a shock absorber (dashpot) and is therefore characterized by the model developed in this exercise. By characterizing the random excitation of the system due to imperfect road surface in terms of the measured power spectral density of the randomly fluctuating elevation of a typical road surface, the power spectral density of the vibrations of the vehicle can be determined by multiplication with the squared magnitude of the system transfer function. Then the natural frequency and damping factor can be adjusted to minimize undesirable effects such as peaks at critical frequencies or simply to minimize the total power of the vibrations. As a simple example, suppose that in some normalized units, the mass is $m = 1$ and that it is desired to make the natural frequency $f_n = 1$ Hz and the damping ratio $\xi = \frac{1}{4}$. Determine appropriate values for the spring constant k and the damping constant c. (Consult Figure 3-6 to see the resultant frequency response of the system.)

11. Verify (76) (see (60) in Chapter 2).

12. Derive the limit-autocorrelation relation for filters (84) from (79) and (82b) (see exercise 12(a) in Chapter 2).

13. (a) The transfer function $\tilde{H}(f)$ of a linear time-invariant discrete-time system can always be obtained by determining the response $y(n) = \tilde{H}(f)e^{i2\pi fn}$ to the excitation

$x(n) = e^{i2\pi fn}$ and then dividing $y(n)/x(n)$. Use this technique to derive (90) from (87) for the MA model.

(b) For the AR model, use the technique in (a) to derive (94) from (93).

(c) For the ARMA model, use the technique in (a) to derive (97) from (96).

14. *Second-order AR model:* The discrete-time analog of the second-order all-pole continuous-time model studied in exercise 10 has transfer function given by

$$\begin{aligned} \tilde{H}(f) &= [1 + a_1 e^{-i2\pi f} + a_2 e^{-i4\pi f}]^{-1} \\ &= [(1 - \alpha e^{-i2\pi f})(1 - \alpha^* e^{-i2\pi f})]^{-1}. \end{aligned} \tag{138}$$

Determine α in terms of a_1 and a_2, and discuss the relationships between the values of a_1 and a_2 and the magnitude and phase of α. Note that $|\alpha| > 1$ yields an unstable system, and $|\alpha| = 1$ yields a marginally stable system, whereas $|\alpha| < 1$ yields a stable system. The impulse-response (inverse Fourier-series transform of $\tilde{H}(f)$) for this system is given by (assuming complex α)

$$h(n) = kr^n \sin(\omega_d n + \phi), \qquad n \geq 0, \tag{139a}$$

where

$$r = \sqrt{a_2} \tag{139b}$$

$$\omega_d = \cos^{-1}(-a_1/2\sqrt{a_2}) = \phi. \tag{139c}$$

Determine an effective damping ratio by analogy with the impulse-response in exercise 10.

15. **(a)** Use trigonometric identities to prove that if $u(t)$ and $v(t)$ are given by (21) in Appendix 3-1 for *any* $y(t)$, then $x(t)$ is given by (23) in Appendix 3-1.

(b) Let $x(t)$ be given by (23) in Appendix 3-1, where $u(t)$ and $v(t)$ are bandlimited to $[-B, B]$ with $B < f_0$, and show that $u(t)$ and $v(t)$ are *uniquely* (for all f_0) determined by (21) in Appendix 3-1, where $y(t)$ is the Hilbert transform of $x(t)$. *Hint:* Use the fact that when $u(t)$ and $v(t)$ are bandlimited as described, then the Hilbert transforms of the two terms

$$\begin{matrix} u(t)\cos(2\pi f_0 t), \\ v(t)\sin(2\pi f_0 t) \end{matrix} \tag{140a}$$

are simply

$$\begin{matrix} u(t)\sin(2\pi f_0 t), \\ -v(t)\cos(2\pi f_0 t), \end{matrix} \tag{140b}$$

respectively.

16. **(a)** Use formulas (24) (in Appendix 3-1) and (67) to derive (25) (in Appendix 3-1) from (21) (in Appendix 3-1).

(b) Use the same approach as in (a) to derive (30) from (21) in Appendix 3-1.

17. Derive (7) in Appendix 3-2 from (4) and (5) in Appendix 3-2. *Hint:* See the note in exercise 9(c) in Chapter 4 regarding changes of variables in double integrals in order to reexpress (4)–(5) as

$$y(t) = \int_{-T}^{T} \hat{R}_s(\tau) R_{x_T}(t, \tau)\, d\tau.$$

Then use Parseval's relation, the Wiener relation, and the periodogram-correlogram relation.

18. Acceptable detectability of random signals typically requires a value of deflection d that is no smaller than 10. For a random signal with bandwidth of $B = 10{,}000$ Hz and an effective SNR of $S_0/N_0 = 1$, how long of an observation time T is needed?

If the bandwidth is doubled, how much smaller can T be? If the SNR is doubled, how much smaller can T be? (Refer to Appendix 3-2.)

19. Consider two detection problems, one involving a broadband random signal (i.e., no sine wave component) and the other involving a pure sine wave signal (i.e., no random component). Compare the dependence of deflection on the observation time T for these two problems. (Refer to Appendix 3-2.)

20. Determine the power spectral density of the response voltage for each of the circuits shown in Figure 3-8 for a white-noise excitation voltage with unity intensity, $N_0 = 1$. Also determine the rms value of the response voltage, $\sqrt{R_v(0)}$. *Hint:* Use the results of exercise 8.

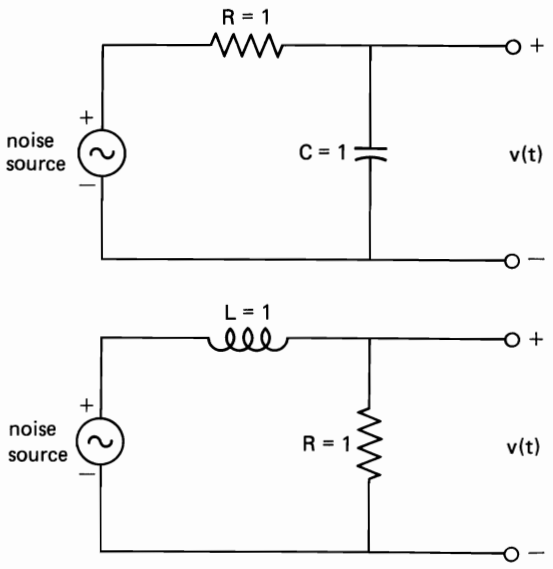


Figure 3-8 Two electrical circuits.

21. It is desired to approximate a time-averaged periodogram of tapered data using a frequency-smoothed periodogram of untapered data. The data-tapering window is a triangle with total base width of 10 μs and unity height. The periodogram is averaged over a time interval of length 1 ms. Determine the data-segment length and appropriate smoothing window for the spectrally smoothed periodogram. Would you expect this to be an accurate approximation? Why? Can this approximation be made exact (in principle)? If so, how?

APPENDIX 3-1
Band-Pass Time-Series

Many signals encountered in signal-processing systems are of the band-pass type; that is, their spectrum is concentrated about some frequency removed from zero. For example, the response of a resonant system such as a band-pass filter is a

band-pass signal. Also, a major purpose for many types of signal modulations is to convert a low-pass signal into a band-pass signal in an appropriate frequency range in order to minimize attenuation and distortion due to propagation through a particular medium. Furthermore, a basic operation in the typical spectrum analyzer is that of decomposing the time-series to be analyzed into narrow-band band-pass signals whose passbands partition the frequency range of interest. Thus, the study of specific spectrum analyzers taken up in the next chapter requires a thorough understanding of band-pass signals. However, for purposes of analysis and computation, it is often more convenient to work with low-pass representations of band-pass signals. These representations are the subject of this appendix.

Let us begin by considering the Fourier transform of a finite-energy function. Since every real finite-energy function $x(t)$ has a Fourier transform $X(f)$ with Hermitian (conjugate) symmetry,

$$X(-f) = X^*(f), \tag{1}$$

then the function $x(t)$ can be completely recovered from the positive-frequency portion of its transform, call it $(\frac{1}{2})\Psi(f)$,

$$\left(\frac{1}{2}\right)\Psi(f) \triangleq \begin{cases} X(f), & f > 0 \\ 0, & f < 0, \end{cases} \tag{2}$$

as described pictorially in Figure 3-1-1(a) and (b). Specifically, the transform of the function $x(t)$ is recovered by

$$X(f) = \left(\frac{1}{2}\right)[\Psi(f) + \Psi^*(-f)]. \tag{3}$$

Furthermore, $X(f)$ and, therefore, $x(t)$ can be recovered from the function $\Gamma(f)$ obtained by translating $\Psi(f)$ by any amount, say f_0,

$$\Gamma(f) \triangleq \Psi(f + f_0), \tag{4}$$

as depicted in Figure 3-1-1(c). In fact, $x(t)$ can be recovered from the inverse transform of $\Gamma(f)$, which is denoted by $\gamma(t)$. Since $x(t)$ can always be recovered from $\gamma(t)$, then $\gamma(t)$ can be interpreted as a representation for $x(t)$. This type of representation is particularly interesting when $x(t)$ is a bandpass function, as depicted in Figure 3-1-2(a), because then $\gamma(t)$ is a low-pass function if f_0 is chosen to be within the passband of $x(t)$, as depicted in Figure 3-1-2(b).

Although $x(t)$ is taken to be a real function, its representation $\gamma(t)$ is in general complex because its Fourier transform $\Gamma(f)$ does not in general exhibit Hermitian symmetry. But $\gamma(t)$ can, of course, be represented in terms of two real functions, namely, its real and imaginary parts, say $u(t)$ and $v(t)$:

$$\gamma(t) = u(t) + iv(t). \tag{5}$$

By letting $w(t)$ be the imaginary process

$$w(t) = iv(t), \tag{6}$$

the transform of $\gamma(t)$ is represented by the sum of transforms

$$\Gamma(f) = U(f) + W(f), \tag{7}$$

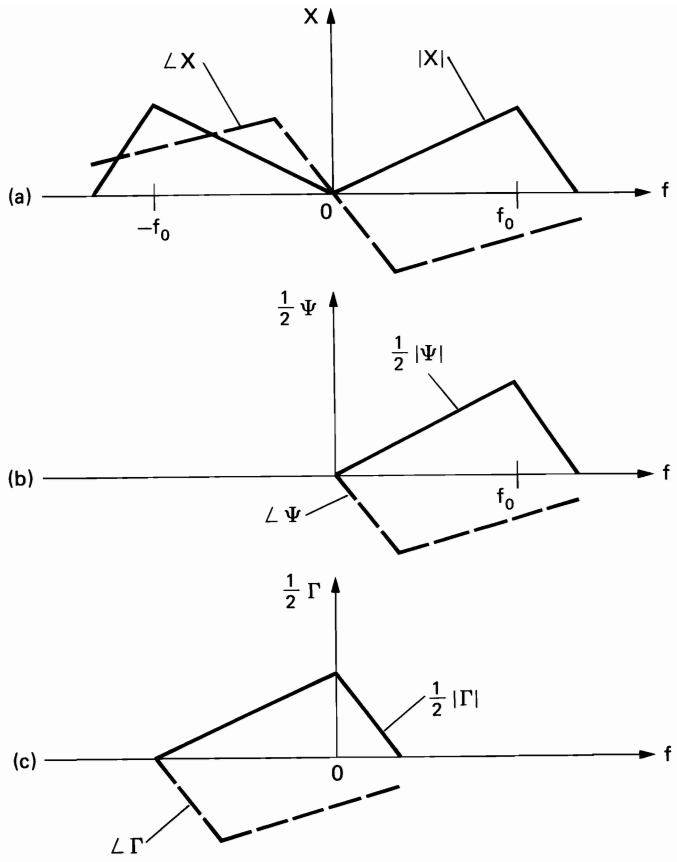


Figure 3-1-1 Equivalent signal representations in the frequency domain.

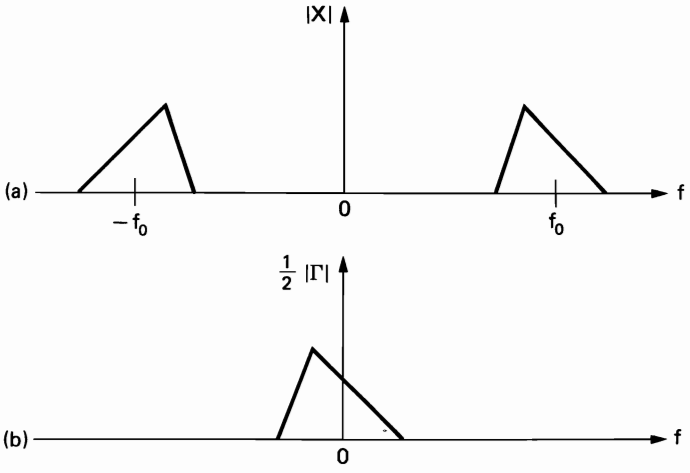


Figure 3-1-2 Low-pass representation of a band-pass signal in the frequency domain.

and it can be shown that these components exhibit the symmetries

$$\begin{aligned} U(-f) &= U^*(f) \\ W(-f) &= -W^*(f). \end{aligned} \tag{8}$$

Thus, $U(f)$ has even conjugate symmetry, but $W(f)$ has odd conjugate symmetry. Hence $\Gamma(f)$, which has no conjugate symmetry in general, is represented by the sum of its even and odd conjugate-symmetric parts, $U(f)$ and $W(f)$. This is illustrated in Figure 3-1-3 for the special case in which Γ, U, and W are real.

Now, let us characterize the operations just described for obtaining the real low-pass representors $u(t)$ and $v(t)$ from a real band-pass function $x(t)$ in terms of operations defined in the time-domain. It follows from (2) that

$$\Psi(f) = X(f) + \text{sgn}(f)X(f), \tag{9}$$

where

$$\text{sgn}(f) \triangleq \begin{cases} +1, & f > 0 \\ -1, & f < 0. \end{cases} \tag{10}$$

Furthermore, the function

$$H(f) \triangleq (-i)\text{sgn}(f) \tag{11}$$

is the transfer function corresponding to the impulse-response function

$$h(t) = \frac{1}{\pi t}, \tag{12}$$

and the transformation

$$y(t) = h(t) \otimes x(t) \tag{13}$$

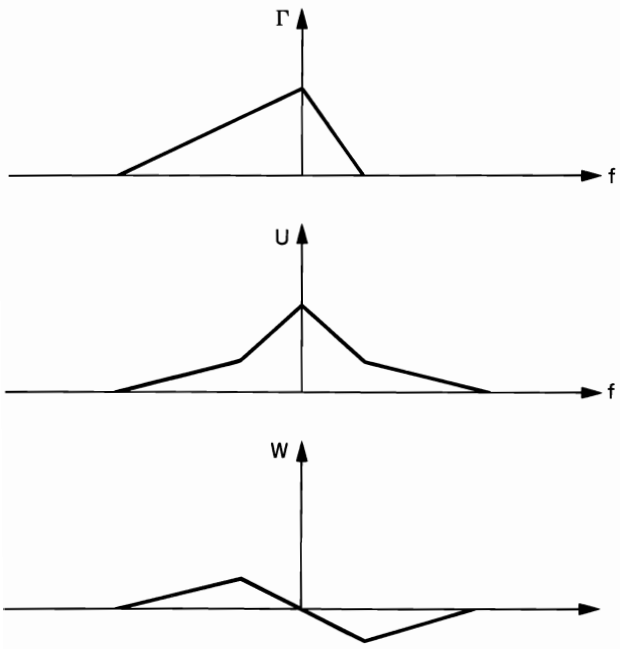


Figure 3-1-3 Even and odd symmetric components U and W of the asymmetric representation Γ (only real parts shown).

is called the *Hilbert transformation*. It follows from the facts

$$|H(f)| \equiv 1$$

$$\arg\{H(f)\} = \begin{cases} -\dfrac{\pi}{2}, & f > 0 \\ \dfrac{\pi}{2}, & f < 0 \end{cases} \tag{14}$$

that this transformation simply shifts the phase of every positive (negative) frequency component by $-(+)\pi/2$ radians $(-(+)90$ degrees). It follows from (9)–(13) that

$$\Psi(f) = X(f) + iY(f) \tag{15}$$

and, therefore,

$$\psi(t) = x(t) + iy(t). \tag{16}$$

Furthermore, it follows from (4) that

$$\gamma(t) = \psi(t)e^{-i2\pi f_0 t}. \tag{17}$$

Equations (16) and (17) yield the result

$$\gamma(t) = [x(t) + iy(t)]e^{-i2\pi f_0 t}, \tag{18}$$

where $y(t)$ is the Hilbert transform of $x(t)$. Now, to obtain the real and imaginary parts of $\gamma(t)$, Euler's formula,

$$e^{-i2\pi f_0 t} = \cos(2\pi f_0 t) - i\sin(2\pi f_0 t), \tag{19}$$

is used in (18) to obtain

$$\begin{aligned} \gamma(t) = {} & [x(t)\cos(2\pi f_0 t) + y(t)\sin(2\pi f_0 t)] \\ & + i[y(t)\cos(2\pi f_0 t) - x(t)\sin(2\pi f_0 t)]. \end{aligned} \tag{20}$$

Thus,

$$\begin{aligned} u(t) &= x(t)\cos(2\pi f_0 t) + y(t)\sin(2\pi f_0 t) \\ v(t) &= y(t)\cos(2\pi f_0 t) - x(t)\sin(2\pi f_0 t). \end{aligned} \tag{21}$$

In order to recover $x(t)$ from $\gamma(t)$, (18) can be used to obtain

$$x(t) = \mathrm{Re}\{\gamma(t)e^{i2\pi f_0 t}\}, \tag{22}$$

where $\mathrm{Re}\{\cdot\}$ denotes the *real part*. Substitution of (5) into (22) yields

$$x(t) = u(t)\cos(2\pi f_0 t) - v(t)\sin(2\pi f_0 t). \tag{23}$$

Formulas (21) and (23) comprise the desired explicit representation for a real band-pass function $x(t)$ in terms of two real low-pass functions $u(t)$ and $v(t)$, which are called the *in-phase* and *quadrature components* of the *complex representation* $\gamma(t)$. This is called *Rice's representation* in honor of Stephen O. Rice's pioneering work in signal representation theory [Rice 1944, 1945, 1948]. It follows from (2) and (4) that if $x(t)$ is bandlimited to the band $[f_0 - B, f_0 + B]$ (and its negative-frequency image), then $\gamma(t)$—and, therefore, $u(t)$ and $v(t)$—are bandlimited to the band $[-B, B]$. Furthermore, given *any* finite-energy function in the form (23), for which $u(t)$ and $v(t)$ are bandlimited to $[-B, B]$ for

$B < f_0$, it can be shown (exercise 15) that $u(t)$ and $v(t)$ are *uniquely* determined by (21).

Although Rice's representation is derived here for transient finite-energy time-series, which are directly Fourier transformable, it is valid for persistent finite-power time-series as well. This raises the question of how the limit spectral densities of $u(t)$, $v(t)$, and $x(t)$ are related. It follows (exercise 16) from Part 5 of Section *D*, Chapter 3, using

$$\hat{S}_{xy}(f) = -\hat{S}_{yx}(f) = i \operatorname{sgn}(f)\hat{S}_x(f) \tag{24a}$$

$$\hat{S}_y(f) = \hat{S}_x(f) \tag{24b}$$

for $x(t)$ and its Hilbert transform $y(t)$ (see Chapter 7 for cross spectra), that

$$\hat{S}_u(f) = \hat{S}_v(f) = \hat{S}_x(f + f_0)U(f_0 + f) + \hat{S}_x(f - f_0)U(f_0 - f), \tag{25}$$

where U is the unit step function

$$U(f) = \begin{cases} 1, & f > 0 \\ 0, & f < 0. \end{cases} \tag{26}$$

Formula (25) is valid provided that $x(t)$ does not contain certain forms of periodicity associated with the frequency f_0, as mentioned in Part 5 of Section D, Chapter 3. Otherwise (25) must be generalized as explained in Part II, Chapter 11 (for example, $\hat{S}_u$ and $\hat{S}_v$ need not be identical). It follows from (25) that if $x(t)$ has limit spectral density that is bandlimited to $[f_0 - B, f_0 + B]$,

$$\hat{S}_x(f) = 0, \qquad ||f| - f_0| \geq B, \tag{27}$$

then $u(t)$ and $v(t)$ have limit spectra that are bandlimited to $[-B, B]$,

$$\hat{S}_u(f) = \hat{S}_v(f) = 0, \qquad |f| \geq B. \tag{28}$$

Furthermore, if $B \leq f_0$, then

$$\hat{S}_u(f) = \hat{S}_v(f) = \begin{cases} \hat{S}_x(f - f_0) + \hat{S}_x(f + f_0), & |f| < f_0 \\ 0, & |f| \geq f_0. \end{cases} \tag{29}$$

It can also be shown that the limit cross-spectral densities for $u(t)$ and $v(t)$ are given by

$$\hat{S}_{uv}(f) = -\hat{S}_{vu}(f) = i[\hat{S}_x(f + f_0)U(f_0 + f) - \hat{S}_x(f - f_0)U(f_0 - f)], \tag{30}$$

and if $B \leq f_0$, then

$$\hat{S}_{uv}(f) = -\hat{S}_{vu}(f) = \begin{cases} i[\hat{S}_x(f + f_0) - \hat{S}_x(f - f_0)], & |f| < f_0 \\ 0, & |f| \geq f_0. \end{cases} \tag{31}$$

Again, formula (30) is valid with the earlier-mentioned provision regarding periodicity. The PSD for $x(t)$ can be expressed in terms of the PSDs for $u(t)$ and $v(t)$ by direct substitution of $u(t)$ and $v(t)$ for $c(t)$ and $s(t)$, respectively, into (67).

APPENDIX 3-2
Random-Signal Detection

In problems of detecting the presence of random signals buried in noise, the concepts and methods of spectral analysis are particularly useful for both detector design and performance analysis. To illustrate this, we consider the problem of detecting (in each interval $[t - T/2, t + T/2]$ indexed by t) the presence of a random signal s hidden in noise n, using the data

$$x(u) = \delta s(u) + n(u), \qquad -\frac{T}{2} \leqslant u - t \leqslant \frac{T}{2}, \tag{1}$$

for which δ is the indicator variable

$$\delta = \begin{cases} 1, & \text{signal present} \\ 0, & \text{signal absent.} \end{cases} \tag{2}$$

A well-known signal detector is the threshold test

$$y(t) \underset{no}{\overset{yes}{\gtrless}} \gamma, \tag{3}$$

for which $y(t)$ is the quadratic detection statistic

$$y(t) = \int_{t-T/2}^{t+T/2} \int_{t-T/2}^{t+T/2} k(u, v)x(u)x(v)\, du\, dv, \tag{4}$$

and γ is the threshold level, which is set to provide the desired *false alarm rate* and/or *detection rate*. If the kernel k that specifies the detection statistic is the limit autocorrelation function (scaled by T),

$$k(u, v) = \frac{1}{T}\hat{R}_s(u - v), \tag{5}$$

then this detector is optimum for a signal in additive white Gaussian[15] noise in the sense[16] of maximizing the following measure of signal-to-noise ratio (SNR), called *deflection* [Gardner 1985],

$$d = \left[\frac{\left|\langle y(t)|\delta = 1\rangle - \langle y(t)|\delta = 0\rangle\right|^2}{\langle |y(t)|^2|\delta = 0\rangle - \left|\langle y(t)|\delta = 0\rangle\right|^2}\right]^{1/2}, \tag{6}$$

where $\langle\cdot\rangle$ denotes limit time-average (see (22) in Chapter 1). Substitution of (5) into (4) and application of a change of variables and Parseval's relation yields the revealing characterization

$$y(t) = \int_{-\infty}^{\infty} \hat{S}_s(f)S_{x_T}(t, f)\, df, \tag{7}$$

[15] A Gaussian time-series is defined in Chapter 5.

[16] This detector is also optimum in other senses—see [Gardner 1985]—but maximization of deflection is especially appropriate for weak signals.

for which $\hat{S}_s(f)$ is the limit spectrum of the signal s and $S_{x_T}(t, f)$ is the time-variant periodogram of the data x on the interval $[t - T/2, t + T/2]$. Thus, the detector (3)–(5) measures the periodogram and correlates (multiplies and integrates) it with a stored replica of the limit spectrum of the signal to be detected. The detector decides *yes*, the signal is present, if and only if this correlation exceeds the threshold level γ.

The maximum value of the deflection (6) is given by

$$d = \left\{ \frac{1}{2N_0^2} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} [\hat{R}_s(u - v)]^2 \, du \, dv \right\}^{1/2} \tag{8}$$

[Gardner 1985], which can be reexpressed as

$$d = \left\{ \frac{T}{2N_0^2} \int_{-T}^{T} [\hat{R}_s(\tau)]^2 (1 - |\tau|/T) \, d\tau \right\}^{1/2}. \tag{9}$$

Application of Parseval's relation and the convolution theorem to (9) yields the alternative formula in terms of the limit spectrum

$$d = \left\{ \frac{T^2}{2N_0^2} \int_{-\infty}^{\infty} [\hat{S}_s(f) \otimes \hat{S}_s(f)] \left[\frac{\sin(\pi f T)}{\pi f T} \right]^2 df \right\}^{1/2}. \tag{10}$$

For a long observation time, that is, T much greater than the width of the limit autocorrelation of the signal (which is necessary for good detection performance for weak signals), (9) is closely approximated by

$$d \cong \left\{ \frac{T}{2N_0^2} \int_{-\infty}^{\infty} [\hat{R}_s(\tau)]^2 \, d\tau \right\}^{1/2}. \tag{11}$$

Application of Parseval's relation to (11) yields the alternative close approximation

$$d \cong \left\{ \frac{T}{2N_0^2} \int_{-\infty}^{\infty} [\hat{S}_s(f)]^2 \, df \right\}^{1/2}. \tag{12}$$

Thus, performance can be determined solely from the limit spectra of the signal to be detected and the noise.

These characterizations, (7) and (12), of the optimum detector and its performance can be generalized for nonwhite noise (see [Gardner 1985]).

As an example of the deflection formula (12), if $s(t)$ is a band-pass signal with effective spectral intensity S_0 and (positive-frequency) bandwidth B, then

$$\int_{-\infty}^{\infty} [\hat{S}_s(f)]^2 \, df = 2BS_0^2, \tag{13}$$

and

$$d \cong \frac{S_0}{N_0} \sqrt{BT}. \tag{14}$$

Thus, the deflection is determined by the input signal-to-noise ratio and the duration-bandwidth product.

As another example, if $s(t)$ consists of a random component with spectral density $\hat{S}(f)$ plus a sine wave with frequency f_0 and amplitude a, then

$$\hat{S}_s(f) = \frac{a^2}{4} [\delta(f + f_0) + \delta(f - f_0)] + \hat{S}(f), \tag{15}$$

and therefore

$$\hat{S}_s(f) \otimes \hat{S}_s(f) = \frac{a^4}{16}[\delta(f + 2f_0) + \delta(f - 2f_0) + 2\delta(f)] + \frac{a^2}{2}[\hat{S}(f + f_0) + \hat{S}(f - f_0)] + \hat{S}(f) \otimes \hat{S}(f). \tag{16}$$

Substitution of (16) into (10) yields

$$d = \left[\frac{T^2}{2N_0^2}\left\{\left(\frac{a^4}{8}\right)\left[1 + \left(\frac{\sin 2\pi f_0 T}{2\pi f_0 T}\right)^2\right] + a^2\int_{-\infty}^{\infty} \hat{S}(f + f_0)\left(\frac{\sin \pi f T}{\pi f T}\right)^2 df + \int_{-\infty}^{\infty} \hat{S}(f) \otimes \hat{S}(f)\left(\frac{\sin \pi f T}{\pi f T}\right)^2 df\right\}\right]^{1/2}. \tag{17}$$

If many cycles of the sine wave are observed, $f_0 T >> 1$, and if T is much greater than the width of the limit autocorrelation of the random component (so that $1/T$ resolves $\hat{S}(f)$), then (17) yields the close approximation

$$d \cong \left[\frac{T}{2N_0^2}\left\{\frac{a^4}{8}T + a^2\hat{S}(f_0) + \int_{-\infty}^{\infty} [\hat{S}(f)]^2 df\right\}\right]^{1/2}. \tag{18}$$

If the random component of the signal is band-pass with effective spectral intensity S_0 and (positive-frequency) bandwidth B, then (18) yields

$$d \cong \left[\frac{T}{N_0^2}\left(\frac{a^4}{16}T + \frac{a^2}{2}S_0 + BS_0^2\right)\right]^{1/2}. \tag{19}$$

It follows that the sine wave dominates the performance and

$$d \cong \frac{Ta^2}{4N_0} \tag{20}$$

when

$$\frac{a^2}{4} > S_0\sqrt{\frac{B}{T}}. \tag{21}$$

In this case for which the signal contains an additive sine wave component, the detection statistic (7) decomposes into two terms that can be determined by substitution of (15) into (7),

$$y(t) = \frac{a^2}{2}S_{x_T}(t, f_0) + \int_{-\infty}^{\infty} \hat{S}(f)S_{x_T}(t, f)\, df. \tag{22}$$

If the random component of the signal is band-pass with effective spectral intensity S_0, center frequency f_*, and (positive-frequency) bandwidth B, then (22) yields

$$y(t) = \frac{a^2}{2}S_{x_T}(t, f_0) + 2S_0\int_{f_*-B/2}^{f_*+B/2} S_{x_T}(t, f)\, df. \tag{23}$$

If follows that for a sufficiently large sine wave component, the single-frequency sample $S_{x_T}(t, f_0)$ of the periodogram tends to dominate the detection statistic.

As discussed in Chapter 9, the periodogram is also an optimum detection statistic for a sine wave (in white Guassian noise) with unknown amplitude, phase, and frequency. Nevertheless, it is also explained in Chapter 9 that the periodogram can be considerably improved upon for detection of multiple sine waves in noise when the observation time T is short by using *parametric spectral analysis* methods.

4

ANALOG METHODS

In Chapter 3 it is established that a statistical spectrum can be obtained from a periodogram by either the temporal-smoothing method or the spectral-smoothing method and that these two methods yield approximately the same statistical spectrum when a substantial amount of smoothing is done ($\Delta t \Delta f >> 1$). In this chapter it is shown that a variety of alternative methods yield approximately or exactly the same statistical spectrum, but it is emphasized that differences can be quite important in practice. These alternatives include the methods of temporal or spectral smoothing of the pseudospectrum, hopped temporal smoothing of the periodogram and pseudospectrum, Fourier transformation of the tapered correlogram and finite-average autocorrelation, real and complex wave-analysis, real and complex demodulation, and swept-frequency-demodulation wave-analysis. The methods are referred to as *analog methods* because they process the continuous-time waveforms directly. The actual form of implementation of such methods can employ conventional resistive-capacitive-inductive passive electrical circuits, more modern active electrical circuits, microwave devices, various optical, acousto-optical, and electro-acoustical devices, or mechanical devices. The particular form of implementation depends on available technology, economic constraints, environmental constraints (e.g., temperature, mechanical vibration, humidity, and so on), and frequency ranges of interest. Chapter 6 presents *digital methods*, so called because they process discrete-time data and because digital electrical forms of implementation (both hardware and software) are the primary means for discrete-time processing.

Before proceeding, a word about notation is in order. As in Chapter 3, $S_{x_T}(t, f)_{\Delta t}$ and $S_{x_T}(t, f)_{\Delta f}$ denote time- and frequency-smoothed spectra obtained from the periodogram $S_{x_T}(t, f)$,

$$S_{x_T}(t,f)_{\Delta t} \triangleq S_{x_T}(t,f) \otimes g_{\Delta t}(t) \tag{1a}$$

$$S_{x_T}(t,f)_{\Delta f} \triangleq S_{x_T}(t,f) \otimes H_{\Delta f}(f), \tag{2a}$$

for which $g_{\Delta t}$ and $H_{\Delta f}$ are arbitrary pulselike smoothing functions with width parameters Δt and Δf, respectively. Similarly, $S_x(t, f)_{T,\Delta t}$ and $S_x(t, f)_{T,\Delta f}$ denote time- and frequency-smoothed spectra obtained from the pseudospectrum, $S_x(t, f)_T$,

$$S_x(t,f)_{T,\Delta t} \triangleq \mathrm{S}_x(t,f)_T \otimes g_{\Delta t}(t) \tag{3a}$$

$$S_x(t,f)_{T,\Delta f} \triangleq \mathrm{S}_x(t,f)_T \otimes H_{\Delta f}(f). \tag{4a}$$

As in Chapter 3, there is no explicit denotation of the data-tapering window a_T used in the periodogram. In (1a) and (3a), T is the reciprocal of the spectral resolution width, $T = 1/\Delta f$, and in (2a) and (4a), T is the temporal resolution width, $T = \Delta t$; thus, we have the following four distinct combinations of subscripts in Δt and Δf,

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} = S_{x_{1/\Delta f}}(t,f) \otimes g_{\Delta t}(t) \tag{1b}$$

$$S_{x_{\Delta t}}(t,f)_{\Delta f} = S_{x_{\Delta t}}(t,f) \otimes H_{\Delta f}(f) \tag{2b}$$

$$S_x(t,f)_{1/\Delta f,\Delta t} = S_x(t,f)_{1/\Delta f} \otimes g_{\Delta t}(t) \tag{3b}$$

$$S_x(t,f)_{\Delta t,\Delta f} = S_x(t,f)_{\Delta t} \otimes H_{\Delta f}(f). \tag{4b}$$

In all four cases, the rightmost subscript denotes the smoothing parameter, and the type of smoothing is identified by the dimension of the parameter (T and Δt have temporal dimension, and $1/T$ and Δf have spectral dimension).

A. TEMPORAL AND SPECTRAL SMOOTHING

In the derivation (11)–(17) in Chapter 3, approximation (15) can be deleted to obtain the alternative approximation

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} \cong S_x(t,f)_{\Delta t,\Delta f}, \qquad \Delta t \Delta f >> 1, \tag{5a}$$

for which $g_{\Delta t}$ in (1b) is the rectangle

$$g_{\Delta t}(t) = u_{\Delta t}(t) \tag{5b}$$

and $H_{\Delta f}$ in (4b) is given by

$$H_{\Delta f}(f) = \Delta f |A_{\Delta f}(f)|^2 \tag{5c}$$

$$A_{\Delta f}(\cdot) = F\{a_{1/\Delta f}(\cdot)\}, \tag{5d}$$

where $a_{1/\Delta f}$ is the data-tapering window in (1b) (see (3) and (11) in Chapter 2).

Motivated by approximation (5a), which was obtained by manipulating (1b),

we consider the following manipulation of (3b):

$$S_x(t,f)_{T,\Delta t} \triangleq \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} S_x(t-u,f)_T\,du \tag{6}$$

$$= \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2}\int_{-T}^{T} R_x(t-u,\tau)_T e^{-i2\pi f\tau}\,d\tau\,du \tag{7}$$

$$= \int_{-T}^{T}\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} R_x(t-u,\tau)_T\,du\,e^{-i2\pi f\tau}\,d\tau \tag{8}$$

$$= \int_{-T}^{T}\frac{1}{T}\int_{-T/2}^{T/2} R_x(t-u,\tau)_{\Delta t}\,2Tu_{2T}(\tau)\,du\,e^{-i2\pi f\tau}\,d\tau \tag{9}$$

$$= \frac{1}{T}\int_{-T/2}^{T/2}\int_{-T}^{T} R_x(t-u,\tau)_{\Delta t}\,2Tu_{2T}(\tau)e^{-i2\pi f\tau}\,d\tau\,du \tag{10}$$

$$= \frac{1}{T}\int_{-T/2}^{T/2}\int_{-\infty}^{\infty} S_x(t-u,f-\nu)_{\Delta t}\,w_{1/2T}(\nu)\,d\nu\,du \tag{11}$$

$$= \frac{1}{T}\int_{-T/2}^{T/2} S_x(t-u,f)_{\Delta t,1/T}\,du. \tag{12}$$

Equation (7) is definition (22) in Chapter 2, (9) is identity (26) in Chapter 2, (11) is the convolution theorem and definition (22) in Chapter 2, and (12) is definition (4b) with[1]

$$H_{1/T}(f) = w_{1/2T}(f). \tag{13}$$

If approximation (27) in Chapter 2 were used following (10), then we would obtain (using $\Delta f = 1/T$) the approximation

$$S_x(t,f)_{1/\Delta f,\Delta t} \cong S_x(t,f)_{\Delta t,\Delta f}, \qquad \Delta t\Delta f >> 1, \tag{14a}$$

for which $g_{\Delta t}$ in (3b) is the rectangle

$$g_{\Delta t}(t) = u_{\Delta t}(t), \tag{14b}$$

and $H_{\Delta f}$ in (4b) is the sinc

$$H_{\Delta f}(f) = w_{\Delta f/2}(f). \tag{14c}$$

Approximation (14a) also follows from (12) by noting that for $\Delta t >> T$, the integral in (12) can be omitted with negligible effect since the temporal resolution of the integrand is Δt. In view of approximations (8) in Chapter 3, (5), and (14), we see that we also have the approximation

$$S_x(t,f)_{\Delta t,\Delta f} \cong S_{x_{\Delta t}}(t,f)_{\Delta f}, \qquad \Delta t\Delta f >> 1. \tag{15a}$$

[1] The factor of $\frac{1}{2}$ in $1/2T$ is due to the particular definition (14) in Chapter 2 of the width parameter for the sinc function. If the width were taken to be the distance between the first zero crossings to the left and right of the central peak, then the factor of $\frac{1}{2}$ would vanish.

To verify (15a) directly, we manipulate (4b) as follows (using $\Delta f = 1/T$):

$$S_x(t,f)_{\Delta t,\Delta f} \triangleq \int_{-\infty}^{\infty} S_x(t,f-\nu)_{\Delta t}\, H_{\Delta f}(\nu)\, d\nu \tag{16}$$

$$= \int_{-T}^{T} R_x(t,\tau)_{\Delta t}\, h_T(\tau) e^{-i2\pi f\tau}\, d\tau \tag{17}$$

$$\cong \int_{-T}^{T} R_{x_{\Delta t}}(t,\tau) h_T(\tau) e^{-i2\pi f\tau}\, d\tau, \qquad \Delta t >> T \tag{18}$$

$$= \int_{-\infty}^{\infty} S_{x_{\Delta t}}(t,f-\nu) H_{\Delta f}(\nu)\, d\nu \tag{19}$$

$$\triangleq S_{x_{\Delta t}}(t,f)_{\Delta f}. \tag{20}$$

Equation (17) is the convolution theorem and definition (22) in Chapter 2, for which it is assumed that

$$h_T(\tau) = 0, \qquad |\tau| > T. \tag{15b}$$

Approximation (18) is approximation (30) in Chapter 2, and (19) is the convolution theorem. Thus, (15a) is valid for $H_{\Delta f}$ given by the Fourier transform of any duration-limited aperture (15b),

$$H_{\Delta f}(\cdot) = F\{h_{1/\Delta f}(\cdot)\}. \tag{15c}$$

In summary, all four of the statistical spectra (1b)–(4b) are approximately equal, with appropriate choices for smoothing windows, and the accuracy of approximation increases as the resolution product $\Delta t \Delta f$ increases. In fact, all four statistical spectra become identical[2] to the limit spectrum $\hat{S}_x$ as $\Delta t \to \infty$ and $\Delta f \to 0$. Thus, for sufficiently large Δt and $1/\Delta f$, the particular shapes of the data-tapering aperture and the spectral-smoothing window are irrelevant. However, it should be emphasized that in applications where Δt must be relatively small, because of limited data, the desire to track time-variant phenomena, or implementational limitations, the differences among the statistical spectra obtained from different smoothing methods and different windows can be substantial, and the particular choice then becomes an important component of the *design problem*. This is especially so because of *spectral leakage effects*, which are described in Section C. The window-design problem is considered in Chapter 5, Section D.

Example

Let us consider a hypothetical situation in which we know that the limit spectrum to be estimated has resolution width $\Delta f^* = 1$ KHz. Let us also assume that a resolution product of $\Delta t \Delta f = 100$ is desired for reduction of random effects in our spectrum estimate. In order to ensure that our spectrum estimate will provide adequate spectral resolution, we select its resolution-width parameter to be $\Delta f =$

[2] Except possibly for the scale factor γ defined by (25) in Chapter 3.

$\Delta f^*/2 = 500$ Hz. Then we conclude that the parameter Δt must be given by $\Delta t = 100/\Delta f = 200$ ms. If we choose to use a time-averaged periodogram as our spectrum estimate, then we must select appropriate values for the length of the data segment to be Fourier transformed and for the length of the time-interval over which the periodogram will be averaged. It follows directly from the definition of the notation $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ that the data segment length is $T = 1/\Delta f = 1/500$ Hz $=$ 2 ms and that the averaging time-interval length is $\Delta t = 200$ ms. On the other hand, if we choose to use a frequency-smoothed periodogram, then we can conclude from the definition of the notation $S_{x_{\Delta t}}(t, f)_{\Delta f}$ that the length of the data segment to be Fourier transformed is $T = \Delta t = 200$ ms and that the width of the spectral smoothing window is $\Delta f = 500$ Hz.

B. FOURIER TRANSFORMATION OF TAPERED AUTOCORRELATION

It follows directly from the time-frequency dual of the convolution theorem, the periodogram-correlogram relation (18) in Chapter 2, and its analog for the pseudospectrum, (22) in Chapter 2, that the two spectrally smoothed statistical spectra (2b) and (4b) can each be obtained by *Fourier transformation of tapered autocorrelations*,

$$S_{x_{\Delta t}}(t, \cdot)_{\Delta f} = F\{R_{x_{\Delta t}}(t, \cdot)h_{1/\Delta f}(\cdot)\} \tag{21}$$

$$S_x(t, \cdot)_{\Delta t, \Delta f} = F\{R_x(t, \cdot)_{\Delta t}h_{1/\Delta f}(\cdot)\}. \tag{22}$$

In both cases (21) and (22), the spectrum has undergone substantial spectral smoothing ($\Delta t \Delta f >> 1$) if and only if the *autocorrelation-tapering aperture* h_T is substantially narrower than the lag-width parameter Δt of the autocorrelation, $T << \Delta t$ (recall from Chapter 2 that $R_{x_{\Delta t}}(t, \tau) = R_x(t, \tau)_{\Delta t} = 0$ in general only if $|\tau| > \Delta t$; thus, $2\Delta t$ is the lag-width).

Alternative terms for distinguishing between data tapering introduced in Chapter 2, Section B, and autocorrelation tapering introduced here, are *linear tapering* and *quadratic tapering*, respectively. Although both tapering methods have an impact on the *effective spectral smoothing window* as explained in the sequel, their effects are different. For example, it follows directly from (21) and (2b) and (22) and (4b) that the autocorrelation-tapering window has a strong primary influence on the effective spectral smoothing window, and it follows directly from (11)–(17) in Chapter 3 that the data-tapering window used with a temporally smoothed periodogram can have a strong primary influence on the effective spectral-smoothing window. However, it can be shown that the data-tapering window used with a spectrally smoothed periodogram has only a secondary influence on the effective spectral smoothing window. These comments are supported by derivations of effective spectral smoothing windows in Chapter 5, Section D.

The spectrally smoothed periodogram is sometimes called a *spectrograph estimate* [Grenander and Rosenblatt 1984], and therefore the equivalence (21) reveals that the Fourier transform of the tapered correlogram is also a spectrograph estimate. However, the term *spectrograph* is most commonly used to denote a graph of a time-variant spectrum estimate in which the magnitude is represented by the intensity of the graph, and time and frequency are plotted along the abscissa and ordinate, respectively [Koenig, et al. 1946]. The corresponding graph of a time-variant correlation estimate has been called a *correlatograph* [Bennett 1953], but will be referred to here as a *correlograph*.

C. SPECTRAL LEAKAGE AND PREWHITENING

The eight alternative methods for obtaining the four different types of statistical spectra, (1b)–(4b), that have been discussed so far are summarized in block-diagram form in Figures 4-1 and 4-2. For each of these eight methods, either a spectral smoothing window is used directly or an effective (exact or approximate) spectral smoothing window results from another operation (data-tapering or data-

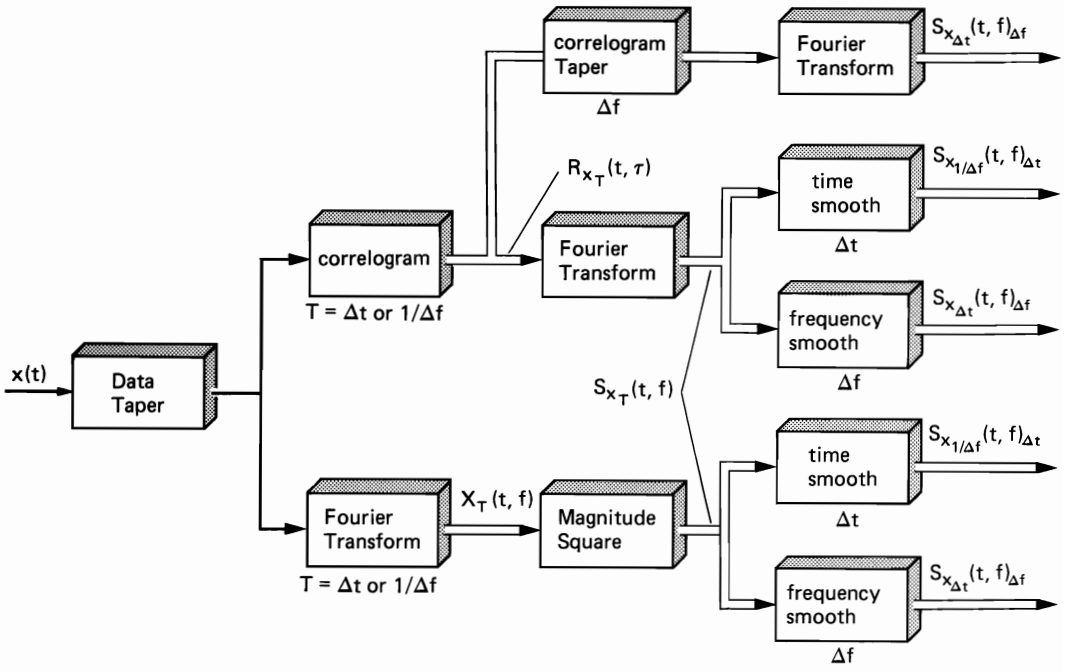


Figure 4-1 Five methods for obtaining statistical spectra based on the periodogram or correlogram ($T = \Delta t$ or $1/\Delta f$, depending on the method).

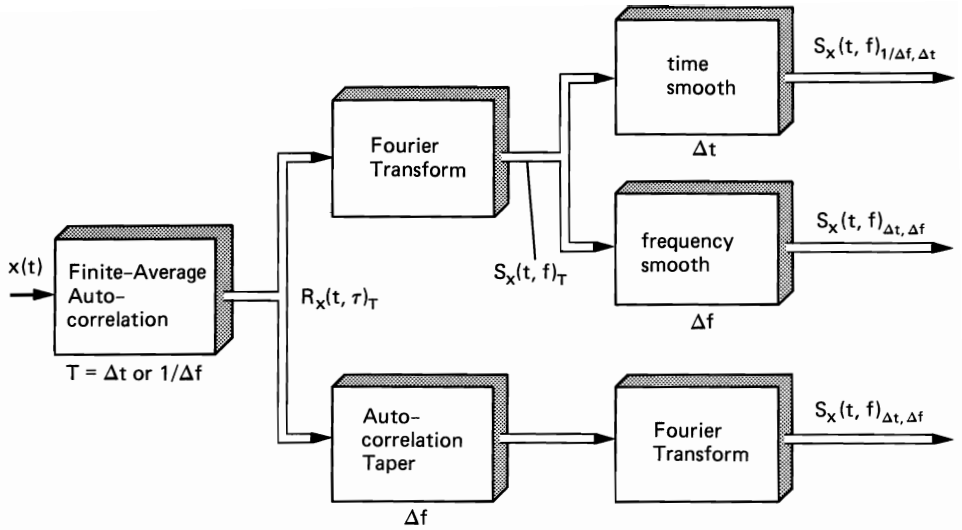


Figure 4-2 Three methods for obtaining statistical spectra based on the pseudospectrum or finite-average auotcorrelation ($T = \Delta t$ or $1/\Delta f$, depending on the method).

windowing and/or autocorrelation-tapering or autocorrelation-windowing[3]). For each of these eight methods, the amount of smoothing is determined (exactly or approximately) by the resolution product $\Delta t \Delta f$, and substantial smoothing is effected if and only if the width Δf of the (effective) spectral smoothing window greatly exceeds the reciprocal of the total amount Δt of data used,

$$\Delta t \Delta f >> 1. \tag{23}$$

However, when Δt is limited so that (23) requires a Δf that is not small compared with the widths of the spectral features to be measured, then the *shape* as well as the width of the spectral-smoothing aperture is important. Moreover, for the methods in which spectral smoothing is accomplished indirectly, this shape is determined, as discussed in the preceding paragraph, by temporal-aperture shapes, namely, 1) the data-tapering aperture a_T in (1b), for which the effective spectral smoothing window is (5c), or 2) the autocorrelation-tapering aperture h_T in (21), for which the effective spectral smoothing window is approximately (15c). (Exact effective spectral smoothing windows are determined in Chapter 5.) For example, as shown in Figure 4-3, the sinc and squared-sinc spectral windows, corresponding to the rectangle and triangle autocorrelation-tapering apertures, respectively, both exhibit a desirable *main lobe*, whose width determines the spectral resolution, and both exhibit a number of potentially nonnegligible undesirable *sidelobes*. But since the triangle aperture has a more gradual taper than the rectangle aperture (which tapers from no attenuation to complete attenuation in an infinitesimal interval), then its Fourier transform (the corresponding spectral window) has a

[3] To make a distinction, if the aperture is the rectangle, then the term *windowing* can be used; otherwise, the term *tapering* can be used.

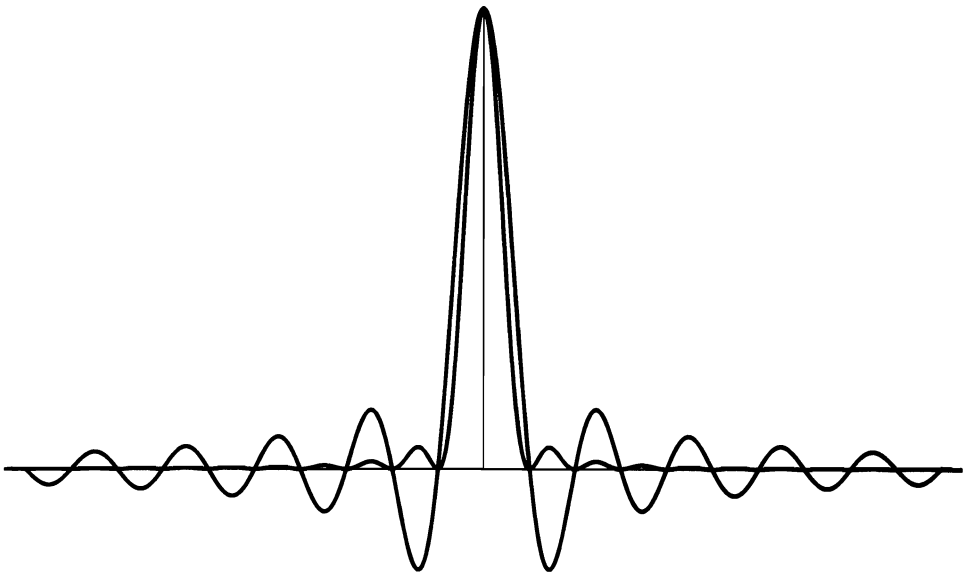


Figure 4-3 Comparison of sidelobes of sinc and squared-sinc windows (with unity width parameter).

less gradual taper, that is, its side lobes are smaller. This can be important because large side lobes can result in the undesirable phenomenon known as *spectral leakage*, whereby spectral content in a spectral band that is coincident with a side lobe contributes to the measured spectral content in the desired spectral band which is coincident with the main lobe. Depending on the sign of the sidelobe, this spectral leakage phenomenon can either increase or decrease the measured spectral content at the main-lobe frequency. This spectral leakage effect is especially problematic when spectral lines are present. As an illustration of this, refer to the graphs of the periodogram for two sine waves shown in Figure 2-6. In Figure 2-6(c) and (d), leakage effects are minimal, but they are significant in Figure 2-6(b), where the peak due to each sine wave has been substantially reduced due to (negative) leakage from the other. Thus, it is generally advisable whenever possible to remove finite additive sine wave components from otherwise random data prior to measurement of the spectrum. Methods for doing this are described in Chapter 9.

It follows from the relationships (5c) and (15c) between the data-tapering or autocorrelation-tapering aperture and the effective spectral-smoothing window that spectral leakage can be controlled by appropriate design of the tapering window.[4] (The topic of window design is treated in Chapter 5, Section D.)

[4] For time-series from periodic phenomena (see, for example, Chapter 3, Section D, Parts 3–5), there is another spectral leakage phenomenon referred to as *cycle leakage* that has apparently not previously been studied. The effects of this phenomenon on spectral measurements are described in Part II, Chapter 15.

However, in some applications where the desired spectrum $\hat{S}_x$ has a wide range (high peaks and/or low valleys), an alternative to data-tapering or autocorrelation-tapering, called *prewhitening*, can sometimes be superior in the sense of yielding a lower degree of randomness for a given resolution product, $\Delta t \Delta f$ (see [Priestley 1981, Section 7.4.1]). Specifically, if a wide spectral range is discovered from a preliminary spectral analysis, then this range can be reduced (heights of peaks lowered and depths of valleys heightened) by filtering the data in an attempt to *whiten* it, that is, to flatten its spectrum. After whitening the data (approximately), a statistical spectrum with the desired resolution can be obtained while incurring only minimal spectral leakage. This statistical spectrum must then be corrected to account for prewhitening. The spectrum relations for filters, (39) and (40) in Chapter 2, can be used to show that the corrected statistical spectrum for x should be obtained by dividing the spectrum for the prewhitened data y by the squared magnitude of the transfer function of the whitening filter; for example,

$$S_{x_{\Delta t}}(t, f)_{\Delta f} \triangleq \frac{S_{y_{\Delta t}}(t, f)_{\Delta f}}{|H(f)|^2}. \tag{24}$$

The prewhitening method suggests another quite distinct method for obtaining a statistical spectrum, which can be especially well suited to the case of sharp spectral peaks. If a particularly effective whitening method is used, then the whitened data y will have a very nearly flat statistical spectrum, say

$$S_{y_{\Delta t}}(t, f)_{\Delta f} \cong N_0, \tag{25}$$

and therefore (24) reduces to

$$S_{x_{\Delta t}}(t, f)_{\Delta f} \cong \frac{N_0}{|H(f)|^2}. \tag{26}$$

Consequently, there is no need to do spectral analysis after particularly effective whitening. There do exist methods for whitening that do not require a preliminary spectral analysis. When the *least squares autoregressive method* of whitening discrete-time data is used, (26) yields the *autoregressive spectrum estimate* originally proposed by George Udny Yule [Yule 1927]. Discussion of this method of spectrum estimation and other methods that minimize spectral leakage is postponed until Chapters 6 and 9 since they typically require digital implementations.

D. HOPPED TEMPORAL SMOOTHING

In some situations, the implementation of a temporally smoothed statistical spectrum can be simplified if the smoothing time index is hopped along discretely rather than slid along continuously. For example, for periodograms of data segments of length T, the smoothing time index can be hopped along at integer multiples of $T/2$, so that subsequent data segments that are Fourier transformed overlap by 50%. For a fixed amount of data, it can be shown that for 50% or more overlap, the effect of smoothing on reduction of degree of randomness is only slightly degraded by hopping rather than sliding. Even less than 50% overlap is sometimes acceptable. This motivates inquiry into the specific relationships between hopped temporal smoothing and other smoothing methods. For example,

it can be shown that hopped-temporal smoothing of the pseudospectrum is exactly equivalent to spectral smoothing, when the hop time interval is equal to the integration time T of the finite-average autocorrelation,

$$S_x(t,f)_{1/\Delta f} \otimes \eta_{\Delta t,\Delta f}(t) = S_x(t,f)_{\Delta t} \otimes H_{\Delta f}(f) \tag{27a}$$

in which

$$H_{\Delta f}(f) = w_{\Delta f/2}(f) \tag{27b}$$

and $\eta_{\Delta t,\Delta f}$ is the *rectangle comb window* defined by

$$\eta_{\Delta t,\Delta f}(t) \triangleq \frac{1}{\Delta t \Delta f} \sum_{m=-M}^{M} \delta\left(t + \frac{m}{\Delta f}\right) \tag{28}$$

with overall width

$$\Delta t \triangleq \frac{(2M+1)}{\Delta f}.$$

Convolution with $\eta_{\Delta t,\Delta f}$ is a hopped time-smoothing operation, as revealed by substitution of (28) into the left member of (27a) to obtain

$$S_x(t,f)_T \otimes \eta_{\Delta t,1/T}(t) = \frac{1}{2M+1} \sum_{m=-M}^{M} S_x(t + mT, f)_T. \tag{29}$$

Equivalence (27) can be derived as follows (using $T = 1/\Delta f$):

$S_x(t,f)_{\Delta t} \otimes H_{\Delta f}(f)$

$$= \int_{-\infty}^{\infty} S_x(t, f-\nu)_{\Delta t} H_{\Delta f}(\nu)\, d\nu \tag{30}$$

$$= \int_{-T}^{T} R_x(t,\tau)_{\Delta t} h_T(\tau) e^{-i2\pi f\tau}\, d\tau \tag{31}$$

$$= \int_{-T}^{T} \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} x\left(u + \frac{\tau}{2}\right) x\left(u - \frac{\tau}{2}\right) du\, h_T(\tau)\, e^{-i2\pi f\tau}\, d\tau \tag{32}$$

$$= \frac{1}{2M+1} \sum_{m=-M}^{M} \int_{-T}^{T} \frac{1}{T} \int_{t+mT-T/2}^{t+mT+T/2} x\left(u + \frac{\tau}{2}\right) x\left(u - \frac{\tau}{2}\right) du\, [2Tu_{2T}(\tau)] e^{-i2\pi f\tau}\, d\tau \tag{33}$$

$$= \frac{1}{2M+1} \sum_{m=-M}^{M} \int_{-T}^{T} R_x(t + mT, \tau)_T\, e^{-i2\pi f\tau}\, d\tau \tag{34}$$

$$= \frac{1}{2M+1} \sum_{m=-M}^{M} S_x(t + mT, f)_T \tag{35}$$

$$= S_x(t,f)_{1/\Delta f} \otimes \eta_{\Delta t,\Delta f}(t). \tag{36}$$

Equation (31) is the convolution theorem (since $h_T(\tau) = 2Tu_{2T}(\tau) = 0$ for $|\tau| > T$) and definition (22) in Chapter 2, and (32) is definition (21) in Chapter 2. Equation (33) is simply the expression of an integral over an interval as the sum of integrals over subintervals that partition the interval. Equation (34) is definition (21) in Chapter 2, and (35) is definition (22) in Chapter 2. Equation (36) follows from (29). If the spectral-smoothing window $H_{\Delta f}$ is not the sinc window, then (27) is an approximation.

The hopped temporally smoothed periodogram is studied in exercises 1 and 2 and for discrete-time data in Chapter 6.

E. WAVE ANALYSIS

One of the most attractive analog methods of spectral analysis is based on the use of electrical wave filters, and is called the wave-analysis method. It is shown in this section that, with certain wave-filter characteristics, this method is exactly equivalent to the temporally smoothed periodogram method.

1. Complex Implementation

The temporally smoothed periodogram for tapered data is defined by (see (3) in Chapter 2 and (1b))

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} \triangleq \Delta f|[x(t)e^{-i2\pi ft}] \otimes a_{1/\Delta f}(t)|^2 \otimes g_{\Delta t}(t) \tag{37}$$

but can be reexpressed (exercise 3) as

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} = \Delta f|x(t) \otimes a^f_{1/\Delta f}(t)|^2 \otimes g_{\Delta t}(t), \tag{38}$$

for which

$$a^f_{1/\Delta f}(t) \triangleq a_{1/\Delta f}(t)e^{i2\pi ft}. \tag{39}$$

The smoothing functions $a^f_{1/\Delta f}$ and $g_{\Delta t}$ in (38) can be interpreted as the impulse-response functions of filters with transfer functions

$$A^f_{\Delta f}(\cdot) = F\{a^f_{1/\Delta f}(\cdot)\} \tag{40}$$

and

$$G_{1/\Delta t}(\cdot) = F\{g_{\Delta t}(\cdot)\}, \tag{41}$$

and it is easily shown that

$$A^f_{\Delta f}(\nu) = A_{\Delta f}(\nu - f), \tag{42}$$

where

$$A_{\Delta f}(\cdot) = F\{a_{1/\Delta f}(\cdot)\}. \tag{43}$$

For a typical data-tapering aperture $a_{1/\Delta f}$, $A_{\Delta f}$ is a pulselike function centered at the origin, with width on the order of Δf. Thus, $A^f_{\Delta f}$ is a pulselike function centered at f, with width on the order of Δf. Hence, $A^f_{\Delta f}$ is the transfer function of a *band-pass filter* (BPF) that passes primarily the frequency components of $x(t)$ in the spectral band of width Δf centered at f. Similarly, $G_{1/\Delta t}$ is the transfer function of a *low-pass filter* (LPF) with bandwidth $1/\Delta t$ centered at zero frequency. Thus, we see that the statistical spectrum $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ can be obtained by band-pass filtering, magnitude-squaring, and low-pass filtering the waveform $x(t)$, as illustrated in Figure 4-4. For this statistical spectral analysis method, which is called *wave analysis*, the condition $\Delta t \Delta f >> 1$ for substantial smoothing requires that the bandwidth Δf of the input BPF (typically called the *predetection*, or *resolution*, bandwidth) greatly exceed the bandwidth $1/\Delta t$ of the output LPF

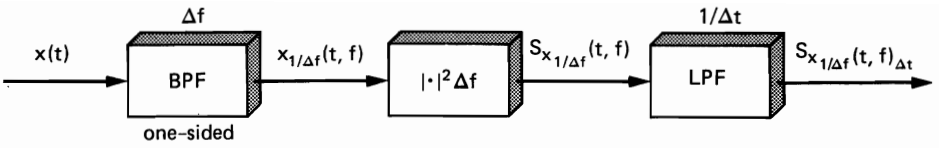


Figure 4-4 Wave analysis method for time-variant statistical spectral analysis. (For $\Delta t \Delta f \gg 1$ and $|f|/\Delta f \gg 1$, the complex one-sided BPF can be replaced with a real two-sided BPF with negligible effect on the measured spectrum.)

(typically called the *postdetection*, or *video*, bandwidth).[5] It follows from approximation (11)–(17) in Chapter 3 that the squared magnitude of the BPF transfer function is the effective spectral smoothing window, $H_{\Delta f}$ in (2b).

The output of the BPF, denoted by $x_T(t, f)$, is given by

$$x_T(t,f) \triangleq x(t) \otimes a_T^f(t) \tag{44}$$

$$= X_T(t,f)e^{i2\pi ft}. \tag{45}$$

Generalizing on interpretations discussed in Part 2 of Section B in Chapter 1, the waveform $x_T(t, f)$ can be interpreted as the *local sine wave component* of $x(t)$ at frequency f and time-locale $[t - T/2, t + T/2]$. The low-pass waveform $X_T(t, f)$ obtained from the band-pass waveform $x_T(t, f)$ by the frequency-shifting operation (45) is called the *complex envelope*[6] or *complex demodulate* of $x_T(t, f)$. By expanding (45) into real and imaginary parts, denoted by subscripts r and i, respectively, we obtain

$$x_T(t,f)_r = X_T(t,f)_r\cos(2\pi ft) - X_T(t,f)_i\sin(2\pi ft) \tag{46}$$

$$= |X_T(t,f)|\cos[2\pi ft + \arg\{X_T(t,f)\}] \tag{47}$$

$$x_T(t,f)_i = X_T(t,f)_i\cos(2\pi ft) + X_T(t,f)_r\sin(2\pi ft) \tag{48}$$

$$= |X_T(t,f)|\sin[2\pi ft + \arg\{X_T(t,f)\}]. \tag{49}$$

Thus, $|X_T(t, f)|$ represents the envelope (amplitude) of either of two sine waves in quadrature at frequency f, with slowly varying (for $1/T << |f|$) amplitude and phase. Substitution of (44) into (38) yields the alternative expression for the output of the wave analyzer,

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} = \Delta f|x_{1/\Delta f}(t,f)|^2 \otimes g_{\Delta t}(t). \tag{50}$$

It is interesting that the basic time-frequency uncertainty principle described in Chapter 2, which establishes the approximate relation $\Delta t^\circ \Delta f^\circ \cong 1$ on the temporal and spectral resolution widths of a nonstatistical spectral measurement, can be interpreted here as a constraint on the response-time Δt° of a filter with bandwidth Δf°. Specifically, it follows directly from the Fourier transform re-

[5] The terms *detection* and *video bandwidth* originated in the fields of radar and radio, where detection used to mean extraction of the amplitude of an amplitude-modulated sine wave and video referred to a signal that is to be visually displayed on the screen of a cathode ray tube.

[6] If $x_T(t, f)$ is sufficiently narrow band that it is negligibly small for $f < 0$, then the complex envelope $X_T(t, f)$ is a close approximation to the complex envelope $\gamma(t)$ of $x_T(t, f)$ as defined in Rice's representation for band-pass waveforms with center frequency $f_0 = f$ (Appendix 3-1).

lationship between the transfer function and impulse-response function that a filter with bandwidth Δf° takes on the order of $\Delta t^\circ \cong 1/\Delta f^\circ$ units of time to respond fully to an excitation at a given instant. Furthermore, if an input BPF with bandwidth $\Delta f^\circ = \Delta f$ is used, then the condition $\Delta t >> 1/\Delta f$ on the integration time Δt of the output LPF requires that a long segment of the input (much longer than the input-filter-response time) be processed. Moreover, when the relative bandwidth of a BPF is much less than 1, that is, $\Delta f/|f| << 1$, then the relation $\Delta t^\circ \Delta f^\circ \cong 1$ when applied to such a BPF yields $\Delta t^\circ >> 1/|f|$, which reveals that it takes many cycles of excitation before the BPF fully responds. Consequently, if an input BPF with narrow relative bandwidth $\Delta f << |f|$ is used, then the condition $\Delta t >> 1/\Delta f$ on the integration time of the output LPF results in $\Delta t >> 1/\Delta f >> 1/|f|$, which requires that integration be carried out over very many cycles of the local sine wave component.

2. Real Implementation

The BPF defined by (40) has a passband at f but none at the image frequency $-f$. This one-sided BPF requires a complex impulse-response function (39). In order to obtain an implementation involving only real filters, we simply expand the square in (50) to obtain

$$S_{x_{1/\Delta f}}(t,f)_{\Delta t} = \Delta f\,[x_{1/\Delta f}(t,f)_r]^2 \otimes g_{\Delta t}(t) + \Delta f\,[x_{1/\Delta f}(t,f)_i]^2 \otimes g_{\Delta t}(t), \tag{51}$$

in which

$$\begin{aligned} x_T(t,f)_r &\triangleq x(t) \otimes a_T^f(t)_r \\ x_T(t,f)_i &\triangleq x(t) \otimes a_T^f(t)_i, \end{aligned} \tag{52}$$

where

$$\begin{aligned} a_T^f(t)_r &\triangleq a_T(t)\cos(2\pi f t) \\ a_T^f(t)_i &\triangleq a_T(t)\sin(2\pi f t). \end{aligned} \tag{53}$$

It can be shown (exercise 4) that the two terms in (51) are approximately equal for

$$\frac{|f|}{\Delta f} >> 1 \tag{54}$$

and

$$\Delta t \Delta f >> 1, \tag{55}$$

and therefore

$$\begin{aligned} S_{x_{1/\Delta f}}(t,f)_{\Delta t} &\cong 2\Delta f\,[x_{1/\Delta f}(t,f)_r]^2 \otimes g_{\Delta t}(t) \\ &\cong 2\Delta f\,[x_{1/\Delta f}(t,f)_i]^2 \otimes g_{\Delta t}(t), \qquad \frac{|f|}{\Delta f} >> 1,\ \Delta t \Delta f >> 1. \end{aligned} \tag{56}$$

F. DEMODULATION

A direct implementation of a spectrum analyzer using the wave-analysis method requires a bank of typically many BPF-squarer-LPF devices, one for each frequency f of interest in the spectral band of interest. For a spectral resolution width of

Δf and a total bandwidth of B, this amounts to $B/\Delta f$ individual values of f. This number is often as large as several orders of magnitude. A means for simplifying such a highly complex spectrum analyzer is based on the method of demodulation, by which spectral components at different frequencies are shifted to a common frequency such as $f = 0$. This method is described here.

1. Complex Implementation

An alternative to the wave-analysis method of implementation of the statistical spectrum $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ is suggested by (37). Specifically, the operation

$$X_T(t,f) \triangleq [x(t)e^{-i2\pi ft}] \otimes a_T(t) \tag{57}$$

on $x(t)$, which consists of frequency-shifting down by an amount f and low-pass filtering, is called *complex demodulation*, and it produces the complex demodulate (envelope) of the local sine wave component $x_T(t, f)$. It follows from (45) that the complex demodulate is simply the local sine wave component shifted down to zero frequency, as illustrated in Figure 4-5.

It follows directly from (37) and (57) that the temporally smoothed statistical spectrum for tapered data $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ can be obtained by low-pass filtering the squared magnitude of the complex demodulate, as illustrated in Figure 4-6(a). The condition $\Delta t \Delta f >> 1$ for substantial smoothing requires that the bandwidth Δf of the input LPF greatly exceed the bandwidth $1/\Delta t$ of the output LPF. It follows from approximation (11)–(17) in Chapter 3 that the squared magnitude of the input LPF transfer function is the effective spectral-smoothing window, $H_{\Delta f}$ in (2b).

It is explained in exercises 3 and 5 that the demodulation spectrum analyzer can be derived from the wave analyzer simply by using band-pass to low-pass transformations on the filters in the latter analyzer, as illustrated in Figures 4-5 and 4-8. (Observe that the multiplication by $\exp(-i2\pi ft)$ just preceding the squared-magnitude operation has no effect after this operation.)

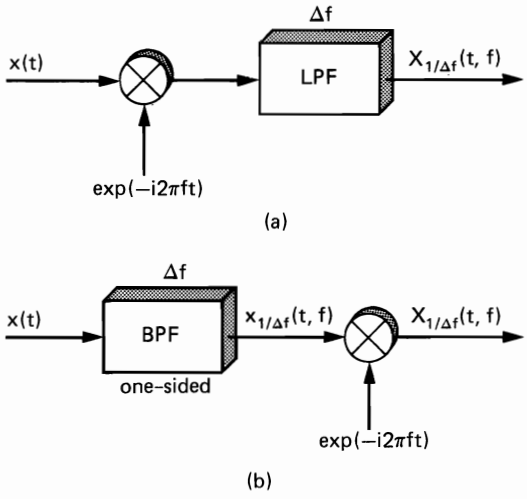


Figure 4-5 (a) Direct generation of complex demodulate. (b) Indirect generation of complex demodulate from local sine wave component.

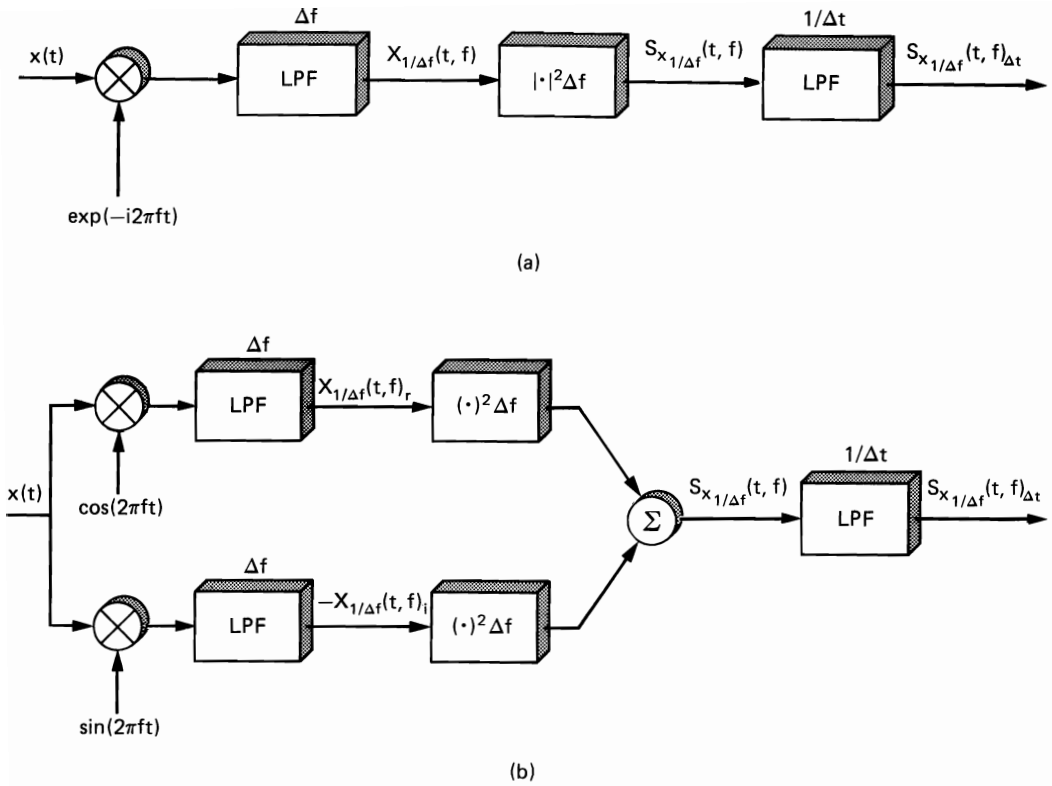


Figure 4-6 (a) Demodulation method for time-variant statistical spectral analysis. (b) Real implementation of demodulation method for time-variant statistical spectral analysis. (Either of the two parallel paths can be deleted with negligible effect on the measured spectrum provided that (i) $\Delta t \Delta f \gg 1$ and (ii) either $x(t)$ contains no finite additive sine wave components in the lag product or the swept-frequency implementation in Figure 4-7 is used.)

2. Real Implementation

The demodulation operation (57) involves a complex sine wave. In order to obtain an implementation involving only real sine wave generators, we simply expand the squared magnitude of (57) to obtain from (37)

$$S_{x_{1/\Delta f}}(t, f)_{\Delta t} = \Delta f\, [X_{1/\Delta f}(t, f)_r]^2 \otimes g_{\Delta t}(t) + \Delta f\, [X_{1/\Delta f}(t, f)_i]^2 \otimes g_{\Delta t}(t), \tag{58}$$

in which

$$\begin{aligned} X_T(t, f)_r &\triangleq [x(t)\cos(2\pi f t)] \otimes a_T(t) \\ X_T(t, f)_i &\triangleq [x(t)\sin(2\pi f t)] \otimes a_T(t). \end{aligned} \tag{59}$$

This real implementation is illustrated in Figure 4-6(b).

It can be shown (exercise 4) that the two terms in (58) are approximately equal for

$$\Delta t \Delta f >> 1, \tag{60}$$

provided that $x(t)$ contains no periodicities[7] that would violate the condition

$$\left|\mathrm{Re}\left\{\int_{-T}^{T} R_x^{2f}(t,\tau)_{\Delta t} T v_T(\tau)\, d\tau\right\}\right| \ll S_{x_T}(t,f)_{\Delta t}, \tag{61}$$

for which (using $\alpha = 2f$)

$$R_x^{\alpha}(t,\tau)_{\Delta t} \triangleq \frac{1}{\Delta t}\int_{t-\Delta t/2}^{t+\Delta t/2} x\left(u+\frac{\tau}{2}\right)x\left(u-\frac{\tau}{2}\right) e^{-i2\pi\alpha u}\, du\, [2\Delta t\, u_{2\Delta t}(\tau)]. \tag{62}$$

Therefore, (58) is approximated by

$$\begin{aligned} S_{x_{1/\Delta f}}(t,f)_{\Delta t} &\cong 2\Delta f\,[X_{1/\Delta f}(t,f)_r]^2 \otimes g_{\Delta t}(t) \\ &\cong 2\Delta f\,[X_{1/\Delta f}(t,f)_i]^2 \otimes g_{\Delta t}(t), \qquad \Delta t\Delta f >> 1 \end{aligned} \tag{63}$$

provided that (61) holds. In this case, either of the two parallel signal-flow paths in Figure 4-6(b) can be deleted to obtain a single-demodulator implementation.

3. Swept-Frequency Implementation

An economical way to construct a spectrum analyzer that covers a broad range of frequencies is to use the demodulation method and sweep the frequency f of the sine wave used for demodulation throughout the desired spectral band. Let us consider a linear sweep for which the *instantaneous frequency* (see Appendix 2-1) is

$$f(t) = f_0 + \beta t. \tag{64}$$

That is, the wave used for demodulation is

$$d(t) \triangleq \cos[2\pi \textstyle\int f(t)\, dt] = \cos(2\pi f_0 t + \pi\beta t^2), \tag{65}$$

which is a *linear frequency modulated sine wave*. In order to obtain the desired spectral resolution, it is required that $f(t)$ not change more than Δf during the integration time Δt,

$$|f(t+\Delta t) - f(t)| \leq \Delta f, \tag{66}$$

which results in the condition

$$|\beta| \leq \frac{\Delta f}{\Delta t}. \tag{67}$$

Since the linear FM sine wave (65) can be reexpressed as

$$d(t) = \cos[2\pi f(t_0)t - \theta_0(t)], \tag{68}$$

for which $f(t_0)$ is given by (64), evaluated at $t = t_0$, and

$$\theta_0(t) \triangleq 2\pi\beta\left(t_0 t - \frac{t^2}{2}\right), \tag{69}$$

then in any time interval of length Δt, centered at $t = t_0$, we can *interpret* $d(t)$ as having fixed frequency $f(t_0)$ and time-varying phase $\theta_0(t)$. Furthermore, if we choose the maximum sweep rate consistent with condition (67), $|\beta| = \Delta f/\Delta t$,

[7] The statistical parameter (62) is a measure of the strength of any possible additive sine wave component at frequency α contained in the lag product and is studied in Part II (see also Chapter 7, Section C).

then this phase sweeps through a range of

$$\max[\theta_0(t)] - \min[\theta_0(t)] = \frac{\pi}{4}\Delta t \Delta f \tag{70}$$

during the time interval $[t_0 - \Delta t/2,\ t_0 + \Delta t/2]$. Thus, in all cases for which $\Delta t \Delta f > 2$, the phase $\theta_0(t)$ sweeps through more than $\pi/2$ radians for each frequency of interest. It follows that the simplified method of implementation that uses only one demodulator, (63), rather than both the *in-phase* and *quadrature* demodulators, (58)-(59), effectively realizes both an in-phase and a quadrature demodulator (since $\cos(2\pi f_0 t + \pi/2) = \sin(2\pi f_0 t)$). Hence, the single-demodulator swept-frequency implementation can be used even if the data contains periodicity that violates condition (61).

Most practical swept-frequency spectrum analyzers use a combination of the wave-analysis and demodulation methods, by demodulating all frequencies of interest down to a fixed *intermediate frequency* $f_1 > 0$ (rather than to $f_1 = 0$), as illustrated in Figure 4-7. However, this creates a problem that requires *prefiltering* to solve, because the real sine wave used for demodulation not only shifts spectral components down from frequencies near $f = f_0$ to frequencies near f_1 (i.e., a negative shift by $f - f_1$) but also shifts spectral components up from frequencies near $2f_1 - f$ to frequencies near f_1 (i.e., a positive shift by $f - f_1$). Let us assume that all frequencies f of interest for spectral analysis are greater than f_1. Then the undesired *image* components at frequencies near $2f_1 - f$ can be eliminated by prefiltering with a high-pass filter with cutoff frequency f_* that satisfies the inequalities

$$-f_* < 2f_1 - f - \frac{\Delta f}{2} \tag{71a}$$

and

$$f_* < f - \frac{\Delta f}{2}. \tag{71b}$$

For example, if f_* is put in the middle of the range specified by (71) then (within $\pm \Delta f/2$)

$$f_* = f - f_1, \tag{72}$$

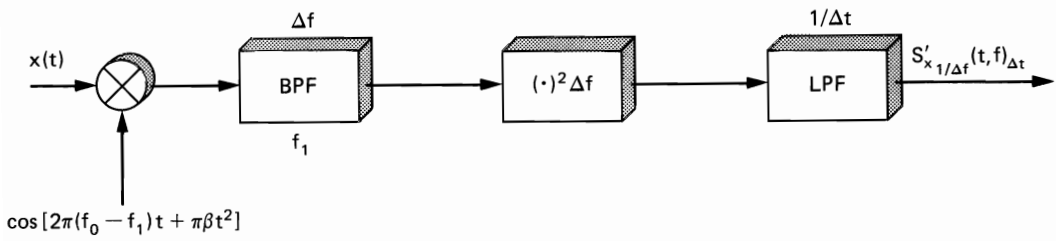


Figure 4-7 Swept-frequency implementation of spectrum analyzer. (The analysis frequency at time t is $f(t) = f_0 + \beta t$.)

which indicates that the cutoff frequency must track the swept frequency f. As an alternative, the cutoff frequency can be increased in discrete steps as the swept frequency f moves from one range to another. In particular, each time f increases to the point where (71a) will be violated, f_* can be increased instantaneously to the point where (71b) is just satisfied.

G. A GENERAL REPRESENTATION

Every method of spectral analysis considered in this chapter (except the swept-frequency wave analyzer) can be mathematically represented by quadratic time-invariant transformations of the data to be analyzed, $x(t)$. (The swept-frequency wave analyzer is a time-variant quadratic transformation.) Specifically, if $y_f(t)$ denotes a time-variant statistical spectrum at a given frequency f, such as

$$y_f(t) = S_{x_{\Delta t}}(t, f)_{\Delta f},$$

then $y_f(t)$ can be represented by

$$y_f(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_f(u, v)x(t - u)x(t - v)\, du\, dv \tag{73}$$

for some appropriate kernel k_f. Furthermore, the kernel is conveniently represented by

$$k_f(u, v) = m\left(\frac{u + v}{2}, v - u\right)e^{-i2\pi f(v-u)}, \tag{74}$$

where the function $m(t, \tau)$ is independent of f and has width Δt in the variable t and width $1/\Delta f$ in the variable τ. Furthermore, for every method considered (except the hopped temporal-smoothing method), the function is a two-dimensional pulselike function. For $\Delta t \Delta f >> 1$, it has oblong contours of constant elevation of its main lobe. (For the hopped temporal-smoothing method without overlap, the function $m(t, \tau)$ is a series versus t of uniformly spaced pulses in τ.) Moreover, for each of these methods this function can be approximated by a separable function,

$$m(t, \tau) \cong g_{\Delta t}(t)h_{1/\Delta f}(\tau). \tag{75}$$

This approximation is either exact, or it is a close approximation for $\Delta t \Delta f >> 1$. For example, for the temporally smoothed periodogram $y_f(t) = S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ for tapered data, it follows (exercise 9) from (37) that

$$m(t, \tau) = \Delta f \int_{-\infty}^{\infty} a_{1/\Delta f}\left(t - w + \frac{\tau}{2}\right)a_{1/\Delta f}\left(t - w - \frac{\tau}{2}\right)g_{\Delta t}(w)\, dw \tag{76a}$$

$$\cong g_{\Delta t}(t)h_{1/\Delta f}(\tau), \qquad \Delta t \Delta f >> 1, \tag{76b}$$

where

$$h_{1/\Delta f}(\tau) = r_a(\tau)\Delta f. \tag{76c}$$

Because of the equivalences described in Sections E and F, (76) also applies to the wave-analysis and demodulation methods. As another example, if $y_f(t)$ is the spectrally smoothed periodogram $S_{x_{\Delta t}}(t, f)_{\Delta f}$, then it follows (exercise 9) from

(21) that

$$m(t, \tau) = \frac{1}{\Delta t} a_{\Delta t}\left(t + \frac{\tau}{2}\right) a_{\Delta t}\left(t - \frac{\tau}{2}\right) h_{1/\Delta f}(\tau) \tag{77a}$$

$$\cong g_{\Delta t}(t) h_{1/\Delta f}(\tau), \qquad \Delta t \Delta f >> 1, \tag{77b}$$

where

$$g_{\Delta t}(t) = \frac{1}{\Delta t} [a_{\Delta t}(t)]^2. \tag{77c}$$

Because of the equivalence described in Section B, (77) also applies to the method based on Fourier transformation of a tapered correlogram. As a final example, if $y_f(t)$ is the Fourier transform of the tapered finite-average autocorrelation $S_x(t, f)_{\Delta t, \Delta f}$ (or, equivalently, the spectrally smoothed pseudospectrum), then it follows (exercise 9) from (22) that

$$m(t, \tau) = \begin{cases} g_{\Delta t}(t) h_{1/\Delta f}(\tau), & |\tau| \leq T \\ 0, & |\tau| > T, \end{cases} \tag{78a}$$

where

$$g_{\Delta t}(t) = u_{\Delta t}(t). \tag{78b}$$

This general representation, (73)–(74), provides a unifying basis for describing the similarities and differences among the various methods of spectral analysis and for design of alternative methods of spectral analysis. Furthermore, (73)–(74) provides a unifying basis for the study of the resolution, leakage, and reliability properties of spectrum estimates, which is taken up in the next chapter.

As one example of the use of (73)–(74) for design, consider the kernel representation

$$m(t, \tau) = g_{\Delta t}\left(t + \frac{\tau}{2}\right) a_{1/\Delta f}(\tau) \tag{79a}$$

$$\cong g_{\Delta t}(t) h_{1/\Delta f}(\tau), \qquad \Delta t \Delta f >> 1. \tag{79b}$$

Substitution of (79a) into (74) and then (74) into (73) yields (exercise 10)

$$y_f(t) = \{x(t)[x(t) \otimes (a_{1/\Delta f}(t) e^{-i2\pi f t})]\} \otimes g_{\Delta t}(t), \tag{80}$$

which is a type of cross-wave analyzer. Because of (79b), we know that the measurement (80) is an appropriate spectrum estimate. This type of spectrum analyzer is considered further in Appendix 4-1.

H. SUMMARY

In this chapter, an introductory comparative study of a variety of analog (continuous-time and continuous-amplitude) methods of measurement of statistical spectra is conducted. In Section A, approximate equivalences among the four methods based on temporal and spectral smoothing of the periodogram and pseudospectrum are derived, and in Section B it is established that the two spectral smoothing

methods are each exactly equivalent to a method consisting of Fourier transformation of a tapered autocorrelation function. The resultant eight distinct methods for obtaining the four distinct (but approximately equivalent) statistical spectra are summarized in Figures 4-1 and 4-2. In Section C, the spectral leakage phenomenon that results from the sidelobes of the effective spectral smoothing window is explained, and the sine-wave-removal, tapering, and prewhitening approaches to reducing spectral leakage are described. Then Section D explains that temporal smoothing based on continuously sliding periodograms or pseudospectra can be modified to obtain hopped periodograms or pseudospectra, and an exact equivalence between a hopped time-averaged pseudospectrum and a spectrally smoothed pseudospectrum is derived. A similar but approximate equivalence for the hopped time-averaged periodogram is derived in exercises 1 and 2.

In Section E, an alternative method for implementing the temporally smoothed periodogram, which is based on filtering is introduced. Both real and complex implementations, called *wave analyzers*, are developed (Figure 4-4). Then in Section F, another alternative implementation based on demodulation is derived. The real and complex implementations of the *demodulation spectrum analyzer* (Figure 4-6) can be obtained directly from the corresponding implementations of the wave analyzer by using band-pass-to-low-pass transformations on the filters (Figure 4-8). It is then explained that an economical way to construct a spectrum analyzer that covers a broad range of frequencies is to use the demodulation method and sweep the frequency of the sine wave used for demodulation. It is also explained that it is often more practical to use swept-frequency demodulation to down-convert all frequencies to a fixed nonzero intermediate frequency and then use the wave-analysis method (Figure 4-7). In addition, an alternative method of swept-frequency spectral analysis that incorporates time compression is described in exercise 14.

Finally in Section G, a general representation for all preceding types of spectrum analyzers (except the swept-frequency wave analyzer) is introduced, and it is explained that the two width parameters Δt and Δf of the kernel $m(t, \tau)$ that prescribes the representation for a particular spectrum analyzer determine the temporal and spectral resolution widths of the statistical spectrum produced by the analyzer. A convenient separable approximation (75) to the kernel is introduced, and it is explained that the resultant approximate and exact general representations provide a unifying basis for the design and analysis of spectrum analyzers. This is demonstrated in the next chapter.

In Appendix 4-1, an alternative wave-analysis method that is equivalent to a method based on Fourier transformation of a tapered autocorrelation is presented.

EXERCISES

1. (a) Let $H_{\Delta f}(\cdot) = F\{h_T(\cdot)\}$ with $h_T(\tau) = 0$ for $|\tau| > T_0/2$ and $T_0 \geq T$.

Verify the following identity for the spectrally smoothed pseudospectrum:

$$S_x(t,f)_{\Delta t} \otimes H_{\Delta f}(f) = S_x(t,f)_{T_0} \otimes \eta_{\Delta t,1/T_0}(t) \otimes H_{\Delta f}(f), \tag{81}$$

where $\eta_{\Delta t,1/T_0}$ is the rectangular comb window defined by (28) and $\Delta t/T_0$ is an integer.

(b) Show that if $T_0 >> T = 1/\Delta f$, then (81) yields the following close approximation for the spectrally smoothed periodogram:

$$S_{x_{\Delta t}}(t,f) \otimes H_{\Delta f}(f) \cong S_{x_{T_0}}(t,f) \otimes \eta_{\Delta t,1/T_0}(t) \otimes H_{\Delta f}(f). \tag{82}$$

Hint: Follow a procedure analogous to (11)–(13) in Chapter 3. Then use (30) in Chapter 2 (with Δt replaced by T_0). (Equivalence (82) reveals that before spectral smoothing, the length of the data segment that must be Fourier transformed to obtain a periodogram can be reduced from Δt to T_0 by partitioning the time interval of length Δt into $\Delta t/T_0$ nonoverlapping adjacent intervals of length T_0; then these periodograms [$\Delta t/T_0$ in number] are simply averaged. In practice, the spectral-smoothing operation is usually omitted, as discussed in exercise 2.)

(c) Is the identity (81) valid if the convolution with $H_{\Delta f}$ is deleted on both sides? Why?

2. (a) If $x(t)$ contains no *second-order periodicities*, in the sense that

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x\left(t - \frac{\tau}{2}\right) e^{-i2\pi\alpha t}\, dt = 0$$

for all $\alpha \neq 0$ (especially for α equal to integer multiples of $1/T_*$), then it can be shown, as it is in Part II (*synchronized averaging identity*), that the limit autocorrelation defined by (6) in Chapter 1 can be obtained with hopped time-averaging,

$$\hat{R}_x(\tau) = \lim_{M\to\infty} \frac{1}{2M+1} \sum_{m=-M}^{M} x\left(t + mT_* + \frac{\tau}{2}\right) x\left(t + mT_* - \frac{\tau}{2}\right) \tag{83}$$

for all t. Use (83) to prove that the limit spectrum can be obtained with hopped time-averaging,

$$\lim_{T\to\infty} \lim_{M\to\infty} \frac{1}{2M+1} \sum_{m=-M}^{M} S_{x_T}(t + mT_*, f) = \hat{S}_x(f). \tag{84}$$

Hint: First use the periodogram-correlogram relation; then substitute the definition of the correlogram; then evaluate the limit, $M \to \infty$; then evaluate one of the integrals and then perform the remaining Fourier transformation using the convolution theorem. Finally, evaluate the limit $T \to \infty$.

(b) To verify that hopped temporal smoothing of the periodogram, $S_{x_T}(t, f) \otimes \eta_{\Delta t,1/T_*}(t)$, is approximately equivalent to spectral smoothing of the pseudospectrum (for appropriate values of T/T_*), modify the procedure used in part (a) by replacing the step $M \to \infty$ with a step that introduces the approximation

$$\frac{1}{2M+1} \sum_{m=-M}^{M} R_{x_T}(t + mT_*, \tau) \cong R_x(t, \tau)_{(2M+1)T_*} \frac{1}{T} r_a(\tau),$$

which is useful for $T/T_* \geq 2$ and $\Delta t = (2M+1)T_* >> T$. Since the spectrally smoothed periodogram approximates the spectrally smoothed pseudospectrum for $\Delta t \Delta f >> 1$ (see (15a)), then the result obtained reveals that the hopped time-averaged periodogram approximates the spectrally smoothed periodogram with

smoothing window $(1/T)|A_{1/T}(f)|^2$, provided that $T/T_* \geq 2$ and $\Delta t = (2M + 1)T_* >> T = 1/\Delta f$.

(c) A method for computation of statistical spectra proposed by P. D. Welch [Welch 1967] (for discrete-time data) averages the periodograms of tapered data. The data segments overlap each other by 50%. Thus $T = 2T_*$ and there are $1 + \Delta t/T_*$ data segments each of length T. Consider the triangular tapering window $a_T(t) = T_* v_{T_*}(t)$, and use the result of (b) to show that the effective spectral-smoothing window is $(1/T)[z_{2/T}(f)]^2$.

3. (a) For the wave analyzer, derive (38) from (37).

(b) Verify relation (45) between the complex demodulate and the local sine wave component.

(c) Verify the equivalence illustrated in Figure 4-8 for which the BPF is a complex one-sided filter with impulse-response function

$$a^f(t) = a(t)e^{i2\pi ft},$$

where $a(t)$ is the impulse-response function of the LPF.

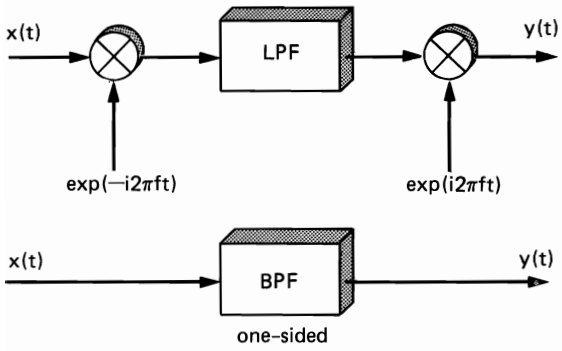


Figure 4-8 Equivalent filters.

4. (a) Verify that the two terms in (56), which describe the wave analyzer, are approximately equal for $\Delta t \Delta f >> 1$ and $|f| >> \Delta f$. *Hint:* Use the following trigonometric identities in the order given:

(i)
$$\begin{aligned} \cos(2\pi f[t - u]) &= \cos(2\pi ft)\cos(2\pi fu) + \sin(2\pi ft)\sin(2\pi fu) \\ \sin(2\pi f[t - u]) &= \sin(2\pi ft)\cos(2\pi fu) - \cos(2\pi ft)\sin(2\pi fu), \end{aligned} \tag{85}$$

(ii)
$$\begin{aligned} [\cos(2\pi ft)]^2 &= \tfrac{1}{2} + \tfrac{1}{2}\cos(4\pi ft) \\ [\sin(2\pi ft)]^2 &= \tfrac{1}{2} - \tfrac{1}{2}\cos(4\pi ft), \end{aligned} \tag{86}$$

(iii)
$$\cos(2\pi ft)\sin(2\pi ft) = \tfrac{1}{2}\sin(4\pi ft). \tag{87}$$

Also use the fact that the terms (using $T = 1/\Delta f$)

$$\begin{aligned} &\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} [X_T(t,f)_r]^2 \cos(4\pi ft)\, dt, \\ &\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} [X_T(t,f)_i]^2 \cos(4\pi ft)\, dt, \\ &\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} X_T(t,f)_r X_T(t,f)_i \sin(4\pi ft)\, dt \end{aligned} \tag{88}$$

can be neglected. (This can be seen as follows. Since $X_T(t, f)_r$ and $X_T(t, f)_i$ have bandwidths on the order of $1/T$ centered at zero, then $[X_T(t, f)_r]^2$, $X_T(t, f)_r X_T(t, f)_i$, and $[X_T(t, f)_i]^2$ have bandwidths on the order of $2/T$ centered at zero, and therefore both the cosine- and sine-Fourier transforms of each of these, at frequencies $|\nu|$ exceeding $2/T$, are relatively small and become negligible for $|\nu|T/2 >> 1$. For Δt substantially exceeding the reciprocal bandwidth $T/2$, the three integrals (88) are finite cosine- and sine-transforms (scaled by $1/\Delta t$) that accurately reflect the spectral characteristics of the time-functions $[X_T(t, f)_r]^2$, $X_T(t, f)_r X_T(t, f)_i$, and $[X_T(t, f)_i]^2$, and these Fourier transforms are evaluated at $\nu = 2f$. Hence for $|f| >> 1/T$ and $\Delta t >> T$, these integrals become negligible compared with remaining terms identified in (89) below.) This yields the close approximations (with $T = 1/\Delta f$)

$$\begin{aligned} S_{x_T}(t, f)_{\Delta t} &\cong \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} \frac{2}{T} \left[\int_{-\infty}^{\infty} a_T^f(t - v - u)_r x(u)\, du \right]^2 dv \\ &\cong \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} \frac{2}{T} \left[\int_{-\infty}^{\infty} a_T^f(t - v - u)_i x(u)\, du \right]^2 dv. \end{aligned} \tag{89}$$

(b) Verify that the two terms in (63) (and (58)), which describe the demodulation spectrum analyzer, are approximately equal for $\Delta t \Delta f >> 1$. *Hint:* Define the function

$$R_{x_T}^{\alpha}(t, \tau) \triangleq \frac{1}{T} \int_{t-(T-|\tau|)/2}^{t+(T-|\tau|)/2} x\left(u + \frac{\tau}{2}\right) x\left(u - \frac{\tau}{2}\right) e^{-i2\pi\alpha u}\, du\, [2Tu_{2T}(\tau)],$$

and show that the difference d of the two terms in (63) is given by

$$d = \mathrm{Re}\left\{ \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} \int_{-T}^{T} R_{x_T}^{\alpha}(t - u, \tau)\, d\tau\, du \right\}$$

with $\alpha = 2f$. Then use a slight generalization (to $\alpha \neq 0$) of the result of exercise 6, Chapter 2 (namely, (23) in Chapter 2) to obtain the identity

$$d = \mathrm{Re}\left\{ \int_{-T}^{T} [R_x^{\alpha}(t, \tau)_{\Delta t} \otimes u_{T-|\tau|}(t)][Tv_T(\tau)]\, d\tau \right\},$$

for which $R_x^{\alpha}(t, \tau)_{\Delta t}$ is defined by (62). Now, if $x(t)$ contains no second-order periodicity, then (as explained in Part II; see also Chapter 7, Section C),

$$\lim_{\Delta t \to \infty} R_x^{\alpha}(t, \tau)_{\Delta t} = 0,$$

and therefore

$$\lim_{\Delta t \to \infty} d = 0.$$

Hence, d is negligible for sufficiently large Δt. (An alternative approach to this problem is to observe that the two terms in (63) converge ($\Delta t \to \infty$) to the frequency-smoothed limit spectra for the in-phase and quadrature low-pass components of the real local sine wave component (52); and then use the fact that these limit spectra are identical if $x(t)$ contains no second-order periodicity (in which case the real local sine wave component (52) contains no second-order periodicity), as established in Appendix 3-1.)

5. For the real demodulation spectrum analyzer, derive (58) directly from (50) using nothing more than (45). This emphasizes the fact that the demodulation method is simply the low-pass implementation of the bandpass wave-analysis method.

6. **(a)** Consider the problem of using the swept-frequency method of implementation for an audio spectrum analyzer that covers the band from 10 Hz to 10,000 Hz, with a frequency resolution of $\Delta f = 10$ Hz and an integration time of $\Delta t = 1$ s ($\Delta t \Delta f = 10$). Verify that it would take $16\frac{2}{3}$ min to analyze the whole band from 10 Hz to 10,000 Hz. (This reveals why the swept-frequency method is used primarily for higher frequency ranges with broader resolution widths.)
   (b) Consider a swept-frequency spectrum analyzer that covers the band from f_0 to $10^3 f_0$, with a spectral resolution of $\Delta f = f_0$ and an integration time that satisfies $\Delta t \Delta f = 10$. Show that if it is desired to analyze the whole band from f_0 to $10^3 f_0$ in 0.1 s, then f_0 must be at least 100 KHz.
   (c) Show that for a swept-frequency spectrum analyzer, only 0.1% of the total data used to obtain an analysis of the band from f_0 to $10^3 f_0$ is used to measure the spectrum in the minimum resolvable band of width Δf, if $\Delta f = f_0$.
   (d) Verify that approximately 100 cycles at frequency f are averaged by the output LPF if $\Delta t \Delta f = 10$ and the input BPF has relative bandwidth $\Delta f / |f| = 1/10$.

7. As explained in Chapter 2 (see exercise 14, Chapter 2), the resolution width Δf^* of the limit spectrum $\hat{S}_x(f)$ is on the order of the reciprocal of the overall width $\Delta\tau^*$ of the limit autocorrelation $\hat{R}_x(\tau)$,

$$\Delta f^* \cong \frac{1}{\Delta\tau^*}.$$

(For example, an oscillatory time-series gives rise to a limit autocorrelation with decaying oscillation, which dies away for $|\tau| > \Delta\tau^* \cong 1/\Delta f^*$, where Δf^* is the width of the spectral peak that corresponds to the oscillatory behavior.) Thus, if it is desired to resolve all fine structure in $\hat{S}_x(f)$, then the spectral resolution capability Δf of the measurement method must satisfy

$$\Delta f \leqslant \Delta f^* \cong \frac{1}{\Delta\tau^*},$$

and if it is desired to obtain a reliable measurement, then the amount Δt of data analyzed must satisfy $\Delta t \Delta f >> 1$, and therefore

$$\Delta t >> \frac{1}{\Delta f} \geqslant \Delta\tau^*.$$

Assume that $\Delta\tau^* = 1$ ms, or $\Delta f^* = 1$ KHz, and $\Delta t \Delta f \cong 100$.

Select appropriate values for the parameters for the following measurement methods:
(a) Temporally smoothed periodogram of tapered data: tapering window width and time-smoothing interval.
(b) Spectrally smoothed periodogram: length of data segment Fourier transformed and width of spectral smoothing window.
(c) Fourier transformed tapered correlogram: length of data segment correlated and width of correlogram-tapering window.
(d) Hopped temporally smoothed periodograms of half-overlapped tapered data segments: width of tapering window and number of periodograms averaged.
(e) Wave analyzer: bandwidths of input BPF and output LPF and total amount of data needed.
(f) Demodulation method: bandwidths of input and output LPFs, and total amount of data needed.
(g) Swept-frequency method: bandwidths of input BPF and output LPF and total

amount of data needed for a sweep rate of $\beta = \Delta f/\Delta t$ Hz/s and an overall frequency range of 100 KHz.

Answers: (a) width = 2 ms, smoothing interval = 200 ms; (d) width = 2 ms, number = 200; (g) input bandwidth = 500 Hz, output bandwidth = 5 Hz, sweep rate = 2500 Hz/s, amount of data = 40 s.

8. Consider an LTI system with unknown transfer function $H(f)$, which is subjected to random excitation $n(t)$ and random measurement noise $m(t)$, as depicted in Figure 4-9. The random waveforms $n(t)$ and $m(t)$ are stationary (i.e., their limit autocorrelations exist) and broadband with limit spectra that are flat (with values of N_0 and M_0, respectively) for $|f| \leq B$, where B is the bandwidth of the LTI system. Also, $n(t)$ and $m(t)$ are uncorrelated with each other. Only $x(t)$ is accessible, but N_0 and M_0 are known.

(a) Propose and describe in detail two distinct methods for measuring $|H(f)|$ with a degree of accuracy that can (theoretically) be made as high as desired. Identify the *effective* spectral smoothing window for each method, and compare them in terms of potential leakage.

(b) Assume that $x(t)$ is available for only Δt seconds, and describe in detail how this limits the accuracy (resolution and reliability) with which $|H(f)|$ can be measured. Discuss how you might go about choosing specific values for parameters (to specify spectral resolution Δf) for your measurement methods.

9. The purpose of this exercise is to develop a systematic approach to solving for the kernel representor $m(t, \tau)$ in the general representation (73)–(74) for time-variant statistical spectra. This is accomplished by first reviewing the basic method for changing variables in a double integral in part (a). Then in part (b), this method is used to solve for the effect of time-averaging and frequency-smoothing on the kernel representor. In parts (c) and (d), the kernel representors for the time-variant periodogram and time-variant pseudospectrum are solved for. Finally in part (e) the results of parts (b) through (d) are combined to obtain explicit formulas for the kernel representors for various specific statistical spectra.

(a) When changing variables in a double integral,

$$K = \int_a^b \int_c^d I[u, v]\, du\, dv,$$

according to the transformation

$$u = p(t, \tau)$$
$$v = q(t, \tau)$$

with inverse

$$t = r(u, v)$$
$$\tau = s(u, v),$$

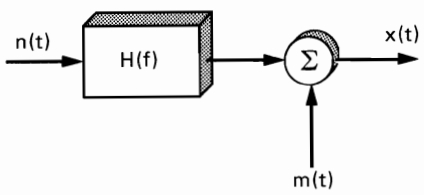


Figure 4-9 System subject to random excitation and random measurement noise.

the area of the incremental volume $dudv$ gets transformed into the area $J(t, \tau)dtd\tau$, where

$$J(t, \tau) = |A(t, \tau)|$$

and $A(t, \tau)$ is the Jacobian determinant

$$A(t, \tau) = \det\begin{bmatrix} \frac{\partial p}{\partial t} & \frac{\partial p}{\partial \tau} \\ \frac{\partial q}{\partial t} & \frac{\partial q}{\partial \tau} \end{bmatrix}.$$

Thus, K becomes

$$K = \int_{a'}^{b'} \int_{c'}^{d'} I[p(t, \tau), q(t, \tau)]\, J(t, \tau)\, dt\, d\tau.$$

For example, for the linear transformation

$$u = p(t, \tau) = \alpha t + \beta\tau + \mu$$
$$v = q(t, \tau) = \gamma t + \sigma\tau + \eta,$$

we have

$$J(t, \tau) = \alpha\sigma - \beta\gamma.$$

Also, since the inverse transformation is given by

$$t = r(u, v) = \alpha' u + \beta' v + \mu'$$
$$\tau = s(u, v) = \gamma' u + \sigma' v + \eta',$$

where α', β', γ', σ', μ', and η' are determined from α, β, γ, σ, μ, and η by matrix inversion, then the limits of integration a', b', c', and d' can be determined from a, b, c, and d. For example, we can first choose the limits a' and b' for τ according to

$$a' = \min_{\substack{u \in [a, b] \\ v \in [c, d]}} s(u, v)$$
$$b' = \max_{\substack{u \in [a, b] \\ v \in [c, d]}} s(u, v),$$

and then solve for c' and d' as functions of τ, using the same minimization and maximization methods on $r[u, v]$ but subject to the constraint $s(u, v) = \tau$ for each $\tau \in [a', b']$. As an exercise, show that for $a = c$, $b = d$, $t = (u + v)/2$, $\tau = v - u$, we obtain $J = 1$, $a' = c - b$, $b' = d - a$, $c' = c - |\tau|/2$, $d' = d + |\tau|/2$. *Hint:* Draw a picture that describes this change of variables as a transformation of coordinates in a plane and shows the region of integration as a rectangle.

(b) Consider the quadratic time-invariant transformation

$$w_f(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_f'(u, v)x(t - u)x(t - v)\, du\, dv,$$

where

$$k_f'(u, v) = m'\left(\frac{u + v}{2}, v - u\right)e^{-i2\pi f(v-u)}.$$

Use the change of variables

$$s = t - \frac{u + v}{2}$$

$$\tau = v - u$$

to show that

$$w_f(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m'(t - s, \tau) e^{-i2\pi f\tau} x\left(s + \frac{\tau}{2}\right) x\left(s - \frac{\tau}{2}\right) ds\, d\tau.$$

Let

$$y_f(t) = w_f(t) \otimes g(t)$$

and show that $y_f(t)$ has the same form of representation as $w_f(t)$ with kernel representor given by

$$m(t, \tau) = m'(t, \tau) \otimes g(t).$$

Similarly let

$$y_f(t) = w_f(t) \otimes H(f)$$

and show that

$$m(t, \tau) = m'(t, \tau) h(\tau)$$

where $h = F^{-1}\{H\}$.

(c) Show that the time-variant periodogram with data-tapering window a_T can be put into the form of $w_f(t)$ in (b) with kernel representor

$$m'(t, \tau) = \frac{1}{T} a_T\left(t + \frac{\tau}{2}\right) a_T\left(t - \frac{\tau}{2}\right).$$

Hint: Use (18) and (20) in Chapter 2.

(d) Show that the time-variant pseudospectrum can be put into the form of $w_f(t)$ in (b) with kernel representor

$$m'(t, \tau) = u_T(t) 2T u_{2T}(\tau).$$

(e) Use the results of (b), (c), and (d) to derive the kernel representors (76), (77), and (78).

10. Show that the kernel specified by (74) and (79) yields the spectrum estimate (80).

11. Sketch the kernel representation $m(t, \tau)$ as a surface above the (t, τ) plane for the following spectrum estimates (assume $\Delta t \Delta f >> 1$):

(a) Temporally smoothed periodogram (1b) with $a_{1/\Delta f} = \frac{1}{\Delta f} u_{1/\Delta f}$ and $g_{\Delta t} = u_{\Delta t}$.

(b) Spectrally smoothed periodogram (2b) with $a_{\Delta t} = \Delta t\, u_{\Delta t}$ and $H_{\Delta f} = u_{\Delta f}$.

(c) The spectrum estimate (80) with $a_{1/\Delta f} = \frac{1}{\Delta f} u_{1/\Delta f}$ and $g_{\Delta t} = u_{\Delta t}$.

(d) Fourier transformed tapered correlogram (21) with $h_{1/\Delta f} = \frac{1}{\Delta f} v_{1/\Delta f}$.

12. Derive the kernel representation $m(t, \tau)$ for the following spectrum estimates:

(a) Temporally smoothed pseudospectrum, (3b).

(b) Spectrally smoothed pseudospectrum, (4b).

(c) Hopped time-averaged periodogram.

(d) Hopped time-averaged pseudospectrum.

(e) Fourier transformed tapered correlogram.

Hint: Wherever possible use exact equivalences to be able to apply results from exercise 9.

13. As an alternative to (11)–(17) in Chapter 3, use the general representation method of Section G to show that

$$S_{x_{1/\Delta f}}(t, f)_{\Delta t} \cong S_{x_{\Delta t}}(t, f)_{\Delta f}$$

for $\Delta t \Delta f >> 1$, by showing that the representor $m(t, \tau)$ is approximately the same for these two methods.

14. As explained in exercise 6(c), although the swept-frequency method of spectral analysis is particularly attractive from an implementation standpoint, it does not use the data efficiently, since for each portion of width Δf of the total spectral band of width B, only a fraction $\Delta f/B$ of the total data segment analyzed is used to measure the spectral content. This is especially problematic for detection of a narrow-band feature that is only intermittently present, since the fraction-of-time probability of measuring the spectrum in the appropriate part of the spectral band when the narrow-band feature is present is proportional to the small fraction $\Delta f/B$. An alternative to the swept-frequency method that avoids this drawback is the following compressive spectral analysis method, which uses a modified swept-frequency analyzer in which the output is compressed in time so that only T units of time are needed to sweep across the total spectral band for a data segment of length T. Specifically, a finite segment of $x(t)$ is both down-converted and up-converted by multiplication with the chirp signal $e^{-i\pi\beta t^2}$ (the instantaneous frequency $f(t) = -\beta t$ is negative for $t > 0$ and positive for $t < 0$), filtered with a quadratic-phase filter, which has impulse-response function $q(t) = e^{i\pi\beta t^2}$, and then again both down-converted and up-converted. Thus, the response of this analyzer is given by

$$y(t) = e^{-i\pi\beta t^2}\{e^{i\pi\beta t^2} \otimes [Tu_T(t)x(t)e^{-i\pi\beta t^2}]\}.$$

Show that this reduces to

$$y(t) = \int_{-T/2}^{T/2} x(u)e^{-i2\pi\beta tu}\, du = X_T(\beta t),$$

which is the complex spectrum evaluated at frequency $f = \beta t$. Hence, if the bandwidth of interest is B and $\beta = B/T$, then the total spectral band is swept across during an interval of length T. A statistical spectrum can be obtained by the method of spectral smoothing, which translates here to temporal smoothing since $f = \beta t$. Specify the temporal-smoothing window width needed for a resolution product of $\Delta t \Delta f = 10$, where $\Delta t = T$. Evaluate this for $T = 1$ ms and B = 1MHz.

15. Consider the statistical spectrum measurement specified by the quadratic time-invariant transformation

$$S(t, f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k_f(u, v)x(t - u)x(t - v)\, du\, dv,$$

where

$$k_f(u, v) = m\left(\frac{u + v}{2}, v - u\right)e^{-i2\pi f(v-u)}$$

$$m(t, \tau) = b\left(t + \frac{\tau}{2}\right)b\left(t - \frac{\tau}{2}\right) \otimes c(t).$$

Show how to implement this spectrum analyzer using only two real BPFs, two squarers, a summor, and one LPF. Specify the impulse-response functions of the filters in terms of $b(t)$ and $c(t)$. Explain how the widths of the impulse-response functions must be related in order to provide a reliable spectrum estimate.

APPENDIX 4-1
Other Wave-Analysis Methods

1. The Fano Identity

Motivated by the simplicity of analog implementation of a first-order (1-pole) LPF, with impulse-response function

$$g_T(t) \triangleq \begin{cases} e^{-2t/T}, & t \geq 0 \\ 0, & t < 0, \end{cases} \tag{1}$$

and second-order (2-pole) BPFs, with impulse-response functions

$$a_T^f(t)_r = \begin{cases} \sqrt{2}\, e^{-t/T}\cos(2\pi ft), & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{2}$$

$$a_T^f(t)_i = \begin{cases} \sqrt{2}\, e^{-t/T}\sin(2\pi ft), & t \geq 0 \\ 0, & t < 0, \end{cases} \tag{3}$$

Robert M. Fano [Fano 1950] proposed the following method of computation of an autocorrelation and a spectrum:

$$R'_x(t, \tau)_T \triangleq g_T(t) \otimes [x(t)x(t - |\tau|)] \tag{4}$$

$$S'_{x_T}(t, f) \triangleq \frac{1}{T}|x'_T(t, f)|^2, \tag{5}$$

where $x'_T(t, f)$ is the local sine wave component

$$x'_T(t, f) \triangleq a_T^f(t) \otimes x(t) \tag{6}$$

$$a_T^f(t) \triangleq \begin{cases} \sqrt{2}\, e^{-t/T}e^{i2\pi ft}, & t \geq 0 \\ 0, & t < 0. \end{cases} \tag{7}$$

Moreover, Fano proved that, analogous to the tapered-autocorrelation methods of Section B, (4) and (5) are related by

$$S'_{x_T}(t, \cdot) = F\{R'_x(t, \cdot)_T e^{-|\cdot|/T}\}, \tag{8}$$

which we shall call the *Fano identity*.

2. The Schroeder-Atal Identity

Generalizing on Fano's work, M. R. Schroeder and B. S. Atal [Schroeder and Atal 1962] defined a statistical spectrum by

$$S'_x(t, \cdot)_{\Delta t \Delta f} \triangleq F\{R'_x(t, \cdot)_{\Delta t} h_{1/\Delta f}(|\cdot|)\}, \tag{9}$$

in which $R'_x(t, \cdot)_{\Delta t}$ is defined by (4), with *arbitrary causal aperture*,

$$g_{\Delta t}(t) = 0, \qquad t < 0, \tag{10}$$

and in which $h_{1/\Delta f}$ is an arbitrary causal aperture,

$$h_{1/\Delta f}(t) = 0, \qquad t < 0. \tag{11}$$

Moreover, they proved that the spectrum (9) is a kind of temporally smoothed

cross periodogram (see Chapter 7),

$$S'_x(t,f)_{\Delta t,\Delta f} = g_{\Delta t}(t) \otimes [x(t)x'_{1/\Delta f}(t,f)_r], \tag{12}$$

in which $x'_T(t, f)_r$ is the real local sine wave component

$$x'_T(t,f)_r \triangleq h^f_T(t)_r \otimes x(t), \tag{13}$$

where

$$h^f_T(t)_r = h_T(t)\cos(2\pi ft). \tag{14}$$

Characterization (12) reveals that the statistical spectrum (9) can be measured by low-pass filtering the product of the data and the band-pass-filtered data instead of the square of the band-pass-filtered data, as in the wave-analysis method of Section E (compare with (80)).

5

FRACTION-OF-TIME PROBABILISTIC ANALYSIS

In this chapter it is shown that the notion of the *degree of randomness* or *variability* of a statistical spectrum can be quantified solely in terms of time-averages, without resort to the popular but more abstract probabilistic model based on a hypothetical ensemble and an associated probability measure. Nevertheless, it is also shown that the mechanics of this quantification are equivalent to those based on an abstract probabilistic model, because the time-averages employed here can be reinterpreted as ensemble averages, or expected values, based on fraction-of-time probability. Thus, the actual techniques developed here are conventional, but the interpretation is unconventional. The results obtained from the fraction-of-time probabilistic analysis are used to obtain general formulas for an effective spectral smoothing window and for variability. It is then explained how these formulas can be used as a basis for evaluating design trade-offs among resolution, leakage, and reliability.

A. MOTIVATION

As explained in Chapter 3, Section A, the fundamental reason for interest in statistical (average) spectra is a belief that interesting aspects of a phenomenon being investigated have spectral influences on the data that are masked by uninteresting random effects and an additional belief that these spectral influences can be revealed by averaging out the random effects, thereby obtaining a statistical spectrum. As an alternative to the temporal-smoothing and spectral-smoothing methods of averaging described in Chapters 3 and 4, for the moment let us consider ensemble averaging as a means for reducing the degree of randomness

of a spectrum. Specifically, consider an ensemble $\{x(t, s)\}$ of *random samples* of time-series from a given phenomenon, for which the ensemble index s is a positive-integer-valued variable. For each ensemble member, we can obtain a nonstatistical spectrum (periodogram)

$$S_{x_T}(t, f, s) \triangleq \frac{1}{T}|X_T(t, f, s)|^2, \tag{1}$$

where

$$X_T(t, f, s) \triangleq \int_{t-T/2}^{t+T/2} x(u, s)e^{-i2\pi fu}\, du. \tag{2}$$

Then a statistical spectrum can be obtained by averaging over the ensemble

$$S_{x_T}(t, f)_M \triangleq \frac{1}{M}\sum_{s=1}^{M} S_{x_T}(t, f, s). \tag{3}$$

An *idealized statistical spectrum* is defined by the limit

$$E\{S_{x_T}(t, f)\} \triangleq \lim_{M\to\infty} S_{x_T}(t, f)_M \tag{4}$$

and can be interpreted (via the *law of large numbers* in probability theory; see [Gardner 1985]) as the *expected value* of the time-variant periodogram. Moreover, by analogy with the definition of the deterministic limit spectrum in Chapter 3, Section C, the *probabilistic limit spectrum* is defined by the limit

$$\mathscr{S}_x(f) \triangleq \lim_{T\to\infty} E\{S_{x_T}(t, f)\}, \tag{5}$$

as the spectral resolution width $\Delta f = 1/T$ becomes infinitesimal.

There are two fundamental motives for interest in ensemble averaging and resultant probabilistic spectra. The first motive is a desire to reduce random effects without forfeiting temporal or spectral resolution. That is, since the ensemble-averaged statistical spectrum $S_{x_T}(t, f)_M$ is obtained without smoothing over either t or f, then the fundamental time-frequency uncertainty principle (Chapter 2, Section C) is the only constraint on temporal and spectral resolution. Specifically, $\Delta t = T$ and $\Delta f \cong 1/T$ and, therefore,

$$\Delta t \Delta f \cong 1, \tag{6}$$

regardless of the amount M of averaging. The second motive is a desire to apply probabilistic methods of analysis to spectral analysis problems. For example, it might be desirable to obtain a probabilistic quantification of the *degree of randomness* of a statistical spectrum or the accuracy of approximation between each of various temporally or spectrally smoothed statistical spectra and the limit spectrum.

In the case of the first motive, it is required that an ensemble of time-series from the phenomenon under study be physically available. There are many applications in which ensembles are (or can be made) available, because the underlying experiment is repeatable. For example, in exploration seismology, a given region of earth can be repeatedly shaken. However there are also many applications for which ensembles are not (and cannot be made) available. For example, in studies of dynamic ocean-wave spectra from a volcanic explosion at a given spatial and temporal locale, no ensemble of oceans or explosions is

available. Similarly, in studies of dynamic seismographic spectra from an earthquake phenomenon at a given temporal and spatial locale, no ensemble of earths or quakes is available. Also, for meteorological time-series analysis of weather dynamics associated with a particular storm at a given temporal and spatial locale, no ensemble of atmospheres is available.

In the case of the second motive, it is not required that an ensemble be physically available. All that is needed is the *concept* of an ensemble and a mathematical model for its probabilistic description. Nevertheless, it is far more intuitively satisfying if the ensemble can be conceived of as corresponding to time-series that can be made physically available. It is this situation that is pursued in this chapter, namely, the situation for which probabilistic analysis is the motive and for which an ensemble of time-series can be made physically available. However, this ensemble will be derived from a single time-series. As briefly explained in Part 2 of Section B, Chapter 1, and expanded upon in the following section, this can be accomplished for a time-series representing data from a time-invariant phenomenon.

In Chapter 8, the complementary situation of statistical spectral analysis for time-variant phenomena based on ensemble averaging, according to the first motive, is pursued.

Although the general approach taken here is equally valid for both continuous time and discrete time, the results in these two cases are analogous, and it is therefore more efficient to treat only one case. In keeping with the preceding chapters, the continuous-time case is chosen, but the necessary modifications for discrete time are explained in the following chapter.

B. FRACTION-OF-TIME PROBABILISTIC MODEL

A *probabilistic model* of a time-series is a mathematical specification of a probability law governing a hypothetical ensemble of random samples of time-series and is called a *stochastic process* or a *random process*. We adopt the empirical approach proposed for signal-processing problems involving time-invariant phenomena by Donald G. Brennan [Brennan 1961] and developed by Edward M. Hofstetter [Hofstetter 1964]. In this approach, the conceptualization of an appropriate ensemble of random samples is a trivial matter for constant phenomena. One simply envisions the members of the ensemble to be all time-translated versions of a single persistent time-series,

$$x(t, s) = x(t + s), \tag{7}$$

as discussed in Part 2 of Section B, Chapter 1. The probability law can be envisioned as being derived from relative frequencies of occurrence of events, so that probability distributions are really *fraction-of-time distributions*.

For example, the *probability distribution* for the amplitude of a time-series is formally defined by

$$F_x(y) \triangleq \textit{probability that } x(t) < y \tag{8}$$

$$= \lim_{T_0 \to \infty} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} U[y - x(t)]\, dt, \tag{9}$$

for which $U(\cdot)$ is the unit-step function

$$U(y) \triangleq \begin{cases} 1, & y > 0 \\ 0, & y \leq 0. \end{cases} \tag{10}$$

In (8), t is interpreted as a random sample of time. In (9), $U[y - x(t)]$ is the *indicator* of the event $x(t) < y$, and (9) is the relative frequency (fraction of time) of occurrence of this event. That is, the integral adds up the lengths of time intervals for which the event $x(t) < y$ occurs and this is divided by the length of the overall interval.[1]

When probability law is conceived of in this way, it is unquestionably *stationary* and *ergodic*, as discussed in Part 2 of Section B, Chapter 1. That is, expected values (ensemble averages) are time-invariant and are equal to time averages. For example, the *expected value* of the amplitude is defined by

$$E\{x(t)\} \triangleq \int_{-\infty}^{\infty} y \, dF_x(y) = \int_{-\infty}^{\infty} y \frac{dF_x(y)}{dy} \, dy, \tag{11}$$

for which $dF_x(y)/dy$ is the *probability density* of the amplitude [Gardner 1985]. Substitution of (9) into (11) and application of the identity

$$\frac{dU(y)}{dy} = \delta(y) \tag{12}$$

yields

$$E\{x(t)\} = \langle x(t) \rangle \triangleq \lim_{T_0 \to \infty} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t) \, dt. \tag{13}$$

Similarly, the *probabilistic autocorrelation* defined (in terms of a second-order joint probability density) by

$$\mathcal{R}_x(\tau) \triangleq E\left\{x\left(t + \frac{\tau}{2}\right)x\left(t - \frac{\tau}{2}\right)\right\} \tag{14}$$

is time-invariant and identical (exercise 11) to the deterministic (time-average) autocorrelation

$$\hat{R}_x(\tau) = \left\langle x\left(t + \frac{\tau}{2}\right)x\left(t - \frac{\tau}{2}\right)\right\rangle \triangleq \lim_{T_0 \to \infty} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x\left(t + \frac{\tau}{2}\right)x\left(t - \frac{\tau}{2}\right) dt. \tag{15}$$

Thus, it is clear that the mapping (7) from an individual time-series to an ensemble corresponding to a stochastic process, known as *H. O. A. Wold's isomorphism* for the discrete-time case [Wold 1948], can be viewed (at least heuristically) as being responsible for the duality between deterministic and probabilistic theories

[1] In order for this double-sided limit to yield an appropriate model that properly reflects time-invariance, it must be assumed that the two single-sided limits,

$$\lim_{T_0 \to \infty} \frac{1}{T_0} \int_{0}^{T_0} (\cdot) \, dt = \lim_{T_0 \to \infty} \frac{1}{T_0} \int_{-T_0}^{0} (\cdot) \, dt,$$

are identical. Otherwise, the theory can be based on a single-sided limit (e.g., for phenomena with finite starting times).

of time-series from time-invariant phenomena. (For a development of this duality, see [Brillinger 1975] for discrete time and [Gardner 1985] for continuous time.)

The most commonly used probabilistic model for time-series is the Gaussian random process, which is defined to be a random process for which all time-samples are jointly Gaussian random variables. Specifically, for every positive integer N and every N time-points $t_1, t_2, \ldots, t_N$, the samples $x(t_1), x(t_2), \ldots, x(t_N)$ are jointly Gaussian in the sense that every linear combination of these samples is a Gaussian random variable (see [Rao 1973; Gardner 1985]). That is, in the fraction-of-time framework adopted here, for every N real numbers $\omega_1, \omega_2, \ldots, \omega_N$, the variable

$$z(t) \triangleq \omega_1 x(t + t_1) + \omega_2 x(t + t_2) + \cdots + \omega_N x(t + t_N)$$

has a fraction-of-time probability density of the form

$$\frac{dF_z(y)}{dy} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right], \tag{16}$$

for some real number μ (the *mean*) and some positive real number σ (the *standard deviation*, or square root of the *variance*). Based on this definition, it is a simple matter to show that the *joint characteristic function* for the N samples $\boldsymbol{x}(t) \triangleq [x(t + t_1), x(t + t_2), x(t + t_3), \ldots, x(t + t_N)]'$, which is defined by

$$\Psi_{\boldsymbol{x}}(\boldsymbol{\omega}) \triangleq \langle \exp\{i\boldsymbol{\omega}'\boldsymbol{x}(t)\}\rangle, \tag{17}$$

is given by (exercise 18)

$$\Psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \exp\{i\boldsymbol{\omega}'\hat{\boldsymbol{m}}_x - \tfrac{1}{2}\boldsymbol{\omega}'\hat{\boldsymbol{K}}_x\boldsymbol{\omega}\}, \tag{18a}$$

where

$$\hat{\boldsymbol{m}}_x \triangleq \langle \boldsymbol{x}(t)\rangle \tag{18b}$$

$$\hat{\boldsymbol{K}}_x \triangleq \langle [\boldsymbol{x}(t) - \hat{\boldsymbol{m}}_x][\boldsymbol{x}(t) - \hat{\boldsymbol{m}}_x]'\rangle \tag{18c}$$

and $\boldsymbol{\omega} \triangleq [\omega_1, \omega_2, \omega_3, \ldots, \omega_N]'$. From this joint characteristic function, all probabilistic parameters of interest can be obtained, including the joint fraction-of-time probability density (exercise 18). An important property of a zero-mean Gaussian random process, which is used in the next section, is *Isserlis' formula* for the fourth joint moment in terms of the second joint moments [Isserlis 1918] (see also [Gardner 1985]). The fraction-of-time version of this formula is

$$\begin{aligned}
&\langle x(t + t_1)x(t + t_2)x(t + t_3)x(t + t_4)\rangle \\
&\quad \triangleq \lim_{T_0\to\infty} \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t + t_1)x(t + t_2)x(t + t_3)x(t + t_4)\, dt \\
&\quad = \langle x(t + t_1)x(t + t_2)\rangle\langle x(t + t_3)x(t + t_4)\rangle \\
&\qquad + \langle x(t + t_1)x(t + t_3)\rangle\langle x(t + t_2)x(t + t_4)\rangle \\
&\qquad + \langle x(t + t_1)x(t + t_4)\rangle\langle x(t + t_2)x(t + t_3)\rangle.
\end{aligned} \tag{19}$$

Although no particular probabilistic model is needed for the study of the mean and variance of the complex spectrum and the mean of the spectrum and

statistical spectrum, the study of the variance of the spectrum and the statistical spectrum is substantially more tractable for the Gaussian model than for other models. However, as discussed in the sequel, the results obtained for the Gaussian model are indeed representative for many other models and in fact are asymptotically ($\Delta t \Delta f \to \infty$) equivalent to the results obtainable for a large class of non-Gaussian models.

C. BIAS AND VARIABILITY

In the motivating example in Chapter 3, Section A, it is shown that the periodogram $S_{x_T}(t, f)$ is an erratic function of frequency f, and as the time interval of analysis is made longer by increasing T, the periodogram becomes even more erratic. Furthermore, the particular erratic function does not represent spectral influences of the underlying time-invariant phenomenon, because this particular function also changes erratically with time t (for time changes exceeding T). In addition, in Chapter 3, Section C, it is asserted that the limit of $S_{x_T}(t, f)$ as $T \to \infty$ does not even exist. However, it is also shown that not only does the temporally smoothed or spectrally smoothed periodogram, $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ or $S_{x_{\Delta t}}(t, f)_{\Delta f}$, become a *less* erratic function of f as the amount of smoothing (as measured by the resolution product $\Delta t \Delta f$) is increased, but it also becomes a less erratic function of t for arbitrarily large time changes (e.g., changes greatly exceeding Δt), and the limit spectrum obtained from $\Delta t \to \infty$ is a completely nonrandom function that represents only the spectral influences of the phenomenon, that is, all random effects are averaged out in the limit. The purpose of this section is to prove these assertions by means of a type of probabilistic analysis that is based entirely on time averages, as explained in the previous two sections.

This study of *degree of randomness* or *variability* of spectral density measurements is based on two probabilistic parameters, the *coefficient of variation* (squared), which is the variance normalized by the squared mean,

$$r_S \triangleq \frac{\text{var}\{S_{x_T}(t, f)\}}{[\text{mean}\{S_{x_T}(t, f)\}]^2}, \tag{20}$$

and the *correlation coefficient*, which is the covariance normalized by the variances,

$$\rho_S \triangleq \frac{\text{cov}\{S_{x_T}(t, f_1), S_{x_T}(t, f_2)\}}{[\text{var}\{S_{x_T}(t, f_1)\}\text{var}\{S_{x_T}(t, f_2)\}]^{1/2}}. \tag{21}$$

Analogous definitions of r_S and ρ_S apply to the statistical spectra $S_{x_T}(t, f)_{\Delta t}$ and $S_{x_T}(t, f)_{\Delta f}$. We shall say that a spectral density measurement is *nonrandom as a function of t* if and only if $r_S = 0$ and that it is *highly random as a function of t* if r_S is not much smaller than unity. We shall also say that it is *completely random as a function of f* if and only if $\rho_S = 0$ for all $f_1 \neq \pm f_2$. In general, we shall say that the *reliability* of a spectral density measurement increases as its degree of randomness in both t and f decreases.

Because of the fraction-of-time type of probabilistic model that we have adopted, the probabilistic parameters (20) and (21) can be expressed in terms of

time-averages only. Specifically,

$$\text{mean}\{S_{x_T}(t,f)\} \triangleq \langle S_{x_T}(t,f)\rangle = \lim_{T_0\to\infty}\frac{1}{T_0}\int_{-T_0/2}^{T_0/2} S_{x_T}(t,f)\,dt, \tag{22}$$

$$\text{var}\{S_{x_T}(t,f)\} \triangleq \langle |S_{x_T}(t,f) - \langle S_{x_T}(t',f)\rangle|^2\rangle$$

$$= \lim_{T_0\to\infty}\frac{1}{T_0}\int_{-T_0/2}^{T_0/2} |S_{x_T}(t,f) - \langle S_{x_T}(t',f)\rangle|^2\,dt, \tag{23}$$

$$\text{cov}\{S_{x_T}(t,f_1), S_{x_T}(t,f_2)\} \triangleq \langle [S_{x_T}(t,f_1) - \langle S_{x_T}(t',f_1)\rangle][S_{x_T}(t,f_2) - \langle S_{x_T}(t',f_2)\rangle]^*\rangle$$

$$= \lim_{T_0\to\infty}\frac{1}{T_0}\int_{-T_0/2}^{T_0/2} [S_{x_T}(t,f_1) - \langle S_{x_T}(t',f_1)\rangle]$$

$$\times [S_{x_T}(t,f_2) - \langle S_{x_T}(t',f_2)\rangle]^*\,dt. \tag{24}$$

(In (23) and (24), t and t' represent distinct dummy variables over which averaging is carried out.) But we shall also consider a few frequency-averages, to augment the argument based on (20) and (21).

For the purposes of this section, it is assumed that the phenomenon of interest is *time-invariant* in the sense that 1) the limit autocorrelation (15) exists and is not identically zero and 2) the phenomenon exhibits no periodicity in the sense that the time-series $x(t)$ and the lag product time-series $x(t + \tau/2) \times x(t - \tau/2)$ (for each τ) contain no finite additive sine wave components, so that the following periodicity parameters (see Part II, Chapter 10) vanish:

$$\lim_{T_0\to\infty}\frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)e^{-i2\pi\alpha t}\,dt = 0, \qquad \alpha \neq 0 \tag{25}$$

$$\lim_{T_0\to\infty}\frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x\left(t + \frac{\tau}{2}\right)x\left(t - \frac{\tau}{2}\right)e^{-i2\pi\alpha t}\,dt \equiv 0, \qquad \alpha \neq 0. \tag{26}$$

It is assumed that the convergence in (25) is in the temporal mean square sense. In this case it can be (and is in Part II) shown that the limit autocorrelation $\hat{R}_x(\tau)$ contains no finite additive sine wave components. Furthermore, it is assumed that $\hat{R}_x(\tau)$ is absolutely integrable so that its Fourier transform $\hat{S}_x(f)$ exists and that (25) holds for $\alpha = 0$ so that $x(t)$ has zero mean value and, therefore, $\hat{S}_x(f)$ contains no dirac deltas.

As explained in Part II, Chapter 15, when (25) or (26) is violated the results on variability presented in this section must be modified. In fact the Gaussian model assumed here cannot be valid if (26) is violated (as proven in Chapter 15).

1. The Finite-Time Complex Spectrum

We consider first the normalized time-variant finite-time complex spectrum for tapered data,

$$\overline{X}_T(t,f) \triangleq \frac{1}{\sqrt{T}}X_T(t,f), \tag{27}$$

where $X_T(t, f)$ is defined by (11) in Chapter 2. It is easily shown using (25)

(exercise 1) that the temporal mean is zero for all T and all $f \neq 0$:

$$\text{mean}\{\overline{X}_T(t,f)\} = 0, \qquad f \neq 0, \tag{28}$$

regardless of the mean of $x(t)$. It can also be shown (exercise 1) that the temporal variance is given by

$$\text{var}\{\overline{X}_T(t,f)\} = \hat{S}_x(f) \otimes \frac{1}{T}|A_{1/T}(f)|^2, \tag{29}$$

where $A_{1/T}$ is the Fourier transform of the data-tapering window a_T. As T is increased without bound, we obtain the limit

$$\lim_{T\to\infty} \text{var}\{\overline{X}_T(t,f)\} = \gamma \hat{S}_x(f), \tag{30}$$

where γ is the limiting area of the smoothing window in (29) (see (24)–(25) in Chapter 3), which is typically on the order of unity for a unity-height tapering window a_T. Hence, the erratic fluctuation in t of the normalized complex spectrum does not decrease (i.e., the variance does not decrease) as T is increased. Furthermore, the temporal correlation coefficient of $\overline{X}_T(t, f_1)$ and $\overline{X}_T(t, f_2)$ can be shown using (26) (exercise 1) to vanish for all T and $f_1 \neq f_2$, since

$$\text{cov}\{\overline{X}_T(t,f_1), \overline{X}_T(t,f_2)\} = 0, \qquad f_1 \neq f_2. \tag{31}$$

Hence, the normalized complex spectrum is completely random as a function of f. Moreover, for a rectangle data-tapering window $a_T = Tu_T$, the finite autocorrelation in f is (exercise 1)

$$\int_{-\infty}^{\infty} \overline{X}_T\left(t, f + \frac{\nu}{2}\right) \overline{X}_T^*\left(t, f - \frac{\nu}{2}\right) df = \frac{1}{T}\int_{t-T/2}^{t+T/2} x^2(u) e^{-i2\pi\nu u}\, du, \tag{32}$$

which vanishes in the limit as $T \to \infty$ (because of (26)) for all $\nu \neq 0$. (The same is true for any window a_T.)

2. The Finite-Time Spectrum

It can be shown (exercise 2) that the temporal mean of the time-variant finite-time spectrum

$$S_{x_T}(t,f) = |\overline{X}_T(t,f)|^2 \tag{33}$$

for tapered data is given by

$$\text{mean}\{S_{x_T}(t,f)\} = \hat{S}_x(f) \otimes \frac{1}{T}|A_{1/T}(f)|^2. \tag{34}$$

Also, for a zero-mean Gaussian time-series $x(t)$, Isserlis' formula (19) can be used to show (exercise 2) that the temporal variance of the spectrum is given by

$$\begin{aligned}\text{var}\{S_{x_T}(t,f)\} &= [\text{mean}\{S_{x_T}(t,f)\}]^2 \\ &\quad + \left[\frac{1}{T}\int_{-\infty}^{\infty} \hat{S}_x(\nu) A_{1/T}(f-\nu) A_{1/T}(f+\nu)\, d\nu\right]^2.\end{aligned} \tag{35}$$

As T is increased without bound, (20), (34), and (35) yield (exercise 2)

$$\lim_{T\to\infty} r_S = 1 + \delta_f, \tag{36}$$

in which δ_f is the Kronecker delta

$$\delta_f \triangleq \begin{cases} 1, & f = 0 \\ 0, & f \neq 0. \end{cases} \tag{37}$$

Hence, the erratic fluctuation in t of the spectrum does not decrease as T is increased. In fact the limit

$$\lim_{T\to\infty} S_{x_T}(t, f) \tag{38}$$

does not even exist in the temporal mean square sense. This can be established as follows. The mean of $S_{x_T}(t, f)$ converges ($T \to \infty$) to $\hat{S}_x(f)$, as revealed by (34), but the limit ($T \to \infty$) of the temporal mean-square deviation of $S_{x_T}(t, f)$ about its limit mean is given by the nonzero quantity

$$\begin{aligned}
\lim_{T\to\infty} \langle [S_{x_T}(t,f) - \hat{S}_x(f)]^2 \rangle &= \lim_{T\to\infty} \langle [S_{x_T}(t,f)]^2 \rangle - [\hat{S}_x(f)]^2 \\
&= \lim_{T\to\infty} \langle [S_{x_T}(t,f)]^2 \rangle - [\lim_{T\to\infty} \langle S_{x_T}(t',f) \rangle]^2 \\
&= \lim_{T\to\infty} \{ \langle [S_{x_T}(t,f)]^2 \rangle - [\langle S_{x_T}(t',f) \rangle]^2 \} \\
&= \lim_{T\to\infty} \langle [S_{x_T}(t,f) - \langle S_{x_T}(t',f) \rangle]^2 \rangle \\
&= \lim_{T\to\infty} \operatorname{var}\{S_{x_T}(t,f)\} \neq 0.
\end{aligned} \tag{39}$$

Isserlis' formula (19) can again be used to show (exercise 2) that the temporal covariance of $S_{x_T}(t, f_1)$ and $S_{x_T}(t, f_2)$ is given by[2]

$$\begin{aligned}
\operatorname{cov}\{S_{x_T}(t,f_1), S_{x_T}(t,f_2)\} = &\left[\frac{1}{T}\int_{-\infty}^{\infty} \hat{S}_x(\nu) A_{1/T}(f_1 - \nu) A_{1/T}(\nu - f_2)\, d\nu\right]^2 \\
&+ \left[\frac{1}{T}\int_{-\infty}^{\infty} \hat{S}_x(\nu) A_{1/T}(f_1 - \nu) A_{1/T}(\nu + f_2)\, d\nu\right]^2.
\end{aligned} \tag{40}$$

It follows (exercise 2) from (40) and (35) that the temporal correlation coefficient is very small for $|f_1 - f_2| >> 1/T$ and $|f_1 + f_2| >> 1/T$,

$$\rho_S << 1, \qquad |f_1 \pm f_2| >> \frac{1}{T}, \tag{41}$$

and vanishes for $f_1 \neq \pm f_2$ in the limit as $T \to \infty$:

$$\lim_{T\to\infty} \rho_S = 0, \qquad f_1 \neq \pm f_2. \tag{42}$$

[2] Within the conventional probabilistic framework, the results (35) and (40) are good approximations for a broad class of *non-Gaussian* time-series, provided that T is sufficiently large. This is a well-known result from asymptotic distribution theory [Grenander and Rosenblatt 1984; Hannan 1970]. For example, (35) is asymptotically correct for any non-Gaussian time-series for which $x(t_1)$ and $x(t_2)$ are statistically independent for all t_1 and t_2 such that $|t_1 - t_2| > T_*$ for some finite T_*. It follows from Wold's isomorphism that equivalent results can be obtained for a single time-series.

Hence, the spectrum approaches a completely random function of f as T is increased without bound.

The finite-autocorrelation in f is

$$\int_{-\infty}^{\infty} S_{x_T}\left(t, f + \frac{\nu}{2}\right) S_{x_T}\left(t, f - \frac{\nu}{2}\right) df = \int_{-\infty}^{\infty} [R_{x_T}(t, \tau)]^2 e^{-i2\pi\nu\tau}\, d\tau, \tag{43}$$

but its limit as $T \to \infty$ apparently does not exist by analogy with (38), which can be reexpressed as

$$\lim_{T\to\infty} \int_{-\infty}^{\infty} R_{x_T}(t, \tau) e^{-i2\pi f\tau}\, d\tau. \tag{44}$$

3. Statistical Spectra

In direct contrast to the increasingly erratic behavior with increasing T of the nonstatistical spectrum $S_{x_T}(t, f)$ just described, we shall find in this section that the spectrally smoothed statistical spectrum $S_{x_T}(t, f)_{\Delta f}$ exhibits a degree of randomness that *decreases* as the resolution product $\Delta t \Delta f = T\Delta f$ is increased by increasing Δt and that *vanishes* in the limit as $\Delta t \Delta f \to \infty$ for any $\Delta f > 0$ no matter how small. We shall also find that the same is true for the temporally smoothed spectrum $S_{x_T}(t, f)_{\Delta t}$ for which $\Delta t \Delta f = \Delta t/T$. In fact this result will be obtained for a large class of statistical spectra.

There are actually two statistical characteristics of spectrum estimates that are of interest: the bias and the variability. If the mean of an estimate does not equal the quantity being estimated (the limit spectrum in this case), the estimate is said to be *biased*. The bias is a particularly convenient characterization of both the spectral resolution and spectral leakage properties of a given spectrum estimation method (when variability is sufficiently small). In particular, we shall see that the temporal means of the time-variant spectrum estimates, to be denoted by $y_f(t)$, can be expressed in the form

$$\text{mean}\{y_f(t)\} = \hat{S}_x(f) \otimes E(f), \tag{45}$$

where $E(f)$ is interpreted as an *effective spectral smoothing window* whose mainlobe width determines the spectral resolution, and whose sidelobe heights determine the spectral leakage (when variability is sufficiently small). We shall also find that in the limit as $\Delta f \to 0$, the effective window $E(f)$ approaches an impulse which reveals that resolution becomes perfect and leakage vanishes, yielding an asymptotically unbiased estimate (assuming $E(f)$ has unity area in the limit). The term *fidelity* is sometimes used for the degree to which the mean of a spectrum estimate approximates the ideal limit spectrum. Thus, high fidelity means low bias.

The most concise and transparent approach to the study of both bias and variability of the variety of statistical spectra obtained from the various methods described in Chapter 4 is based on the general representation for statistical spectra introduced in Chapter 4, Section G,

$$y_f(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_f(u, v) x(t - u) x(t - v) du\, dv, \tag{46}$$

where the kernel is represented by

$$k_f(u, v) = m\left(\frac{u + v}{2}, v - u\right)e^{-i2\pi f(v-u)}, \tag{47}$$

in which $m(t, \tau)$ is independent of f. Substitution of (47) into (46) and use of a change of variables yields

$$y_f(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(w, \tau)x\left(t - w + \frac{\tau}{2}\right)x\left(t - w - \frac{\tau}{2}\right)e^{-i2\pi f\tau}\, dw\, d\tau, \tag{48}$$

which is the starting point for our study.

Effective spectral smoothing window

The temporal mean of a time-variant statistical spectrum denoted by $y_f(t)$ and represented by (48) is easily shown (exercise 6) to be[3]

$$\text{mean}\{y_f(t)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(t, \tau)\, dt\, \hat{R}_x(\tau)e^{-i2\pi f\tau}\, d\tau \tag{49}$$

$$= \hat{S}_x(f) \otimes E(f), \tag{50}$$

where

$$E(f) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(t, \tau)\, dt\, e^{-i2\pi f\tau}\, d\tau = M(0, -f), \tag{51}$$

in which $M(\nu, \mu)$ is the double Fourier transform of $m(t, \tau)$,

$$M(\nu, \mu) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(t, \tau)e^{-i2\pi(\nu t - \mu\tau)}\, dt\, d\tau. \tag{52}$$

The effective spectral window $E(f)$ can be expressed explicitly in terms of the windows that define a particular statistical spectrum, as illustrated in Chapter 4. Specifically, for the spectrally smoothed periodogram for tapered data and for the Fourier transform of the tapered correlogram, it is shown in Chapter 4, Section G, that

$$m(t, \tau) = \frac{1}{\Delta t}\, a_{\Delta t}\left(t + \frac{\tau}{2}\right)a_{\Delta t}\left(t - \frac{\tau}{2}\right)h_{1/\Delta f}(\tau), \tag{53}$$

from which it follows that

$$E(f) = \frac{1}{\Delta t}\,|A_{1/\Delta t}(f)|^2 \otimes H_{\Delta f}(f) \tag{54a}$$

$$\cong \gamma\, H_{\Delta f}(f), \qquad \text{for } \Delta t\Delta f \gg 1, \tag{54b}$$

where $A_{1/\Delta t}$ is the Fourier transform of the data-tapering window, $H_{\Delta f}$ is the spectral smoothing window, or $h_{1/\Delta f}$ is the autocorrelation-tapering window, and γ is the data-tapering window parameter defined by (33) in Chapter 2. For the temporally smoothed periodogram for tapered data and the statistical spectra obtained from the wave-analysis and demodulation methods, it is shown in Chapter

[3] This result, (49)–(51), does not require (25). It is equally valid if $x(t)$ contains finite additive sine wave components, in which case $\hat{S}_x(f)$ contains spectral lines.

4, Section G, that

$$m(t, \tau) = \Delta f \int_{-\infty}^{\infty} a_{1/\Delta f}\left(t - w + \frac{\tau}{2}\right) a_{1/\Delta f}\left(t - w - \frac{\tau}{2}\right) g_{\Delta t}(w)\, dw, \tag{55}$$

from which it follows that

$$E(f) = \Delta f |A_{\Delta f}(f)|^2 G_{1/\Delta t}(0), \tag{56}$$

where $A_{\Delta f}$ is the Fourier transform of the data-tapering window or the predetection filter transfer function for the demodulation method and $G_{1/\Delta t}$ is the Fourier transform of the temporal smoothing window or the postdetection filter transfer function. If this latter window has unity area, then $G_{1/\Delta t}(0) = 1$. For any statistical spectrum that is representable by (48) and admits the separable approximation

$$m(t, \tau) \cong g_{\Delta t}(t) h_{1/\Delta f}(\tau), \tag{57}$$

(which is discussed in Chapter 4, Section G), the effective spectral smoothing window is given by

$$E(f) \cong G_{1/\Delta t}(0) H_{\Delta f}(f). \tag{58}$$

This includes (54b) and (56) as examples.

Coefficient of variation

The temporal variance of the time-series $y_f(t)$ is equal to the temporal mean-squared value of the centered time-series

$$z_f(t) \triangleq y_f(t) - \text{mean}\{y_f(t)\}, \tag{59}$$

and it follows from the inverse of the Wiener relation,

$$\hat{R}_z(\tau) = \int_{-\infty}^{\infty} \hat{S}_z(\nu) e^{i2\pi\nu\tau}\, d\nu,$$

that the temporal mean-squared value of $z_f(t)$, namely $\hat{R}_z(0)$, is given by the integral over all frequencies of the limit spectrum $\hat{S}_z$. The limit spectrum of the time-series $z_f(t)$ specified by (59) and (48), which represents a large class of statistical spectra, takes on a particularly tractable form for zero-mean Gaussian time-series $x(t)$. Specifically, it can be shown (exercise 7) that

$$\hat{S}_{z_f}(\nu) = \int_{-\infty}^{\infty} [|M(\nu, \mu - f)|^2 + M(\nu, \mu - f) M^*(\nu, -\mu - f)] \times \hat{S}_x\left(\mu + \frac{\nu}{2}\right) \hat{S}_x\left(\mu - \frac{\nu}{2}\right) d\mu, \tag{60}$$

where $M(\nu, \mu)$ is the transform (52) of the kernel $m(t, \tau)$ in the representation (48). The transformed kernels for the various spectrum analyzers described in Chapter 4 are given in Table 5-1. In those cases for which $m(t, \tau)$ can be approximated by the separable form (57) (e.g., this is typically a close approximation for $\Delta t \Delta f >> 1$; see Chapter 4, Section G), the transform $M(\nu, \mu)$ can also be approximated by the separable form

$$M(\nu, \mu) \cong G_{1/\Delta t}(\nu) H_{\Delta f}(-\mu). \tag{61}$$

TABLE 5-1 SPECTRAL ANALYSIS KERNELS*

Spectrum estimate	Kernel	Kernel transform	Spectrum analyzer
$S_{x_T}(t, f)$	$m_T(t, \tau) = \frac{1}{T}a_T\left(t + \frac{\tau}{2}\right)a_T\left(t - \frac{\tau}{2}\right)$	$M_{1/T}(\nu, \mu) = \frac{1}{T}A_{1/T}\left(\frac{\nu}{2} + \mu\right)A_{1/T}\left(\frac{\nu}{2} - \mu\right)$	Periodogram ($T = 1/\Delta f$ or Δt)
$S_x(t, f)_T$	$m(t, \tau)_T = u_T(t)2Tu_{2T}(\tau)$	$M(\nu, \mu)_{1/T} = \frac{1}{T}w_{1/T}(\nu)w_{1/2T}(\mu)$	Pseudospectrum ($T = 1/\Delta f$ or Δt)
$S_{x_{1/\Delta f}}(t, f)_{\Delta t}$	$m_{1/\Delta f}(t, \tau) \otimes g_{\Delta t}(t)$	$M_{\Delta f}(\nu, \mu)G_{1/\Delta t}(\nu)$	Temporally smoothed periodogram, wave analyzer, demodulation wave analyzer, hopped time-averaged periodogram ($g_{\Delta t} = \eta_{\Delta t,\Delta f}$).
$S_{x_{\Delta t}}(t, f)_{\Delta f}$	$m_{\Delta t}(t, \tau)h_{1/\Delta f}(\tau)$	$M_{1/\Delta t}(\nu, \mu) \otimes H_{\Delta f}(-\mu)$	Spectrally smoothed periodogram, Fourier transformed tapered correlogram.
$S_x(t, f)_{1/\Delta f,\Delta t}$	$m(t, \tau)_{1/\Delta f} \otimes g_{\Delta t}(t)$	$M(\nu, \mu)_{\Delta f}G_{1/\Delta t}(\nu)$	Temporally smoothed pseudospectrum, hopped time-averaged pseudospectrum ($g_{\Delta t} = \eta_{\Delta t,\Delta f}$).
$S_x(t, f)_{\Delta t,\Delta f}$	$m(t, \tau)_{\Delta t}h_{1/\Delta f}(\tau)$	$M(\nu, \mu)_{1/\Delta t} \otimes H_{\Delta f}(-\mu)$	Spectrally smoothed pseudospectrum, Fourier transformed tapered finite-average autocorrelation.

* $g_{\Delta t}$ is an arbitrary temporal smoothing window; $\eta_{\Delta t,\Delta f}$ is the comb window given by (28) in Chapter 2; $H_{\Delta f}$ is an arbitrary spectral smoothing window; a_T is an arbitrary data-tapering window; u_T and $w_{1/T}$ are the rectangle and sinc windows given by (12) and (14) in Chapter 2.

In such cases, (60) can be approximated by (exercise 7)

$$\hat{S}_{z_f}(\nu) \cong |G_{1/\Delta t}(\nu)|^2 \int_{-\infty}^{\infty} [|H_{\Delta f}(f - \mu)|^2 + H_{\Delta f}(f - \mu)H^*_{\Delta f}(f + \mu)] \times \hat{S}_x\left(\mu + \frac{\nu}{2}\right)\hat{S}_x\left(\mu - \frac{\nu}{2}\right) d\mu. \tag{62}$$

Furthermore, if the width $1/\Delta t$ is small enough and the sidelobes of $G_{1/\Delta t}$ are low enough to resolve the fine structure in $\hat{S}_x$, then (62) yields (exercise 7) the close approximation

$$\hat{S}_{z_f}(\nu) \cong |G_{1/\Delta t}(\nu)|^2 \int_{-\infty}^{\infty} [|H_{\Delta f}(f - \mu)|^2 + H_{\Delta f}(f - \mu)H^*_{\Delta f}(f + \mu)][\hat{S}_x(\mu)]^2 d\mu. \tag{63}$$

Moreover, if Δf is also small enough and the sidelobes of $H_{\Delta f}$ are low enough to resolve the fine structure in $\hat{S}_x$, and if $H_{\Delta f}$ is real and even, then (63) yields

(exercise 7) the close approximation

$$\hat{S}_{z_f}(\nu) \cong |G_{1/\Delta t}(\nu)|^2 \, [\hat{S}_x(f)]^2 \, [r_{H_{\Delta f}}(0) + r_{H_{\Delta f}}(2f)], \tag{64}$$

where $r_{H_{\Delta f}}(f)$ is the finite autocorrelation of $H_{\Delta f}(f)$,

$$r_{H_{\Delta f}}(f) \triangleq \int_{-\infty}^{\infty} H_{\Delta f}\left(\mu + \frac{f}{2}\right) H_{\Delta f}\left(\mu - \frac{f}{2}\right) d\mu. \tag{65}$$

It follows from (64) that $z_f(t)$ is a low-pass process whose spectral shape is determined primarily by the output LPF transfer function $G_{1/\Delta t}$ and whose spectral intensity is determined primarily by the gains of the output LPF and input BPF and the spectral density $\hat{S}_x(f)$ being estimated. The spectral density of the spectrum estimate $y_f(t)$ is the same as that for $z_f(t)$ except for the presence of a spectral line at zero frequency, which has area $[\text{mean}\{y_f(t)\}]^2$ determined primarily by $\hat{S}_x(f)$.

The temporal variance of the spectrum estimate $y_f(t)$ is obtained by integration of the spectral density for $z_f(t) = y_f(t) - \text{mean}\{y_f(t)\}$, and it follows from (60) that this temporal variance is given by

$$\begin{aligned} \text{var}\{y_f(t)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} & [|M(\nu, \mu - f)|^2 + M(\nu, \mu - f) M^*(\nu, -\mu - f)] \\ & \times \hat{S}_x\left(\mu + \frac{\nu}{2}\right) \hat{S}_x\left(\mu - \frac{\nu}{2}\right) d\mu \, d\nu. \end{aligned} \tag{66}$$

Since $m(t, \tau)$ is in general a two-dimensional oblong pulselike function with widths of Δt and $1/\Delta f$ in t and τ, respectively, then its double Fourier transform $M(\nu, \mu)$ is also, in general, a two-dimensional oblong pulselike function, and its widths are on the orders of $1/\Delta t$ and Δf in ν and μ, respectively. Thus, if the width $1/\Delta t$ is small enough and the corresponding sidelobes are low enough to resolve the fine structure in $\hat{S}_x$, then (66) yields (exercise 7) the close approximation

$$\begin{aligned} \text{var}\{y_f(t)\} \cong \int_{-\infty}^{\infty} \Bigg[\int_{-\infty}^{\infty} & |M(\nu, \mu - f)|^2 \\ & + M(\nu, \mu - f) M^*(\nu, -\mu - f) \, d\nu \Bigg] [\hat{S}_x(\mu)]^2 \, d\mu. \end{aligned} \tag{67}$$

Furthermore, if the width Δf is also small enough and the corresponding sidelobes are low enough to resolve the fine structure in $\hat{S}_x$, then (67) yields (exercise 7) the close approximation

$$\text{var}\{y_f(t)\} \cong L(f)[\hat{S}_x(f)]^2, \tag{68}$$

where

$$L(f) \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [|M(\nu, \mu)|^2 + M(\nu, \mu - f) M^*(\nu, -\mu - f)] \, d\nu \, d\mu. \tag{69}$$

Moreover, if $M(\nu, \mu)$ can be approximated by the separable form (61) in which $H_{\Delta f}$ is real and even, then (69) is approximated by

$$L(f) \cong \int_{-\infty}^{\infty} |G_{1/\Delta t}(\nu)|^2 \, d\nu \, [r_{H_{\Delta f}}(0) + r_{H_{\Delta f}}(2f)], \tag{70}$$

where $r_{H_{\Delta f}}$ is given by (65).

To obtain the temporal coefficient of variation, (66) is divided by the square of (50). By use of approximation (68) and the similar approximation to (50),

$$\text{mean}\{y_f(t)\} \cong \hat{S}_x(f) \int_{-\infty}^{\infty} E(\mu)\, d\mu, \tag{71}$$

the following approximation to the coefficient of variation for $y_f(t)$ is obtained:

$$r_{y_f} \triangleq \frac{\text{var}\{y_f(t)\}}{|\text{mean}\{y_f(t)\}|^2} \cong \frac{L(f)}{\left[\int_{-\infty}^{\infty} E(\mu)\, d\mu\right]^2} \triangleq R(f), \qquad \frac{1}{\Delta t} < \Delta f < \Delta f^*, \tag{72}$$

where Δf^* is the resolution width of $\hat{S}_x(f)$. This approximation, which is specified by

$$R(f) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [|M(\nu, \mu)|^2 + M(\nu, \mu - f)M^*(\nu, -\mu - f)]\, d\nu\, d\mu}{\left[\int_{-\infty}^{\infty} M(0, \mu)\, d\mu\right]^2}, \tag{73}$$

reduces to

$$R(f) \cong \frac{\int_{-\infty}^{\infty} |G_{1/\Delta t}(\nu)|^2\, d\nu\, [r_{H_{\Delta f}}(0) + r_{H_{\Delta f}}(2f)]}{|G_{1/\Delta t}(0)|^2 \left|\int_{-\infty}^{\infty} H_{\Delta f}(\mu)\, d\mu\right|^2} \tag{74}$$

when the separable approximation (61) in which $H_{\Delta f}$ is real and even is used. By introducing the factor

$$\eta \triangleq \left[\frac{\Delta t\, \|G_{1/\Delta t}\|^2}{|G_{1/\Delta t}(0)|^2}\right]\left[\frac{\Delta f\, \|h_{1/\Delta f}\|^2}{[h_{1/\Delta f}(0)]^2}\right], \tag{75a}$$

where

$$\|G_{1/\Delta t}\|^2 \triangleq \int_{-\infty}^{\infty} |G_{1/\Delta t}(\nu)|^2\, d\nu \tag{75b}$$

$$\|h_{1/\Delta f}\|^2 \triangleq \int_{-\infty}^{\infty} h_{1/\Delta f}^2(\tau)\, d\tau, \tag{75c}$$

and using the approximation and inequality

$$r_{H_{\Delta f}}(2f) \begin{cases} \cong r_{H_{\Delta f}}(0), & |f| << \Delta f \\ << r_{H_{\Delta f}}(0), & |f| >> \Delta f, \end{cases} \tag{76}$$

(72) and (74) can be reexpressed as

$$r_{y_f} \cong \frac{\eta}{\Delta t \Delta f}[1 + \delta_{\Delta f}(f)], \qquad \text{for } \Delta t \Delta f >> 1 \quad \text{and} \quad \Delta f < \Delta f^*, \tag{77a}$$

where Δf^* is the resolution width of $\hat{S}_x$, $\delta_{\Delta f}$ is a type of indicator function that is positive and less than or approximately equal to unity (for all f) and satisfies

$$\delta_{\Delta f}(f) \begin{cases} \cong 1, & |f| << \Delta f \\ << 1, & |f| >> \Delta f, \end{cases} \tag{77b}$$

and η is in general on the order of unity, regardless of Δt and Δf (exercise 12).

For example, if $h_{1/\Delta f}$ and $G_{1/\Delta t}$ are rectangle or sinc windows, $\eta = 1$. The reason the coefficient of variation is doubled in the vicinity of $f = 0$ is that when the spectral window $H_{\Delta f}$ is centered at $f = 0$, half of the spectral component magnitudes (those for $f > 0$) are perfectly correlated with the other half (those for $f < 0$) and, therefore, when they are summed by the frequency smoothing operation, their variance is twice as large as it would be if they were all uncorrelated.

The explicit relationships between the window functions in (75) and those used in the various specific statistical spectra can be determined from Table 5-1 (see also Chapter 4, Section G and exercise 9). In conclusion, we see that (under the stated assumptions) the temporal coefficient of variation is independent of $\hat{S}_x(f)$ and is inversely proportional to the resolution product $\Delta t \Delta f$, and the proportionality constant is on the order of unity and is twice as large at f near zero. Moreover, this result applies to all the various methods for spectrum estimation described in Chapter 4. Furthermore, it can be shown that for the discrete-time counterparts to all these continuous-time spectrum estimates, the result is the same as (77), except that in (75) the integral over frequency ranges over $[-\frac{1}{2}, \frac{1}{2}]$ rather than $(-\infty, \infty)$, and the integral over time becomes a discrete sum over time. This is explained further in Chapter 6. Moreover, the result (77) is valid for a large class of non-Gaussian time-series as well as the Gaussian time-series on which its derivation is based. This follows from the discussion in the footnote in Part 2 of Section C. The reader is reminded that if assumption (26) is violated (as it often is when dealing with modulated signals; see Chapter 3, Section D), then the well-known result (77) must be modified to incorporate additional terms as explained in Chapter 15 of Part II.

Since the temporal variance approaches zero as $\Delta t \to \infty$ for any $\Delta f > 0$, then the statistical spectrum $y_f(t)$ approaches its own temporal mean value, in the temporal-mean-square sense:

$$\underset{\Delta t \to \infty}{\text{l.i.m.}}\ y_f(t) = \text{mean}\{y_f(t)\}. \tag{78}$$

Therefore, the erratic behavior in t of the statistical spectrum vanishes in the limit as $\Delta t \to \infty$. Thus, regardless of how erratic in f the true spectral influences of the phenomenon as reflected in $\hat{S}_x(f)$ might be, the statistical spectrum (e.g., $y_f(t) = S_{x_{\Delta t}}(t, f)_{\Delta f}$) closely approximates it for *all* t in the temporal-mean-square sense, provided that Δt is sufficiently large and Δf is sufficiently small. In contrast, the nonstatistical spectrum $S_{x_T}(t, f)$ becomes increasingly erratic as T is increased, in the sense that $S_{x_T}(t_1, f)$ and $S_{x_T}(t_2, f)$ do not closely approximate each other for $|t_1 - t_2| > T$ and neither closely approximates $\hat{S}_x(f)$, and furthermore $S_{x_T}(t, f_1)$ and $S_{x_T}(t, f_2)$ are essentially uncorrelated for $|f_1 - f_2| > 1/T$.

Explanation of variability

A popular approach to explaining why the variance of the smoothed periodogram $S_{x_{\Delta t}}(t, f)_{\Delta f}$ approaches zero as $\Delta t \to \infty$, when the variance of the raw (unsmoothed) periodogram $S_{x_{\Delta t}}(t, f)$ does not approach zero as $\Delta t \to \infty$, is based on the formulas

$$S_{x_{\Delta t}}(t, f) = \int_{-\Delta t}^{\Delta t} R_{x_{\Delta t}}(t, \tau) e^{-i2\pi f\tau}\, d\tau \tag{79}$$

$$S_{x_{\Delta t}}(t, f)_{\Delta f} = \int_{-\Delta t}^{\Delta t} R_{x_{\Delta t}}(t, \tau) h_{1/\Delta f}(\tau) e^{-i2\pi f\tau}\, d\tau \tag{80}$$

$$\cong \int_{-1/2\Delta f}^{1/2\Delta f} R_{x_{\Delta t}}(t, \tau) e^{-i2\pi f\tau}\, d\tau \tag{81}$$

$$R_{x_{\Delta t}}(t, \tau) = \frac{1}{\Delta t} \int_{-(\Delta t - |\tau|)/2}^{(\Delta t - |\tau|)/2} x\left(t + u + \frac{\tau}{2}\right) x\left(t + u - \frac{\tau}{2}\right) du. \tag{82}$$

Specifically, it is argued (e.g., [Koopmans 1974, p. 266; Papoulis 1977, pp. 380, 383; Papoulis 1984, p. 491]) that no matter how large Δt is, $R_{x_{\Delta t}}(t, \tau)$ in (82) is highly unreliable for $|\tau|$ near Δt because the length $\Delta t - |\tau|$ of the averaging interval $[-(\Delta t - |\tau|)/2, (\Delta t - |\tau|)/2]$ is small; furthermore, the highly unreliable portion of $R_{x_{\Delta t}}(t, \tau)$ is omitted from the Fourier transform integral (81) that produces $S_{x_{\Delta t}}(t, f)_{\Delta f}$ (for $\Delta t >> 1/\Delta f$) but is included in the integral (79) that produces $S_{x_{\Delta t}}(t, f)$. Unfortunately, this explanation is insufficient. One reason is that although the mean-squared error between $\hat{R}_x(\tau)$ and $R_{x_{\Delta t}}(t, \tau)$ approaches the large value $\hat{R}_x(\tau)^2$ as $|\tau|$ approaches Δt, the variance of $R_{x_{\Delta t}}(t, \tau)$ is proportional to $1/\Delta t$ for all $|\tau| \leq \Delta t$ and therefore does approach zero as $|\Delta t| \to \infty$. The real reason that the variance of $S_{x_{\Delta t}}(t, f)$ approaches a nonzero constant is that the function $R_{x_{\Delta t}}(t, \tau)$ for $|\tau| \leq \Delta t$ can be interpreted as $\Delta t/\Delta\tau^*$ contiguous segments of length $\Delta\tau^*$ in τ, and only adjacent segments and mirror images about $\tau = 0$ are nonnegligibly correlated ($\Delta\tau^*$ is the width of $\hat{R}_x(\tau)$); thus when these segments are summed in the Fourier transform integral (79), their variances, each of which is proportional to $\Delta\tau^*/\Delta t$, add up to produce a variance that does not decrease as Δt increases (exercise 3). The fact that the poor quality of the estimate $R_{x_{\Delta t}}(t, \tau)$ of $\hat{R}_x(\tau)$ for $|\tau|$ near Δt is not responsible for the large variance of the periodogram $S_{x_{\Delta t}}(t, f)$ is confirmed by the fact that the variance of the pseudospectrum

$$S_x(t, f)_{\Delta t} = \int_{-\Delta t}^{\Delta t} R_x(t, \tau)_{\Delta t} e^{-i2\pi f\tau}\, d\tau \tag{83}$$

$$R_x(t, \tau)_{\Delta t} = \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} x\left(t + u + \frac{\tau}{2}\right) x\left(t + u - \frac{\tau}{2}\right) du, \tag{84}$$

is just as large, and the estimate $R_x(t, \tau)_{\Delta t}$ is of no poorer quality for $|\tau|$ near Δt than it is for $|\tau| << \Delta t$; that is, the mean-squared error (and variance) of $R_x(t, \tau)_{\Delta t}$ is proportional to $1/\Delta t$ for all $|\tau| \leq \Delta t$, but the coefficient of variation for $S_x(t, f)_{\Delta t}$ is asymptotically ($\Delta t \to \infty$) the same as it is for $S_{x_{\Delta t}}(t, f)$, namely (36) (see exercise 5).

Example: ***White Noise***

To illustrate the decrease in degree of randomness of a statistical spectrum with an increase in the resolution product, the spectra of simulated white Gaussian noise for three analysis intervals of lengths $T = \Delta t = 128, 512, 2048$, are shown in Figure 5-1. In Figure 5-1(a)–(c), $\Delta f = 1/T = 1/\Delta t$, yielding the nonstatistical periodogram, and the result is an increase in degree of randomness with an increase in Δt. But in Figures 5-1(d)–(f), Δf is fixed at $\frac{1}{32}$, yielding a spectrally smoothed periodogram with resolution products of $\Delta t \Delta f = 4$, 16, and 64, and the result is a decrease in degree of randomness with an increase in Δt. To illustrate the facts that the nonstatistical

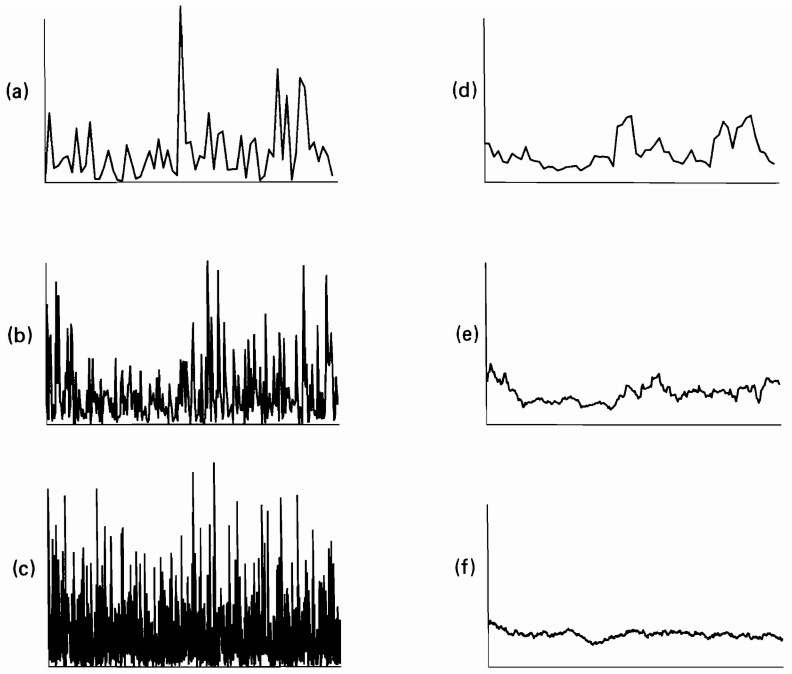


Figure 5-1 Illustration of dependence of degree of randomness on length Δt of data segment and on length of frequency-smoothing interval Δf. (a) $\Delta t = 128$, $\Delta t \Delta f = 1$. (b) $\Delta t = 512$, $\Delta t \Delta f = 1$. (c) $\Delta t = 2048$, $\Delta t \Delta f = 1$. (d) $\Delta t = 128$, $\Delta t \Delta f = 4$. (e) $\Delta t = 512$, $\Delta t \Delta f = 16$. (f) $\Delta t = 2048$, $\Delta t \Delta f = 64$.

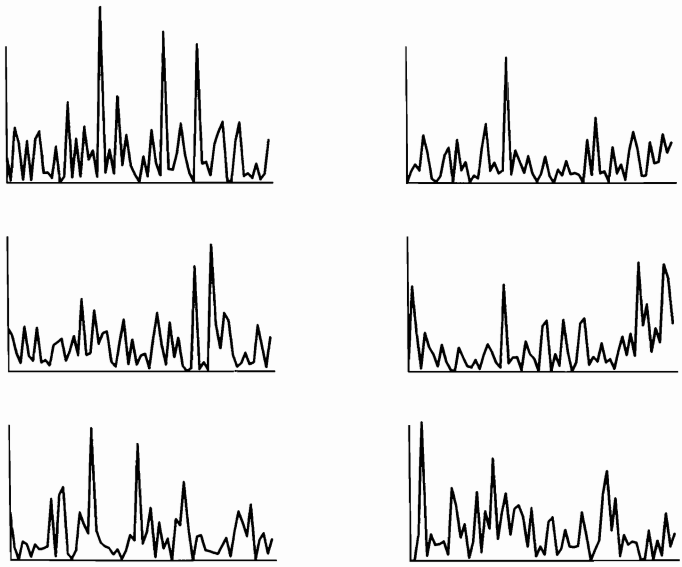


Figure 5-2 Periodograms of various segments of white Gaussian noise.

periodogram fluctuates randomly with time as well as frequency and that the degree of randomness in frequency is reduced by time averaging, the nonstatistical periodograms for six adjacent segments of white Gaussian noise data are shown in Figure 5-2, and their averages are shown in Figure 5-3(a)–(h).

Example: ***Hidden Periodicities and Disturbed Harmonics***

To demonstrate the lack of reliability of nonstatistical spectra of random data, consider the problem of trying to distinguish between two possible phenomena, one of which gives rise to periodic components masked by additive white noise, the other of which gives rise to white noise filtered by a narrow-band BPF. The former phenomenon yields data with *hidden periodicities* (its limit spectrum contains spectral lines). The latter yields data that might be called *disturbed harmonics*—the data does not contain any periodicities (its limit spectrum does not contain spectral lines). Two segments of data $y(t)$ from *one* of these two phenomena are shown in Figure 5-4(a) and (b). The nonstatistical spectra (periodograms) for these two data segments are shown in Figure 5-4(c) and (d). The several significant peaks in these

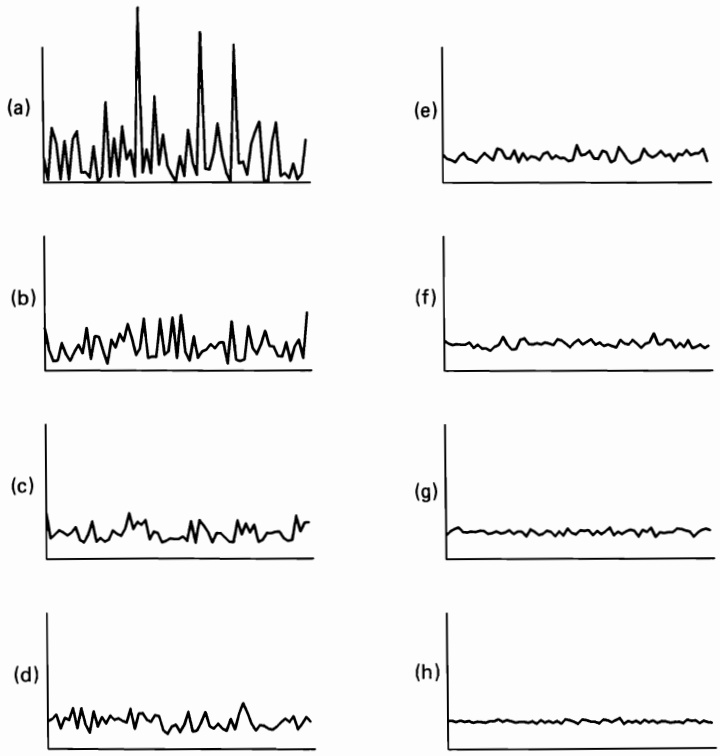
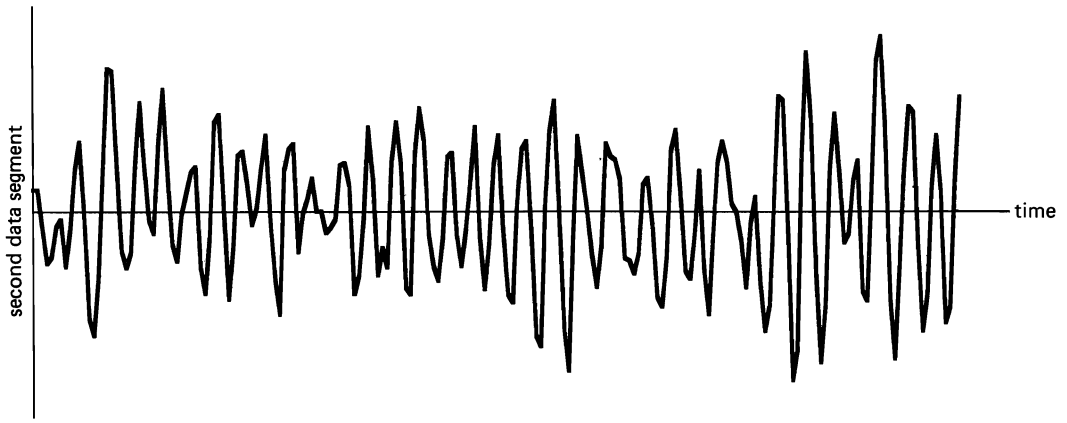


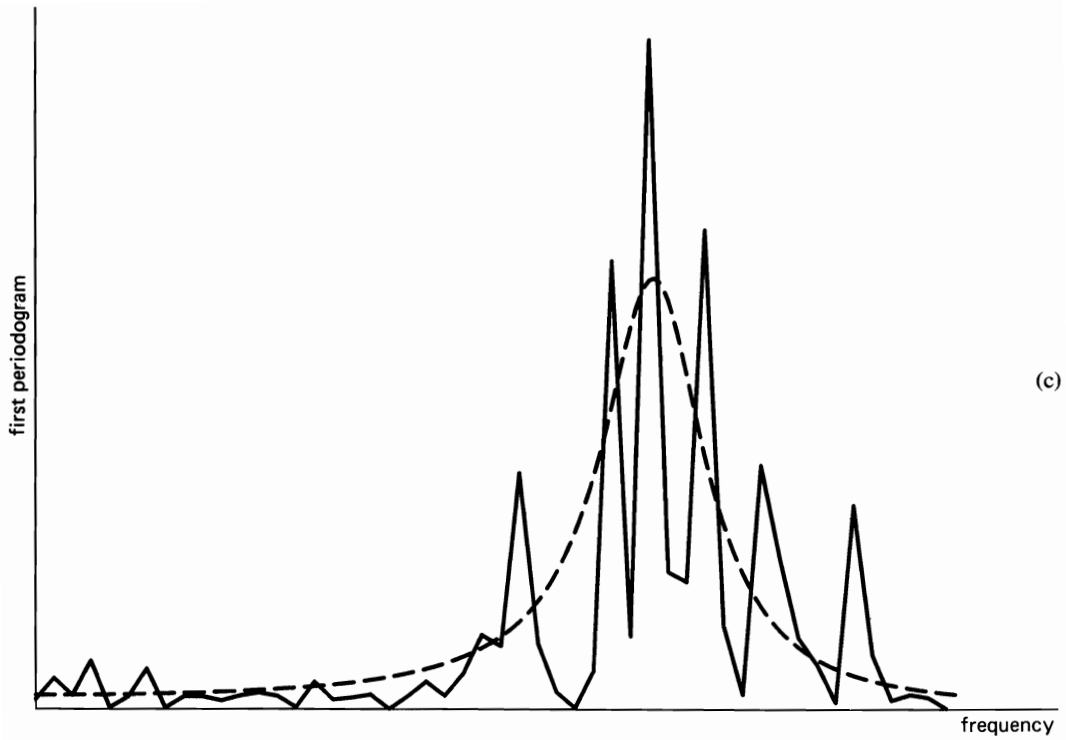
Figure 5-3 Averages of periodograms from various segments of white Gaussian noise (number of periodograms averaged is N): (a) $N = 1$, (b) $N = 4$, (c) $N = 8$, (d) $N = 16$, (e) $N = 32$, (f) $N = 64$, (g) $N = 128$, (h) $N = 256$.

(a)

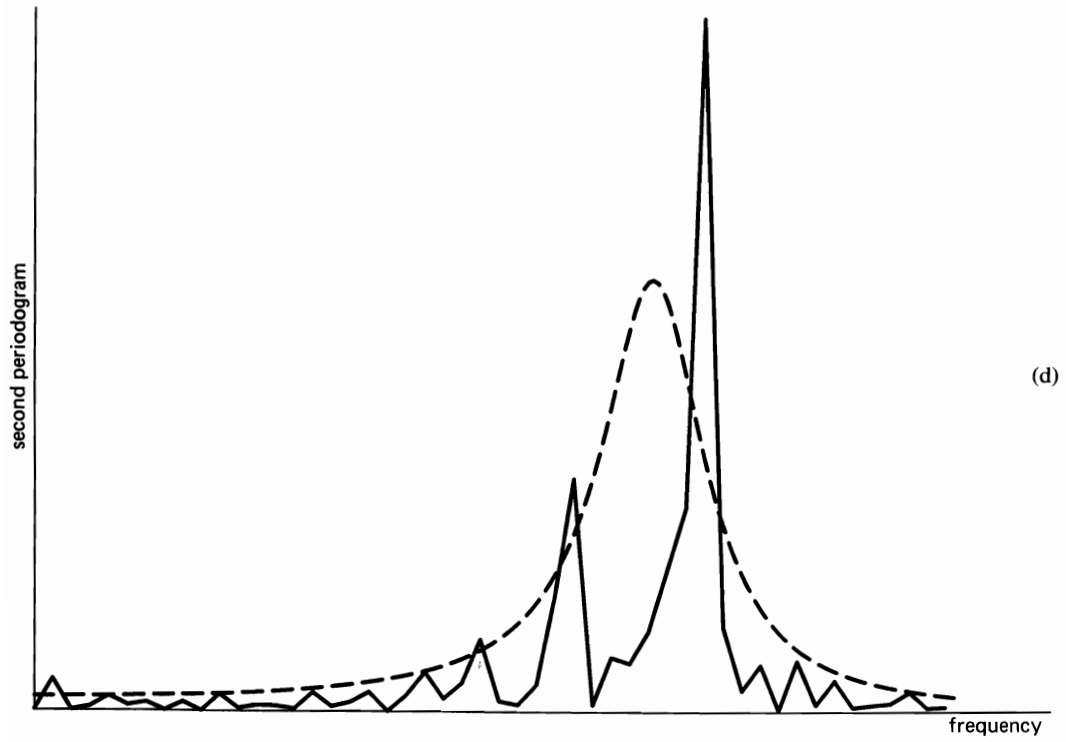


(b)

Figure 5-4 (a), (b) Two segments of data from one phenomenon. (c), (d) Periodograms of the two data segments shown in (a) and (b) (broken curve is the limit spectrum).



(c)



(d)

Figure 5-4 (continued)

two spectra would seem to suggest that there are indeed hidden periodicities; however, the frequencies of these hypothetical periodicities suggested by one spectrum are different from those suggested by the other spectrum. As a matter of fact, there are *no* hidden periodicities. The data is band-pass filtered white noise. The magnitude-squared transfer function $|H(\cdot)|^2$ of the BPF is shown superimposed on the spectra in Figure 5-4(c) and (d), and this is the limit spectrum that is approached by a statistical spectrum with sufficiently fine spectral resolution Δf and sufficiently long averaging time $\Delta t >> 1/\Delta f$ (see Figures 3-1 to 3-3). On the other hand, if it is known that the data consists of periodic components in additive white noise, then the normalized complex spectrum is indeed useful if enough data is available. This can be seen from the fact that for the model

$$x(t) = \sum_{\alpha} m_\alpha e^{i2\pi\alpha t} + n(t) \tag{85}$$

where $n(t)$ contains no periodic components, the temporal mean of the normalized complex spectrum is given by (exercise 1)

$$\text{mean}\{\overline{X}_T(t,f)\} = \sqrt{T} \sum_{\alpha} m_\alpha \delta_{f-\alpha} \tag{86}$$

where $\delta_{f-\alpha}$ is the Kronecker delta, and the temporal variance is given by (exercise 1)

$$\text{var}\{\overline{X}_T(t,f)\} = \hat{S}_n(f) \otimes z_{1/T}(f) + \sum_{\alpha} |m_\alpha|^2 \left[\frac{1}{T} w_{1/T}(f-\alpha) - \delta_{f-\alpha}\right]^2. \tag{87}$$

Thus, as the length T of the data segment is increased, the coefficient of variation (20) decreases inversely proportional to T for $f = \alpha$ for each of the periodicity frequencies α.

4. Time-Frequency Uncertainty Condition

In Chapter 2, Section C, it is established that for a nonstatistical spectrum, that is, a time-variant periodogram, the product of temporal and spectral resolutions is on the order of unity,

$$\Delta t \Delta f \cong 1, \tag{88}$$

regardless of the particular data-tapering aperture. This principle imposes *a limit on the simultaneous resolvability of temporal and spectral characteristics of the DATA*. By comparison, in this section it is established that for a statistical spectrum obtained, for example, from either temporal or spectral smoothing of the time-variant periodogram, the product of temporal and spectral resolutions must greatly exceed unity,

$$\Delta t \Delta f >> 1, \tag{89}$$

regardless of the particular (effective) spectral smoothing window, if the statistical spectrum is to be *reliable,* that is, if it is to exhibit a low degree of randomness as reflected in the size of its coefficient of variation (77). This condition[4] imposes

[4] This condition (89) is referred to as *Grenander's uncertainty condition,* in honor of Ulf Grenander's pioneering work in probabilistic analysis for spectral analysis of time-series [Grenander 1951, 1958].

a limit on the reliable simultaneous resolvability of temporal and spectral characteristics of the PHENOMENON (which are masked in the data by random effects). If the phenomenon is time-invariant so that temporal resolution is irrelevant, then Δt simply represents the total amount of data available for analysis, and (89) indicates that the spectrum is reliable only if the spectral resolution width greatly exceeds the reciprocal of the length of the total amount of data analyzed. If higher spectral resolution is desired for a fixed amount of data, reliability must be forfeited.

As an alternative to the term *reliability* for the degree of variability of a statistical spectrum, the term *stability* is sometimes used. Also, the term *degrees of freedom* is commonly used in discussions of stability because for $\Delta t \Delta f >> 1$, it can be shown that for a large class of time-series models a statistical spectrum has an approximate (exact for Gaussian time-series) chi-squared probability density (fraction-of-time density in the nonprobabilistic framework adopted in this book) with $\nu = 2\Delta t \Delta f / \eta$ degrees of freedom. Therefore, there is a one-to-one correspondence between the coefficient of variation, r_{y_f}, as given by (77), and the number of degrees of freedom, ν, that is,

$$\nu = 2r_{y_f} \quad \text{for} \quad |f| >> \Delta f.$$

The results in Chapter 4 show that when condition (89) is satisfied, most of the many alternative methods for continuous-time measurement of statistical spectra yield approximately the same spectrum. Consequently, the spectral analysis design problem is particularly challenging when (89) cannot be satisfied because the desired spectral resolution width does not greatly exceed the reciprocal of the total amount of data available (or the desired temporal resolution width). This is a problem to which considerable research effort has been devoted during the last several decades and is still being devoted under the title of *high-resolution spectral analysis*, as discussed in Chapter 9.

In Chapter 2, Section G, it is shown that the periodogram $S_{x_T}(t, f)$ is the local-average power spectral density of $x(t)$ in the sense that the instantaneous power in the spectral band $[f - \Delta f/2, f + \Delta f/2]$, averaged over the time-interval $[t - \Delta t/2, t + \Delta t/2]$ at locale t, is given to a close approximation by

$$P_{\Delta t \Delta f}(t, f) \cong \int_{f-\Delta f/2}^{f+\Delta f/2} S_{x_{\Delta t}}(t, \nu)\, d\nu \tag{90}$$

if (89) holds. Moreover, we now see that if (89) holds, then the spectrally smoothed periodogram (90) closely approximates the correspondingly smoothed limit spectrum $\hat{S}_x(f)$, which is shown in Chapter 3, Section C, to be precisely the average (over all time) power spectral density of $x(t)$. Thus, the condition (89), which assures that $P_{\Delta t \Delta f}(t, f)$ can be appropriately interpreted as the local-average power in the spectral band $[f - \Delta f/2, f + \Delta f/2]$, also assures that $P_{\Delta t \Delta f}(t, f)$ is—to a close approximation—the *global*-average power in the spectral band $[f - \Delta f/2, f + \Delta f/2]$ for a constant phenomenon. Furthermore, (89) is not only a sufficient condition for this latter interpretation of $P_{\Delta t \Delta f}(t, f)$, it is also a necessary condition for a broad class of time-series data.

D. RESOLUTION, LEAKAGE, AND RELIABILITY: DESIGN TRADE-OFFS

As revealed in Chapter 4, each method for continuous-time measurement of a statistical spectrum has an effective spectral smoothing window, and as explained in Part 3 of Section C, this window can be identified from the formula for the mean of the measured spectrum (see exercise 9). Furthermore, both the spectral resolution and the potential spectral leakage for each measurement method can be determined from the main-lobe width and the sidelobe heights of the effective spectral smoothing window provided that variability is sufficiently low. Moreover, the multiplicative coefficient η in formula (77) for the coefficient of variation for the statistical spectrum is determined by the effective spectral smoothing window (and the effective temporal window) through (75). Consequently, all three of the major performance parameters—resolution, leakage, and reliability—are determined at least in part (to a close approximation for $\Delta t \Delta f >> 1$) by the effective spectral smoothing window. For some applications, such as those for which spectral lines are present, there are actually two spectral leakage performance parameters of interest, namely, the heights of sidelobes close to the main lobe (which are typically the highest) and the rate of decay of the heights of the sidelobes when their distance from the main lobe is large.

Many different windows (more than 25) have been proposed and evaluated in terms of these four (and related) performance parameters. A few of these are described here. Since special window designs are most conveniently implemented digitally, the windows described here are defined for discrete time. Both the effective spectral smoothing windows and their inverse Fourier transforms, the effective correlation-tapering windows, are described—the latter first.

Rectangle (*Dirichlet*):

$$h(\tau) = \begin{cases} 1, & |\tau| \leq T/2 \\ 0, & |\tau| > T/2 \end{cases} \tag{91a}$$

Triangle (*Bartlett* or *Fejér*):

$$h(\tau) = \begin{cases} 1 - \dfrac{2|\tau|}{T}, & |\tau| \leq T/2 \\ 0, & |\tau| > T/2 \end{cases} \tag{92a}$$

Raised Cosine (*von Hann*):

$$h(\tau) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\dfrac{2\pi\tau}{T}\right)\right], & |\tau| \leq T/2 \\ 0, & |\tau| > T/2 \end{cases} \tag{93a}$$

Raised Cosine on a Platform (*Hamming*):

$$h(\tau) = \begin{cases} 0.54 + 0.46\cos\left(\dfrac{2\pi\tau}{T}\right), & |\tau| \leq T/2 \\ 0, & |\tau| > T/2 \end{cases} \tag{94a}$$

Blackman:

$$h(\tau) = \begin{cases} 0.42 + 0.50\cos\left(\frac{2\pi\tau}{T}\right) + 0.08\cos\left(\frac{4\pi\tau}{T}\right), & |\tau| \leq T/2 \\ 0, & |\tau| > T/2. \end{cases} \tag{95a}$$

These windows are identical for discrete and continuous time. The corresponding Fourier-series transforms (using the sampling increment $T_s = 1$) are as follows:

Dirichlet (*Rectangle*):

$$H(f) = \frac{\sin(\pi f T)}{\sin(\pi f)}, \qquad |f| \leq \frac{1}{2} \tag{91b}$$

Bartlett or *Fejér* (*Triangle*):

$$H(f) = \frac{2}{T}\left[\frac{\sin(\pi f T/2)}{\sin(\pi f)}\right]^2, \qquad |f| \leq \frac{1}{2} \tag{92b}$$

von Hann (*Raised Cosine*):

$$H(f) = \frac{1}{2}\frac{\sin(\pi f T)}{\sin(\pi f)} + \frac{1}{4}\frac{\sin \pi T(f - 1/T)}{\sin \pi(f - 1/T)} + \frac{1}{4}\frac{\sin \pi T(f + 1/T)}{\sin \pi(f + 1/T)}, \qquad |f| \leq \frac{1}{2} \tag{93b}$$

Hamming (*Raised Cosine on a Platform*):

$$H(f) = 0.54\frac{\sin(\pi f T)}{\sin \pi f} + 0.23\frac{\sin \pi T(f - 1/T)}{\sin \pi(f - 1/T)} + 0.23\frac{\sin \pi T(f + 1/T)}{\sin \pi(f + 1/T)}, \qquad |f| \leq \frac{1}{2} \tag{94b}$$

Blackman:

$$H(f) = 0.42\frac{\sin(\pi f T)}{\sin(\pi f)} + 0.25\frac{\sin \pi T(f - 1/T)}{\sin \pi(f - 1/T)} + 0.25\frac{\sin \pi T(f + 1/T)}{\sin \pi(f + 1/T)} + 0.04\frac{\sin[(\pi T/2)(f + 2/T)]}{\sin \pi(f + 2/T)} + 0.04\frac{\sin[(\pi T/2)(f - 2/T)]}{\sin \pi(f - 2/T)}, \qquad |f| \leq \frac{1}{2}. \tag{95b}$$

For $T >> 1$, these spectral windows are essentially the same as the Fourier transforms of the continuous-time counterparts for $|f| \leq 1/2$.

In addition to these five spectral windows, there is the rectangle spectral window,

$$H(f) = \begin{cases} T, & |f| \leq 1/2 \\ 0, & |f| > 1/2, \end{cases} \tag{96}$$

which is referred to as the *Daniell window*. The four spectral windows (91b)–(94b) are shown in Figure 5-5. Observe how much smaller the sidelobes are for the second two compared with the first two. The Hamming window is designed

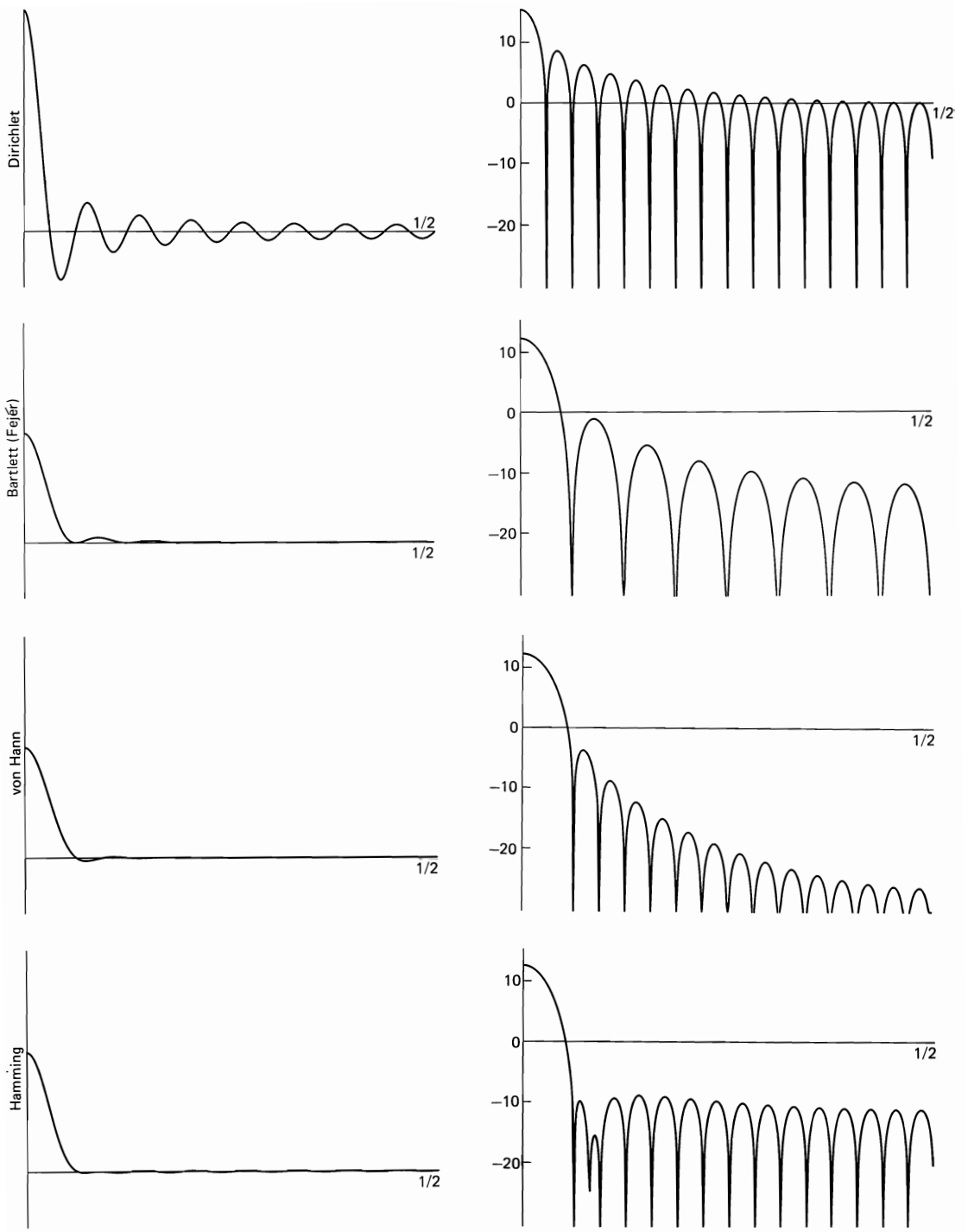


Figure 5-5 Correlation tapering windows (left) displayed with linear ordinate and spectral smoothing windows (right) displayed with log ordinate (from (91)–(95) with $T = 32$).

TABLE 5-2 SPECTRAL WINDOW PARAMETERS

Effective spectral window	Highest sidelobe (dB)	Asymptotic decay rate (dB/octave)	3-dB bandwidth ($\times T$)	Reliability factor η_0
Dirichlet (rectangle)	−6.5	−3	1.21	1
Bartlett or Fejér (triangle)	−13.5	−6	1.78	0.333
von Hann (raised cosine)	−16	−9	2.00	0.375
Hamming (raised cosine on a platform)	−22.5	−3	1.81	0.397
Blackman	−29	−9	2.35	0.305

to produce a zero in the spectral window near the peak of the first sidelobe of the Dirichlet window and thereby approximately minimize the height of the highest sidelobe [Blackman and Tukey 1958; Harris 1978]. The Blackman window is designed to produce two zeros in the spectral window [Blackman and Tukey 1958; Harris 1978], and it also removes the discontinuities at the temporal window edges that are present in the Hamming window, which results in an improvement in the sidelobe decay rate for the Blackman window relative to the Hamming window. Table 5-2 presents the following parameters for these windows: the height of the highest sidelobe relative to the main lobe (in decibels, i.e., 10 $\log_{10}$ of the ratio), the rate of decay of the sidelobes (in decibels per octave), the 3-dB bandwidth of the main lobe, and the reliability factor

$$\eta_0 = \frac{\sum_{|\tau| \leq T/2} h_T^2(\tau)}{T\, h_T^2(0)}, \tag{97}$$

which occurs in the discrete-time counterpart to the proportionality coefficient η (75) that determines the coefficient of variation (77). The factor η_0 in (97) depends not only on the particular window but also on the number of time samples per window width T; however, this latter dependence becomes negligible as the number of time-samples increases ($T >> 1$). Therefore, the values given in Table 5-2 are asymptotic values for (97) obtained from the continuous-time counterparts of the windows.[5] Observe that if the temporal windows in Table 5-2, (91a)–(95a), are data-tapering windows rather than effective autocorrelation-tapering windows, then $|H(f)|^2$ rather than $H(f)$ is the effective spectral smoothing window for a temporally smoothed periodogram, and therefore the decibel values

[5] Since $\Delta f/\eta_0$ in (77) with $\eta = \eta_0$ can be interpreted as an effective bandwidth, then $1/\eta_0$ is sometimes called the *standardized bandwidth* but should not be confused with the resolution bandwidth (e.g., the 3-dB bandwidth in Table 5-2).

in the first two columns of Table 5-2 need to be doubled and the values in the last two columns need to be recalculated.

Example:

Consider the problem of measuring the squared magnitude of the transfer function, here denoted by $K(f)$, of the resonant system driven by white Gaussian noise described in Section A, Chapter 3. Let B denote the bandwidth of the resonance peak centered at f_0, as depicted in Figure 5-6, and assume that it is desired to have a spectral resolution of $\Delta f = B/5$. Also assume that the coefficient of variation of the spectrum estimate must be no larger than $\frac{1}{10}$ in order to obtain a sufficient reduction of random effects, and choose a time-averaged periodogram as the spectrum estimate. Let us determine the minimum amount Δt of data that must be analyzed, and then evaluate the result for a resonance bandwidth of $B = 1$ KHz. A coefficient of variation of $\frac{1}{10}$ suggests that $\Delta t \Delta f = \Delta t/T$ is on the order of 10, in which case approximation (77) can be used; that is, the coefficient of variation is approximated by

$$r_{y_f} \cong \eta \frac{T}{\Delta t} \tag{98}$$

for frequencies sufficiently far removed from zero ($|f| > 1/T$). For a time-averaged periodogram, the factor η in (98), which is defined by (75), reduces to the reliability factor

$$\eta = \eta_0 \triangleq \frac{\int_{-T/2}^{T/2} h_T^2(\tau)\, d\tau}{T h_T^2(0)}, \tag{99}$$

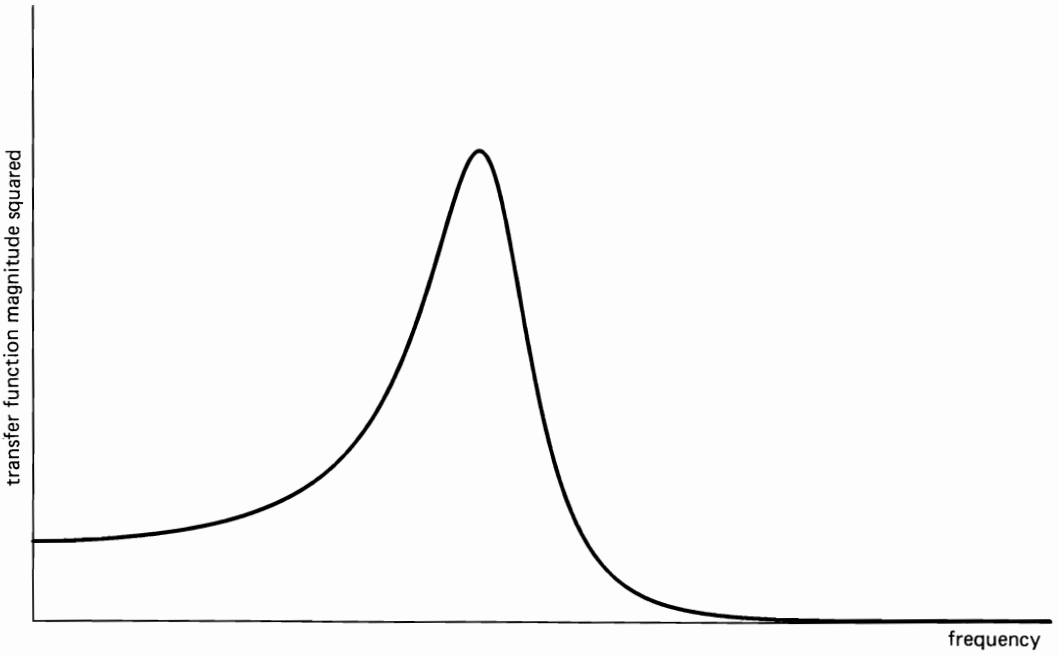


Figure 5-6 Transfer function for resonant system.

assuming that a rectangle time-averaging window $g_{\Delta t}(t) = u_{\Delta t}(t)$ is used. It follows from (55)–(57) that

$$h_T(\tau) = \frac{1}{T} a_T(\tau) \otimes a_T(-\tau), \tag{100}$$

where $a_T(\tau)$ is the data-tapering window. It follows from (98) and (99) that for a coefficient of variation of $\frac{1}{10}$, we need a data segment of length

$$\Delta t = 10\eta_0 T. \tag{101}$$

For example, it follows from (100) and Table 5-2 that for no data tapering (a_T = rectangle $\Rightarrow$ h_T = triangle), the 3-dB resolution bandwidth is $\Delta f = 1.78/T$ and the reliability factor is $\eta_0 = 0.333$, and therefore (101) becomes

$$\Delta t = \frac{(10)(0.333)(1.78)}{\Delta f} = \frac{5.9}{\Delta f}.$$

Thus, for $\Delta f = B/5$ and $B = 1$ KHz, we need $\Delta t \cong 30$ ms of data. For comparison, if we use a data-tapering window that yields an effective autocorrelation tapering window (100) that is the raised cosine (namely, the inverse transform of the square root of the magnitude of the transform of the raised cosine) to obtain better leakage behavior at the cost of moderately larger resolution bandwidth ($\Delta f = 2/T$) and moderately larger reliability factor ($\eta_0 = 0.375$), then (101) becomes

$$\Delta t = \frac{(10)(0.375)(2)}{\Delta f} \cong \frac{7.5}{\Delta f},$$

and therefore we need $\Delta t = 37.5$ ms of data. This is only a 25% increase, but it reduces the highest sidelobe of the spectral window by 2.5 dB and reduces the decay rate of the sidelobes by 3 dB/octave, as shown in Table 5-2. Alternatively, the performance obtained in this second case can be realized with a spectrally smoothed periodogram rather than a temporally smoothed periodogram, provided that the spectral smoothing window is the transform of the raised cosine. Although the effective smoothing window is the convolution of the spectral window actually used for smoothing with the squared magnitude of the transform of the data tapering window (see (54a)), which is a sinc-squared (Fejér) spectral window if no tapering is used, nearly the same length of data segment will be required (since nearly the same 3-dB bandwidth and reliability factor will result—see (54b)), and the same decay rate of the sidelobes will result.

As another alternative, a spectrally smoothed periodogram with the rectangle smoothing window can be used. Then since $\Delta t \Delta f >> 1$, approximation (54b) can be used for the effective spectral window to obtain the approximate 3-dB bandwidth ($\Delta f = 1/T$) and approximate reliability factor ($\eta_0 = 1$); however, the exact formula (54a) must be used to determine the sidelobe behavior. For example, with no data tapering, the asymptotic decay rate will be that of the Fejér spectral window, which is the same as it was for the time-averaged periodogram with no data tapering. However, the highest sidelobe will be lower (and will continually decrease as $\Delta t \Delta f$ is increased). The cost of this improvement is that $\Delta t = 50$ ms of data is needed, which is an increase of 67% above the 30 ms needed for the time-averaged periodogram with no data tapering.

Observe that if the desired spectral resolution width were cut in half to $\Delta f = B/10$, then for the same coefficient of variation, the amount of data required is simply doubled in each of the cases considered.

Example:

To illustrate the particular importance of spectral window sidelobe behavior when the data contains additive periodicity, which gives rise to spectral lines, we reconsider the previous example but with modified data that is contaminated with sine wave interference of power P_* at frequency f_*, as illustrated in Figure 5-7. Let us evaluate and compare the amount of spectral leakage at the frequency $f = f_0$ of the resonance peak, due to the sine wave power at $f = f_*$, for a time-averaged periodogram with no data tapering and with triangle data tapering. For this purpose, we let $f_* = f_0 + 3B/10$ and $P_* = 2B|K(f_0)|^2$. (The proportionality to the bandwidth B renders the leakage performance independent of the bandwidth parameter when the resolution is also proportional to B.)

If we assume a reliability factor of $\frac{1}{10}$ as in the previous example, then the leakage behavior is accurately characterized by the mean of the measured spectrum. It follows from (50) and (56) that the mean of the time-averaged periodogram is given by

$$\text{mean}\{y_f(t)\} = \hat{S}_x(f) \otimes \frac{1}{T}|A_{1/T}(f)|^2, \tag{102}$$

assuming that a rectangle time-averaging window $g_{\Delta t}(t) = u_{\Delta t}(t)$ is used. Formula (102) together with

$$\hat{S}_x(f) = \begin{cases} |K(f)|^2, & \text{interference absent} \\ |K(f)|^2 + \dfrac{P_*}{2}[\delta(f - f_*) + \delta(f + f_*)], & \text{interference present} \end{cases} \tag{103}$$

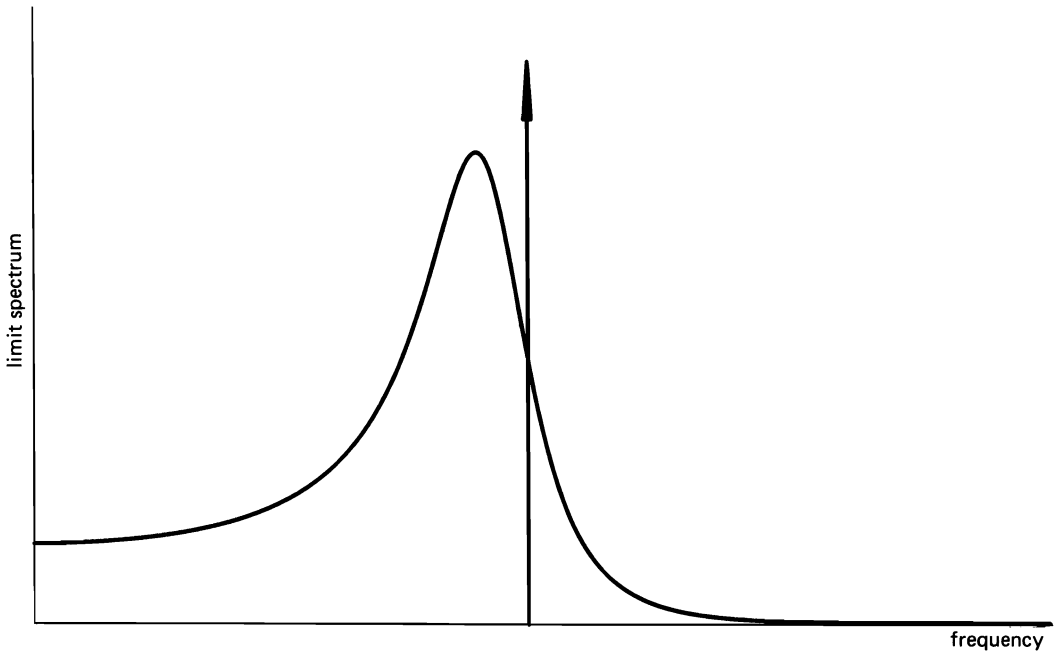


Figure 5-7 Limit spectrum of response of resonant system contaminated with sine wave interference.

(in which we have let $N_0 = 1$ for the intensity of the white noise excitation of the resonant system) yields

$$\text{mean}\{y_f(t)\} = |K(f)|^2 \otimes \frac{1}{T}|A_{1/T}(f)|^2, \tag{104}$$

for interference absent and

$$\begin{aligned}\text{mean}\{y_f(t)\} = |K(f)|^2 \otimes \frac{1}{T}|A_{1/T}(f)|^2 \\ + \frac{P^*}{2}\left[\frac{1}{T}|A_{1/T}(f - f_*)|^2 + \frac{1}{T}|A_{1/T}(f + f_*)|^2\right]\end{aligned} \tag{105}$$

for interference present. If we assume that $\Delta f = 1/T$ is small enough to resolve $|K(f)|^2$, say $1/T = B/5$, and that $|A_{1/T}(f_0 + f_*)| << |A_{1/T}(f_0 - f_*)|$ (which means that the image spectral line at $f = -f_*$ has negligible effect in the measurement at $f = f_0$), then we obtain simply

$$\text{mean}\{y_f(t)\} \cong |K(f_0)|^2 \tag{106}$$

for interference absent and

$$\text{mean}\{y_f(t)\} \cong |K(f_0)|^2 + \frac{P_*}{2T}|A_{1/T}(f_0 - f_*)|^2 \tag{107}$$

for interference present. Given that $P^* = 2B|K(f_0)|^2$ and $B = 5/T$, the *fractional leakage* FL at $f = f_0$ is

$$\text{FL} = \frac{500}{T^2}|A_{1/T}(f_0 - f_*)|^2 \quad \text{percent.} \tag{108}$$

Using

$$A_{1/T} = \frac{\sin(\pi f T)}{\pi f} \tag{109a}$$

for no data tapering and

$$A_{1/T} = \frac{1}{T}\left[\frac{\sin(\pi f T)}{\pi f}\right]^2 \tag{109b}$$

for triangle tapering together with $f_* = f_0 + 3B/10$ yields

$$\text{FL} = 500\left[\frac{\sin(3\pi/2)}{3\pi/2}\right]^n \quad \text{percent,} \tag{110}$$

where $n = 2$ for no data tapering and $n = 4$ for triangle tapering. It follows that FL = 22.5% for no data tapering and FL = 1.0% for triangle tapering. Thus, although the use of data tapering requires only a modest increase in the amount Δt of data in order to maintain a desired spectral resolution and coefficient of variation (as illustrated in the previous example), it greatly reduces spectral leakage.

The preceding two examples illustrate that although there is a fundamental trade-off between resolution and reliability, in that for a given amount of data Δt, the spectral resolution can be made finer (Δf smaller) only by decreasing the reliability (increasing the coefficient of variation) in inverse proportion, and vice versa, much can be gained in terms of improving spectral leakage performance, with only modest losses in either spectral resolution or reliability. Consequently,

the use of carefully designed data-tapering windows or autocorrelation-tapering windows to obtain good effective spectral smoothing windows is generally advisable. Sophisticated window design is particularly feasible in practice when digital methods of implementation are used, as discussed in the next chapter.

This discussion of window design is concluded with an anecdote on the lighter side. Recall that the Fourier transform of the rectangle is the sinc. Because of the common interpretation of the word *window* as a rectangular aperture in a structure such as a house, the subject of window design has long been referred to as *window carpentry*. This led one of my students, Louis W. Botsford, to remark that in order to Fourier transform a house, you need only change all the windows to sinks and all the sinks to windows.

E. SUMMARY

In this chapter the concept of fraction-of-time probabilistic analysis is introduced and used to quantify the resolution, leakage, and reliability properties of statistical spectra. In Section A it is explained that probabilistic analysis can be carried out without relying on the abstract notion of a probability space and an associated ensemble of random samples by using the concept of fraction-of-time probability. Then in Section B, the general fraction-of-time probabilistic model is defined and the particularly important special case, the Gaussian model, is defined. In Section C, the two temporal probabilistic measures of performance called *bias* and *variability* are defined and characterized in terms of the temporal mean, temporal coefficient of variation, and temporal correlation coefficient. These temporal probabilistic parameters are evaluated for the complex spectrum, periodogram, and various statistical spectra specified by the general representation introduced in Chapter 4, Section G. A general formula (50)–(51) for the effective spectral smoothing window is obtained and evaluated for various specific types of statistical spectra. A general formula (72)–(73) for the coefficient of variation is obtained, and it is simplified ((74)–(77)) by using the separable approximation to the kernel in the general representation (48), and the variability phenomenon is explained. Then two examples that illustrate the effects of variability are presented, and a time-frequency uncertainty principle for statistical spectra is described. Finally, the utility of the explicit formula for the effective spectral smoothing window is brought to light by explaining how it can be used in design to tradeoff resolution, leakage, and reliability performance (see Table 5-2). Two examples are presented to illustrate these design trade-offs. For situations in which the amount of data available is severely restricted or the range of the spectrum is large, such that the conditions required for the approximate formula for the coefficient of variation to be accurate are violated, the exact formulas for the mean (50)–(51) and variance (66) can be used simply by substituting in the kernel $M(\nu, \mu)$ that specifies the particular spectrum estimate of interest (see Table 5-1). This is important because leakage effects that do not show up in the effective spectral smoothing window can be revealed in the variability when the exact formulas are used.

EXERCISES

1. **(a)** Use (25) to show that the mean of the normalized time-variant finite-time complex spectrum (27) is zero for all nonzero frequencies, $f \neq 0$.

(b) Use (26) to show that the covariance of the normalized spectrum (27), at frequencies $f + \nu/2$ and $f - \nu/2$, is zero for all $\nu \neq 0$. *Hint:* In the integral

$$\int_{-T_0/2}^{T_0/2} x(t + u)x(t + v)e^{-i\pi\nu(2t+u+v)}\, dt,$$

let $t + u = t' + \tau/2$ and $t + v = t' - \tau/2$ to obtain the form in (26).

(c) Show that the variance of the normalized complex spectrum (27) is given by (29) for $f \neq 0$ and also for $f = 0$ if the mean of $x(t)$ is zero. *Hint:* Use the identity

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_T(u)a_T(v)\hat{R}_x(u - v)e^{-i2\pi f(u-v)}\, du\, dv = \int_{-\infty}^{\infty} r_a(\tau)\hat{R}_x(\tau)e^{-i2\pi f\tau}\, d\tau. \tag{112}$$

(d) Show that the finite autocorrelation in f of the normalized complex spectrum is given by (32) and is therefore zero in the limit, as $T \to \infty$, for $\nu \neq 0$. *Hint:* Use the identity

$$\int_{-\infty}^{\infty} e^{-i2\pi f(u-v)}\, df = \delta(u - v). \tag{113}$$

(e) Using methods similar to those in (a)–(c), show that the mean and variance of the normalized complex spectrum for the model (85) are given by (86) and (87), respectively.

2. **(a)** Show that the mean of the periodogram is the variance of the normalized complex spectrum for $f \neq 0$.

(b) Show that the covariance of the periodogram for tapered data, at frequencies f_1 and f_2, is given by (40) for a zero-mean Gaussian time-series $x(t)$. *Hint:* Use Isserlis' formula (19) to show that the correlation is given by

$$\langle S_{x_T}(t, f_1)S_{x_T}(t, f_2)\rangle = \frac{1}{T^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_T(u)a_T(v)a_T(r)a_T(s)$$
$$[\hat{R}_x(u - v)\hat{R}_x(r - s) + \hat{R}_x(u - r)\hat{R}_x(v - s) + \hat{R}_x(u - s)\hat{R}_x(v - r)]$$
$$\times e^{-i2\pi(f_1u - f_1v + f_2r - f_2s)}\, du\, dv\, dr\, ds. \tag{114}$$

Then show that the three terms in this expression are given by

$$\left[\frac{1}{T}\hat{S}_x(f_1) \otimes |A_{1/T}(f_1)|^2\right]\left[\frac{1}{T}\hat{S}_x(f_2) \otimes |A_{1/T}(f_2)|^2\right], \tag{115a}$$

$$\left[\frac{1}{T}\int_{-\infty}^{\infty} \hat{S}_x(\nu)A_{1/T}(f_1 - \nu)A_{1/T}(\nu + f_2)\, d\nu\right]^2, \tag{115b}$$

$$\left[\frac{1}{T}\int_{-\infty}^{\infty} \hat{S}_x(\nu)A_{1/T}(f_1 - \nu)A_{1/T}(\nu - f_2)\, d\nu\right]^2. \tag{115c}$$

In order to accomplish this, use the identity (112) and the identity

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_T(u)a_T(v)\hat{R}_x(u - v)e^{-i2\pi(gu+hv)}\, du\, dv$$
$$= \int_{-\infty}^{\infty} \hat{S}_x(\nu)A_{1/T}(g - \nu)A_{1/T}(h + \nu)\, d\nu, \tag{116}$$

where $A_{1/T}$ is the Fourier transform of a_T. Finally, use the formula

$$\text{cov}\{S_{x_T}(t, f_1), S_{x_T}(t, f_2)\} = \langle S_{x_T}(t, f_1) S_{x_T}(t, f_2)\rangle - \langle S_{x_T}(t, f_1)\rangle\langle S_{x_T}(t, f_2)\rangle$$

for the covariance, together with (34) and (115).

(c) Show that the variance of the periodogram is given by (35) for a zero-mean Gaussian time-series $x(t)$. *Hint:* Use the result of (b).

(d) With the aid of a graph, verify that

$$\lim_{T\to\infty} \frac{1}{T} A_{1/T}(f - \nu) A_{1/T}(\nu + f) = \lim_{T\to\infty} \frac{1}{T} |A_{1/T}(\nu)|^2 \delta_f = \gamma\delta(\nu)\delta_f, \tag{117}$$

where γ is the limit area defined by (25) in Chapter 3. Then use this result to derive (36), for the limiting coefficient of variation, from (35) and to derive (42), for the limiting correlation coefficient, from (40).

(e) Use the periodogram-correlogram relation to verify (43) for the finite autocorrelation of the periodogram.

3. Assume that $x(t)$ is a zero-mean Gaussian time-series and

$$\hat{R}_x(\tau) \begin{cases} > 0, & |\tau| \leq \Delta\tau^* \\ = 0, & |\tau| > \Delta\tau^*. \end{cases} \tag{118}$$

(a) Show that the variance of the correlogram $R_{x_T}(t, \tau)$ is proportional to $\Delta\tau^*/T$ for all $|\tau| \leq T$. *Hint:* Use Isserlis' formula to show that

$$\langle [R_{x_T}(t, \tau)]^2\rangle = \frac{1}{T^2} \int_{t-(T-|\tau|)/2}^{t+(T-|\tau|)/2} \int_{t-(T-|\tau|)/2}^{t+(T-|\tau|)/2} \{\hat{R}_x(u - v + \tau)\hat{R}_x(u - v - \tau) + [\hat{R}_x(u - v)]^2 + [\hat{R}_x(\tau)]^2\}\, du\, dv. \tag{119}$$

Then subtract the squared mean and use the identity (112) to show that

$$\text{var}\{R_{x_T}(t, \tau)\} = \frac{T - |\tau|}{T^2} \int_{-(T-|\tau|)}^{T-|\tau|} \left(1 - \frac{|w|}{T - |\tau|}\right) \times \{\hat{R}_x(w + \tau)\hat{R}_x(w - \tau) + [\hat{R}_x(w)]^2\}\, dw \tag{120}$$

$$\geqslant \frac{T - |\tau|}{T^2} \int_{-(T-|\tau|)}^{T-|\tau|} \left(1 - \frac{|w|}{T - |\tau|}\right) [\hat{R}_x(w)]^2\, dw \tag{121}$$

$$= \frac{T - |\tau|}{T^2} \int_{-\Delta\tau^*}^{\Delta\tau^*} \left(1 - \frac{|w|}{T - |\tau|}\right) [\hat{R}_x(w)]^2\, dw, \qquad \Delta\tau^* \leq T - |\tau| \tag{122}$$

$$\cong \sigma_x^4 \left(1 - \frac{|\tau|}{T}\right)\left(\frac{\Delta\tau^*}{T}\right), \qquad \text{for } \Delta\tau^* \ll T - |\tau|, \tag{123}$$

and also that

$$\text{var}\{R_{x_T}(t, \tau)\} \leq \frac{2(T - |\tau|)}{T^2} \int_{-\Delta\tau^*}^{\Delta\tau^*} [\hat{R}_x(w)]^2\, dw \qquad \Delta\tau^* \leq T - |\tau| \tag{124}$$

$$\cong 2\sigma_x^4 \left(1 - \frac{|\tau|}{T}\right)\left(\frac{\Delta\tau^*}{T}\right). \tag{125}$$

Thus, the variance of $R_{x_T}(t, \tau)$ is (to a close approximation) between the bounds (123) and (125).

(b) Show that the covariance of $R_{x_T}(t, \tau_1)$ and $R_{x_T}(t, \tau_2)$ is zero for $|\tau_1 + \tau_2| > 2\Delta\tau^*$ and $|\tau_1 - \tau_2| > 2\Delta\tau^*$. *Hint:* Use Isserlis' formula to show that

$$\langle R_{x_T}(t, \tau_1)R_{x_T}(t, \tau_2)\rangle = \frac{1}{T^2}\int_{t-(T-|\tau_1|)/2}^{t+(T-|\tau_1|)/2}\int_{t-(T-|\tau_2|)/2}^{t+(T-|\tau_2|)/2}\left\{\hat{R}_x(\tau_1)\hat{R}_x(\tau_2)\right.$$
$$+ \hat{R}_x\left(u - v + \frac{\tau_1 + \tau_2}{2}\right)\hat{R}_x\left(u - v - \frac{\tau_1 + \tau_2}{2}\right)$$
$$\left.+ \hat{R}_x\left(u - v + \frac{\tau_1 - \tau_2}{2}\right)\hat{R}_x\left(u - v - \frac{\tau_1 - \tau_2}{2}\right)\right\} du\, dv. \quad (126)$$

Then use the identity

$$I \triangleq \int_{t-(T-|\tau_1|)/2}^{t+(T-|\tau_1|)/2}\int_{t-(T-|\tau_2|)/2}^{t+(T-|\tau_2|)/2} f(u - v)\, du\, dv = \int_{-(T-a)}^{-(T-b)} (T - a - |w|)f(w)\, dw$$
$$+ \int_{T-b}^{T-a}(T - a - |w|)f(w)\, dw + \int_{-(T-b)}^{(T-b)} (T - b)f(w)\, dw \quad (127)$$

where $a \triangleq \frac{1}{2}(|\tau_1| + |\tau_2|)$ and $b \triangleq \frac{1}{2}\min\{|\tau_1|, |\tau_2|\}$, which yields

$$0 \leqslant \mathrm{I} \leqslant \int_{-T}^{T} Tf(w)\, dw, \quad (128)$$

for $f(w) \geqslant 0$ and $\Delta\tau^* < T$, to obtain

$$|\mathrm{cov}\{R_{x_T}(t, \tau_1), R_{x_T}(t, \tau_2)\}| \leqslant \frac{1}{T}\int_{-T}^{T}\left\{\hat{R}_x\left(w + \frac{\tau_1 + \tau_2}{2}\right)\hat{R}_x\left(w - \frac{\tau_1 + \tau_2}{2}\right)\right.$$
$$\left.+ \hat{R}_x\left(w + \frac{\tau_1 - \tau_2}{2}\right)\hat{R}_x\left(w - \frac{\tau_1 - \tau_2}{2}\right)\right\} dw. \quad (129)$$

It follows from (118) and (129) (or more directly from (118) and (126)) that for $|\tau_1 + \tau_2| > 2\Delta\tau^*$ and $|\tau_1 - \tau_2| > 2\Delta\tau^*$, the covariance is zero. It also follows from this result that if $\{R_{x_T}(t, \tau) : |\tau| \leqslant T\}$ is divided into $T/\Delta\tau^*$ contiguous segments, then only adjacent segments (which can satisfy $|\tau_1 - \tau_2| < 2\Delta\tau^*$) and mirror images about $\tau = 0$ (which can satisfy $|\tau_1 + \tau_2| < 2\Delta\tau^*$) are correlated. All others are mutually uncorrelated, and their variances therefore add when they are summed in an integral such as

$$S_{x_T}(t, f) = \int_{-T}^{T} R_{x_T}(t, \tau)e^{-i2\pi f\tau}\, d\tau.$$

Thus, the variance of such an integral is essentially independent of T and therefore does not approach zero as $T \to \infty$. This is true in spite of the fact that the variance of the integrand approaches zero as $T \to \infty$ for each and every value of τ.

4. (a) For a zero-mean Gaussian time-series, evaluate the mean-squared error in the approximation

$$R_x(t, \tau)_{\Delta t} \cong R_{x_{\Delta t}}(t, \tau), \qquad |\tau| \leqslant T << \Delta t, \quad (130)$$

and show that the error can be made arbitrarily small by selecting $\Delta t/T$ to be sufficiently large. *Hint:* See the hint in exercise 3(a).

(b) For a zero-mean Gaussian time-series, evaluate the mean-squared error in the approximation

$$R_{x_T}(t, \tau) \otimes u_{\Delta t}(t) \cong R_x(t, \tau)_{\Delta t}\frac{1}{T}r_a(\tau), \qquad T << \Delta t, \quad (131)$$

for tapered data (see (20) in Chapter 2), and show that the error can be made arbitrarily small by selecting $\Delta t/T$ to be sufficiently large.

5. Show that the limit of the variance of the pseudospectrum for a zero-mean Gaussian time-series is given by

$$\lim_{T\to\infty} \operatorname{var}\{S_x(t,f)_T\} = [\hat{S}_x(f)]^2(1 + \delta_f). \tag{132}$$

Hint: Use Isserlis' formula to show that the mean-squared value is given by

$$\langle [S_x(t,f)_T]^2 \rangle = \int_{-T}^{T}\int_{-T}^{T} \left\{ \hat{R}_x(u)\hat{R}_x(v) + \left(\frac{1}{T}\right)^2 \int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \left[\hat{R}_x\left(r - s + \frac{u-v}{2}\right) \right.\right.$$

$$\left.\left. \times\, \hat{R}_x\left(r - s - \frac{u-v}{2}\right) + \hat{R}_x\left(r - s + \frac{u+v}{2}\right)\hat{R}_x\left(r - s - \frac{u+v}{2}\right)\right] dr\, ds \right\} e^{-i2\pi f(u+v)}\, du\, dv. \tag{133}$$

Then use the identity (112) to show that

$$\operatorname{var}\{S_x(t,f)_T\} = \int_{-T}^{T}\int_{-T}^{T} \frac{1}{T}\int_{-T}^{T} \left(1 - \frac{|\tau|}{T}\right)\left[\hat{R}_x\left(\tau + \frac{u+v}{2}\right)\hat{R}_x\left(\tau - \frac{u+v}{2}\right) \right.$$

$$\left. + \hat{R}_x\left(\tau + \frac{u-v}{2}\right)\hat{R}_x\left(\tau - \frac{u-v}{2}\right)\right] d\tau\, e^{-i2\pi f(u+v)}\, du\, dv, \tag{134}$$

and then use the condition $T >> \Delta\tau^*$ to justify omission of the factor $(1 - |\tau|/T)$ and evaluate the integral with respect to τ. Finally, use the identity (116) to obtain

$$\operatorname{var}\{S_x(t,f)_T\} \cong \frac{1}{T}\int_{-\infty}^{\infty} [\hat{S}_x(\nu)]^2\{w_{1/T}(\nu - f)w_{1/T}(\nu + f) + [w_{1/T}(\nu - f)]^2\}\, d\nu,$$

$$T >> \Delta\tau^*. \tag{135}$$

Now, since $1/T << 1/\Delta\tau^* = \Delta f^*$, where Δf^* is the resolution width of $\hat{S}_x(f)$, and $z_{1/T} = (1/T)w_{1/T}^2$, then the desired result (132) follows (see (117)).

6. **(a)** Derive formula (50)–(51) for the mean of a statistical spectrum of the form (48).
(b) Simplify (51) for the case in which the kernel is separable, (57).

7. Let $x(t)$ be a zero-mean Gaussian time-series.
(a) Use Isserlis' formula (19) to derive the formula (62) for the limit spectrum of the centered spectrum estimate $z_f(t)$ specified by (59) and (48) with a separable kernel (57). *Hint:* Show that the limit autocorrelation of $y_f(t)$ (based on the conjugate lag product $y_f(t + \tau/2)y_f^*(t - \tau/2)$) is given by

$$\hat{R}_{y_f}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{\Delta t}(w_1)g_{\Delta t}^*(w_2)h_{1/\Delta f}(\tau_1)h_{1/\Delta f}^*(\tau_2)e^{-i2\pi f(\tau_1-\tau_2)}$$

$$\times \left[\hat{R}_x(\tau_1)\hat{R}_x(\tau_2) + \hat{R}_x\left(\tau + w_1 - w_2 + \frac{\tau_1-\tau_2}{2}\right)\hat{R}_x\left(\tau + w_1 - w_2 - \frac{\tau_1-\tau_2}{2}\right) \right.$$

$$\left. + \hat{R}_x\left(\tau + w_1 - w_2 + \frac{\tau_1+\tau_2}{2}\right)\hat{R}_x\left(\tau + w_1 - w_2 - \frac{\tau_1+\tau_2}{2}\right)\right] dw_1\, dw_2\, d\tau_1\, d\tau_2. \tag{136}$$

(The generalization of Isserlis' formula for complex-valued time-series is given in Chapter 15.) Then evaluate the Fourier transform. For example, the second term denoted by $S_2(\nu)$ is evaluated as follows. Application of the convolution theorem to the Fourier transform of the second term of (136) yields

$$S_2(\nu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{\Delta t}(w_1)g^*_{\Delta t}(w_2)h_{1/\Delta f}(\tau_1)h^*_{1/\Delta f}(\tau_2)e^{-i2\pi f(\tau_1-\tau_2)}$$
$$\times \int_{-\infty}^{\infty} \hat{S}_x(\nu-\mu)e^{i2\pi[w_1-w_2+(\tau_1-\tau_2)/2](\nu-\mu)}\hat{S}_x(\mu)e^{i2\pi[w_1-w_2-(\tau_1-\tau_2)/2]\mu}\,d\mu\,dw_1\,dw_2\,d\tau_1\,d\tau_2. \tag{137}$$

Factoring the five integrals into four Fourier transformation integrals and one remaining integral yields

$$S_2(\nu) = \int_{-\infty}^{\infty} G(\nu)G^*(\nu)H\left(f+\mu-\frac{\nu}{2}\right)H^*\left(f+\mu-\frac{\nu}{2}\right)\hat{S}_x(\nu-\mu)\hat{S}_x(\mu)\,d\mu, \tag{138}$$

which can be reexpressed as the first term in (62). Proceed similarly for the first and third terms in (136), and use (50) together with (57) to identify the first term in (136) with the squared mean term which is present in S_{y_f} but not in S_{z_f}. Then use (59) to express S_{y_f} in terms of S_{z_f}.

(b) When a window function, say A, is narrow enough to resolve accurately another function, say B, then the following approximation is accurate:

$$A(\nu)B\left(\mu+\frac{\nu}{2}\right)B\left(\mu-\frac{\nu}{2}\right) \cong A(\nu)B^2(\mu). \tag{139}$$

Use this approximation to derive (63) from (62).

(c) Use the fact that when a window function A is narrow enough to resolve accurately another function B then the approximation

$$\int_{-\infty}^{\infty} A(f-\mu)B(\mu)\,d\mu \cong \int_{-\infty}^{\infty} A(\mu)\,d\mu\; B(f) \tag{140}$$

is accurate to show that (64)–(65) accurately approximates (63). Note: A more accurate approximation than (64) can be obtained by replacement of $[\hat{S}_x(f)]^2\, r_{H_{\Delta f}}(2f)$ with $[\hat{S}_x(0)]^2\, r_{H_{\Delta f}}(2f)$. However, as long as $H_{\Delta f}$ resolves $\hat{S}_x$, then (64) is accurate.

(d) Finally, use (50), (58), (64), and (65) to derive (72) and (74), and then manipulate this result into the form (75) and (77).

(e) Using the same techniques as those described in (a)–(d), derive the general formulas (60) and (66) and the approximations (67)–(69). Then show that the separable approximation (57)–(58) can be used to obtain (62) from (60), (70) from (69), and (74) from (73).

8. (a) Sketch the mean of the time-averaged periodogram $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ for a sine wave in additive white noise,

$$x(t) = \cos(2\pi f_0 t + \theta) + n(t), \tag{141}$$

and a rectangle data-tapering window.

(b) Sketch the mean of the spectrally smoothed periodogram $S_{x_{\Delta t}}(t, f)_{\Delta f}$ for the same data as in (a) and a rectangle spectral smoothing window with $\Delta t \Delta f \gg 1$.

(c) If one had an unlimited amount Δt of data, how might one proceed to determine if the data is narrow-band-pass filtered white noise with bandwidth B in additive white noise or a sine wave in additive noise?

9. Determine the kernel transform $M(\nu, \mu)$ and the effective spectral smoothing window $E(f)$ for the following measurement methods:

(a) Temporally smoothed periodogram with a triangle data-tapering window $Tv_T(t)$.

(b) Spectrally smoothed periodogram with a rectangle smoothing window, $u_{\Delta f}(\nu)$.

(c) Fourier-transformed tapered correlogram with a triangle correlation-tapering window $(1/\Delta f)v_{1/\Delta f}(\tau)$.

(d) Hopped temporally smoothed pseudospectrum with hop time-interval of length T.

(e) Hopped temporally smoothed periodogram of half-overlapped data segments each of length T, with a triangle data-tapering window, $Tv_T(t)$.

(f) Wave analyzer with input BPF transfer function magnitude given by a sinc-squared function $z_{\Delta f}(\nu - f)$ for each f.

(g) Demodulation method with input LPF transfer function given by a sinc-squared function $z_{\Delta f}(\nu)$.

Observe that $E(f)$ is the same for (a), (e), (f), and (g). (Do not use the separable approximation (57) until exact formulas for $M(\nu, \mu)$ and $E(f)$ have been obtained.) *Hint:* See exercise 9 in Chapter 4.

10. (a) Show that the effective autocorrelation tapering window, which is defined to be the inverse Fourier transform of the effective spectral smoothing window, is given by

$$e(\tau) = \int_{-\infty}^{\infty} m(t, \tau)\, dt.$$

(b) Determine the effective autocorrelation tapering window for the statistical spectra prescribed in exercise 9(a), (b), (e), (f), and (g).

11. Use the definition of probabilistic autocorrelation,

$$E\{x(t_1)x(t_2)\} \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 f_{x(t_1)x(t_2)}(y_1, y_2)\, dy_1\, dy_2, \tag{142}$$

where $f_{x(t_1)x(t_2)}$ is the joint fraction-of-time probability density defined by the second mixed partial derivative of the joint fraction-of-time distribution [Gardner 1985],

$$f_{x(t_1)x(t_2)}(y_1, y_2) \triangleq \frac{\partial^2}{\partial y_1 \partial y_2} \lim_{T_0 \to \infty} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} U[y_1 - x(t_1 + t)]U[y_2 - x(t_2 + t)]\, dt, \tag{143}$$

to verify the equivalence between (14) and (15).

12. (a) Use the continuous-time counterpart of (97),

$$\eta_0 = \frac{\int_{-T/2}^{T/2} h_T^2(\tau)\, d\tau}{T h_T^2(0)}, \tag{144}$$

to verify the values given in Table 5-2 for the five windows (91)–(95).

(b) Evaluate η_0 for the following windows (see Chapter 2, Section B):

(i) Sinc: $h_T(\tau) = Tw_T(\tau)$

(ii) Sinc^2: $h_T(\tau) = Tz_T(\tau)$.

Hint: Use Parseval's relation.

13. In order to illustrate that the general formula (77) for the coefficient of variation of a statistical spectrum $y_f(t)$ can be obtained without using the separable approximation, proceed as follows for the spectrally smoothed periodogram (with no data tapering). Express the variance in the form

$$\text{var}\{S_{x_{\Delta t}}(t, f)_{\Delta f}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_S(\nu_1, \nu_2) H_{\Delta f}(f - \nu_1) H_{\Delta f}(f - \nu_2)\, d\nu_1\, d\nu_2 \tag{145}$$

where $K_S(\nu_1, \nu_2)$ is the temporal covariance

$$K_S(\nu_1, \nu_2) \triangleq \text{cov}\{S_{x_{\Delta t}}(t, \nu_1), S_{x_{\Delta t}}(t, \nu_2)\}$$

and $H(f)$ is assumed to be real. Show that for Δt large enough for $1/\Delta t$ to resolve $\hat{S}_x$, formula (40) (with T replaced by Δt) yields the close approximation

$$K_S(\nu_1, \nu_2) \cong \left[\frac{1}{\Delta t}\hat{S}_x(\nu_1)w_{1/\Delta t}(\nu_1 - \nu_2)\right]^2 + \left[\frac{1}{\Delta t}\hat{S}_x(\nu_1)w_{1/\Delta t}(\nu_1 + \nu_2)\right]^2. \tag{146}$$

Hint: $w_{1/\Delta t} \otimes w_{1/\Delta t} = w_{1/\Delta t}$. Then substitute (146) into (145) to obtain the close approximation

$$\text{var}\{S_{x_{\Delta t}}(t, f)_{\Delta f}\} \cong \frac{1}{\Delta t}[\hat{S}_x(f)]^2 \otimes [H_{\Delta f}(f)]^2 + \frac{1}{\Delta t}\int_{-\infty}^{\infty}[\hat{S}_x(\nu)]^2 H_{\Delta f}(f - \nu)H_{\Delta f}(f + \nu)\, d\nu, \qquad \Delta t\Delta f \gg 1. \tag{147}$$

Finally, show that it follows from (147) that the temporal coefficient of variation of $S_{x_{\Delta t}}(t, f)_{\Delta f}$ is closely approximated by

$$r_S \cong \frac{\xi}{\Delta t \Delta f}[1 + \delta_{\Delta f}(f)], \qquad \Delta t\Delta f \gg 1, \tag{148}$$

where the coefficient ξ in (148) is defined by

$$\xi \triangleq \frac{[\hat{S}_x(f)]^2 \otimes [\Delta f H_{\Delta f}^2(f)]}{[\hat{S}_x(f) \otimes H_{\Delta f}(f)]^2}. \tag{149}$$

Let Δf be small enough to resolve $\hat{S}_x$ and show that (149) and (75) are then approximately equivalent. *Hint:* Since $a_{\Delta t}$ in (53) is a rectangle, then $g_{\Delta t}$ in (75) is a rectangle.

14. Consider the hopped time-averaged periodogram of triangle-tapered data as a statistical spectrum.

(a) Determine the 3-dB bandwidth of the sinc^4 spectral window, which results from a triangle data-tapering window. That is, determine the value of f at which

$$\left[\frac{\sin(\pi f T)}{\pi f T}\right]^4 = \frac{1}{2},$$

and then double this value of f. Compare the result with that for the triangle autocorrelation tapering window (which corresponds to no data tapering) from Table 5-2.

(b) Determine the reliability factor η_0 (99) for the sinc^4 spectral window, and compare the result with that for the triangle autocorrelation tapering window. *Hint:* The corresponding effective autocorrelation tapering window for sinc^4 is

$$h_{2T}(\tau) = T v_T(\tau) \otimes v_T(\tau),$$

where $v_T(\tau)$ is the unit-area triangle window with base of width $2T$. Use this convolution characterization to show that

$$h_{2T}(\tau) = \begin{cases} 1 - \frac{3}{2}\left(\frac{\tau}{T}\right)^2 + \frac{3}{4}\left|\frac{\tau}{T}\right|^3, & |\tau| \leq T \\ 2\left(1 - \frac{1}{2}\left|\frac{\tau}{T}\right|\right)^3, & T \leq |\tau| \leq 2T \\ 0, & |\tau| > 2T. \end{cases}$$

(c) Consider the first example in Section D, and use the results of (a) (3-dB bandwidth $= 0.636/T$) and (b) ($\eta_0 = 0.269$) to determine the length Δt of data segment needed for a coefficient of variation of $r_{y_f} = \frac{1}{10}$ and a spectral resolution width

of $\Delta f = B/5$, when triangle data tapering is used. Compare the result with the results in the example.

(d) Consider the second example in Section D, and evaluate the fractional leakage at $f = f_0$ for triangle data tapering, when the sine wave frequency is $f_* = f_0 + B/5$, $f_0 + 3B/10$, $f_0 + 2B/5$, and the sine wave power is $P_* = 2B|K(f_0)|^2$. Explain any unexpected results and discuss practical implications.

15. Consider the problem of designing a wave analyzer using conventional analog electrical circuitry, that is, passive resistive-inductive-capacitive networks. The simplest BPF to implement would be a second-order resonant circuit with impulse-response function

$$k(t) = e^{-t/T}\cos(2\pi ft), \qquad t \geqslant 0,$$

and the simplest LPF to implement would be the first-order circuit with impulse response function

$$g(t) = e^{-t/\Delta t}, \qquad t \geqslant 0.$$

Determine the following characteristics of this spectrum analyzer:

(a) Effective data-tapering window.

(b) Effective autocorrelation-tapering window.

(c) Effective spectral smoothing window and its 3-dB bandwidth ($E(\Delta f/2) = E(0)/2$).

(d) Time-averaging window.

(e) Reliability factor (75).

(f) Highest sidelobe (if any) of the effective spectral smoothing window.

(g) Rate of decay of spectral window sidelobes (skirts).

Hint: Use the fact established in Chapter 4, Section E, that the real implementation described here is essentially equivalent to the complex implementation with BPF

$$k(t) = e^{-t/T}e^{i2\pi ft},$$

provided that $\Delta t \Delta f \gg 1$ and $|f|/\Delta f \gg 1$.

16. Consider the problem of designing a swept-frequency wave analyzer for audio spectral analysis. Assume that the spectral band to be analyzed ranges from 300 Hz to 15,000 Hz, the desired resolution is 100 Hz, and the desired coefficient of variation is $\frac{1}{10}$. Also assume that at each frequency f in the band that is swept across, one can treat the swept frequency analyzer as an unswept wave analyzer with the particular filters described in exercise 15. In order to specify design parameters for this spectrum analyzer, determine the following characteristics:

(a) The time constant T required for a 3-dB resolution bandwidth of $\Delta f = 100$ Hz.

(b) The time constant Δt required for a coefficient of variation of $\frac{1}{10}$.

(c) The sweep rate $\beta = \Delta f/\Delta t$.

(d) The analysis time AT (AT = period of sweep).

(e) If it is desired to detect a very brief audio event that occupies a band of width 500 Hz, what is the fraction-of-time probability of detection using this spectrum analyzer?

(f) Assume that there is a constraint to cut the analysis time to half that found in (d). Propose a modification to the above design to meet this constraint; that is, adjust the requirements on sweep rate, resolution, and reliability.

Answers: (a) $T = 1/100\pi$ s. (b) $\Delta t = 5/100\pi$ s. (c) $\beta = 2000\pi$ Hz/s. (d) AT = $7.35/\pi$ s. (e) Probability = 0.034. (f) Since β must be doubled and since $\Delta t \Delta f = (\Delta f)^2/\beta$, then if reliability is held fixed Δf must be increased by the factor $\sqrt{2}$, or if Δf is held fixed then the coefficient of variation is doubled.

17. Evaluate the coefficient of variation $R(f)$ given by (73) for the spectrum estimates specified in exercise 9(a)–(g).

18. **(a)** The characteristic function for $z(t)$ is defined by

$$\Psi_z(\omega) \triangleq \langle \exp\{i\omega z(t)\} \rangle \tag{150a}$$

$$= \int_{-\infty}^{\infty} f_z(y) \exp\{i\omega y\}\, dy, \tag{150b}$$

where

$$f_z(y) = \frac{dF_z(y)}{dy}$$

is the fraction-of-time probability density for $z(t)$. Here (150b) is a conjugate Fourier transform, and it can be shown, by use of (16), that for a Gaussian time-series $z(t)$, (150b) yields

$$\Psi_z(\omega) = \exp\{i\omega\hat{m}_z - \tfrac{1}{2}\omega^2\hat{\sigma}_z^2\}. \tag{151}$$

To verify that for a Gaussian time-series $x(t)$ the joint characteristic function (17) is given by (18), use (151) with $\omega = 1$ and

$$z(t) = \boldsymbol{\omega}'\boldsymbol{x}(t).$$

(b) Use the result of (a) with $N = 2$, $t_1 = \tau/2$, and $t_2 = -\tau/2$ to show that the joint characteristic function for $x(t + \tau/2)$ and $x(t - \tau/2)$ is given by

$$\Psi_x(\omega_1, \omega_2) = \exp\{i(\omega_1 + \omega_2)\hat{m}_x - \tfrac{1}{2}(\omega_1^2 + \omega_2^2)[\hat{R}_x(0) - \hat{m}_x^2] - \tfrac{1}{2}\omega_1\omega_2[\hat{R}_x(\tau) - \hat{m}_x^2]\}. \tag{152}$$

(c) From (17), we have

$$\Psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{x}}(\boldsymbol{y}) \exp\{i\boldsymbol{\omega}'\boldsymbol{y}\}\, d\boldsymbol{y}, \tag{153}$$

which is the N-dimensional conjugate Fourier transform of the Nth-order joint fraction-of-time probability density. Consequently, $f_{\boldsymbol{x}}(\boldsymbol{y})$ for a Gaussian time-series can be obtained by Fourier transformation of (18). If the inverse matrix $\hat{\boldsymbol{K}}_x^{-1}$ exists, then the result is

$$f_{\boldsymbol{x}}(\boldsymbol{y}) = [(2\pi)^N |\hat{\boldsymbol{K}}_x|]^{-1/2} \exp\left\{ -\tfrac{1}{2}(\boldsymbol{y} - \hat{\boldsymbol{m}}_x)'\hat{\boldsymbol{K}}_x^{-1}(\boldsymbol{y} - \hat{\boldsymbol{m}}_x) \right\}, \tag{154}$$

where $|\hat{\boldsymbol{K}}_x|$ denotes the determinant of $\hat{\boldsymbol{K}}_x$. Express (153) as explicitly as possible in terms of $\hat{m}_x$ and $\hat{R}_x(\tau)$ for $N = 2$, $t_1 = \tau/2$, and $t_2 = -\tau/2$. (When the inverse matrix $\hat{\boldsymbol{K}}_x^{-1}$ does not exist, then $f_{\boldsymbol{x}}(\boldsymbol{y})$ contains impulse fences that reflect the linear dependence among the N variables in $\boldsymbol{x}(t)$; see [Gardner 1985].)

6

DIGITAL METHODS

A. INTRODUCTION

Modern general-purpose spectral analysis instruments are typically implemented using primarily analog technology for frequencies above 100 KHz and digital technology for frequencies below 100 Hz, and both technologies are used in the midrange. The swept-frequency method described in Chapter 4 is the most commonly used analog method for general-purpose instruments, whereas the *fast Fourier transform* (FFT), with the discrete-time and discrete-frequency counterparts of the frequency smoothing and/or hopped time-averaging methods described in Chapter 4, is used for most digital implementations. Digital methods are especially attractive for low frequencies because the most attractive analog method (swept frequency) requires long measurement times compared with the simultaneous analysis methods based on Fourier transformation of the data. Analog methods are especially appropriate for high frequencies because of technological limitations on switching times, which limit the speed of digital computation. When the required speed is not a limiting factor, digital implementations are generally attractive because of economy as well as high accuracy, stability, and flexibility, including programmability. Furthermore, spectral analysis at frequencies far above 100 KHz can be accomplished digitally by down-converting spectral bands (of width less than 100 KHz) from higher frequency ranges (e.g., megahertz to gigahertz) to lower frequency ranges (below 100 KHz), and this band-selective approach can be used to obtain very high spectral resolution. Moreover, the flexibility of digital methods is an attractive feature for many special-purpose spectral analysis tasks, where general-purpose instruments are inappropriate. An example of this flexibility is the fact that digital methods can be directly implemented in software so that both the convenience of personal computers and the immense data-handling capabilities of supercomputers are available for

spectral analysis. Finally, because of the increasing amount of data that is digitally encoded for storage and transmission, digital methods of spectral analysis that can be directly applied to digital data are especially appropriate.

Unfortunately, the study of digital methods of spectral analysis is somewhat more complicated than the study of analog methods for several reasons. These include 1) the spectral aliasing phenomenon that results from time-sampling, 2) the discrete nature of the frequency parameter in FFT and other *discrete Fourier transform* (DFT) algorithms, and 3) the block format for data that is required by DFT algorithms. All three of these items are sources of conceptual complication that can lead to complications in practice, including erroneous procedures and misinterpretation of results. Fortunately, many of the fundamentals of spectral analysis can be understood, as explained in the other chapters of this book, without introducing the complications associated with digital methods of implementation. This applies especially to the digital methods of spectral analysis that are simply discrete-time and discrete-frequency counterparts of the analog methods studied in Chapter 4.

In Section B, the DFT is introduced and its properties and relationships with other Fourier transformations are studied. Then in Section C, various digital counterparts of the analog methods developed in Chapter 4 are described. Finally in Section D, the applicability to discrete-time spectrum estimates of the results on fraction-of-time probabilistic analysis obtained in Chapter 5 for continuous time is explained.

B. THE DFT

1. Resolution and Zero-Padding

Consider a finite sequence of numbers $\{x_n\} = \{x_0, x_1, x_2, \ldots, x_{N-1}\}$. The DFT of this sequence is denoted by $\{X_m\} = \{X_0, X_1, X_2, \ldots, X_{N-1}\}$ and is defined by

$$X_m \triangleq \sum_{n=0}^{N-1} x_n e^{-i(2\pi/N)mn}, \qquad m = 0, 1, 2, \ldots, N-1, \tag{1a}$$

and the inverse DFT is defined by

$$x_n \triangleq \frac{1}{N} \sum_{m=0}^{N-1} X_m e^{i(2\pi/N)nm}, \qquad n = 0, 1, 2, \ldots, N-1. \tag{1b}$$

It is easily verified (exercise 2) that (1a) and (1b) are indeed a transform pair in the sense that if (1a) is substituted into (1b), an identity is obtained. It is clear from the inverse DFT (1b) that the DFT $\{X_m\}$ is the complex-valued discrete density (scaled by N) of sine wave components contained in the sequence $\{x_n\}$. The frequencies of these sine waves are $0, 1/N, 2/N, \ldots, (N-1)/N$, which indicates that the DFT cannot resolve frequency any finer than the increment $1/N$ between adjacent frequencies. However, there is a way to interpolate between the N points in the DFT. Specifically, if it is desired to add N interpolating points between the N original points, then one just adds N zeros to the end of the time sequence to obtain a new sequence, say $\{y_n\}$, of length $2N$,

$$y_n = \begin{cases} x_n, & n = 0, 1, 2, \ldots, N-1 \\ 0, & n = N, N+1, N+2, \ldots, 2N-1. \end{cases} \tag{2}$$

The DFT of $\{y_n\}$ is given by

$$Y_m = \sum_{n=0}^{2N-1} y_n e^{-i(2\pi/2N)mn}, \qquad m = 0, 1, 2, \ldots, 2N-1, \tag{3}$$

which upon substitution of (2) becomes

$$Y_m = \sum_{n=0}^{N-1} x_n e^{-i(2\pi/N)nm/2}, \qquad m = 0, 1, 2, \ldots, 2N-1. \tag{4}$$

Comparison of (1a) and (4) reveals that

$$Y_{2m} = X_m, \qquad m = 0, 1, 2, \ldots, N-1. \tag{5}$$

Hence, the even-indexed frequency points in the new DFT are identical to the N points in the original DFT, but there are N additional points between these, namely, the odd-indexed frequencies. To show that the additional N frequency points are nothing more than an interpolation from the original N frequency points, we can substitute (1b) into (4) to obtain (exercise 4)

$$Y_m = \sum_{p=0}^{N-1} X_p I\left(p - \frac{m}{2}\right), \tag{6a}$$

where $I(q/2)$ is an interpolating sequence with magnitude given by

$$\left|I\left(\frac{q}{2}\right)\right| = \frac{1}{N}\left|\frac{\sin(\pi q/2)}{\sin(\pi q/2N)}\right| \tag{6b}$$

and phase (in radians) given by

$$\arg\left\{I\left(\frac{q}{2}\right)\right\} = \left(\frac{\pi q}{2}\right)\left(1 - \frac{1}{N}\right) + \pi U\left[-\frac{\sin(\pi q/2)}{\sin(\pi q/2N)}\right], \tag{6c}$$

where $U(\cdot)$ is the unit-step function.

If $(K-1)N$ zeros are appended to the time-sequence $\{x_n\}$, then the DFT (4) and the interpolation formula (6) are modified by replacement of $m/2$ and $q/2$ with m/K and q/K, and m ranges from 0 to $NK-1$. Thus, the DFT of the K-fold zero-padded time-sequence $\{x_n\}$ is given by

$$Y_m = \sum_{n=0}^{NK-1} x_n e^{-i(2\pi/N)nm/K}, \qquad m = 0, 1, 2, \ldots, NK-1, \tag{7}$$

which can be reexpressed as

$$Y_m = \sum_{p=0}^{N-1} X_p I\left(p - \frac{m}{K}\right), \tag{8a}$$

where

$$\left|I\left(\frac{q}{K}\right)\right| = \frac{1}{N}\left|\frac{\sin(\pi q/K)}{\sin(\pi q/KN)}\right| \tag{8b}$$

$$\arg\left\{I\left(\frac{q}{K}\right)\right\} = \left(\frac{\pi q}{K}\right)\left(1 - \frac{1}{N}\right) + \pi U\left[-\frac{\sin(\pi q/K)}{\sin(\pi q/KN)}\right]. \tag{8c}$$

A graph of $|I(q/K)|$ for large K is shown in Figure 6-1. To illustrate the utility of frequency interpolation by zero-padding the time sequence, the DFT of a sequence consisting of 16 uniformly spaced time samples of a sum of three sine waves is shown in Figure 6-2 for various amounts of zero-padding ($K = 1, 2, 4, 16$).

It follows directly from (1a) that the DFT with frequency parameter m extended to include all the integers is periodic with period N. As a result, the N numbers (N is assumed to be even)

$$\{X_m : m = 0, \pm 1, \pm 2, \ldots, \pm (N/2 - 1), N/2\}$$

that are centered (except for the rightmost point) at zero frequency, $m = 0$, can be used in place of the N numbers

$$\{X_m : m = 0, 1, 2, \ldots, N - 1\}$$

since the latter can be obtained from the former by using the identity

$$X_{m+N} = X_m.$$

In Section C, the centered DFT is used exclusively.

There are various numerically efficient algorithms for computation of the DFT and its inverse. The most well known of these are collectively referred to as *fast Fourier transform* (FFT) algorithms (see [Nussbaumer 1982; Blahut 1985; Burrus and Parks 1985]. However, a closely related but distinct class of numerically

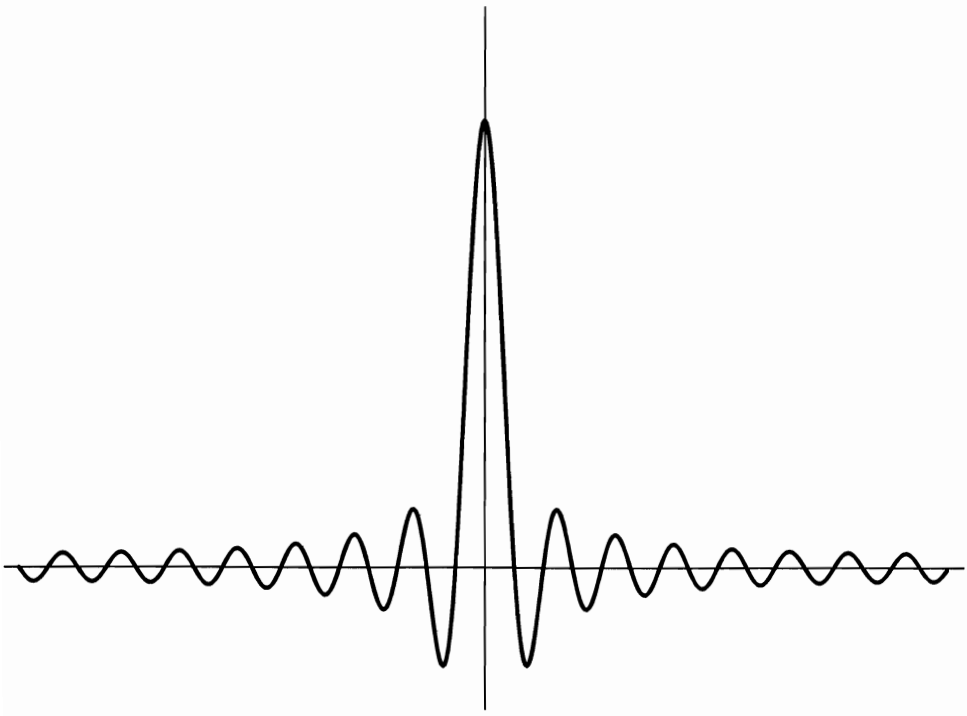
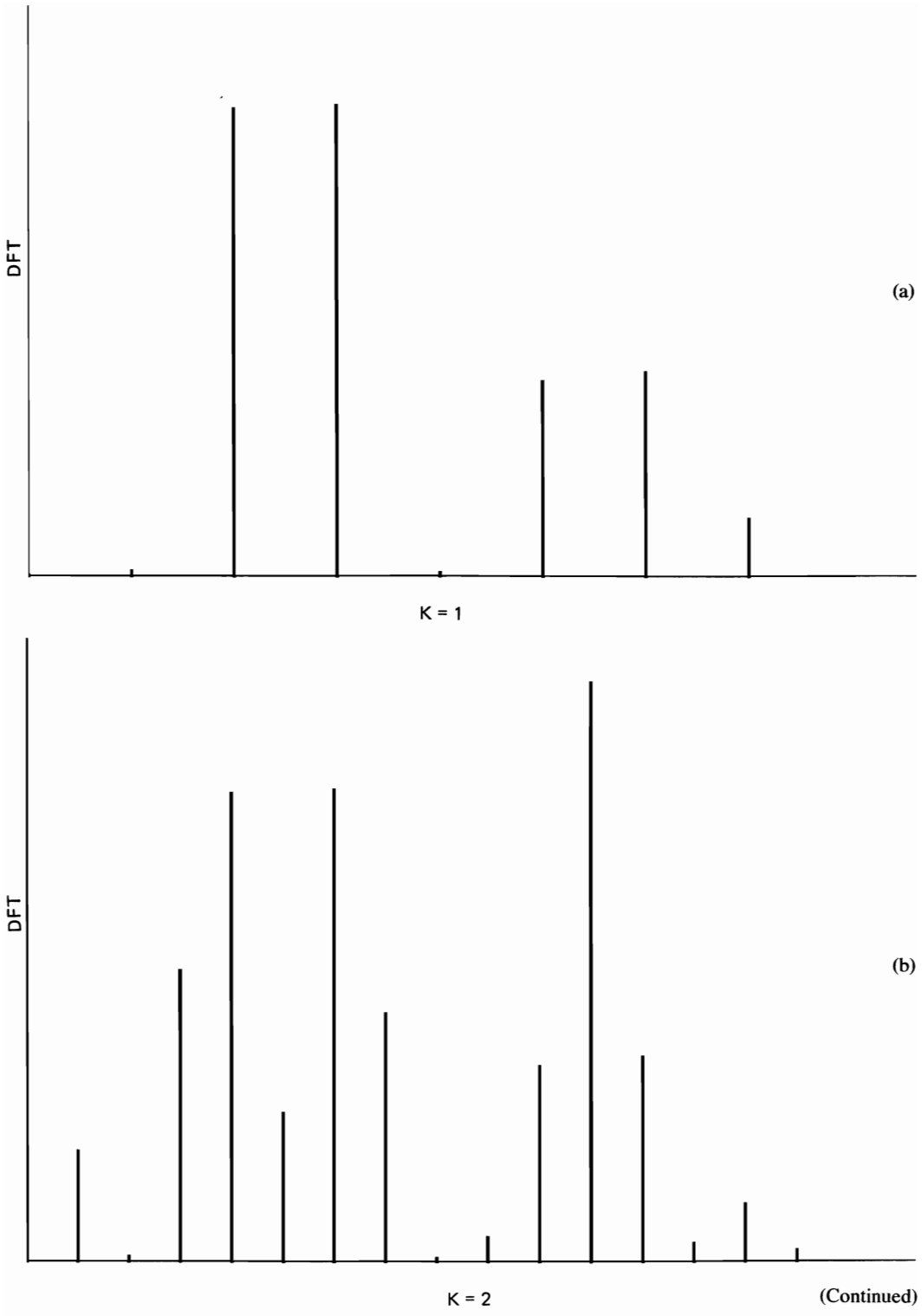
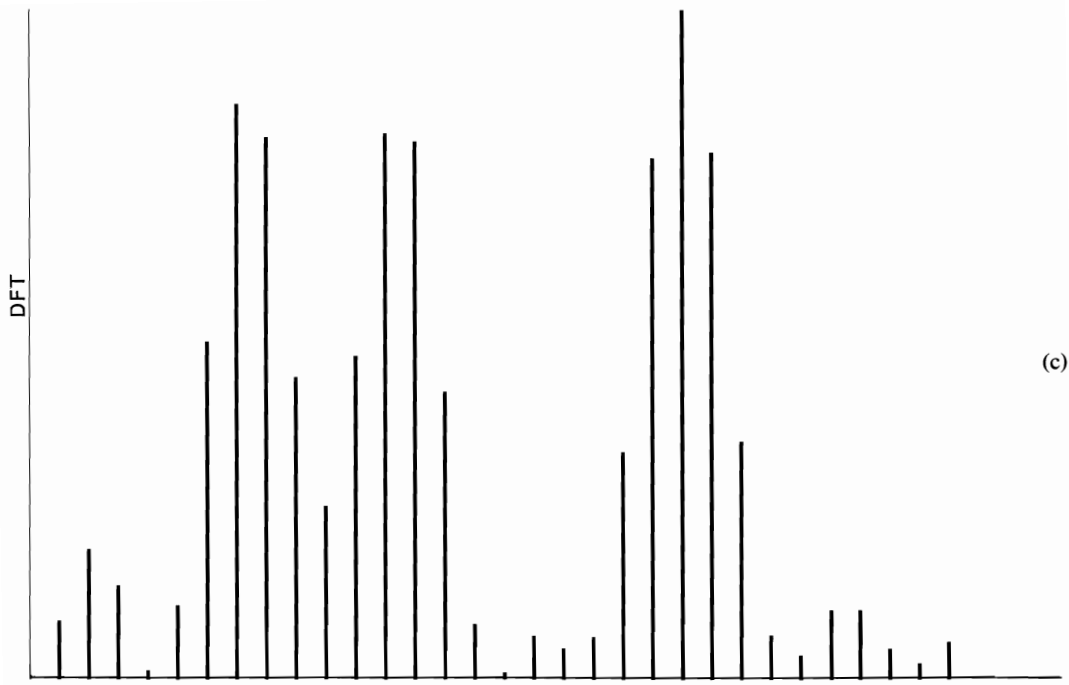


Figure 6-1 Interpolation sequence $\sin(\pi q/K)/\sin(\pi q/KN)$ for $K = 16$, $N = 32$, $|q| \leq 256$. (Sequence for smaller values of K obtained by subsampling in q.)



(Continued)

Figure 6-2 Effect of zero-padding on spectral resolution: (a) $K = 1$. (b) $K = 2$.



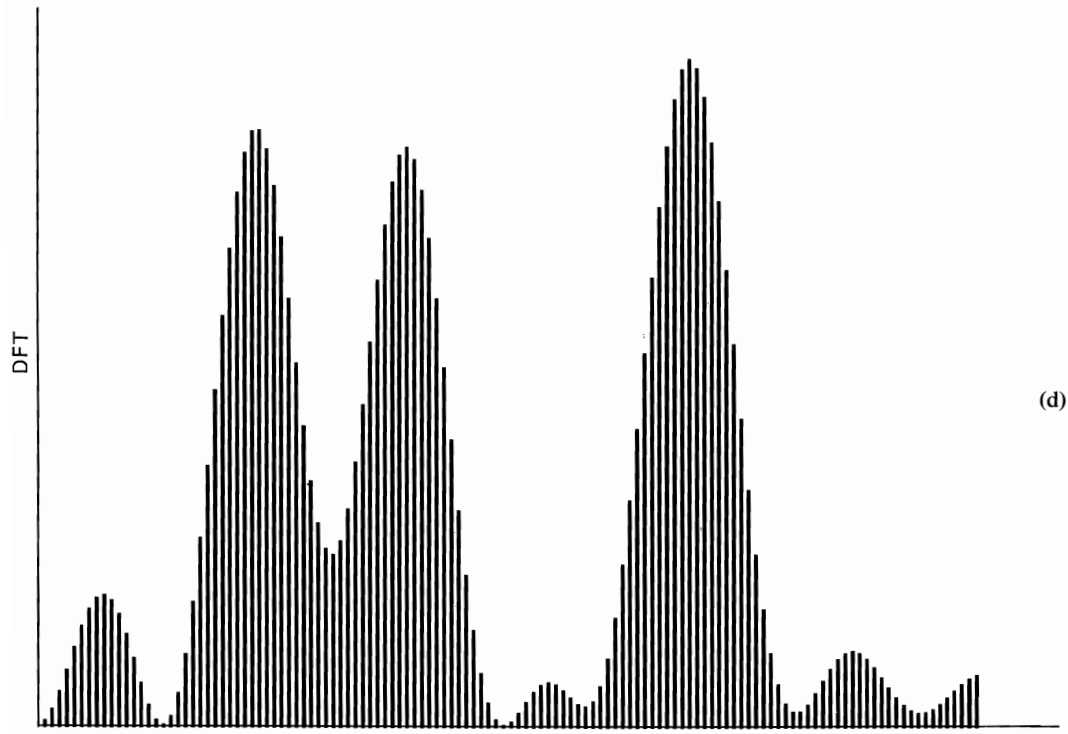


Figure 6-2 (continued) (c) $K = 4$. (d) $K = 16$.

efficient algorithms, collectively referred to as *fast Hartley transform* (FHT) algorithms, have recently been developed by Ronald N. Bracewell [Bracewell 1985, 1986]. The *discrete Hartley transform* (DHT) is simply the real part of the DFT minus the imaginary part, whereas the DFT is the even part of the DHT minus i times the odd part. The FFT and FHT algorithms require that the number of points to be transformed be an integer multiple of 2. For real data, the most efficient FHT algorithm has some advantages in terms of program running time and storage requirements relative to the most efficient FFT algorithms, whereas for complex data they are apparently equivalent (see [Sorenson et al. 1985]).

2. Circular Convolution

Recall from Chapter 2 that the Fourier series transform (FST) of the convolution of two sequences, say

$$z_n = \sum_{k=-\infty}^{\infty} x_k y_{n-k} \triangleq x_n \otimes y_n, \tag{9a}$$

is given by the product of FSTs

$$\bar{Z}(f) = \bar{X}(f)\bar{Y}(f). \tag{9b}$$

If the sequences $\{x_n\}$ and $\{y_n\}$ are both nonzero for only $0 \leq n \leq N-1$, then the convolution (9a) reduces to

$$z_n = \sum_{k=0}^{n} x_k y_{n-k} \tag{9c}$$

but (9b) remains unchanged. One might expect an analogous result to hold for the DFT of the convolution of two finite-length sequences, but there is a complication in that the length of the convolution of two sequences each of length N is in general $2N$, and therefore it would appear that the DFT in the left member of the analog of (9b) would have to be of length $2N$ when the DFTs in the right member of the analog of (9b) are of length N. However, if the sequences $\{x_n\}$ and $\{y_n\}$ of length N are first zero-padded out to length $2N$, then $\{z_n\}$ given by (9a) would still have length $2N$ (for its nonzero portion), and therefore a DFT of length $2N$ could be used for both members in the analog of (9b). In fact, in this case the analog of (9b) for the DFT would simply be the frequency-sampled version of (9b) as explained in the following subsection on the FST.

Nevertheless, there is a modified convolution theorem that holds for DFTs of length N for sequences of length N (without zero-padding), but the type of convolution operation in this theorem is a modification of the standard convolution (9a). Specifically, (9a) can be called a *linear convolution* because it can be visualized as consisting of a sequence of three operations carried out on a straight line representing the domain of the sequences being convolved, namely, the set of all integers. The three operations are those of reflecting one of the sequences, say $\{y_n\}$, about the origin, shifting the reflected version by integer amounts k, and summing the product of the reflected and shifted sequence $\{y_{k-n}\}$ with the

other sequence $\{x_n\}$, as indicated by (9a). On the other hand, if $\{x_n\}$ and $\{y_n\}$ are both finite-length (N) sequences, then their domain can be taken to be the N integers 0, 1, 2, . . ., $N - 1$ arranged uniformly on a circle, with an arbitrary location on this circle identified as the origin 0. In this case, the sequence of three operations described above can be taken as the definition of what is called *circular convolution*. As a matter of fact, the circular convolution of two sequences of length N can also be obtained by periodically repeating these sequences to obtain

$$x_n = x_{n+N}$$

$$y_n = y_{n+N}$$

for all integers n, and then forming the finite linear convolution

$$z_n = \sum_{k=0}^{N-1} x_k y_{n-k}. \tag{10}$$

It follows from the preceding descriptions that the circular convolution of the two finite-length sequences $\{x_n\}$ and $\{y_n\}$ is given by

$$z_n = \sum_{k=0}^{n} x_k y_{n-k} + \sum_{k=n+1}^{N-1} x_k y_{N+n-k}, \qquad n = 0, 1, 2, \ldots, N-1. \tag{11a}$$

Observe that the indexes on x and y in (11a) always remain within the set {0, 1, 2, . . ., $N - 1$}, regardless of the value of the index of summation k, for all values of n in this same set, given that the second sum is defined to be zero for $k > N - 1$. Comparison of (11a) with (9c) reveals that the difference between linear convolution and circular convolution is the presence of the second sum in (11a). This sum contains what are called the *wraparound* terms; that is, the $N - n - 1$ terms at the end of the sequence $\{y_k\}$ that are wrapped around (due to reflection and shifting by n around the circle) and therefore overlap the $N - n - 1$ terms at the end of the sequence $\{x_k\}$. It is easily shown that if $\{x_n\}$ and $\{y_n\}$ are zero-padded out to length $2N$ and N in (11a) is replaced with $2N$, then the second sum vanishes (since $2N + n - k > N$ for $n + 1 \leq k \leq N - 1$ and $0 \leq n \leq 2N - 1$), and therefore circular convolution is equivalent to linear convolution in this case.

The usefulness of the concept of circular convolution results not only from the fact that it preserves the length of sequences but also from the fact that the DFT of a circular convolution is given by the product of DFTs

$$Z_m = X_m Y_m, \qquad m = 0, 1, 2, \ldots, N-1. \tag{11b}$$

This circular convolution theorem is proved in exercise 5.

The time-frequency dual of the circular convolution theorem is also useful. This theorem states that the DFT of the product of two time-series segments

$$z_n = x_n y_n, \qquad n = 0, 1, 2, \ldots, N-1 \tag{12a}$$

is given by the circular convolution of their DFTs

$$Z_m = \sum_{p=0}^{m} X_p Y_{m-p} + \sum_{p=m+1}^{N-1} X_p Y_{N+m-p}, \tag{12b}$$

where the second sum is defined to be zero for $p > N - 1$. It follows from the circular convolution theorem that there is an analog of the periodogram-correlogram relation for the FST that holds for the DFT, provided that the correlogram is of the circular type (see exercise 6). However, if the data is zero padded with $K \geq 2$, then the circular correlogram is identical to the linear correlogram (see exercise 6).

3. The FST and CFT

In the limit as the number of zeros appended to $\{x_n\}$ approaches infinity ($K \to \infty$), the DFT approaches the FST for finite-length sequences

$$\tilde{X}(f) \triangleq \sum_{n=0}^{N-1} x_n e^{-i2\pi nf}, \tag{13}$$

which is defined for all real values of the frequency parameter f. It follows from (1a) and (13) that the N-point DFT is simply the frequency-sampled version of the FST,

$$X_m = \tilde{X}\left(\frac{m}{N}\right), \qquad m = 0, 1, 2, \ldots, N - 1. \tag{14}$$

This is illustrated in Figure 6-2(d), where the envelope of the DFT lines is the FST. The fact that the FST for a finite segment can be exactly recovered from its sampled version, the DFT, by interpolation is a direct result of the time-frequency dual of the sampling theorem for bandlimited time-series (see exercise 15, Chapter 2). This follows from the fact that a finite-segment time-series is duration-limited and is therefore the time-frequency dual of a bandlimited time-series.

If the time-sequence $\{x_n\}$ is obtained from a waveform $x(t)$ by time-sampling,

$$x_n = x(nT_s), \tag{15a}$$

or if it is desired to interpret $\{x_n\}$ in this way, then it is of interest to determine the relationship between the FST (and therefore the DFT) and the *continuous Fourier transform* (CFT). For this purpose, we consider the CFT of the finite segment of a waveform

$$X_T(\hat{f}) \triangleq \int_0^T x(t) e^{-i2\pi \hat{f} t}\, dt, \tag{16}$$

where $T = (N - 1)T_s$, as illustrated in Figure 6-3. As shown in Section H of Chapter 2, we have the relationship

$$\tilde{X}(f) = \frac{1}{T_s} \sum_{q=-\infty}^{\infty} X_T\left(\frac{f}{T_s} - \frac{q}{T_s}\right), \tag{15b}$$

which describes the spectral aliasing phenomenon for time-sampling as in (15a). It follows from (15b) that the dimensionless frequency parameter f and the dimensioned frequency parameter $\hat{f}$ are related by the sampling increment,

$$\hat{f} = \frac{f}{T_s}. \tag{17}$$

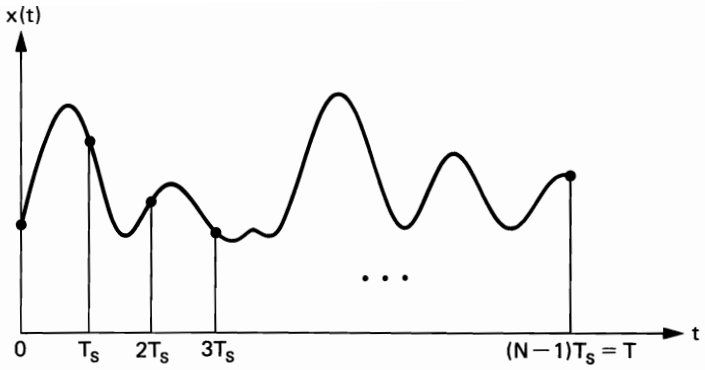


Figure 6-3 Illustration of N samples in interval of length T with sampling increment T_s.

Whereas the symbol f is used in all other chapters for the dimensioned frequency parameter, the symbol $\hat{f}$ is used instead in this chapter so that f can be used for the dimensionless frequency, which is the most prevalent of the two in this chapter.

There is a fundamental theorem of Fourier analysis that establishes that a finite segment of a waveform cannot be strictly bandlimited. Thus, there will always be some nonzero aliasing error involved in approximating the CFT $X_T(\hat{f})$ by the FST $\tilde{X}(f)$ of a finite segment. However, it can be argued that if the waveform $x(t)$ is sufficiently bandlimited (e.g., if a conservative measure of bandwidth, say B, satisfies $B < 1/2T$) before $x(t)$ is truncated to the finite time interval $[0, T]$, then the error in the approximation from (15b),

$$\tilde{X}(f) \cong \frac{1}{T_s} X_T\left(\frac{f}{T_s}\right), \qquad |f| \leq \frac{1}{2}, \tag{18}$$

will be negligible. This suggests that in applications where there is a desire to relate the time sequence $\{x_n\}$ to a continuous-time waveform, a highly interpolated DFT, which closely approximates the FST, which in turn approximates a CFT, is more appropriate than the DFT with no zero-padding. However, this appears to be at odds with the arguments that 1) since a CFT of a segment of length T has spectral resolution width on the order of $1/T$, then samples of the frequency $\hat{f}$ (or f) separated by $1/T = 1/(N-1)T_s \cong 1/NT_s$ (or $1/N$) should yield an adequate spectral representation of the CFT and 2) since the FST approximates the CFT and samples in f separated by $1/N$ of the FST are given by the DFT with no zero padding, then the DFT without zero padding should yield an adequate representation of the CFT. The problem with this argument is that we can say only that the spectral resolution of the CFT is *on the order of* $1/T$. Thus, one might need samples of the frequency $\hat{f}$ separated by $1/2T$ or $1/3T$ or less to represent the FST adequately (without additional interpolation). The following example illustrates that the DFT can change drastically when the spectral sampling increment is only halved by appending only N zeros to the time sequence, even though the new DFT is just an interpolated version of the original DFT.

Example: ***Pseudonoise***

A *pseudonoise* (PN) sequence is a sequence of numbers, such as +1s and −1s that appears to be random but, in fact, is perfectly predictable given the algorithm that generates it. A particularly important type of ±1 PN sequence is the *maximal-length shift-register sequence,* which when periodically repeated yields a limit autocorrelation sequence that is similar to that for white noise. For a periodic sequence with period N, we have

$$\tilde{R}_x(k) \triangleq \lim_{Q\to\infty} \frac{1}{2Q+1} \sum_{n=-Q}^{Q} x_{n+k}x_n = \tilde{R}_x(k+N) \tag{19a}$$

and

$$\tilde{R}_x(k) = \tilde{R}_{x_N}(k)_c \triangleq \frac{1}{N}\sum_{n=0}^{N-1} x_{n+k}x_n. \tag{19b}$$

The sequence defined by (19b) is the *circular correlogram* of one period of the sequence $\{x_n\}$. For a maximal-length sequence $\{x_n\}$, it can be shown [Golomb 1967] that

$$\tilde{R}_{x_N}(k)_c = \begin{cases} 1, & k = 0 \\ -\dfrac{1}{N}, & 1 \leq |k| \leq N-1, \end{cases} \tag{20}$$

where $N = 2^M - 1$ for some positive integer M. It follows from the periodogram-correlogram relation for the DFT that the N-point DFT of the circular correlogram of this sequence with length N is given by

$$\text{DFT}\{\tilde{R}_{x_N}(k)_c\} = \frac{1}{N}|\text{DFT}\{x_n\}|^2. \tag{21}$$

Since (20) can be used to show that (exercise 7)

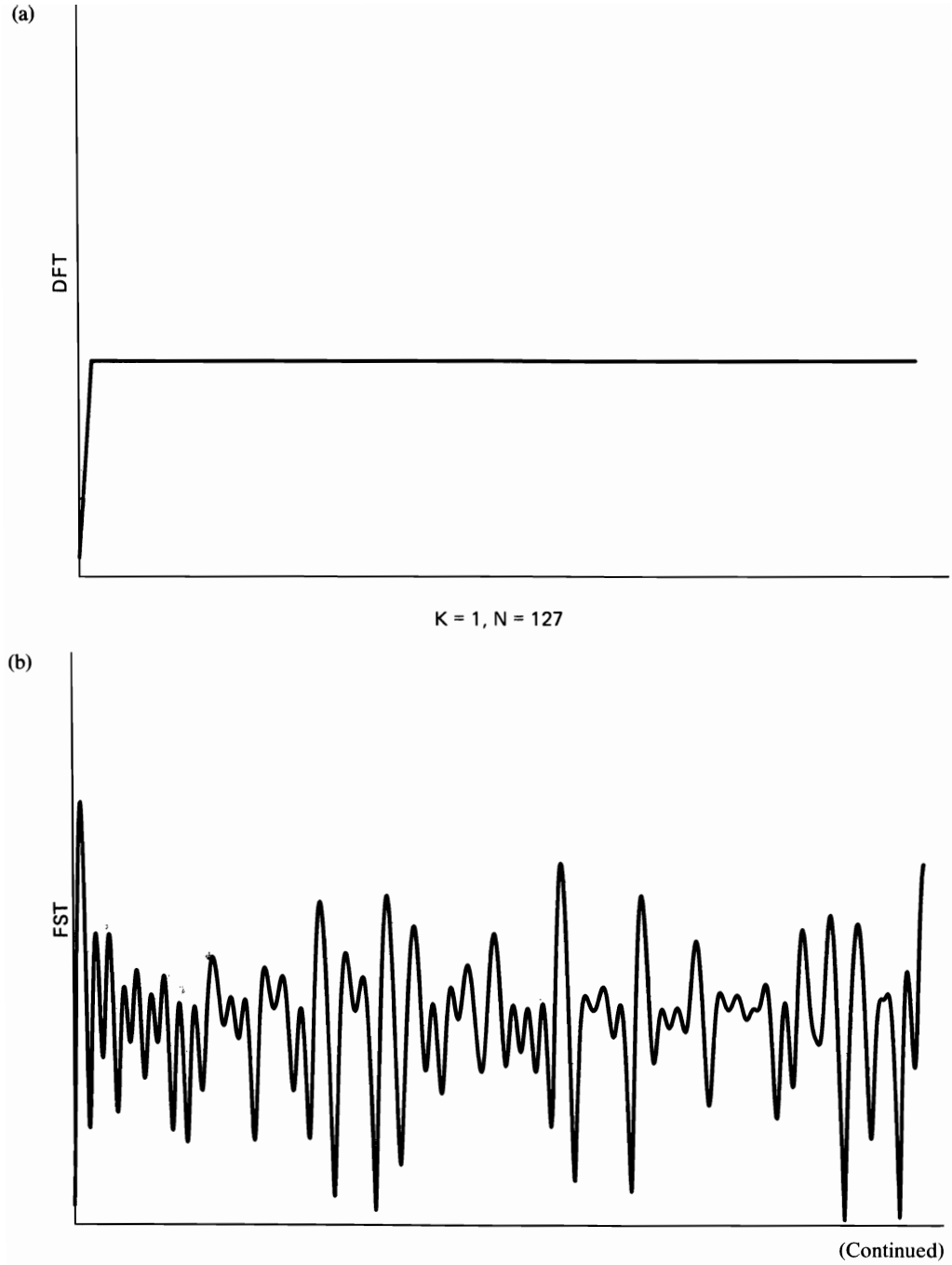
$$\text{DFT}\{\tilde{R}_{x_N}(k)_c\} = \begin{cases} \dfrac{1}{N}, & m = 0 \\ 1 + \dfrac{1}{N}, & 1 \leq |m| \leq N-1, \end{cases} \tag{22}$$

then it follows from (21) that the magnitude of the DFT of a maximal-length sequence $\{x_n\}$ is given by

$$|X_m| = \begin{cases} 1, & m = 0 \\ \sqrt{N+1}, & 1 \leq |m| \leq N-1, \end{cases} \tag{23}$$

which is flat except for a notch at zero frequency, as shown in Figure 6-4(a).

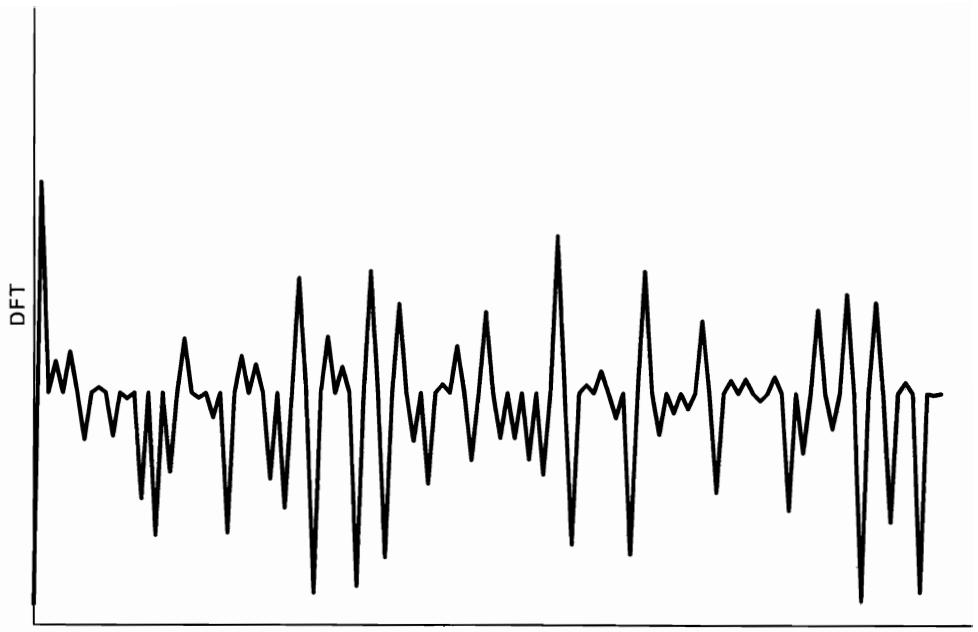
If the spectral sampling increment of the DFT is decreased by appending a large number of zeros to $\{x_n\}$, a close approximation to the FST is obtained, as shown in Figure 6-4(b). The original N frequency points are still given by (23) as they must be, but the interpolated points in between are surprisingly different. At first glance this seems impossible, since the points in between are obtained from the original points by interpolation. However, it is the complex-valued numbers $\{X_m\}$ that are interpolated, not their magnitudes. Evidently, the phase sequence is sufficiently erratic to yield a highly erratic magnitude after interpolation. It can be seen from the FST in Figure 6-4(b) that merely halving the spectral sampling increment of the DFT by appending N zeros yields a drastically different result than the original DFT specified by (23). This is illustrated in Figure 6-4(c). As a matter of fact, appending only one point to obtain a total number of points that is an integer



(Continued)

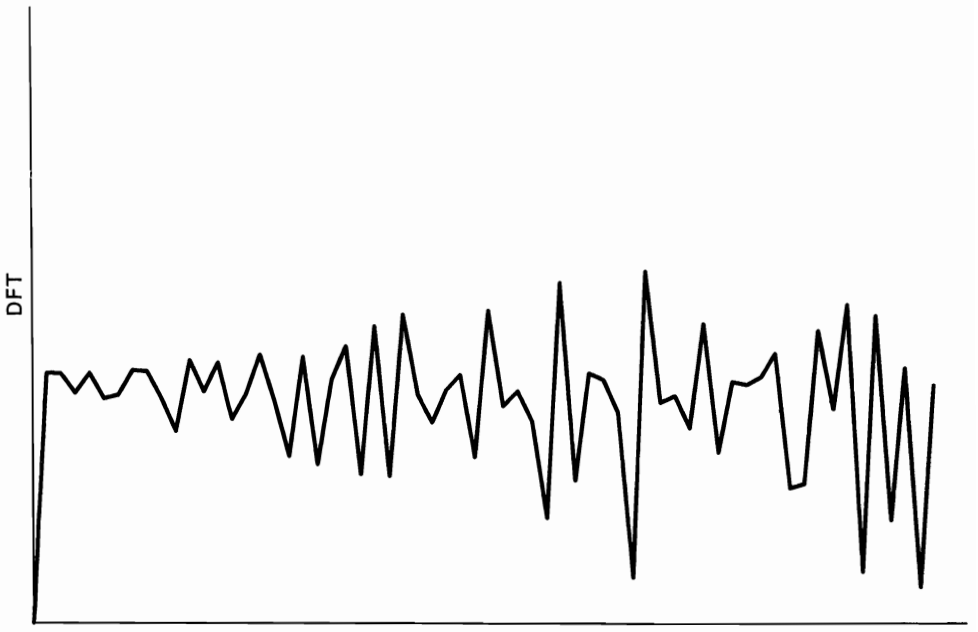
Figure 6-4 (a) DFT magnitude of a PN sequence of length 127 (points connected by straight lines). (b) FST magnitude of the PN sequence whose DFT magnitude is shown in (a). (c) DFT magnitude of the PN sequence whose DFT magnitude is shown in (a) but zero-padded out to length 254. (d) DFT magnitude of PN sequence whose DFT magnitude is shown in (a), but with the digit -1 appended to obtain a length of 128.

(c)



K = 2, N = 127

(d)



K = 1, N = 128

Figure 6-4 (continued)

multiple of two $N = 2^M$, so that a fast DFT algorithm can be used, yields a drastically different result than the original DFT with $N = 2^M - 1$ points (see exercise 9). This is illustrated in Figure 6-4(d). It can be seen from this example that without sufficient zero-padding, the DFT can be a misleading measure of spectral content. This is further illustrated by the previous example of a time-series consisting of three sine waves. As shown in Figure 6-2(a), it is not at all clear from the DFT without zero-padding that there should be only three spectral peaks, but this is confirmed by the DFT with double ($K = 2$) and quadruple ($K = 4$) zero-padding, as shown in Figures 6-2(b) and 6-2(c).

C. METHODS BASED ON THE DFT

In this section, the notation used for discrete-time quantities is very similar to that used for continuous-time quantities in previous chapters. Although this requires a change from the notation used in Section B, which is conventional, the notation to be used here emphasizes the fact that all the methods described in this section are discrete-time analogs of the continuous-time methods described in Chapter 4. The correspondences between the notations in this section and Section B are as follows:

$$x_n = x(t), \qquad n = t = 0, \pm 1, \pm 2, \pm 3, \ldots, \tag{24a}$$

$$X_m = \tilde{X}_T(f), \quad \frac{m}{KN} = f = 0, \pm\frac{1}{KN}, \pm\frac{2}{KN}, \pm\frac{3}{KN}, \ldots, \pm\left(\frac{1}{2} - \frac{1}{KN}\right), \frac{1}{2}, \tag{24b}$$

where K is the zero-padding factor. The sampling increment is taken to be unity ($T_s = 1$), and therefore the number of time points in a finite segment of length T is given by

$$N = T + 1 \tag{25}$$

and the dimensioned and dimensionless frequencies, $\hat{f}$ and f, are identical. Sampling increments other than unity can be accommodated simply by multiplying all discrete-time variables by T_s and all discrete-frequency variables by $1/T_s$.

If the DFT with K-fold zero-padding is applied to the sequence of data blocks

$$\{x(t - u) : u = 0, 1, 2, 3, \ldots, N - 1\},$$

indexed by $t = 0, 1, 2, 3, \ldots$, it produces the sequence of transformed blocks

$$\left\{\tilde{x}_T(t,f) : f = 0, \pm\frac{1}{KN}, \pm\frac{2}{KN}, \pm\frac{3}{KN}, \ldots, \pm\left(\frac{1}{2} - \frac{1}{KN}\right), \frac{1}{2}\right\},$$

indexed by $t = 0, 1, 2, 3, \ldots$, where

$$\tilde{x}_T(t,f) \triangleq \sum_{v=0}^{T} x(t - T + v)e^{-i2\pi fv}. \tag{26}$$

The index t represents the most recent time in a block, or the leading edge of a block. For each value of f, $\tilde{x}_T(t, f)$ is a band-pass time-series with center frequency f and bandwidth $\Delta f^o \cong 1/N$ ($\Delta f^o \cong 1/T$ for $N >> 1$). Thus, $\tilde{x}_T(t, f)$ is the discrete-time-discrete-frequency analog of the continuous-time-continuous-

frequency local sine wave component $x_T(t, f)$ (except that t in $\tilde{x}_T(t, f)$ represents the leading edge of the locale, whereas t in $x_T(t, f)$ represents the midpoint of the locale—see (44) in Chapter 4). The low-pass time-series that is the complex demodulate of $\tilde{x}_T(t, f)$ is obtained by frequency-shifting,

$$\tilde{X}_T(t, f) = \tilde{x}_T(t, f)e^{-i2\pi ft}. \tag{27}$$

Thus, we have (using $u = -v$ in (26))

$$\tilde{X}_T(t, f) = \sum_{u=-T}^{0} x(t - T - u)e^{-i2\pi f(t-u)}, \tag{28}$$

which is the discrete-time-discrete-frequency analog of the time-variant finite-time complex spectrum $X_T(t, f)$—see (2a) in Chapter 2. Consequently all the analog methods introduced in Chapter 4 that are described in terms of $x_T(t, f)$ or $X_T(t, f)$ can be implemented digitally using $\tilde{x}_T(t, f)$ or $\tilde{X}_T(t, f)$ obtained from a DFT. This is explained in the following subsections.

1. Bartlett-Welch Method

One of the most computationally efficient digital methods (especially for long data segments) is referred to as the *Bartlett-Welch method* [Bartlett 1948; Welch 1967]. This method uses hopped time-averaging of periodograms with data-tapering and typically 50% overlap,

$$\tilde{S}_{x_{1/\Delta f}}(t, f)_{\Delta t} \triangleq \frac{1}{2L/N - 1} \sum_{u=0}^{2(L/N-1)} \tilde{S}_{x_T}\left(t - \frac{uN}{2}, f\right), \tag{29}$$

where $\tilde{S}_{x_T}(t, f)$ is the time-variant periodogram (with $T = N - 1$)

$$\tilde{S}_{x_T}(t, f) \triangleq \frac{1}{N}|\tilde{X}_T(t, f)|^2 \tag{30a}$$

$$= \frac{1}{N}|\tilde{x}_T(t, f)|^2 \tag{30b}$$

and

$$\tilde{x}_T(t, f) = \sum_{v=0}^{N-1} a_T(-v)x(t - T + v)e^{-i2\pi fv}. \tag{31}$$

The sequence $a_T(u)$ is the data-tapering window.[1] Figure 6-5 illustrates 50% overlapped triangle windows. The Bartlett-Welch method is sometimes referred to as the *WOSA* (weighted overlapped-segment averaging) *method.*

The number of blocks averaged in (29) is

$$2M \triangleq \frac{2L}{N} - 1.$$

The total time span of data used is $\Delta t = L - 1$, and L is the number of time

[1] The tapering window $a_T(t)$ is defined here for $t \leq 0$ in order to be able to express $\tilde{x}_T(t, f)$ as a convolution with $a_T(t)$, rather than $a_T(-t)$. This is consistent with the convention in earlier chapters.

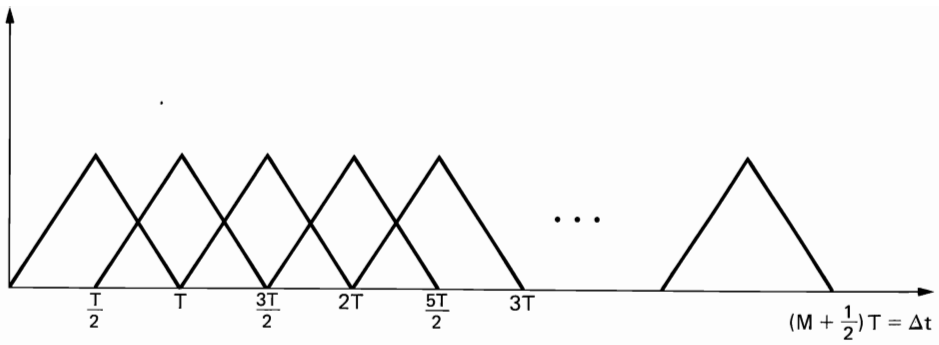


Figure 6-5 Illustration of 50% overlapped data-tapering windows.

points in the total span. Thus, the length of the time span of data analyzed is given by

$$\Delta t = (M + \tfrac{1}{2})N - 1, \tag{32}$$

and the spectral resolution width is on the order of $1/T$, that is, it is on the order of the separation between frequency points without zero padding,

$$\Delta f = \frac{1}{N} \tag{33}$$

as explained in Section B. Consequently, the temporal-spectral resolution product is given by

$$\Delta t \Delta f = M + \frac{1}{2} - \frac{1}{N} \cong M. \tag{34}$$

As explained in Section D and Chapter 5, the data-tapering window determines the effective spectral smoothing window

$$\tilde{E}(f) = \frac{1}{T}|\tilde{A}_{1/T}(f)|^2 \tag{35}$$

for time-averaged periodograms. The most basic tapering windows, including the triangle (Bartlett), raised cosine (von Hann), raised cosine on a platform (Hamming), and Blackman windows, are described in Section D of Chapter 5. Other more sophisticated windows that have certain optimality properties are treated in the literature. These include the family of *Kaiser-Bessel windows,* which are approximations to the *prolate spheroidal wavefunctions* that minimize the window energy outside a specified main-lobe spectral band for a given temporal duration [Kaiser 1966], and the family of *Dolph-Chebyshev windows,* which minimize the main-lobe spectral bandwidth for a specified peak sidelobe level and a given temporal duration [Dolph 1946] (cf. [Harris 1978; Nuttall 1981; Hamming 1983].)

2. Wiener-Daniell Method

As a second example, we consider the spectrally smoothed periodogram, which is perhaps the most popular digital method (especially for analysis of data segments that are not so long as to render the DFT too computationally complex). This method, which is referred to as the *Wiener-Daniell method*[2] [Wiener 1930; Daniell 1946], yields the statistical spectrum

$$\tilde{S}_{x_{\Delta t}}(t,f)_{\Delta f} \triangleq \frac{1}{KM} \sum_{w=-(KM-1)/2}^{(KM-1)/2} \tilde{S}_{x_{\Delta t}}\left(t, f + \frac{w}{KN}\right), \tag{36}$$

where

$$\tilde{S}_{x_{\Delta t}}(t,f) = \frac{1}{N}|\tilde{X}_{\Delta t}(t,f)|^2 \tag{37a}$$

$$= \frac{1}{N}|\tilde{x}_{\Delta t}(t,f)|^2 \tag{37b}$$

and

$$\tilde{x}_{\Delta t}(t,f) = \sum_{v=0}^{\Delta t} a_{\Delta t}(-v)x(t - \Delta t + v)e^{-i2\pi fv}. \tag{38}$$

The data-tapering window $a_{\Delta t}(u)$ is often chosen to be uniform (a rectangle), since its influence on the effective spectral smoothing window,

$$\tilde{E}(f) = \frac{1}{\Delta t}|\tilde{A}_{1/\Delta t}(f)|^2 \otimes u_{\Delta f}(f), \tag{39}$$

is diminished by the spectral smoothing operation in (36). However, this choice is not always advisable because its influence is not always negligible, as shown in Chapter 5. The number of frequency points averaged in (36) is KM, and therefore the spectral resolution is

$$\Delta f = \frac{KM}{KN} = \frac{M}{N}. \tag{40}$$

The total span of data used is

$$\Delta t = N - 1. \tag{41}$$

Consequently, the temporal-spectral resolution product is given by

$$\Delta t \Delta f = \left(1 - \frac{1}{N}\right)M \cong M. \tag{42}$$

Observe that in both the Bartlett-Welch and Wiener-Daniell methods, the resolution product $\Delta t \Delta f \cong M$ is independent of the number KN of points in each block that is transformed and is therefore not affected by zero-padding. Furthermore, the spectral resolution is on the order of the reciprocal of the time-span T of data transformed in the Bartlett-Welch method (29), regardless of the amount of zero padding, and similarly the spectral resolution is equal to the

[2] Because of the recently discovered prior suggestion of this method by Einstein [Einstein 1914], it should be referred to as the *Einstein-Wiener-Daniell method*.

smoothing parameter Δf in the Wiener-Daniell method (36), regardless of how many frequency points are contained in the band Δf and therefore regardless of zero-padding. Nevertheless, some zero-padding (e.g., $K = 2$ to $K = 4$) can be desirable because it ensures that no fine structure is overlooked due to sparse frequency sampling. That is, since we can in general say only that the spectral resolution width is *on the order of* $1/N$, it might actually be more accurately approximated by $1/2N$ or even $1/4N$, thereby requiring $K = 2$ or $K = 4$. However, any advantage gained by zero-padding becomes negligible for sufficiently large resolution product $\Delta t \Delta f$ for the Wiener-Daniell method (in which case the effective spectral smoothing window covers a substantial number of sidelobes of the interpolating function (8)) and can become negligible for the Bartlett-Welch method when the data-tapering window is very smooth (and therefore results in an effective spectral smoothing window with relatively broad main lobe covering the highest sidelobes of the interpolating function (8)).

An alternative approach to spectral smoothing that results in reduced spectral leakage has recently been proposed [Thomson 1982]. The approach is based on an orthogonal series representation of the FST of the data segment, where the orthogonal basis functions are the eigenfunctions of the operator that bandlimits (to a bandwidth of $\Delta f \cong 1/N$) and then duration-limits (to a segment length of $N = \Delta t + 1 = T + 1$) and are known as the *discrete prolate spheroidal wavefunctions*. The approach produces a spectrum estimate that is approximately a sum of weighted shifted periodograms of tapered data (called *eigenspectrum estimates*), each of which is obtained by squaring the magnitude of the FST (or a zero-padded DFT) of the data segment multiplied by a prolate spheroidal wave sequence window. Since each window has the approximate effect of shifting the spectral content of the data by a different amount, the sum of the resultant approximate shifted periodograms is effectively a spectrally smoothed periodogram. However, reduced spectral leakage results from the fact that the equivalent spectral smoothing window exhibits extremely small sidelobes for a given main-lobe width. This approach is potentially attractive in applications where the data segment is short, the range of the spectrum being estimated is large (high peaks or plateaus and/or low notches or valleys), and high computational complexity is acceptable. The details of this method are beyond the scope of this introductory treatment but are given by Thomson [Thomson 1982].

3. Blackman-Tukey Method

A third example of a digital counterpart of an analog method uses the DFT of a tapered linear correlogram,

$$\tilde{S}_{x_{\Delta t}}(t, f)_{\Delta f} \triangleq \sum_{\tau=-\Delta t}^{\Delta t} \tilde{R}_{x_{\Delta t}}(t, \tau) h_{1/\Delta f}(\tau) e^{-i2\pi f\tau} \tag{43a}$$

$$= 2\mathrm{Re}\left\{ \sum_{\tau=0}^{\Delta t} \tilde{R}_{x_{\Delta t}}(t, \tau) h_{1/\Delta f}(\tau) e^{-i2\pi f\tau} \right\} - \tilde{R}_{x_{\Delta t}}(t, 0) h_{1/\Delta f}(0), \tag{43b}$$

where $\tilde{R}_{x_{\Delta t}}(t, \tau)$ is the time-variant linear correlogram defined by

$$\tilde{R}_{x_{\Delta t}}(t, \tau) \triangleq \frac{1}{N} \sum_{u=t-\Delta t}^{t-|\tau|} x(u)x(u + |\tau|), \qquad \tau = 0, \pm 1, \pm 2, \ldots, \pm N - 1 \quad (44)$$

and $N = \Delta t + 1$ is the length of the data-segment $\{x(t - u) : u = 0, 1, 2, \ldots, \Delta t\}$ used. Observe that (43a) can be interpreted as an FST for any real f, whereas (43b) can be implemented with an N-point DFT for $f = 0, \pm 1/N, \pm 2/N, \ldots, +1/2$. It follows from the time-frequency dual of the convolution theorem for the FST that this spectrum estimate is exactly equivalent to a frequency-sampled version of a spectrally smoothed continuous-frequency periodogram. An alternative to this method uses the circular correlogram in place of the linear correlogram,

$$\tilde{S}'_{x_{\Delta t}}(t, f)_{\Delta f} \triangleq 2 \operatorname{Re}\left\{ \sum_{\tau=0}^{N-1} \tilde{R}_{x_{\Delta t}}(t, \tau)_c h_{1/\Delta f}(\tau) e^{-i2\pi f\tau} \right\} - \tilde{R}_{x_{\Delta t}}(t, 0)_c h_{1/\Delta f}(0), \quad (45)$$

where $\tilde{R}_{x_{\Delta t}}(t, \tau)_c$ is the time-variant circular correlogram for the sequence of data blocks $\{x(t - u) : u = 0, 1, 2, 3, \ldots, \Delta t\}$ indexed by t,

$$\tilde{R}_{x_{\Delta t}}(t, \tau)_c \triangleq \frac{1}{N} \sum_{u=t-\Delta t}^{t-\tau} x(\tau + u)x(u) + \frac{1}{N} \sum_{u=t-\tau+1}^{t} x(\tau + u - \Delta t - 1)x(u). \quad (46)$$

This alternative method (45) is exactly equivalent to the modified Wiener-Daniell method that uses a spectrally smoothed discrete-frequency periodogram with smoothing window $\tilde{H}_{\Delta f} = \text{DFT}\{h_{1/\Delta f}\}$, in which circular convolution is used for the spectral smoothing operation (exercise 10). This follows from the periodogram-correlogram relation for the DFT and the time-frequency dual of the circular convolution theorem for the DFT. However, circular convolution is usually not appropriate for spectral smoothing (at frequencies near the ends of the DFT). Furthermore the first method (43), which is often referred to as the *Blackman-Tukey method* [Blackman and Tukey 1959], is not as computationally efficient as the preceding two methods, especially the Bartlett-Welch method, when an FFT or FHT algorithm is used for the DFT and the data segment is long. Nevertheless, in some applications the correlogram must be computed for other purposes, and when it is available the Blackman-Tukey method is very attractive, especially when an FFT or FHT algorithm is used for the DFT of the tapered correlogram. This method is most computationally efficient when the correlogram tapering window is simply a rectangle, although the spectral leakage properties of the corresponding effective spectral smoothing window (and/or the leakage effects in the variability for short data segments) are at their worst in this case.

4. Channelizer Methods

Since the sliding DFT used in the Bartlett-Welch method (and possibly in the Wiener-Daniell method), which produces the bank of $1 + KN/2$ local sine wave components

$$\bar{x}_T(t, f), \qquad f = 0, \frac{1}{KN}, \frac{2}{KN}, \frac{3}{KN}, \ldots, \frac{1}{2},$$

can be interpreted as a bank of band-pass filters with center frequencies of $\{m/KN : m = 0, 1, 2, 3, \ldots, KN/2\}$, then an obvious digital implementation can be based on a bank of computationally efficient band-pass filters, collectively referred to as a *channelizer* (i.e., a spectral channelizer). In fact, this is the digital counterpart of the analog wave-analysis method. Computational savings can be obtained by exploiting the fact that the low-pass time-series to be time-averaged, $|\tilde{x}_T(t, f)|^2 = |\bar{X}_T(t, f)|^2$ for each f, has bandwidth of only $2\Delta f^{\circ} = 2/N$, and $\tilde{x}_T(t, f)$ can therefore be subsampled at a rate of $2/N$ or preferably $4/N$ before it is magnitude-squared and averaged. Also, the time-averaging can be implemented with a computationally efficient digital low-pass filter. The subsampling at a rate of $2/N$ is equivalent to using hopped time-averaging of the periodogram with 50% overlapped blocks, as in the Bartlett-Welch method. The shape of the transfer function of each band-pass filter is equivalent to the DFT of the data-tapering window (frequency-shifted up to passband) in the Bartlett-Welch method (provided that zero-padding is used so that the linear convolution (31) is equivalent to a circular convolution). These equivalences follow by analogy from the explanations of equivalences among analog methods in Chapters 4 and 5. Increasingly efficient and fast computational procedures for realizing channelizers are continually being developed. Thus, this digital wave-analysis generalization of the Bartlett-Welch method is one of the most promising of the existing generic architectures for digital statistical spectral analysis. A variety of approaches to efficient implementation are presented in [Crochiere and Rabiner 1983].

5. Minimum-Leakage Method

A statistical spectrum estimate such as that provided by a wave analyzer at a given frequency f can be interpreted as the time-averaged power at the output of a narrow-band band-pass filter with center frequency f, normalized by the filter bandwidth. The instantaneous power is averaged over a time span of length $\Delta t = L - 1$ and the bandwidth of the filter is $\Delta f = 1/N$. Thus,

$$\bar{S}_{x_{1/\Delta f}}(t, f)_{\Delta t} = \frac{1}{L} \sum_{w=0}^{L-1} \frac{1}{\Delta f} P_{\Delta f}(t - w, f), \tag{47a}$$

where $P_{\Delta f}(t, f)$ is the instantaneous power of the filter output,

$$P_{\Delta f}(t, f) = \left| \sum_{u=0}^{N-1} a^f_{1/\Delta f}(u) x(t - u) \right|^2, \tag{47b}$$

and $a^f_{1/\Delta f}(u)$ is the impulse-response sequence of the filter. (In terms of the data-tapering window in (31), $a^f_{1/\Delta f}(u) = a_{1/\Delta f}(T - u)\, e^{i2\pi f(u-T)}$.) One approach to optimizing this type of spectrum estimate in order to minimize spectral leakage is to optimize the band-pass filter for each value of f to minimize the average power $S_{x_{1/\Delta f}}(t, f)_{\Delta t}$ at its output subject to the constraint that the strength of its response to an input sine wave of frequency f is equal to unity,

$$\sum_{u=0}^{N-1} a^f_{1/\Delta f}(u) e^{i2\pi f(t-u)} = e^{i2\pi ft}. \tag{48}$$

It is shown in exercise 14 that the solution for the optimum filter is given by the N-vector[3]

$$\boldsymbol{g} = [a^f_{1/\Delta f}(0), a^f_{1/\Delta f}(1), a^f_{1/\Delta f}(2), \ldots, a^f_{1/\Delta f}(N-1)]' \tag{49}$$

that is specified by

$$\boldsymbol{g}(t,f) = \frac{\boldsymbol{R}_x^{-1}(t)\boldsymbol{s}(f)}{\boldsymbol{s}'(f)\boldsymbol{R}_x^{-1}(t)\boldsymbol{s}^*(f)} \tag{50a}$$

for each data block indexed by t, where $\boldsymbol{s}(f)$ is the sine wave N-vector

$$\boldsymbol{s}(f) = [1, e^{i2\pi f}, e^{i2\pi f2}, e^{i2\pi f3}, \ldots, e^{i2\pi f(N-1)}]', \tag{50b}$$

$\boldsymbol{R}_x(t)$ is the $N \times N$ matrix of time-variant finite-average autocorrelations with uvth element defined by

$$[\boldsymbol{R}_x(t)]_{uv} \triangleq \frac{1}{L}\sum_{w=0}^{L-1} x(t-w-u)x(t-w-v) \tag{50c}$$

for $u, v = 0, 1, 2, \ldots, N-1$, and $\boldsymbol{R}_x^{-1}(t)$ is the inverse of this matrix. Substitution of (49) into (47) yields

$$\tilde{S}_{x_{1/\Delta f}}(t,f)_{\Delta t} = \frac{1}{\Delta f}\boldsymbol{g}'(t,f)\boldsymbol{R}_x(t)\boldsymbol{g}^*(t,f), \tag{51}$$

and substitution of (50a) into (51) yields (exercise 14) the *minimum-leakage* (ML) *spectrum estimate* (using the notation $S_x(f)_{\mathrm{ML}} \triangleq \tilde{S}_{x_{1/\Delta f}}(t,f)_{\Delta t}$)

$$S_x(f)_{\mathrm{ML}} = \frac{1/\Delta f}{\boldsymbol{s}'(f)\boldsymbol{R}_x^{-1}(t)\boldsymbol{s}^*(f)}. \tag{52}$$

Observe that the optimum filter (50) adapts to the data and is therefore time-variant. Consequently, its bandwidth Δf is also adaptive. One measure of the bandwidth of a filter with transfer function $H(\nu)$ is its equivalent rectangular bandwidth, defined by

$$\Delta f \triangleq \frac{1}{|H(f)|^2}\int_{-1/2}^{1/2}|H(\nu)|^2\,d\nu, \tag{53}$$

where f is the center frequency. Substitution of the transfer function

$$H(\nu) = A^f_{\Delta f}(\nu) = \sum_{u=0}^{N-1} a^f_{1/\Delta f}(u)e^{-i2\pi\nu u} \tag{54}$$

into (53) and use of (49) and (50a) yields (exercise 14)

$$\Delta f = \frac{\boldsymbol{s}'(f)\boldsymbol{R}_x^{-2}(t)\boldsymbol{s}^*(f)}{[\boldsymbol{s}'(f)\boldsymbol{R}_x^{-1}(t)\boldsymbol{s}^*(f)]^2}, \tag{55}$$

where $\boldsymbol{R}_x^{-2}(t)$ is the square of the inverse matrix $\boldsymbol{R}_x^{-1}(t)$. Finally, substitution of (55) into (52) yields the formula

$$S'_x(f)_{\mathrm{ML}} = \frac{\boldsymbol{s}'(f)\boldsymbol{R}_x^{-1}(t)\boldsymbol{s}^*(f)}{\boldsymbol{s}'(f)\boldsymbol{R}_x^{-2}(t)\boldsymbol{s}^*(f)} \tag{56}$$

for the *modified ML spectrum estimate*.

[3] The notation [ ]′ denotes matrix transposition.

Although the derivation of this optimum spectrum estimate presented here is believed to be novel (in that it does not rely on probabilistic models or infinite amounts of data), the estimate itself, in the form (52) with fixed Δf, was proposed at least as early as the mid-1960s [Capon et al. 1967; Capon 1969; Lacoss 1971], but was only recently modified to the form (56) [Lagunas-Hernandez and Gasull-Llampallas 1984]. This spectrum estimate is known to provide somewhat better resolution of spectral lines than the nonoptimized methods and this is demonstrated in Chapter 9. However, it is also somewhat more computationally complex. Nevertheless its computational complexity can be reduced if the Toeplitz form of $\boldsymbol{R}_x(t)$, in which the uvth element depends on only the difference $u - v$, is exploited, and if the inner products

$$\boldsymbol{s}'(f)\boldsymbol{R}_x^p(t) \quad \text{and} \quad [\boldsymbol{s}'(f)\boldsymbol{R}_x^p(t)]\boldsymbol{s}^*(f)$$

(for $p = -1, -2$) are performed using an FFT algorithm. The Toeplitz form of $\boldsymbol{R}_x(t)$ is obtained when the range of summation in (47a) and, therefore, (50c) is chosen large enough to allow the filter to run completely off the ends of the L-point data segment being analyzed.[4] This form can be exploited for efficient matrix inversion, as explained in Chapter 9 (see also [Musicus 1985]). Whatever algorithm is used for the DFT, zero-padding should be used to obtain a frequency increment smaller than $1/N$, since the spectral resolution capability of this method is typically better than that of the nonoptimized methods, which produce a spectral resolution width on the order of $1/N$. Nevertheless, the resolution capability is still tied to N, and there is still a resolution-reliability trade-off. For a given data-segment length of L, as N is increased, resolution increases; but so too does variability due to the increased variability of $[\boldsymbol{R}_x(t)]_{uv}$ for values of u or v close to N. Furthermore, this variability is enhanced by the matrix inversion operation in (52) and is enhanced even more by the squaring operation and the ratio in (56). Large variability can result in spurious peaks in the spectrum estimate. Some improvement in this trade-off with N might be possible by variability reduction through modification of the correlation matrix estimate (50c). For example, the forward-backward covariance-matrix estimate and/or the singular-value-decomposition method of rank reduction described in Chapter 9 might be used (e.g., for data consisting of sine waves in white noise). The higher variability of the modified estimate (56) can render it less desirable than the unmodified estimate (52) (see Section E in Chapter 9).

Since the filter (50a) with $\boldsymbol{R}_x(t)$ replaced by the limit autocorrelation matrix can be shown (exercise 14) to be the solution to a maximum-likelihood estimation problem for a sine wave in additive Gaussian noise, the spectrum estimation method (52) is often called the *maximum-likelihood method* (MLM), but MLM also stands for the *minimum-leakage method*. The minimum-leakage method is also sometimes referred to as the *Capon method* and the *minimum-variance method*. It should not be confused with the classical maximum-likelihood autoregressive method described in Chapter 9.

[4] The estimate (52) was originally proposed for wave-number analysis of spatial data, in which case the Toeplitz form of $\boldsymbol{R}_x(t)$ is not appropriate.

D. FRACTION-OF-TIME PROBABILISTIC ANALYSIS

Fraction-of-time probabilistic analysis of discrete-time spectrum estimates is analogous to the analysis carried out in Chapter 5 for continuous-time spectrum estimates. The obvious differences are that integrals over continuous time must be replaced with sums over discrete time and CFTs must be replaced with FSTs, in which case integrals over the infinite frequency range $(-\infty, \infty)$ get replaced with integrals over the finite frequency range $[-\frac{1}{2}, \frac{1}{2}]$. Specifically, any of the quadratic time-invariant spectrum estimates (such as the Bartlett-Welch, Wiener-Daniell, Blackman-Tukey, and channelizer estimates) can be put into the general form

$$\bar{y}_f(t) = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \bar{k}_f(u, v)x(t - u)x(t - v) \tag{57}$$

for some appropriate kernel $\bar{k}_f$ which can be represented by its double FST

$$\bar{K}_f(\nu', \mu') \triangleq \sum_{u,v=-\infty}^{\infty} \bar{k}_f(u, v)e^{-i2\pi(u\nu' - v\mu')}, \tag{58}$$

which in turn can be represented by the kernel transform

$$\tilde{M}(\nu, \mu) \triangleq \bar{K}_f(\nu + f + \mu/2, \nu + f - \mu/2). \tag{59}$$

Then, the temporal mean of $\bar{y}_f(t)$ is given by (see (50) in Chapter 5)

$$\text{mean } \{\bar{y}_f(t)\} = \int_{-1/2}^{1/2} \bar{S}_x(f - \nu)\bar{E}(\nu)\, d\nu, \tag{60}$$

where

$$\bar{E}(f) \triangleq \tilde{M}(0, -f). \tag{61}$$

Also, the temporal variance of $\bar{y}_f(t)$ is given by (see (66) in Chapter 5)

$$\text{var}\{\bar{y}_f(t)\} = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} [|\tilde{M}(\nu, \mu - f)|^2 + \tilde{M}(\nu, \mu - f)M^*(\nu, -\mu - f)]$$

$$\times \bar{S}_x\left(\mu + \frac{\nu}{2}\right) \bar{S}_x\left(\mu - \frac{\nu}{2}\right) d\mu\, d\nu. \tag{62}$$

Thus, all that is needed to evaluate the mean and variance is to determine the kernel transform $\tilde{M}(\nu, \mu)$ using (58) and (59) by putting the spectrum estimate into the form of (57). The resultant kernel transform will be analogous to those given in Table 5-1 for continuous-time methods. Various effective spectral smoothing windows $H(f) \equiv \bar{E}(f)$ for discrete-time methods are described in Section D of Chapter 5.

The analogy with continuous time is not complete, and this can be seen by attempting to define the inverse double FST of $\tilde{M}(\nu, \mu)$, namely $\tilde{m}(t, \tau)$, directly in terms of $\bar{k}_f(u, v)$ as done for continuous time in Chapter 5. But since

this intermediate step in reaching the kernel transform $\tilde{M}(\nu, \mu)$ is not necessary, we shall not pursue the matter any further.[5]

E. SUMMARY

In Section A, the complementary nature of analog and digital methods of spectral analysis are discussed. Then in Section B the DFT, on which most digital methods are based, is studied. Topics include the use of zero-padding to control resolution, the distinction between circular and linear convolutions, the circular convolution theorem and the associated wraparound phenomenon, and a circular correlogram-periodogram relation. Also, the relationships among the DFT, FST, and CFT are described, and the importance of zero-padding is discussed and illustrated by example. In Section C various digital methods for statistical spectral analysis that are based on the DFT are described and compared. It is explained that these methods, known by the names Bartlett-Welch, Wiener-Daniell, Blackman-Tukey, and channelizer methods, are all digital counterparts of analog methods studied in Chapter 4. Then the minimum-leakage method, which is an optimized wave analyzer (channelizer), is derived and its interpretation in terms of maximum likelihood is explained. Finally in Section D, it is explained that the formulas derived in Chapter 5 for the mean and variance of continuous-time spectrum estimates apply equally well to discrete-time spectrum estimates, provided only that the range of integration over frequency variables is reduced from $(-\infty, \infty)$ to $[-\frac{1}{2}, \frac{1}{2}]$, to reflect the replacement of the CFT by the FST in the derivation.

EXERCISES

1. **(a)** Use polynomial division to prove that $Q(a) = P(a)$ for $a \neq 1$, where

$$Q(a) \triangleq \frac{1 - a^N}{1 - a} \tag{63}$$

$$P(a) \triangleq \sum_{k=0}^{N-1} a^k. \tag{64}$$

Hint:

$$\begin{array}{r} 1 + a + a^2 + \cdots \\ 1 - a \,\overline{)\, 1 - a^N} \\ \underline{1 - a} \\ a - a^N \\ \underline{a - a^2} \\ a^2 - a^N \\ \underline{a^2 - a^3} \\ a^3 - a^N \\ \vdots \end{array}$$

[5] The reader should be warned that the formulas (60)–(62) were not derived by the author. They were simply written down by analogy with (50) and (66) in Chapter 5. Readers with any question about the analogy should derive (60)–(62) directly by following a procedure that is analogous to that outlined in Chapter 5.

Since $P(1) = N$ and $Q(a) \to N$ as $a \to 1$, it is convenient to *define* $Q(1)$ to be

$$Q(1) \triangleq N. \tag{65}$$

(b) Use the result of (a) to prove that

$$\sum_{k=0}^{N-1} e^{i2\pi rk/N} = \frac{1 - e^{i2\pi r}}{1 - e^{i2\pi r/N}} = \begin{cases} N, & r = pN,\ p = \text{integer} \\ 0, & \text{otherwise.} \end{cases} \tag{66}$$

2. (a) Verify that (1a) and (1b) are a transform pair by substituting (1a) into (1b) and using the result of (b) in exercise (1).

(b) Define an $N \times N$ matrix $\boldsymbol{M}$ with mnth element $M_{mn} \triangleq \frac{1}{\sqrt{N}} W^{mn}$ for $n, m = 0, 1, 2, \ldots, N - 1$, where $W \triangleq e^{-i2\pi/N}$. Then the scaled DFT and inverse scaled DFT can be expressed in terms of $\boldsymbol{M}$ by

$$\boldsymbol{X} = \boldsymbol{M}\boldsymbol{x}$$

$$\boldsymbol{x} = \boldsymbol{M}^{-1}\boldsymbol{X}.$$

Show that $\boldsymbol{M}$ is symmetric,

$$\boldsymbol{M}' = \boldsymbol{M},$$

and that it is *orthonormal* in the sense that

$$\boldsymbol{M}'\boldsymbol{M}^* = \boldsymbol{I},$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix. *Hint:* Use the result of exercise 1(b).

3. (a) Use the result of exercise 1(b) to verify that the FST of the rectangle sequence

$$u_n = \begin{cases} 1, & 0 \leq n \leq M - 1 \\ 0, & \text{otherwise} \end{cases} \tag{67a}$$

is given by

$$\tilde{U}(f) = e^{-i\pi f(M-1)} \frac{\sin(\pi f M)}{\sin(\pi f)}. \tag{67b}$$

(b) Use the aliasing formula (15b) for a time-sampled rectangle,

$$u(t) = \begin{cases} 1, & 1 \leq t < M \\ 0, & \text{otherwise,} \end{cases} \tag{68}$$

to verify that the FST $\tilde{U}(f)$ in (a) is also given by (for M even)

$$\tilde{U}(f) = e^{-i\pi fM} \sum_{q=-\infty}^{\infty} \frac{\sin[\pi(f - q)M]}{\pi(f - q)}. \tag{67c}$$

Thus, (67b) and (67c) form an identity.

4. To verify that the DFT (7) of a zero-padded sequence is simply an interpolation of the DFT (1a) of the unpadded sequence, substitute (1b) into (7) and use the result of exercise 1(b) to prove that (8a) holds with

$$I\left(\frac{q}{K}\right) = \frac{1 - e^{i2\pi(q/K)}}{N(1 - e^{i2\pi(q/K)/N})}. \tag{69}$$

Then factor out $e^{i\pi(q/K)}/e^{i\pi(q/K)/N}$ to verify that the magnitude and phase of $I(q/K)$ are given by (8b) and (8c).

5. (a) Illustrate graphically the equivalence of circular convolution of finite-length sequences (11a) and linear convolution of the periodically repeated versions of the finite-length sequences (10). Also show graphically that zero-padding sequences of length N out to length $2N$ renders linear convolution of the zero-padded sequences equivalent to linear convolution of the periodically repeated (period $= 2N$) sequences.

Finally, show that when one sequence is much shorter than the other, then the difference between linear convolution and circular convolution occurs only near the ends of the resultant sequences.

(b) Verify the *circular convolution theorem,*

$$Z_m = X_m Y_m, \tag{70a}$$

where X_m, Y_m, Z_m are the N-point DFTs of x_n, y_n, z_n and where z_n is the circular convolution of x_n and y_n,

$$z_n = \sum_{k=0}^{n} x_k y_{n-k} + \sum_{k=n+1}^{N-1} x_k y_{N+n-k}, \qquad n = 0, 1, 2, \ldots, N-1, \tag{70b}$$

in which the second sum is defined to be zero for $k > N - 1$. *Hint:* Let $\{x_n : n = 0, 1, 2, \ldots, N-1\}$ and $\{y_n : n = 0, 1, 2, \ldots, N-1\}$ be periodically repeated,

$$x_{n+N} = x_n, \qquad y_{n+N} = y_n, \qquad -\infty < n < \infty,$$

and use the result of (a) to show that the circular convolution (70b) of the original finite segments is identical to the linear convolution

$$z_n = \sum_{k=0}^{N-1} x_k y_{n-k} \tag{70c}$$

of the periodically repeated sequences. Then take the DFT of one period of both sides of (70c) and insert the unity factor $e^{-i(2\pi/N)(mk-mk)}$ to obtain

$$\bar{Y}_m = \sum_{k=0}^{N-1} x_k \, e^{-i(2\pi/N)mk} \sum_{n=0}^{N-1} y_{n-k} \, e^{-i(2\pi/N)m(n-k)}.$$

Finally, use the periodicity of the summand in the second sum to argue that k can be set equal to any fixed value, say $k = 0$, in the second sum. The result (70a) follows immediately.

(c) Verify that if x_n is zero for $n < 0$ and $n > N_x - 1$ and y_n is zero for $n < 0$ and $n > N_y - 1$, then by appending enough zeros to x_n and y_n to obtain sequences of length $N_{xy} = N_x + N_y - 1$, the N_{xy}-point circular convolution of these zero-padded sequences is identical to the linear convolution

$$z_n = \sum_{q=0}^{N_{xy}-1} x_q y_{n-q}. \tag{71}$$

6. (a) Verify that the N-point DFT of the *circular correlogram,*

$$\tilde{R}_{x_N}(q)_c \triangleq \frac{1}{N} \sum_{k=0}^{N-q-1} x_{q+k} x_k + \frac{1}{N} \sum_{k=N-q}^{N-1} x_{q+k-N} x_k, \qquad q = 0, 1, 2, \ldots, N-1, \tag{72a}$$

is given by $1/N$ times the squared magnitude of the N-point DFT of x_n, which is the N-point periodogram of x_n,

$$\mathrm{DFT}\{\tilde{R}_{x_N}(q)_c\} = \frac{1}{N} |\mathrm{DFT}\{x_n\}|^2. \tag{72b}$$

Hint: Let $\{x_n\}$ be periodically repeated and show that (72a) is identical to

$$\tilde{R}_{x_N}(q)_c = \frac{1}{N} \sum_{k=0}^{N-1} x_{k+q} x_k. \tag{72c}$$

Then proceed as described in the hint for exercise 5(b).

(b) Verify that the circular correlogram is symmetrical about the midpoint $N/2$ (N even)

$$\tilde{R}_{x_N}(N/2 + p)_c = \tilde{R}_{x_N}(N/2 - p)_c. \tag{73}$$

(Notice that $N/2$ is the midpoint for the set $[0, N]$ rather than $[0, N - 1]$, which has no integral midpoint for even N. Notice also that $\tilde{R}_{x_N}(q)_c$ peaks at $q = 0$, as does the linear correlogram.)

(c) Derive the DFT relation (15) from the FST relation (14).

7. As a supplement to the periodogram-correlogram relation (72b), verify the following relation:

$$2\,\mathrm{Re}[\mathrm{DFT}\{\tilde{R}_{x_N}(q)\}] - \tilde{R}_{x_N}(0) = \frac{1}{N}|\mathrm{DFT}\{x_n\}|^2, \tag{74a}$$

where $\tilde{R}_{x_N}(q)$ is the *linear correlogram* defined by

$$\tilde{R}_{x_N}(q) \triangleq \frac{1}{N}\sum_{k=0}^{N-1} x_k x_{k+|q|}, \qquad q = 0, \pm 1, \pm 2, \ldots, \pm N - 1. \tag{74b}$$

Hint: Use the periodogram-correlogram relation for the FST ((63) in Chapter 2) to obtain

$$\sum_{q=-N+1}^{N-1} \tilde{R}_{x_N}(q)e^{-i2\pi qf} = \frac{1}{N}\left|\sum_{n=0}^{N-1} x_n e^{-i2\pi fn}\right|^2, \tag{75}$$

and then use the fact that $\tilde{R}_{x_N}(q)$ is even together with $f = m/N$ to obtain the desired result.

8. Show that the DFT of the circular correlogram (20) of a PN sequence is given by (22).

9. Let $\{z_n\}$ be the $(N + 1)$-point sequence obtained by appending x_* to the end of an N-point sequence $\{x_n\}$. Show that the $(N + 1)$-point DFT of $\{z_n\}$ is related to the FST of $\{x_n\}$ by

$$Z_m = \tilde{X}\left(\frac{m}{N} - \frac{m}{N(N + 1)}\right) + x_* e^{-i2\pi mN/(N+1)}. \tag{76}$$

Then let $\{x_n\}$ be a PN sequence and let $x_* = \pm 1$ and show that for $m \neq 0$ and $N \gg 1$, $\{Z_m\}$ is closely approximated by

$$Z_m \cong \tilde{X}\left(\frac{m}{N} - \frac{m}{N(N + 1)}\right) \tag{77a}$$

whereas $\{X_m\}$ is given by

$$X_m = \tilde{X}\left(\frac{m}{N}\right). \tag{88b}$$

10. Show that the FST of the limit autocorrelation of a periodic sequence with period N,

$$\tilde{R}_x(k) \triangleq \lim_{Q\to\infty} \frac{1}{2Q + 1}\sum_{n=-Q}^{Q} x_{n+k}x_n, \tag{78a}$$

is given by

$$\tilde{S}_x(f) = \frac{1}{N}\sum_{p=-\infty}^{\infty}\left|\tilde{X}_N\left(\frac{p}{N}\right)\right|^2 \delta\left(f - \frac{p}{N}\right), \tag{78b}$$

where

$$\tilde{X}_N(f) \triangleq \sum_{n=0}^{N-1} x_n e^{-i2\pi fn} \tag{78c}$$

and therefore $\tilde{X}_N([p + N]/N) = \tilde{X}_N(p/N)$. Consequently, $\tilde{S}_x(f)$ contains the same information as the N-point periodogram

$$\frac{1}{N}\left|\tilde{X}_N\left(\frac{p}{N}\right)\right|^2, \qquad p = 0, 1, 2, \ldots, N - 1.$$

Hint: Take the inverse FST of (78b) and then apply the periodogram-correlogram relation for the DFT.

11. **(a)** Show that the KN-point DFT of the N-point sequence

$$x_n = \cos\left(\frac{2\pi kn}{KN} + \theta\right), \qquad n = 0, 1, 2, \ldots, N - 1 \tag{79a}$$

with $(K - 1)N$ zeros appended is given by

$$X_m = 0 \qquad \text{for } m = \pm k + pK, \tag{79b}$$

for all integers $p \neq 0, \pm N, \pm 2N, \pm 3N, \ldots$, if $k = qK/2$ for any integer q such that k is an integer and $K \neq 1$.

(b) Show that if $K = 1$, then $X_m = 0$ for $m \neq \pm k$, and the DFT therefore exhibits no leakage at all when there is no zero-padding. This nonphysical situation should be avoided when conducting simulations by never choosing a sine wave frequency to be exactly at the center of a bin in the DFT.

(c) Explain (a) and (b) by characterizing the DFT as a frequency-sampled FST and showing the graph of the FST of the sine wave sequence x_n.

12. **(a)** Consider a sampling increment $T_s \neq 1$. Give expressions for the generalizations (from $T_s = 1$) of the Bartlett-Welch equations (29)–(31) and the Wiener-Daniell equations (36)–(38).

(b) Give the equations that relate Δt and Δf to K, L, M, N, and T_s for the Bartlett-Welch method, and the equations that relate Δt and Δf to K, M, $L = N$, and T_s for the Wiener-Daniell method.

(c) Give a formula for the frequency-sampling increment, denoted by F_s, of the DFT in terms of K, N, and T_s.

13. **(a)** Consider a stationary time-series with bandwidth $B = 100$ KHz ($\hat{S}_x(f) = 0$ for $|f| \geq B$) and correlation width $\Delta\tau^* = 1$ ms. It is desired to use a frequency-smoothed periodogram computed with the FFT algorithm without zero-padding to measure the spectral density, and it is desired to have a temporal-spectral resolution product of $\Delta t \Delta f = 100$. Determine the minimum allowable time-sampling rate, $1/T_s$, the number, N, of time-samples to be Fourier transformed (the size of the FFT), the separation, F_s, between frequency points, and the number, M, of frequency points to be averaged for each frequency point in the smoothed spectrum.

(b) Let an average of periodograms of half-overlapped data segments be used in place of the frequency-smoothed periodogram in (a), and determine the number of data segments to be transformed and the number of time-points within each segment.

14. **(a)** Use (49) and (50c) to verify that the spectrum estimate (47a) can be expressed as

$$\tilde{S}_{x_{1/\Delta f}}(t, f)_{\Delta t} = \frac{1}{\Delta f}\boldsymbol{g}'(t, f)\boldsymbol{R}_x(t)\boldsymbol{g}^*(t, f), \tag{80}$$

and use (49) and (50b) to show that the constraint (48) on the filter $g(t, f)$ can be expressed as

$$g'(t,f)s^*(f) = 1. \tag{81}$$

(b) A standard result from optimization theory is that the quadratic form (80) is minimum subject to the linear constraint (81) if and only if $g(t, f)$ satisfies the linear equation

$$R_x(t)g(t,f) = cs(f) \tag{82}$$

for some constant c. Since $R_x(t)$ is assumed to be positive definite, then the solution $g(t, f)$ can be obtained by inverting $R_x(t)$ and then substituting the resultant formula for $g(t, f)$ into (81) to determine the appropriate value for c. Show that this yields (50a).

(c) Show that (50a) and (51) yield the ML spectrum estimate (52).

(d) Use (49) and (50a) to show that (53) can be expressed by (55).

15. To establish the relationship between (52) and the time-averaged power of a sequence of maximum-likelihood estimates of a sine wave component of $x(t)$ with frequency f, proceed as follows. It can be shown that the Gaussian fraction-of-time distribution for the N-vector consisting of a *real* sine wave vector as with unknown amplitude a plus a zero-mean residual vector n with limit autocorrelation matrix $\bar{R}_n$,

$$x = as + n,$$

is given by (cf. Chapter 15)

$$f_x(z|as) = [(2\pi)^{N/2}|\hat{R}_n|^{1/2}]^{-1} \exp\{-\tfrac{1}{2}(z - as)'\hat{R}_n^{-1}(z - as)\}, \tag{83}$$

where $|\hat{R}_n|$ denotes the determinant of $\hat{R}_n$ and $x = [x(t), x(t - 1), \ldots, x(t - N + 1)]'$. It follows from the monotonicity of $\exp\{\cdot\}$ that the value of a that maximizes $f_x(z|as)$ is the value of a that minimizes the quadratic form

$$(z - as)'\hat{R}_n^{-1}(z - as). \tag{84}$$

Expand (84) into a sum of four terms, and show that the minimizing value of a is given by (using $z = x$)

$$\hat{a} = \frac{x'\hat{R}_n^{-1}s}{s'\hat{R}_n^{-1}s} \triangleq x'h = h'x. \tag{85}$$

Now, for each data block $x = [x(t), x(t - 1), \ldots, x(t - N + 1)]'$ indexed by t, we get an estimate $\hat{a} = \hat{a}(t)$. Show that the limit time-averaged power of this sequence of estimates is given by

$$\hat{P}_a(f) = h'\hat{R}_x h = \frac{s'\hat{R}_n^{-1}\hat{R}_x\hat{R}_n^{-1}s}{(s'\hat{R}_n^{-1}s)^2}, \tag{86}$$

where

$$\hat{R}_x = a^2ss' + \hat{R}_n. \tag{87}$$

Finally, use (86) and (87) to show that

$$\hat{P}_a(f) = \frac{1}{s'\hat{R}_x^{-1}s}, \tag{88}$$

which is to be compared with (52). (It can be shown that for a sufficiently large value of N, (88) is, to a close approximation, independent of the phase of the real sine wave vector s (because of the near-Toeplitz form of $\hat{R}_n^{-1}$ for large N) and is therefore

closely approximated by the same form with s replaced by the complex sine wave vector (50b).)

Hint: Use *Woodbury's identity,*

$$[\hat{R}_x - a^2 ss']^{-1} = \hat{R}_x^{-1} + \frac{a^2 \hat{R}_x^{-1} ss' \hat{R}_x^{-1}}{1 - a^2 s' \hat{R}_x^{-1} s}, \tag{89}$$

to show that

$$\hat{R}_n^{-1} s = \hat{R}_x^{-1} s \left[1 + \frac{a^2 s' \hat{R}_x^{-1} s}{1 - a^2 s' \hat{R}_x^{-1} s}\right] = \hat{R}_x^{-1} s \left[\frac{1}{1 - a^2 s' \hat{R}_x^{-1} s}\right] \tag{90}$$

and

$$s' \hat{R}_n^{-1} s = s' \hat{R}_x^{-1} s \left[\frac{1}{1 - a^2 s' \hat{R}_x^{-1} s}\right]. \tag{91}$$

Then substitute (90) and (91) into (86) to obtain (88).

16. Consider the problem of measuring the squared magnitude of the transfer function of a resonant system driven by white noise, as discussed in the first example in Section D of Chapter 5. Assume that the resonance peak occurs at $f_0 = 10$ KHz and has a 3-dB bandwidth of 1 KHz and that it is desired to obtain a spectral resolution of 200 Hz and a coefficient of variation of $\frac{1}{10}$ over the spectral band from dc to 30 KHz.

(a) Determine an appropriate sampling rate for the response of the resonant system, and determine roughly how many time-samples will be needed.

(b) Specify an amount of data and zero-padding, the required DFT size, and the number of DFT blocks to be averaged using the Bartlett-Welch method with 50% overlap and no data tapering. Restrict the DFT size to an integer power of 2 so that FFT or FHT algorithms can be used.

(c) Specify an amount of data and zero-padding, the required DFT size, and the number of DFT bins to be averaged using the Wiener-Daniell method with no data tapering. Restrict the DFT size to an integer power of 2.

(d) How would the specifications in (b) and (c) change if triangle data tapering were used?

(e) Identify the autocorrelation tapering windows in the Blackman-Tukey method that is equivalent to the modified Wiener-Daniell method (that uses circular convolution for spectral smoothing) with and without data tapering as in (d) and (c). Also determine the DFT size needed for transforming the tapered autocorrelation. Restrict the DFT size to an integer power of 2.

(f) How many digital band-pass filters would be needed if a channelizer were used, and what would their 3-dB bandwidths be?

17. An oscillator produces an approximation to an exact sine wave. The distortion exhibited in the oscillator's periodic output is usefully characterized by the strength of the components at the harmonics (integer multiples) of the fundamental frequency of the periodic waveform. Assume that the fundamental frequency is 60 Hz and that the first 10 harmonics are the only nonnegligible ones. Specify an appropriate spectral resolution and the total amount of data to be analyzed in order to assess the amount of harmonic distortion. Discuss the basis for selection of a particular data-tapering window to be used, and determine an appropriate amount of zero-padding and the required DFT size (an integer multiple of 2). Assume that oscillator noise and other random effects are negligible, in which case no time-averaging or frequency-smoothing is required.

18. Voiced human speech can be thought of as being generated by passing a nearly periodic pulse train through a slowly time-varying filter. The resultant waveform can be approximately modeled as a fourth- (or sixth-) order autoregressive time-series with autoregression parameters that are nearly time-invariant over intervals of length 25 to 250 ms. The locations of the two spectral peaks in this model, which are called the (first two) *formant frequencies,* are typically contained within the bands 400–1200 Hz and 1000–2400 Hz. The great majority of the speech energy is contained within the band 300–3000 Hz, but during intervals of length 25–250 ms, the majority of energy is in a fraction (roughly $\frac{1}{3}$) of this band, centered around the two formant frequencies, and the 3-dB bandwidths of these peaks is typically on the order of a few hundred Hertz.

Consider the problem of measuring the time-variant spectrum of a speech waveform and specify the following parameters for the Wiener-Daniell method:

(a) Preanalysis-filter bandwidth (to minimize aliasing effects)
(b) Sampling rate
(c) Segment length to be transformed (N)
(d) Number of zeros for padding ($[K - 1]N$)
(e) Width of spectral smoothing window (M)
(f) Hop interval (number of time samples) for discrete-time tracking of the spectrum
(g) The resolution product $\Delta t \Delta f$ realized with the parameters selected

19. Carry out the discrete-time and discrete-frequency counterpart of exercise 9 in Chapter 4. In part (e) consider the Bartlett-Welch, Wiener-Daniell, and Blackman-Tukey methods.

20. Consider a phenomenon that produces data of the form

$$x(t) = \sum_{k=1}^{K} c_k e^{-t/\tau_k} \cos(\omega_k t + \theta_k), \qquad t \geq 0,$$

where $0 \leq \omega_k \leq 1/\tau_k$. This condition implies that the higher the frequency of oscillation of a given component is, the faster it decays. Consequently as time progresses, the range of frequencies of non-negligible components becomes lower and lower. In such a case, time-sampling can be made more and more sparse as time progresses without significant loss of information. Thus, an exponential time-sampling scheme in which the sampling times are related by $t_{n+1} = ct_n$ for some constant c is appropriate. If these time samples are to be used for spectral analysis over a very broad range of frequencies, say several decades, then an exponential frequency-sampling scheme can also be appropriate, say $f_{m+1} = cf_m$. Use the approximation

$$X(f_m) = \int_{t_0}^{T} x(t) e^{-i2\pi f_m t} dt \cong \sum_{n=0}^{N-1} x(t_n) \int_{t_n}^{t_{n+1}} e^{-i2\pi f_m t} dt \tag{92}$$

together with $t_n = c^n t_0$ and $f_m = c^m f_0$ to show that

$$X(f_m) \cong \frac{1}{f_m} \sum_{n=0}^{N-1} A(m + n) x(t_n), \qquad m = 0, 1, 2, \ldots, M - 1, \tag{93}$$

where

$$N = \log_c(T/t_0)$$

$$M = \log_c(F/f_0)$$

and $[f_0, F]$ is the frequency band of interest. Show that the $M \times N$ matrix with mnth element $A(n + m)$ is completely specified by only $N + M$ numbers, and the matrix product (93) is actually a convolution. Thus, an FFT algorithm can be used

to compute (93). Assume that $c = 2$, $t_0 = 10^{-1}$, $T = 10^5$, $f_0 = 10^{-1}$, $F = 10^5$, and compare the size of the FFT needed using (93) with that needed using uniform time sampling, $t_{n+1} - t_n = 1/2F$, and uniform frequency sampling with $M = N$. In practice a more conservative value of c closer to unity could be used. Also, improved performance at the high frequency end of the domain $[f_0, F]$ can be obtained by using linear interpolation in (92) rather than piecewise constant approximation; the same form (93) results.[6]

[6] This efficient method of spectral analysis, based on exponential sampling, was developed by K. G. Weil and H. Wiese, who brought it to the author's attention prior to its publication.

7

CROSS-SPECTRAL ANALYSIS

In this chapter, the concept of the spectral density of a single real-valued time-series is generalized to the concept of the cross-spectral density of two complex-valued time-series. Complex-valued time-series are considered in order to accommodate complex low-pass representations of real band-pass time-series (see Appendix 3-1). It is established that the cross spectrum, which is a measure of spectral correlation, plays a fundamental role in characterizing the degree to which two time-series are related by a linear time-invariant transformation. Methods for measurement of statistical cross spectra that are straightforward generalizations of the methods described in Chapter 4 for measurement of statistical spectra are described. The chapter concludes with a discussion of the resolution, leakage, and reliability properties of cross-spectrum measurements. Three appendices describe applications of cross-spectral analysis to propagation path identification, distant source detection, and time- and frequency-difference-of-arrival estimation.

A. ELEMENTS OF CROSS-SPECTRAL ANALYSIS

There are many phenomena that give rise to some measurable quantity that is the superposition of two or more constituent quantities. We consider as the most elementary case the superposition of two possibly complex-valued waveforms $x(t)$ and $y(t)$, which forms a composite waveform

$$z(t) = x(t) + y(t), \tag{1}$$

and we inquire into how the time-variant spectrum of $z(t)$ depends on the components $x(t)$ and $y(t)$. It is easily shown (exercise 1) that the time-variant periodogram

of $z(t)$ is given by

$$S_{z_T}(t,f) = S_{x_T}(t,f) + S_{y_T}(t,f) + S_{xy_T}(t,f) + S_{yx_T}(t,f), \tag{2}$$

for which the function $S_{xy_T}(t,f)$ is defined by

$$S_{xy_T}(t,f) \triangleq \frac{1}{T} X_T(t,f) Y_T^*(t,f), \tag{3a}$$

where for example

$$X_T(t,f) \triangleq \int_{t-T/2}^{t+T/2} x(u) e^{-i2\pi fu}\, du \tag{3b}$$

is the time-variant finite-time complex spectrum. It can also be shown that $S_{xy_T}(t,f)$ is characterized by the Fourier transform relation

$$S_{xy_T}(t,\cdot) = F\{R_{xy_T}(t,\cdot)\}, \tag{4}$$

for which the function $R_{xy_T}(t,\tau)$ is defined by

$$R_{xy_T}(t,\tau) \triangleq \frac{1}{T} \int_{t-(T-|\tau|)/2}^{t+(T-|\tau|)/2} x\left(u + \frac{\tau}{2}\right) y^*\left(u - \frac{\tau}{2}\right) du\, [2Tu_{2T}(\tau)]. \tag{5}$$

By analogy with the terminology introduced in Chapter 2, $S_{xy_T}(t,f)$ is called the time-variant *cross periodogram,* or *time-variant finite-time cross spectrum,* of $x(t)$ and $y(t)$, $R_{xy_T}(t,\tau)$ is called the *time-variant cross correlogram* of $x(t)$ and $y(t)$, and relation (4) is referred to as the time-variant *cross-periodogram-correlogram relation*. Parallel to (2), it is easily shown (exercise 1) that the correlogram of $z(t)$ is given by

$$R_{z_T}(t,\tau) = R_{x_T}(t,\tau) + R_{y_T}(t,\tau) + R_{xy_T}(t,\tau) + R_{yx_T}(t,\tau). \tag{6}$$

It follows from the linearity of the averaging operation that (2) gives rise to an identical form of relation for statistical spectra. For example, the spectrally smoothed periodogram of $z(t)$ is given by

$$S_{z_{\Delta t}}(t,f)_{\Delta f} = S_{x_{\Delta t}}(t,f)_{\Delta f} + S_{y_{\Delta t}}(t,f)_{\Delta f} + S_{xy_{\Delta t}}(t,f)_{\Delta f} + S_{yx_{\Delta t}}(t,f)_{\Delta f}, \tag{7}$$

in which the function $S_{xy_{\Delta t}}(t,f)_{\Delta f}$ is defined by

$$S_{xy_{\Delta t}}(t,f)_{\Delta f} \triangleq \frac{1}{\Delta f} \int_{-\Delta f/2}^{\Delta f/2} S_{xy_{\Delta t}}(t, f-\nu)\, d\nu \tag{8}$$

and is called a *statistical time-variant cross spectrum*. The form of relation (7) is valid as well for the statistical spectra $S_{z_{1/\Delta f}}(t,f)_{\Delta t}$, $S_z(t,f)_{\Delta t,\Delta f}$, and $S_z(t,f)_{1/\Delta f,\Delta t}$ defined in Chapter 4, and this gives rise to the following definitions of statistical cross spectra:

$$S_{xy_{1/\Delta f}}(t,f)_{\Delta t} \triangleq \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} S_{xy_{1/\Delta f}}(t-u,f)\, du \tag{9}$$

$$S_{xy}(t,f)_{\Delta t,\Delta f} \triangleq \frac{1}{\Delta f} \int_{-\Delta f/2}^{\Delta f/2} S_{xy}(t,f-\nu)_{\Delta t}\, d\nu \tag{10}$$

$$S_{xy}(t,f)_{1/\Delta f,\Delta t} \triangleq \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} S_{xy}(t-u,f)_{1/\Delta f}\, du, \tag{11}$$

in which the function $S_{xy}(t, f)_T$ is defined by

$$S_{xy}(t, \cdot)_T \triangleq F\{R_{xy}(t, \cdot)_T\} \tag{12}$$

and is called (by analogy with terminology in Chapter 2) the *time-variant pseudo–cross spectrum*. The function $R_{xy}(t, \tau)_T$ is defined by

$$R_{xy}(t, \tau)_T \triangleq \frac{1}{T}\int_{t-T/2}^{t+T/2} x\left(u + \frac{\tau}{2}\right)y^*\left(u - \frac{\tau}{2}\right) du\, [2Tu_{2T}(\tau)] \tag{13}$$

and is called the *time-variant finite-average cross correlation* of $x(t)$ and $y(t)$.

The relation (7) among statistical spectra gives rise to an identical form of relation for limit spectra,

$$\hat{S}_z(f) = \hat{S}_x(f) + \hat{S}_y(f) + \hat{S}_{xy}(f) + \hat{S}_{yx}(f), \tag{14}$$

for which the function $\hat{S}_{xy}(f)$ is defined by

$$\hat{S}_{xy}(f) \triangleq \lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{xy_{1/\Delta f}}(t, f)_{\Delta t} \tag{15}$$

and is called the *limit cross spectrum* of $x(t)$ and $y(t)$. Analogous to results in Chapter 3, Section C, the limit cross spectrum exists only if the *limit cross correlation,* defined by

$$\hat{R}_{xy}(\tau) \triangleq \lim_{T \to \infty} \frac{1}{T}\int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right)y^*\left(t - \frac{\tau}{2}\right) dt, \tag{16}$$

exists. The relation (6) among correlations gives rise to an identical form of relation for limit correlations,

$$\hat{R}_z(\tau) = \hat{R}_x(\tau) + \hat{R}_y(\tau) + \hat{R}_{xy}(\tau) + \hat{R}_{yx}(\tau). \tag{17}$$

Analogous to results in Chapter 2, Section E, the limit cross correlation can be obtained from either of the alternative equivalent limits

$$\hat{R}_{xy}(\tau) = \lim_{T \to \infty} R_{xy_T}(t, \tau) = \lim_{T \to \infty} R_{xy}(t, \tau)_T \tag{18}$$

(but this needs a proof, see [Kampé de Fériet 1954]). Similarly, analogous to results in Chapter 3, Section C, and Chapter 4, Section A, the limit cross spectrum can be obtained from any of the alternative equivalent limits

$$\hat{S}_{xy}(f) = \lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{xy_{1/\Delta f}}(t, f)_{\Delta t} \tag{19a}$$

$$= \lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{xy_{\Delta t}}(t, f)_{\Delta f} \tag{19b}$$

$$\hat{S}_{xy}(f) = \lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{xy}(t, f)_{\Delta t, \Delta f} \tag{20a}$$

$$= \lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} S_{xy}(t, f)_{1/\Delta f, \Delta t}. \tag{20b}$$

Moreover, the limit cross spectrum and limit cross correlation are a Fourier transform pair,

$$\hat{S}_{xy}(\cdot) = F\{\hat{R}_{xy}(\cdot)\}. \tag{21}$$

Furthermore, analogous to results in Chapters 2 and 3, all preceding relations

(with the exception of (10)–(13), (20a), and (20b)) apply for tapered data as well, in which case the definitions of cross correlogram and cross periodogram are generalized to

$$R_{xy_T}(t,\tau) \triangleq \frac{1}{T}\int_{-\infty}^{\infty} a_T\left(v+\frac{\tau}{2}\right)a_T^*\left(v-\frac{\tau}{2}\right)x\left(t-v+\frac{\tau}{2}\right)y^*\left(t-v-\frac{\tau}{2}\right)dv \tag{22}$$

and

$$S_{xy_T}(t,f) \triangleq \frac{1}{T}X_T(t,f)Y_T^*(t,f), \tag{23}$$

where for example

$$X_T(t,f) \triangleq \int_{-\infty}^{\infty} a_T(t-u)x(u)e^{-i2\pi fu}\,du. \tag{24}$$

However, (19a) and (19b) must be modified in this case to incorporate the factor γ, defined by (25) in Chapter 3.

Interchange of x with y in any of the cross spectra simply conjugates the cross spectra; for example,

$$S_{xy_T}(t,f) = S_{yx_T}^*(t,f) \tag{25}$$

(see (3a)). Interchange of x with y in any of the cross correlations simply conjugates the cross correlations and reverses the sign of τ (see (5)); for example,

$$R_{xy_T}(t,\tau) = R_{yx_T}^*(t,-\tau). \tag{26}$$

Analogous to results in Chapters 3 and 5, the limit cross spectrum cannot be obtained directly from the cross periodogram because the limit

$$\lim_{T\to\infty} S_{xy_T}(t,f) \tag{27}$$

does not exist in general. Thus, the smoothing operations in the formulas (19) and (20) cannot be circumvented.

An important interpretation of the cross spectrum can be obtained from definition (15). Substitution of (3a) into (9) into (15) reveals that *the limit cross spectrum* $\hat{S}_{xy}(f)$, *evaluated at frequency* f, *is an idealized measure of the temporal correlation between the spectral components of* $x(t)$ *and* $y(t)$ *at frequency* f. Thus, it can be called a *spectral correlation function*. Like the limit spectrum, which is a spectral *density* of temporal mean square (time-average power), the spectral correlation function is a spectral density of temporal correlation. By normalizing this spectral correlation by the temporal mean-square values $\hat{S}_x(f)$ and $\hat{S}_y(f)$ of the two spectral components, it can be made into a spectral correlation coefficient (recall from Chapter 5 that the mean of each spectral component for $f \neq 0$ is zero, and therefore the mean square and correlation are actually the variance and covariance, respectively). This complex-valued *spectral correlation coefficient,*

$$\frac{\hat{S}_{xy}(f)}{[\hat{S}_x(f)\hat{S}_y(f)]^{1/2}},$$

is studied in the next section.

B. COHERENCE

A major reason for interest in statistical cross spectra is a preoccupation with pairs of time-series that are related by LTI transformations. We have already seen in Section A that the notion of cross spectrum arises naturally when we consider a linear combination of time-series, such as (1). In fact, the theory of cross-spectral analysis originated in Sir Arthur Schuster's study of optical interference phenomena [Schuster 1904, par. 185], which are characterized by linear combinations of optical signals from one or more light sources. If two interfering optical signals x and y are derived from the same light source z and the primary difference between these two signals is that they have traveled different paths to the same point in space, then significant interference effects would occur, and the signals would be said to be *coherent*. Motivated by the physics of this situation, two time-series $x(t)$ and $y(t)$ are said to be *completely coherent* if they are both LTI transformations of a single time-series $z(t)$, say

$$x(t) = g(t) \otimes z(t) \tag{28a}$$

$$y(t) = h(t) \otimes z(t), \tag{28b}$$

for some functions $g(t)$ and $h(t)$ that do not annihilate any spectral components (i.e., $H(f) \neq 0$, $G(f) \neq 0$). More generally, the *degree of coherence* between $x(t)$ and $y(t)$ is measured in terms of the degree of accuracy with which $x(t)$ and $y(t)$ can be related to each other by an LTI transformation. We shall see that the idealized statistical cross spectrum—the limit cross spectrum $\hat{S}_{xy}$—plays a fundamental role in the characterization of an idealized measure of this degree of accuracy.

Since the relationship between two random time-series is generally masked by random effects, a reliable measure of the relationship must involve an averaging operation that removes random effects. A particularly useful idealized measure of the degree of accuracy with which $x(t)$ and $y(t)$ can be related by an LTI transformation is the limit time-average (temporal mean) of the squared error

$$e_{xy} \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(t) - \hat{x}(t)|^2 \, dt \tag{29a}$$

between $x(t)$ and an LTI transformation of $y(t)$,

$$\hat{x}(t) = g(t) \otimes y(t), \tag{30a}$$

minimized over all possible LTI transformations $g(t)$. It can be shown (exercise 3) that the minimum value of e_{xy} is given by

$$\min_g e_{xy} \triangleq e_{xy}^0 = \int_{-\infty}^{\infty} \hat{S}_x(f)[1 - |\hat{C}_{xy}(f)|^2] \, df, \tag{31a}$$

in which the function $\hat{C}_{xy}(f)$ is the complex-valued spectral correlation coefficient

$$\hat{C}_{xy}(f) \triangleq \frac{\hat{S}_{xy}(f)}{[\hat{S}_x(f)\hat{S}_y(f)]^{1/2}}. \tag{32a}$$

Similarly, the minimum value of the mean-squared error

$$e_{yx} \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} |y(t) - \hat{y}(t)|^2 \, dt \tag{29b}$$

between $y(t)$ and an LTI transformation of $x(t)$,

$$\hat{y}(t) = h(t) \otimes x(t), \tag{30b}$$

is given by

$$\min_h e_{yx} \triangleq e_{yx}^0 = \int_{-\infty}^{\infty} \hat{S}_y(f)[1 - |\hat{C}_{yx}(f)|^2] \, df, \tag{31b}$$

in which

$$\hat{C}_{yx}(f) \triangleq \frac{\hat{S}_{yx}(f)}{[\hat{S}_y(f)\hat{S}_x(f)]^{1/2}} . \tag{32b}$$

Although

$$|\hat{C}_{xy}(f)| \equiv |\hat{C}_{yx}(f)|, \tag{33}$$

$\hat{S}_x(f)$ and $\hat{S}_y(f)$ are unequal, in general, and therefore e_{xy}^0 and e_{yx}^0 are generally unequal. Consequently, neither e_{xy}^0 nor e_{yx}^0 alone is an appropriate measure of the degree to which $x(t)$ and $y(t)$ are related by an LTI transformation (without reference to the sense of the relation). However, since both e_{xy}^0 and e_{yx}^0 are characterized by the same function (33), then this function $|\hat{C}_{xy}(f)|$ is indeed an appropriate as well as convenient measure of the degree of coherence between $x(t)$ and $y(t)$, as explained in the following. First of all, $|\hat{C}_{xy}(f)|$ incorporates convenient normalization since it can be shown that

$$|\hat{C}_{xy}(f)| \leq 1 \tag{34}$$

for all $x(t)$ and $y(t)$ (because it is a correlation coefficient). Moreover,

$$|\hat{C}_{xy}(f)| \equiv 1 \tag{35}$$

if and only if there exists an $h(t)$ and $g(t)$ such that (28) is valid (at least in the mean-square sense), which indicates that $x(t)$ and $y(t)$ are *completely coherent*. Furthermore,

$$|\hat{C}_{xy}(f)| \equiv 0 \tag{36}$$

if and only if the minimum-mean-squared-error LTI transformations are identically zero,

$$g_0(t) \equiv h_0(t) \equiv 0, \tag{37}$$

in which case $x(t)$ and $y(t)$ are said to be *completely incoherent*. Because of the fundamental role that the spectral correlation coefficient $|\hat{C}_{xy}(f)|$ plays in characterizing the degree of coherence of $x(t)$ and $y(t)$, it is called the *spectral coherence function* for $x(t)$ and $y(t)$ and is often abbreviated to just *coherence function*. $\hat{C}_{xy}(f)$ is called the *complex coherence function*.[1]

[1] The complex coherence function is sometimes called the *coherency function*.

It can be shown (exercise 3) that the minimum-mean-squared-error LTI transformations denoted by $h_0(t)$ and $g_0(t)$ are specified by the transfer functions

$$G_0(f) = \frac{\hat{S}_{xy}(f)}{\hat{S}_y(f)} \tag{38a}$$

$$H_0(f) = \frac{\hat{S}_{yx}(f)}{\hat{S}_x(f)} \tag{38b}$$

and are called *Wiener filters* in honor of Norbert Wiener's pioneering work [Wiener 1949]. It follows from (32) and (38) that

$$|\hat{C}_{xy}(f)|^2 = G_0(f)H_0(f). \tag{39}$$

The most transparent interpretation of $|\hat{C}_{xy}(f)|$ as a measure of coherence can be obtained by considering the situation for which $x(t)$ and $y(t)$ are processed by an ideal band-pass filter that rejects all frequency content outside a small band of width Δ centered at f to produce the local sine wave components $x_{1/\Delta}(t, f)$ and $y_{1/\Delta}(t, f)$ with limit spectra

$$\hat{S}_{x_{1/\Delta}}(\nu) = \begin{cases} \hat{S}_x(\nu), & |\nu - f| \leq \Delta/2 \\ 0, & |\nu - f| > \Delta/2 \end{cases}$$

and (40)

$$\hat{S}_{y_{1/\Delta}}(\nu) = \begin{cases} \hat{S}_y(\nu), & |\nu - f| \leq \Delta/2 \\ 0, & |\nu - f| > \Delta/2. \end{cases}$$

Then the coherence between $x_{1/\Delta}(t, f)$ and $y_{1/\Delta}(t, f)$ can be shown (using (28) in Chapter 3 and (44)) to be

$$|\hat{C}_{xy_{1/\Delta}}(\nu)| = \begin{cases} |\hat{C}_{xy}(\nu)|, & |\nu - f| \leq \Delta/2 \\ 0, & |\nu - f| > \Delta/2. \end{cases} \tag{41}$$

Consequently, with $\Delta \to 0$ we see that for each value of f the degree of coherence of $x_{1/\Delta}(t, f)$ and $y_{1/\Delta}(t, f)$ is given by $|\hat{C}_{xy}(f)|$. Thus, *the spectral coherence function* $|\hat{C}_{xy}(f)|$ *is a frequency-decomposed measure of the degree to which* $x(t)$ *and* $y(t)$ *are related by an* LTI *transformation.*

System identification

A major application of the coherence function is to the problem of identification of dynamical systems. Consider a dynamical system (LTI transformation) with excitation and response denoted by $x(t)$ and $y(t)$,

$$y(t) = h(t) \otimes x(t). \tag{42}$$

It can be shown (exercise 2) that the limit cross correlation of $x(t)$ and $y(t)$ is

$$\hat{R}_{yx}(\tau) = \hat{R}_x(\tau) \otimes h(\tau). \tag{43}$$

Application of relation (21) to (43) yields the limit cross spectrum

$$\hat{S}_{yx}(f) = \hat{S}_x(f)H(f). \tag{44}$$

It follows from (44) that the transfer function of a system can be determined from the limit spectra,

$$H(f) = \frac{\hat{S}_{yx}(f)}{\hat{S}_x(f)}. \tag{45}$$

We see that this formula is identical to (38b), which specifies the minimum-mean-squared-error LTI transformation for *approximation* of an arbitrary $y(t)$ in terms of an arbitrary $x(t)$. Thus, given a system that might not be linear and/or might not be time-invariant, (45) yields an optimum[2] LTI approximating system.[3] The degree of nonlinearity and/or time-variation is measured by the mean-squared error of approximation (31b), which is specified by the coherence function and the limit spectrum $\hat{S}_y(f)$. Moreover, the integrand in (31b) is the spectral density of the approximation error (exercise 3).

Application of this system identification method to the problem of propagation-path identification is described in Appendix 7-1. Application of the coherence concept to the problem of detecting distant sources of energy propagating through frequency selective media is described in Appendix 7-2.

Partial coherence

When a time-series $y(t)$ is related to more than one other time-series, say $x_1(t), x_2(t), \ldots, x_N(t) \triangleq \{x_i(t)\}_1^N$, there is a problem with the direct use of the coherence function for determining the degree to which $y(t)$ is directly related to one of the N time-series $\{x_i(t)\}_1^N$, say $x_1(t)$, rather than indirectly related through its relationships with the others, $\{x_i(t)\}_2^N$. In particular, the effects of the other $N - 1$ time-series, if not properly removed, can yield either an erroneously high coherence or an erroneously low coherence. To avoid this problem, the effects on $x_1(t)$ and $y(t)$ of the other $N - 1$ time-series can be removed before the coherence is measured. This yields what is called the *partial coherence* between $x_1(t)$ and $y(t)$. In particular, the relationships (coherences) between $x_1(t)$ and $\{x_i(t)\}_2^N$ are removed by subtracting from $x_1(t)$ the best fitting (minimum-mean-squared-error) LTI transformation of $\{x_i(t)\}_2^N$ to obtain

$$x_1^0(t) \triangleq x_1(t) - \sum_{i=2}^{N} h_i^0(t) \otimes x_i(t), \tag{46a}$$

and similarly the relationships between $y(t)$ and $\{x_i(t)\}_2^N$ are removed to obtain

$$y^0(t) \triangleq y(t) - \sum_{i=2}^{N} g_i^0(t) \otimes x_i(t). \tag{46b}$$

[2] This system identification method is optimum only for system excitations that are purely stationary in the sense that they do not exhibit second-order periodicity (no sine waves in the lag products). Otherwise there is a better method that is tolerant to measurement noise on the excitation, as explained in Chapter 14, Part II.

[3] Although the actual excitation and response might be related by an LTI transformation, the measured excitation and response might not be due to additive measurement noise; see exercise 6.

Then the complex partial coherence between $x_1(t)$ and $y(t)$ is given by[4]

$$\hat{C}^0_{x_1y}(f) = \frac{\hat{S}_{x_1^0y^0}(f)}{[\hat{S}_{x_1^0}(f)\hat{S}_{y^0}(f)]^{1/2}}. \tag{47}$$

The vector of transfer functions in (46a) is given by (exercise 7)

$$\boldsymbol{H}_0(f) = [\hat{\boldsymbol{S}}_{\boldsymbol{x}}^{-1}(f)\hat{\boldsymbol{S}}_{\boldsymbol{x}x_1}(f)]^*, \tag{48a}$$

where $\hat{\boldsymbol{S}}_{\boldsymbol{x}}$ is the $(N-1) \times (N-1)$ spectral density matrix for the $N-1$ time-series $\{x_i(t)\}_2^N$, and $\hat{\boldsymbol{S}}_{\boldsymbol{x}x_1}$ is the vector of $N-1$ cross-spectral densities. For example, the ijth element of the matrix $\hat{\boldsymbol{S}}_{\boldsymbol{x}}(f)$ is the Fourier transform of the ijth element of the matrix $\hat{\boldsymbol{R}}_{\boldsymbol{x}}(\tau)$ defined by

$$\hat{\boldsymbol{R}}_{\boldsymbol{x}}(\tau) \triangleq \left\langle \boldsymbol{x}\left(t + \frac{\tau}{2}\right)\boldsymbol{x}'\left(t - \frac{\tau}{2}\right)^*\right\rangle; \tag{49}$$

this ijth element is

$$[\hat{\boldsymbol{R}}_{\boldsymbol{x}}(\tau)]_{ij} = \left\langle x_i\left(t + \frac{\tau}{2}\right)x_j\left(t - \frac{\tau}{2}\right)^*\right\rangle \triangleq \hat{R}_{x_ix_j}(\tau). \tag{50}$$

Similarly, the vector of transfer functions in (46b) is given by

$$\boldsymbol{G}_0(f) = [\hat{\boldsymbol{S}}_{\boldsymbol{x}}^{-1}(f)\hat{\boldsymbol{S}}_{\boldsymbol{x}y}(f)]^*. \tag{48b}$$

Furthermore, the spectral densities in (47) are given by (exercise 7)

$$\hat{S}_{y^0}(f) = \hat{S}_y(f) - \boldsymbol{G}_0'(f)\hat{\boldsymbol{S}}_{\boldsymbol{x}}(f)\boldsymbol{G}_0^*(f) \tag{51a}$$

$$\hat{S}_{x_1^0}(f) = \hat{S}_{x_1}(f) - \boldsymbol{H}_0'(f)\hat{\boldsymbol{S}}_{\boldsymbol{x}}(f)\boldsymbol{H}_0^*(f) \tag{51b}$$

$$\hat{S}_{x_1^0y^0}(f) = \hat{S}_{x_1y}(f) - \boldsymbol{H}_0'(f)\hat{\boldsymbol{S}}_{\boldsymbol{y}\boldsymbol{x}}^*(f) - \hat{\boldsymbol{S}}_{x_1\boldsymbol{x}}'(f)\boldsymbol{G}_0^*(f) + \boldsymbol{H}_0'(f)\hat{\boldsymbol{S}}_{\boldsymbol{x}}(f)\boldsymbol{G}_0^*(f). \tag{52}$$

Substitution of (48), (51), and (52) into (47) yields an explicit formula for the complex partial coherence between $x_1(t)$ and $y(t)$ in terms of only cross-spectral densities and spectral densities of $y(t)$ and $\{x_i(t)\}_1^N$.

As a specific example, consider the problem of determining the partial coherence of $y(t)$ and $x_1(t) = x(t)$, with the effect of $x_2(t) = w(t)$ removed. In this case, the vector quantities in (48) to (52) all reduce to scalars, and (47) to (52) yield the partial coherence

$$\hat{C}^0_{xy}(f) = \frac{\left[\hat{S}_{xy}(f) - \dfrac{\hat{S}_{xw}(f)\hat{S}_{yw}^*(f)}{\hat{S}_w(f)}\right]}{\left[\left(\hat{S}_x(f) - \dfrac{|\hat{S}_{xw}(f)|^2}{\hat{S}_w(f)}\right)\left(\hat{S}_y(f) - \dfrac{|\hat{S}_{yw}(f)|^2}{\hat{S}_w(f)}\right)\right]^{1/2}}. \tag{53}$$

The theory and methods of partial coherence are fundamental in the study of multiple-input-multiple-output systems in science and especially in engineering.

[4] The partial coherence is the spectral counterpart of the temporal partial correlation (PARCOR) coefficient, which arises naturally in the lattice implementation of the linear-prediction-error filter described in Chapter 9.

For example, a turbo-alternator used in a power system can be modeled as an approximately linear two-input, two-output system, where the inputs are the deviations from nominal values of the in-phase (active) power and quadrature (reactive) power, which measure the load on the turbo-alternator, and the outputs are the deviations from nominal values of the amplitude and frequency of the voltage generated at the stator terminals. The system-identification technique described in this section can be used together with the partial coherence method to determine experimentally a model of the turbo-alternator. Actual experimental data for this and other system identification problems are given in [Jenkins and Watts 1968]. It should be clarified that the system identification and modeling methods that employ spectrum estimates directly to estimate transfer functions are often used in practice primarily as a first stage, which is followed by a second stage of parametric model fitting using methods such as the autoregressive modeling methods described in Chapter 9. In this two-stage approach, the first stage is used mostly to determine an appropriate type of model, such as the MA, AR, or ARMA models described in Chapter 9.

In addition to the study of multiple-input-multiple-output systems, partial-coherence methodology is fundamental to many, if not most, experimental sciences where physical phenomena are investigated on the basis of relationships among measurements of a multiplicity of fluctuating quantities, including a particular quantity at various spatial locations, such as electromagnetic field strength, acoustical pressure, sea temperature, air humidity, atmospheric temperature, wind velocity, atmospheric pressure, and radiation intensity. Such techniques are used, for example, in astronomy, meteorology, oceanography, geology, and biology.

C. AUTOCOHERENCE AND PERIODICITY

In a natural development of concepts, this is an appropriate point at which to provide a brief introduction to the subject of Part II, which builds on the concepts of cross correlation, cross spectrum, and coherence for the purpose of studying random data from periodic phenomena. The limit autocorrelation of a time-series $x(t)$ is defined by

$$\hat{R}_x(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_{+\tau/2}(t) x^*_{-\tau/2}(t)\, dt, \tag{54}$$

in which $x_\sigma(t)$ is the time-translate

$$x_\sigma(t) \triangleq x(t + \sigma) \tag{55}$$

for $\sigma = \pm\tau/2$. Thus, $\hat{R}_x(\tau)$ is the correlation of time-translates of $x(t)$. An analogous quantity is the correlation of frequency-translates

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^{+\alpha/2}(t)[x^{-\alpha/2}(t)]^*\, dt, \tag{56}$$

in which $x^\nu(t)$ is the frequency translate

$$x^\nu(t) \triangleq x(t) e^{-i2\pi\nu t} \tag{57}$$

for $\nu = \pm\alpha/2$. Let us combine (54) and (56) to obtain a correlation of time- and frequency-translates defined by

$$\hat{R}_x^{\alpha}(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_{+\tau/2}^{+\alpha/2}(t)[x_{-\tau/2}^{-\alpha/2}(t)]^* \, dt, \tag{58a}$$

in which

$$x_{\sigma}^{\nu}(t) \triangleq x(t + \sigma)e^{-i2\pi\nu t}, \tag{58b}$$

for $\nu = \pm\alpha/2$ and $\sigma = \pm\tau/2$. It is easily shown (exercise 9) that this time-frequency limit autocorrelation is given by

$$\hat{R}_x^{\alpha}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi\alpha t} \, dt. \tag{59}$$

This generalized limit autocorrelation plays a fundamental role in the characterization of periodicity in random data. Specifically, we say that a real time-series $x(t)$ exhibits *second-order periodicity* (or is *cyclostationary*) with *cycle frequency* $\alpha \neq 0$ if and only if there exists a stable[5] *quadratic time-invariant* (QTI) transformation of $x(t)$, say

$$w(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(t - u, t - v)x(u)x(v) \, du \, dv, \tag{60}$$

where

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |k(u, v)| \, du \, dv < \infty, \tag{61}$$

that produces an additive sine wave component (spectral line) in $w(t)$ with frequency α. It can be shown (exercise 10) that $x(t)$ exhibits such second-order periodicity if and only if the time-frequency limit autocorrelation function $\hat{R}_x^{\alpha}$ is not identically zero.

If the frequency-translates in (57) are denoted by

$$\begin{aligned} y(t) &= x^{+\alpha/2}(t) \\ z(t) &= x^{-\alpha/2}(t), \end{aligned} \tag{62}$$

then (58) is recognized as a limit cross-correlation function,

$$\hat{R}_x^{\alpha}(\tau) \equiv \hat{R}_{yz}(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} y\left(t + \frac{\tau}{2}\right) z^*\left(t - \frac{\tau}{2}\right) dt. \tag{63}$$

The corresponding limit cross spectrum is denoted by $\hat{S}_x^{\alpha}(f)$,

$$\hat{S}_{yz}(\cdot) \triangleq F\{\hat{R}_{yz}(\cdot)\} \equiv F\{\hat{R}_x^{\alpha}(\cdot)\} \triangleq \hat{S}_x^{\alpha}(\cdot). \tag{64}$$

The limit spectra of $y(t)$ and $z(t)$ are easily shown to be

$$\begin{aligned} \hat{S}_y(f) &= \hat{S}_x\left(f + \frac{\alpha}{2}\right) \\ \hat{S}_z(f) &= \hat{S}_x\left(f - \frac{\alpha}{2}\right). \end{aligned} \tag{65}$$

[5] The QTI transformation (60) is said to be *stable* if the kernel $k(\cdot, \cdot)$ is absolutely integrable, (61).

It follows from (64)–(65) and definition (32) that the complex coherence function for $y(t)$ and $z(t)$ is given by

$$\hat{C}_{yz}(f) \equiv \frac{\hat{S}_x^{\alpha}(f)}{[\hat{S}_x(f + \alpha/2)\hat{S}_x(f - \alpha/2)]^{1/2}} \triangleq \hat{C}_x^{\alpha}(f). \tag{66}$$

This complex coherence between the two frequency-translates $x^{+\alpha/2}(t)$ and $x^{-\alpha/2}(t)$ of the single time-series $x(t)$ is called the *complex spectral autocoherence* (or *self-coherence*) of $x(t)$ at frequency α. It follows directly from the interpretation of $\hat{C}_{yz}(f)$ as a spectral correlation coefficient that $\hat{C}_x^{\alpha}(f)$ is the correlation coefficient for the spectral components of $x^{+\alpha/2}(t)$ and $x^{-\alpha/2}(t)$ at frequency f, which are precisely the same as the spectral components of $x(t)$ at the two frequencies $f + \alpha/2$ and $f - \alpha/2$. Thus, $f = \frac{1}{2}[(f + \alpha/2) + (f - \alpha/2)]$ is the *location* of the pair of frequencies $f \pm \alpha/2$, and $\alpha = (f + \alpha/2) - (f - \alpha/2)$ is the *separation*. To distinguish the complex autocoherence from the complex coherence $\hat{C}_{uv}(f)$ between two *arbitrary* waveforms $u(t)$ and $v(t)$, we can refine the terminology and call $\hat{C}_{uv}(f)$ the *complex spectral cross coherence* (or *mutual coherence*).

It follows directly from the preceding definitions that the autocoherence function is identically zero for all $\alpha \neq 0$ if and only if $x(t)$ contains no second-order periodicity. Also, since

$$|\hat{C}_x^{\alpha}(f)| \leq 1 \tag{67}$$

(see (34)), the magnitude of the complex autocoherence (often called simply the *autocoherence*) is a conveniently normalized measure of the *degree of second-order periodicity* in $x(t)$, or of the *degree of self-coherence* of $x(t)$. A real waveform $x(t)$ is said to be *completely (self-) coherent* at frequency α over a spectral band, say B, if and only if

$$|\hat{C}_x^{\alpha}(f)| \equiv 1, \qquad f \in B, \tag{68}$$

and it is said to be *completely (self-) incoherent* at frequency α over the band B if and only if

$$|\hat{C}_x^{\alpha}(f)| \equiv 0, \qquad f \in B. \tag{69}$$

Stated another way, (68) (or (69)) means that the pair of spectral bands $\{f \pm \alpha/2 : f \in B\}$ is completely coherent (or incoherent). Furthermore, it can be shown that the autocoherence function is appropriately invariant to LTI transformations of the time-series. That is, if

$$y(t) = x(t) \otimes h(t), \tag{70a}$$

and if

$$H(\nu) \neq 0, \qquad \nu = f \pm \alpha/2 \tag{70b}$$

so that the spectral components at $f \pm \alpha/2$ are not annihilated, then it follows from (28) in Chapter 3, (44), and (64)–(66) that

$$C_y^{\alpha}(f) = C_x^{\alpha}(f). \tag{71}$$

In conclusion, the complex autocoherence function plays a fundamental role in the theory of second-order periodicity that is developed in Part II.

Example

As an illustration of second-order periodicity, we consider the example of a real bandlimited amplitude- and phase-modulated sine wave,

$$\begin{aligned} x(t) &= a(t)\cos[\omega_0 t + \phi(t)] \\ &= c(t)\cos(\omega_0 t) - s(t)\sin(\omega_0 t), \end{aligned} \tag{72}$$

(see Part 5 of Section D, Chapter 3) for which $c(t)$ and $s(t)$ are bandlimited such that

$$\hat{S}_c(f) = \hat{S}_s(f) = 0, \qquad |f| > \frac{\omega_0}{2\pi}. \tag{73}$$

It is shown in Part II that if $c(t)$ and $s(t)$ contain no second-order periodicity, then the squared autocoherence for the waveform $x(t)$ is given by

$$|\hat{C}_x^{\alpha}(f)|^2 = \frac{[\hat{S}_c(f) - \hat{S}_s(f)]^2 + 4[\mathrm{Re}\{\hat{S}_{cs}(f)\}]^2}{[\hat{S}_c(f) + \hat{S}_s(f)]^2 - 4[\mathrm{Im}\{\hat{S}_{cs}(f)\}]^2} \tag{74}$$

for $\alpha = \pm\omega_0/\pi$ and is identically zero for all other $\alpha \neq 0$. It follows from (74) that $x(t)$ contains no second-order periodicity (is completely incoherent) if and only if the spectra of in-phase and quadrature components are identical,

$$\hat{S}_c(f) \equiv \hat{S}_s(f), \tag{75a}$$

and the real part of their cross-spectral density is identically zero,

$$\mathrm{Re}\{\hat{S}_{cs}(f)\} \equiv 0. \tag{75b}$$

At the other extreme, it can be shown that the degree of second-order periodicity at frequency $\alpha = \omega_0/\pi$ is maximum ($x(t)$ is completely coherent within the spectral band $|f| \leq \omega_0/\pi$) if and only if the in-phase and quadrature components are completely mutually coherent,

$$|\hat{C}_{cs}(f)| \equiv 1, \tag{76}$$

and either (75a) or (75b) is violated. As an example, if $s(t)$ is the Hilbert transform (see Appendix 3-1) of $c(t)$, then $x(t)$ is a *single-sideband,* amplitude-modulated sine wave,[6]

$$\hat{S}_x(f) = 0, \qquad |f| < \frac{\omega_0}{2\pi}, \tag{77}$$

and (76) is valid, but neither (75a) nor (75b) is violated. Therefore, $x(t)$ is completely incoherent. However, it can be shown that (76) holds and (75b) is violated if $x(t)$ has no phase modulation,

$$x(t) = a(t)\cos(\omega_0 t + \phi). \tag{78}$$

Thus, a bandlimited (see (73)) *double-sideband* amplitude-modulated sine wave is completely coherent within its spectral band.

D. MEASUREMENT METHODS

As explained in Chapter 4, statistical spectral measurements can be obtained from any of a variety of methods, and these various methods are either exactly

[6] The band below the center frequency $f_0 = \omega_0/2\pi$ is canceled out when $s(t)$ is a 90° phase-lagged version of $c(t)$.

or approximately equivalent to each other. In principle, the approximations can be made as accurate as desired by choosing Δt sufficiently large provided only that the limit autocorrelation exists. In particular, there are more than 10 alternative methods, which are described by diagrams in Figures 4-1, 4-2, 4-4, 4-6, and 4-7. Because of the fact that all the elements of cross-spectral analysis are straightforward generalizations of the elements of spectral analysis, as explained in Section A, all these alternative methods for obtaining statistical spectra generalize in a straightforward way for statistical cross spectra. Some of these generalizations are briefly described in this section. As explained in Chapter 4, Section A, although the approximations relating the spectra obtained from these various methods can in principle be made as accurate as desired by choosing Δt sufficiently large, it should be emphasized that in applications where Δt must be relatively small, the differences among statistical spectra obtained from different smoothing methods or different windows can be substantial, and the particular choice then becomes an important component of the design problem, as illustrated in Chapter 5, Section D. Although only analog methods are described here, the corresponding digital methods can easily be deduced from these and the digital methods based on the DFT described in Chapter 6, Section C.

1. Temporal and Spectral Smoothing

By analogy with the arguments in Chapter 3, Section B, and Chapter 4, Section A, it can be shown that all four of the temporally smoothed and spectrally smoothed statistical cross spectra (8)–(11) are approximately equal for $\Delta t \Delta f >> 1$.

2. Fourier Transformation of Tapered Cross Correlation

By analogy with the argument in Chapter 4, Section B, it is easily shown (exercise 10) that the two spectrally smoothed statistical cross spectra, of which (8) and (10) are examples, can be obtained by Fourier transformation of tapered cross correlations,

$$S_{xy_{\Delta t}}(t, \cdot)_{\Delta f} = F\{R_{xy_{\Delta t}}(t, \cdot)h_{1/\Delta f}(\cdot)\} \tag{79}$$

$$S_{xy}(t, \cdot)_{\Delta t, \Delta f} = F\{R_{xy}(t, \cdot)_{\Delta t}\, h_{1/\Delta f}(\cdot)\}, \tag{80}$$

for which $h_{1/\Delta f}$ is the inverse Fourier transform of the spectral smoothing window $H_{\Delta f}$ (see (2b) and (4b) in Chapter 4). Five alternative methods for obtaining the temporally smoothed cross periodogram or the spectrally smoothed cross periodogram are shown in Figure 7-1.

3. Cross-Wave Analysis

By analogy with the derivation of the expression (38) in Chapter 4 for the temporally smoothed spectrum, the temporally smoothed cross spectrum (9) can be reexpressed as

$$S_{xy_{1/\Delta f}}(t, f)_{\Delta t} = \Delta f\{[x(t) \otimes a^f_{1/\Delta f}(t)][y(t) \otimes a^f_{1/\Delta f}(t)]^*\} \otimes g_{\Delta t}(t), \tag{81}$$

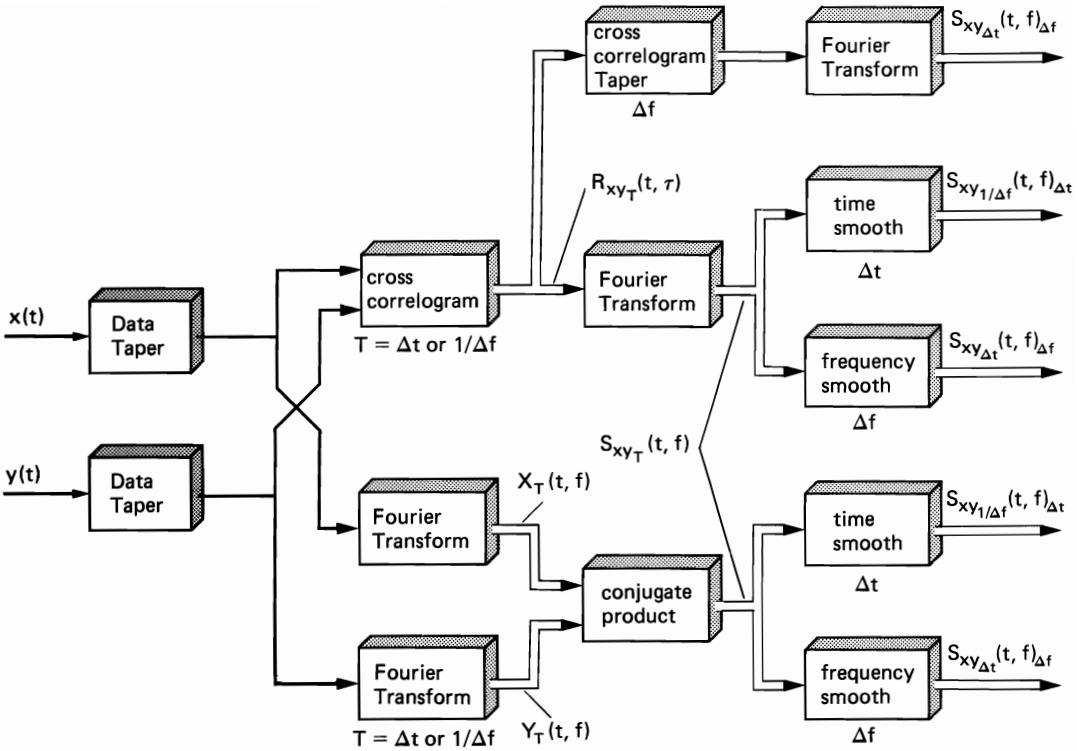


Figure 7-1 Five methods for obtaining statistical cross spectra based on the cross periodogram and cross correlogram.

in which

$$a^f_{1/\Delta f}(t) \triangleq a_{1/\Delta f}(t)e^{i2\pi ft}, \tag{82}$$

where $a_{1/\Delta f}(t)$ is the data-tapering aperture. The smoothing functions $g_{\Delta t}$ and $a^f_{1/\Delta f}$ can be interpreted as the impulse-response functions of filters with transfer functions

$$G_{1/\Delta t}(\cdot) = F\{g_{\Delta t}(\cdot)\}$$

and

$$A^f_{\Delta f}(\nu) = A_{\Delta f}(\nu - f),$$

where

$$A_{\Delta f}(\cdot) = F\{a_{1/\Delta f}(\cdot)\}. \tag{83}$$

$G_{1/\Delta t}$ represents a low-pass filter with bandwidth $1/\Delta t$, and $A^f_{\Delta f}$ represents a band-pass filter with bandwidth Δf centered at f. Consequently, the temporally smoothed statistical spectrum (9) can be obtained by band-pass filtering $x(t)$ and $y(t)$, multiplying the results, and low-pass filtering the product, as depicted in Figure 7-2(a). This cross-spectral analysis method is called *cross-wave analysis*. The condition $\Delta t \Delta f >> 1$ for substantial smoothing requires that the bandwidths of the input BPFs greatly exceed the bandwidth of the output LPF. It follows from

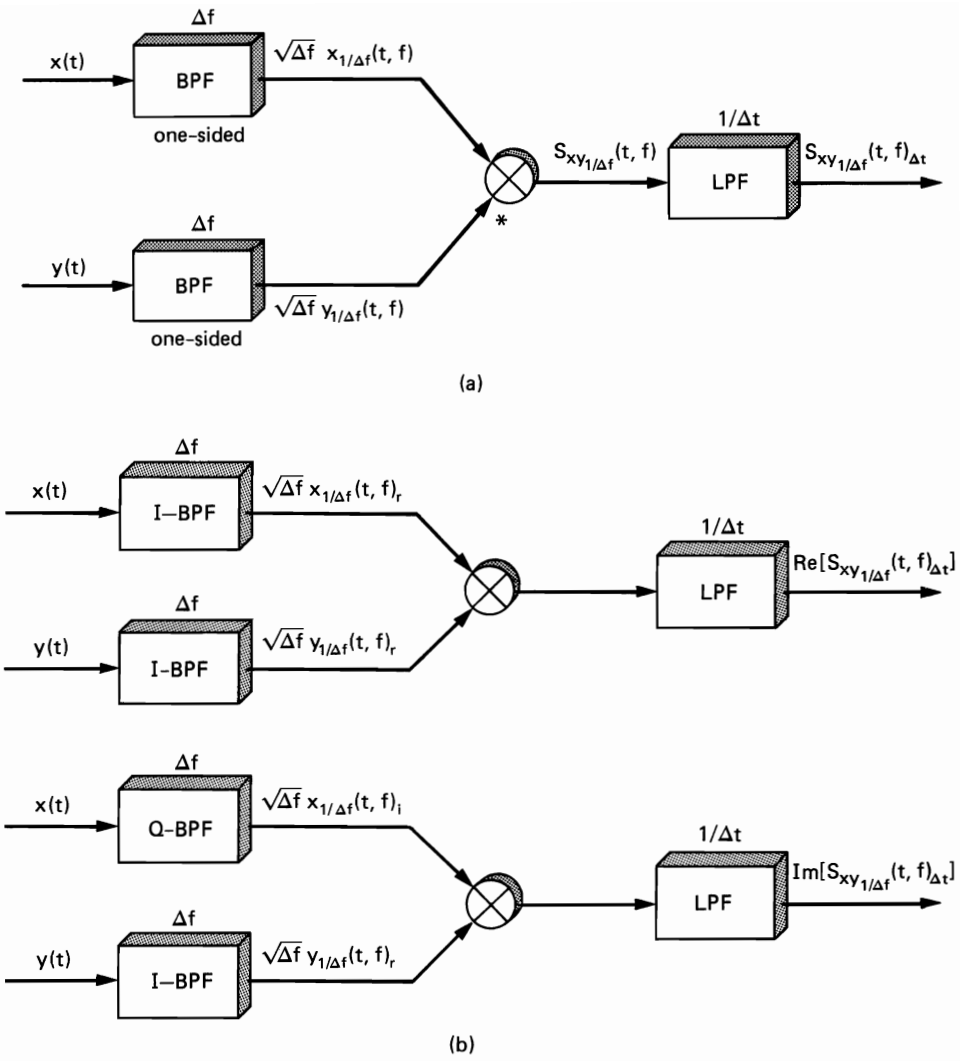


Figure 7-2 (a) Complex implementation of wave-analysis method for time-variant statistical cross-spectral analysis. (b) Approximate real implementation of wave-analysis method for time-variant statistical cross-spectral analysis of real time-series. (Impulse responses of I-BPF and Q-BPF are $a_{1/\Delta f}(t)\cos 2\pi ft$ and $a_{1/\Delta f}(t)\sin 2\pi ft$. Accuracy of approximation is high for $\Delta t \Delta f \gg 1$ and $|f|/\Delta f \gg 1$.)

a temporal-spectral smoothing equivalence, analogous to (11)–(17) in Chapter 3 (or from calculation of the mean, analogous to (56) in Chapter 5) that the squared magnitude of the transfer function of the BPF is the effective spectral smoothing window.

Formula (81) can be reexpressed as

$$S_{xy_{1/\Delta f}}(t, f)_{\Delta t} = \Delta f [x_{1/\Delta f}(t, f)\, y^*_{1/\Delta f}(t, f)] \otimes g_{\Delta t}(t), \tag{84}$$

in which the waveform

$$x_{1/\Delta f}(t,f) \triangleq x(t) \otimes a^{f}_{1/\Delta f}(t) \tag{85}$$

is the *local sine wave component* of $x(t)$.

By analogy with the derivation of approximation (56) in Chapter 4, it can be shown that the cross-wave analyzer for real time-series that employs complex (one-sided) band-pass filters can be approximated by the real implementation depicted in Figure 7-2(b). This is a close approximation for $|f|/\Delta f >> 1$ and $\Delta t \Delta f >> 1$, and the accuracy of approximation increases as these inequalities are strengthened. The statistical cross spectrum obtained from this real cross-wave analyzer can be expressed by

$$\begin{aligned} S_{xy_{1/\Delta f}}(t,f)_{\Delta t} \cong\ & 2\Delta f[x_{1/\Delta f}(t,f)_r y_{1/\Delta f}(t,f)_r] \otimes g_{\Delta t}(t) \\ & + i2\Delta f[x_{1/\Delta f}(t,f)_i y_{1/\Delta f}(t,f)_r] \otimes g_{\Delta t}(t), \qquad |f|/\Delta f >> 1,\ \Delta f \Delta t >> 1, \end{aligned} \tag{86}$$

in which the waveforms

$$x_{1/\Delta f}(t,f)_r = x(t) \otimes [a_{1/\Delta f}(t)\cos(2\pi f t)]$$

and

$$x_{1/\Delta f}(t,f)_i = x(t) \otimes [a_{1/\Delta f}(t)\sin(2\pi f t)] \tag{87}$$

are the real and imaginary parts of the complex local sine wave component (85). Unfortunately, the BPFs required in this real cross-wave analyzer exhibit in-phase and quadrature symmetries that render them problematic for analog implementation. This can be circumvented by using the demodulation method described next.

4. Cross Demodulation

By use of the relationship

$$X_T(t,f) = x_T(t,f)e^{-i2\pi f t} \tag{88}$$

between the *demodulate* $X_T(t,f)$ and the local sine wave component $x_T(t,f)$ of $x(t)$, the formula (84) for the temporally smoothed cross spectrum can be reexpressed as

$$S_{xy_{1/\Delta f}}(t,f)_{\Delta t} = \Delta f[X_{1/\Delta f}(t,f)Y^{*}_{1/\Delta f}(t,f)] \otimes g_{\Delta t}(t), \tag{89}$$

for which

$$X_{1/\Delta f}(t,f) \triangleq [x(t)e^{-i2\pi f t}] \otimes a_{1/\Delta f}(t). \tag{90}$$

We see from (89) and (90) that this statistical cross spectrum can be obtained by low-pass filtering the product of the complex demodulate for $x(t)$ and the conjugate complex demodulate for $y(t)$, and these demodulates can be obtained by low-pass filtering the product of the waveform with a complex sine wave, as depicted in Figure 7-3(a).

By representing the complex sine wave in (90) in terms of the real sine and cosine functions,

$$e^{i2\pi f t} = \cos(2\pi f t) + i\sin(2\pi f t),$$

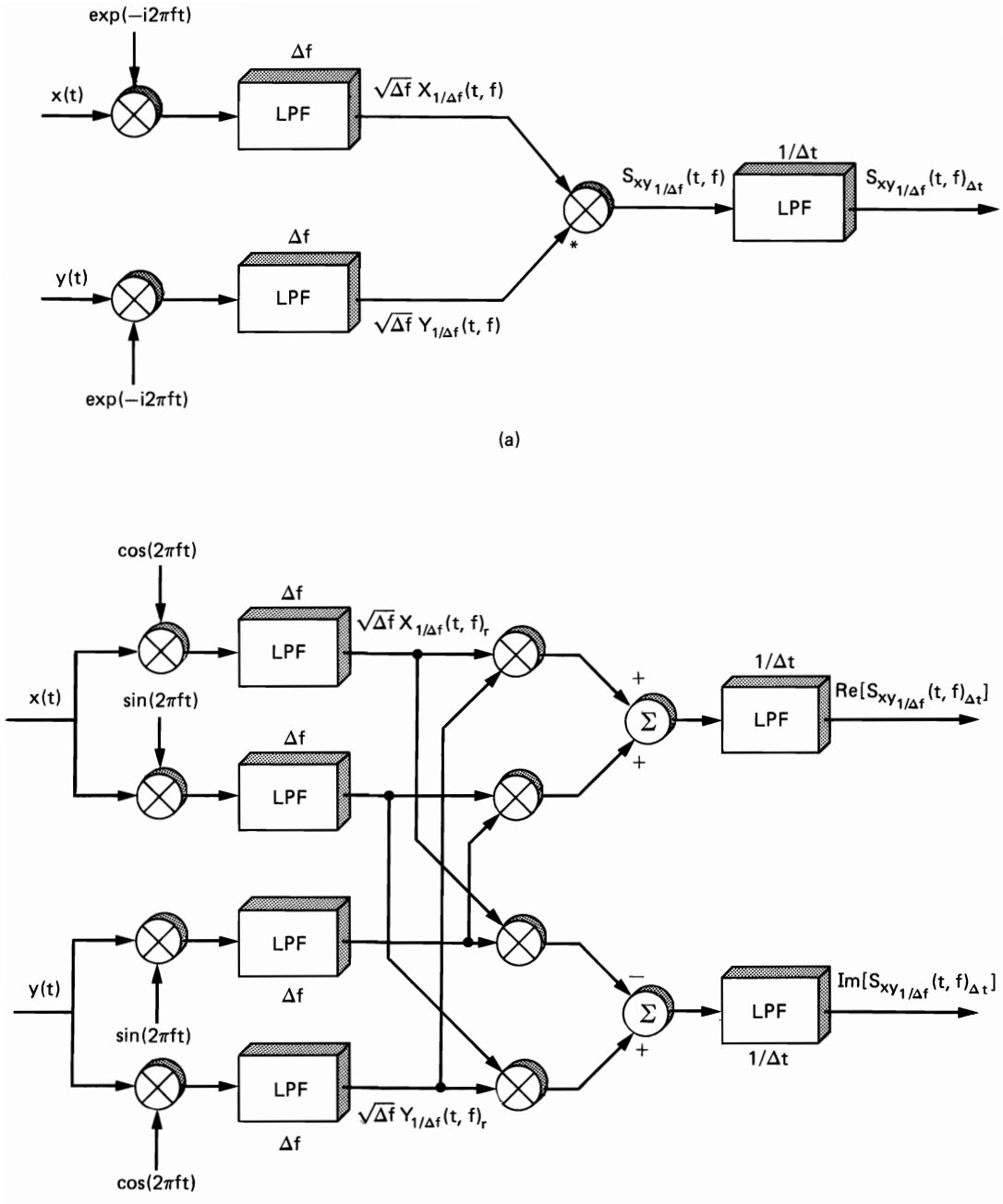


Figure 7-3 (a) Complex implementation of demodulation method for time-variant statistical cross-spectral analysis. (b) Real implementation of demodulation method for time-variant statistical cross-spectral analysis of real time-series.

the statistical cross spectrum (89) for a real time-series can be reexpressed as

$$S_{xy_{1/\Delta f}}(t,f)_{\Delta t} = \Delta f[X_{1/\Delta f}(t,f)_r Y_{1/\Delta f}(t,f)_r + X_{1/\Delta f}(t,f)_i Y_{1/\Delta f}(t,f)_i] \otimes g_{\Delta t}(t) + i\Delta f[X_{1/\Delta f}(t,f)_i Y_{1/\Delta f}(t,f)_r - X_{1/\Delta f}(t,f)_r Y_{1/\Delta f}(t,f)_i] \otimes g_{\Delta t}(t), \tag{91}$$

for which $X_{1/\Delta f}(t, f)_r$ and $X_{1/\Delta f}(t, f)_i$ are the real and imaginary parts of the demodulate,

$$X_{1/\Delta f}(t,f)_r = [x(t)\cos(2\pi ft)] \otimes a_{1/\Delta f}(t)$$
$$X_{1/\Delta f}(t,f)_i = [x(t)\sin(2\pi ft)] \otimes a_{1/\Delta f}(t). \tag{92}$$

It follows from (91) and (92) that the demodulation method has the real implementation shown in Figure 7-3(b).

It can be shown (exercise 12) that for a real time-series, the two terms in the real part of $S_{xy_{1/\Delta f}}(t,f)_{\Delta t}$ in (91) are approximately equal and the two terms in the imaginary part are approximately equal, except for opposite signs, provided that $\Delta t \Delta f >> 1$ and that $x(t)$ and $y(t)$ contain no additive, multiplicative, or other second-order periodicities (see Section C). Thus, we have

$$S_{xy_{1/\Delta f}}(t,f)_{\Delta t} \cong 2\Delta f[X_{1/\Delta f}(t,f)_r Y_{1/\Delta f}(t,f)_r] \otimes g_{\Delta t}(t) + i2\Delta f[X_{1/\Delta f}(t,f)_i Y_{1/\Delta f}(t,f)_r] \otimes g_{\Delta t}(t), \qquad \Delta t\Delta f >> 1, \tag{93}$$

and the accuracy of the approximation increases as $\Delta t \Delta f$ is increased. This reveals that half of the filters and modulators in Figure 7-3(b) can be deleted provided that $x(t)$ and $y(t)$ contain no second-order periodicity.

E. RESOLUTION, LEAKAGE, AND RELIABILITY

The resolution, leakage, and reliability properties of cross-spectrum measurements are very similar to the corresponding properties of spectrum measurements, which are described in Chapter 5. However, as revealed in this section, the coefficient of variation of a cross-spectrum measurement can be much larger than the coefficient of variation of a spectrum measurement, and resolution requirements can be more demanding.

1. Cross Periodogram

We consider first the cross periodogram $S_{xy_T}(t, f)$. It can be shown (exercise 13) that the temporal mean is given by

$$\text{mean}\{S_{xy_T}(t,f)\} = \hat{S}_{xy}(f) \otimes z_{1/T}(f), \tag{94}$$

and for zero-mean jointly Gaussian time-series $x(t)$ and $y(t)$, the temporal variance is given by

$$\text{var}\{S_{xy_T}(t,f)\} = [\hat{S}_x(f) \otimes z_{1/T}(f)][\hat{S}_y(f) \otimes z_{1/T}(f)] + \left| \frac{1}{T}\int_{-\infty}^{\infty} \hat{S}_{xy}(\nu) w_{1/T}(f-\nu) w_{1/T}(f+\nu)\, d\nu \right|^2 \tag{95}$$

and the temporal covariance is given by

$$
\begin{aligned}
K_S(f_1, f_2) &\triangleq \operatorname{cov}\{S_{xy_T}(t, f_1), S_{xy_T}(t, f_2)\} \\
&\triangleq \langle [S_{xy_T}(t, f_1) - \langle S_{xy_T}(t', f_1)\rangle][S_{xy_T}(t, f_2) - \langle S_{xy_T}(t', f_2)\rangle]^*\rangle \\
&= \frac{1}{T^2}\int_{-\infty}^{\infty} \hat{S}_x(\nu) w_{1/T}(f_1 - \nu) w_{1/T}(f_2 - \nu)\, d\nu \\
&\quad \times \int_{-\infty}^{\infty} \hat{S}_y(\nu) w_{1/T}(f_1 - \nu) w_{1/T}(f_2 - \nu)\, d\nu \\
&\quad + \left|\frac{1}{T}\int_{-\infty}^{\infty} \hat{S}_{xy}(\nu) w_{1/T}(f_1 - \nu) w_{1/T}(f_2 + \nu)\, d\nu\right|^2. \qquad (96)
\end{aligned}
$$

If $1/T$ is sufficiently small for $w_{1/T}$ and $z_{1/T}$ to resolve the fine structure in $\hat{S}_x(f)$, $\hat{S}_y(f)$, and $\hat{S}_{xy}(f)$, then it follows from (94) and (95) that the temporal coefficient of variation of the cross periodogram is closely approximated by

$$
r_S \cong \frac{1}{|\hat{C}_{xy}(f)|^2} + \delta_{1/T}(f), \qquad (97)
$$

for which $\delta_{1/T}(f)$ is positive and less than or approximately equal to unity for all f, and satisfies

$$
\delta_{1/T}(f) \begin{cases} \cong 1, & |f| \ll 1/T \\ \ll 1, & |f| \gg 1/T. \end{cases} \qquad (98)
$$

Thus, when the coherence between $x(t)$ and $y(t)$ is small, the coefficient of variation can be much larger than unity, because of the relatively small value of the mean. It follows from (95) and (96) that if $1/T$ is sufficiently small for $w_{1/T}$ to resolve the fine structure in $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_{xy}$, then the temporal correlation coefficient is closely approximated by

$$
\rho_S \cong \delta_{1/T}(f_1 - f_2) + |\hat{C}_{xy}(f_1)|^2 \delta_{1/T}(f_1 + f_2) \qquad (99)
$$

for $|f_1| \gg 1/T$ and $|f_2| \gg 1/T$ and is therefore much smaller than unity for $|f_1 - f_2| \gg 1/T$ and $|f_1 + f_2| \gg 1/T$. Thus, like the periodogram, the cross periodogram is a highly erratic unreliable measurement of cross-spectral density.

2. Statistical Cross Spectra

In order to obtain a unified treatment for the large variety of statistical cross spectra that result from the various methods described in Section D, we shall employ the general representation from Section G of Chapter 4 and Section C of Chapter 5 for statistical spectra but generalized (by replacement of $x(t - u)x(t - v)$ with $x(t - u)y^*(t - v)$) for statistical cross spectra:

$$
w_f(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k_f(u, v) x(t - u) y^*(t - v)\, du\, dv, \qquad (100a)
$$

where

$$
k_f(u, v) = m\left(\frac{u + v}{2}, v - u\right) e^{-i2\pi f(v-u)}. \qquad (100b)
$$

The kernel representor $m(t, \tau)$ and its double Fourier transform $M(\nu, \mu)$ are specified in Table 5-1 for the various methods described in Section D. The temporal mean of the statistical cross spectrum $w_f(t)$ is easily shown (exercise 14) to be given by

$$\text{mean}\{w_f(t)\} = \hat{S}_{xy}(f) \otimes E(f), \tag{101a}$$

where $E(f)$ is an effective spectral smoothing window given by

$$E(f) = M(0, -f), \tag{101b}$$

in which $M(\nu, \mu)$ is the double Fourier transform

$$M(\nu, \mu) \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(t, \tau) e^{-i2\pi(\nu t - \mu\tau)}\, dt\, d\tau. \tag{102}$$

As discussed at length in Chapter 5, Section D, for statistical (auto) spectra, the resolution and leakage properties of a statistical cross spectrum are (assuming low variability) characterized by the effective spectral window $E(f)$ in terms of the width of its main lobe and the heights of its sidelobes. However, the resolution problem can be more crucial for cross spectral analysis because of the effect of time misalignment. For example, if $y(t) = x(t - a)$, then $\hat{S}_{xy}(f) = \hat{S}_x(f)e^{i2\pi af}$. Therefore, a main-lobe width narrower than $\Delta f = 1/a$ is required to avoid substantial attenuation due to averaging of the sinusoidal factor. Consequently, a large time misalignment a will result in a highly attenuated cross-spectrum estimate unless the effective spectral smoothing window has a very narrow main-lobe width.

If $x(t)$ and $y(t)$ are jointly Gaussian zero-mean time-series, then the limit spectral density for the centered time-series

$$z_f(t) \triangleq w_f(t) - \text{mean}\{w_f(t)\} \tag{103}$$

is given by

$$S_{z_f}(\nu) = \int_{-\infty}^{\infty} |M(\nu, \mu - f)|^2\, \hat{S}_x\left(\mu + \frac{\nu}{2}\right)\hat{S}_y\left(\mu - \frac{\nu}{2}\right) d\mu + \int_{-\infty}^{\infty} M(\nu, \mu - f)M^*(\nu, -\mu - f)S_{xy*}\left(\mu + \frac{\nu}{2}\right)S^*_{xy*}\left(-\mu + \frac{\nu}{2}\right) d\mu. \tag{104}$$

This result can be derived by the method outlined in exercise 7, Chapter 5. (A derivation is given in [Brown 1987].) For the sake of generality, the representation (100) and the formulas (101) and (104) are given for complex-valued time-series $x(t)$ and $y(t)$ and analyzers $k_f(t, \tau)$. Thus, the results here apply, for example, to complex envelopes of band-pass time-series.

If the separable approximation

$$M(\nu, \mu) \cong G_{1/\Delta t}(\nu)H_{\Delta f}(-\mu), \tag{105}$$

which is often valid for $\Delta t \Delta f >> 1$ (see Chapter 4, Section G), is used, then the effective spectral smoothing window (101b) is simply

$$E(f) \cong G_{1/\Delta t}(0)H_{\Delta f}(f), \qquad \Delta t \Delta f >> 1, \tag{106}$$

and the limit spectrum of the centered statistical cross spectrum (104) reduces

to

$$S_{z_f}(\nu) \cong |G_{1/\Delta t}(\nu)|^2 \Bigg[\int_{-\infty}^{\infty} |H_{\Delta f}(f-\mu)|^2 \, \hat{S}_x\left(\mu + \frac{\nu}{2}\right) \hat{S}_y\left(\mu - \frac{\nu}{2}\right) d\mu + \int_{-\infty}^{\infty} H_{\Delta f}(f-\mu) H^*_{\Delta f}(f+\mu) \hat{S}_{xy*}\left(\mu + \frac{\nu}{2}\right) \hat{S}^*_{xy*}\left(-\mu + \frac{\nu}{2}\right) d\mu \Bigg],$$

$$\Delta t \Delta f \gg 1. \tag{107}$$

The temporal variance of the statistical cross spectrum $w_f(t)$ is given by the integral of the limit spectrum (104), and if the width $1/\Delta t$ of $M(\nu, \mu)$ in ν is small enough and the corresponding sidelobes are low enough to resolve the fine structure in $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_{xy*}$, then this integral yields the close approximation

$$\text{var}\{w_f(t)\} \cong \int_{-\infty}^{\infty} [L(f, \mu) \hat{S}_x(\mu) \hat{S}_y(\mu) + N(f, \mu) \hat{S}_{xy*}(\mu) \hat{S}^*_{xy*}(-\mu)] \, d\mu, \qquad \frac{1}{\Delta t} < \Delta f^*, \tag{108a}$$

where Δf^* is the minimum of the resolution widths of $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_{xy*}$, and

$$L(f, \mu) \triangleq \int_{-\infty}^{\infty} |M(\nu, \mu - f)|^2 \, d\nu \tag{108b}$$

$$N(f, \mu) \triangleq \int_{-\infty}^{\infty} M(\nu, \mu - f) M^*(\nu, -\mu - f) \, d\nu. \tag{108c}$$

Furthermore, if the width Δf of $M(\nu, \mu)$ in μ is also small enough and the corresponding sidelobes are low enough to resolve the fine structure in $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_{xy*}$, then (108) reduces to the close approximation

$$\text{var}\{w_f(t)\} \cong A \hat{S}_x(f) \hat{S}_y(f) + B(f) \, \text{Re}\{\hat{S}_{xy*}(f) \hat{S}^*_{xy*}(-f)\}, \qquad \frac{1}{\Delta t}, \Delta f < \Delta f^*, \tag{109a}$$

where

$$A \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |M(\nu, \mu)|^2 \, d\nu \, d\mu \tag{109b}$$

$$B(f) \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M(\nu, \mu - f) M^*(\nu, -\mu - f) \, d\nu \, d\mu. \tag{109c}$$

It follows from the fact that $M(\nu, \mu)$ has width on the order of Δf in μ that $B(f)$ is negligible compared with A for $|f| \gg \Delta f$. Thus, for frequencies well beyond one resolution width from zero frequency, the second term in (109a) is negligible. Use of this approximation together with

$$\text{mean}\{w_f(t)\} \cong C \hat{S}_{xy}(f), \qquad \Delta f < \Delta f^*, \tag{110a}$$

where

$$C \triangleq \int_{-\infty}^{\infty} M(0, \mu) \, d\mu, \tag{110b}$$

yields the close approximation

$$r_{w_f} \triangleq \frac{\text{var}\{w_f(t)\}}{|\text{mean}\{w_f(t)\}|^2} \cong \frac{\eta}{\Delta t \Delta f} \frac{1}{|\hat{C}_{xy}(f)|^2}, \qquad \frac{1}{\Delta t}, \Delta f < \Delta f^*, |f| \gg \Delta f \tag{111a}$$

for the temporal coefficient of variation. In (111a), the reliability factor η is given by

$$\eta \triangleq \frac{\Delta t \Delta f \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |M(\nu, \mu)|^2 \, d\nu \, d\mu}{\left| \int_{-\infty}^{\infty} M(0, \mu) \, d\mu \right|^2}. \tag{111b}$$

This factor η is typically on the order of unity (see Chapter 5, Section D). If the separable approximation (105) is used, then η is approximated by

$$\eta \cong \left[\frac{\Delta t \, \|G_{1/\Delta t}\|^2}{|G_{1/\Delta t}(0)|^2} \right] \left[\frac{\Delta f \|h_{1/\Delta f}\|^2}{|h_{1/\Delta f}(0)|^2} \right]. \tag{111c}$$

In conclusion, a statistical cross spectrum is a reliable measurement only if

$$\Delta t \Delta f \gg \frac{1}{|\hat{C}_{xy}(f)|^2}. \tag{112}$$

For $x = y$, $\hat{C}_{xy}(f) \equiv 1$ and (112) reduces to Grenander's uncertainty condition for statistical (auto) spectra discussed in Chapter 5. But for statistical cross spectra for which $x \neq y$, it follows from the inequality $|\hat{C}_{xy}(f)| \leq 1$ that a larger resolution product can be needed. This is a result of the fact that even if the quantity $\hat{S}_{xy}(f)$ that is being estimated is small compared with $[\hat{S}_x(f)\hat{S}_y(f)]^{1/2}$, the variance of the estimate is still determined by the power levels of x and y and is therefore proportional to $\hat{S}_x(f)\hat{S}_y(f)$. Other than this potentially troublesome factor $1/[\hat{C}_{xy}(f)|$ in the coefficient of variation, the design trade-offs among the properties of resolution, leakage, and reliability for statistical cross spectra are the same as they are for statistical (auto) spectra, and the reader is therefore referred to Chapter 5, Section D, for an in-depth discussion of these trade-offs and the associated window-design problem. Nevertheless, when combining auto- and cross-spectrum estimates to obtain an estimate of the coherence function $\hat{C}_{xy}(f)$, additional problems can arise. Specifically, when estimating the coherence function using estimates of the individual spectra and cross spectrum such as

$$C_{xy_{\Delta t}}(t, f)_{\Delta f} \triangleq \frac{S_{xy_{\Delta t}}(t, f)_{\Delta f}}{[S_{x_{\Delta t}}(t, f)_{\Delta f} S_{y_{\Delta t}}(t, f)_{\Delta f}]^{1/2}}, \tag{113}$$

one must exercise considerably more care to minimize the effects of limited resolution, leakage, and variability than would be surmised from the discussions in Chapter 5, Section D. The reasons for this are twofold. In addition to the fact that variability of cross-spectrum estimates can be considerably more problematic, the nonlinear combination of three spectrum estimates in (113) can amplify the effects of leakage and limited resolution. An example that illustrates

this is given in exercise 17. The general problem is treated thoroughly in [Jenkins and Watts 1968], where it is explained that individual analysis of magnitude and phase and real and imaginary parts of cross-spectrum estimates and coherence function estimates provides a more complete picture of the measurement problem.

F. SUMMARY

In Section A, the elements of cross-spectral analysis are introduced, and it is explained that these are all generalizations of the elements of spectral analysis. These elements include the *cross periodogram, cross correlogram, finite-average cross correlation, pseudo–cross spectrum, limit cross correlation* and *limit cross spectrum,* and the various temporally smoothed and spectrally smoothed *statistical cross spectra*. It is also explained that whereas the limit spectrum gives the mean-square strength of spectral components, the limit cross spectrum gives the correlation of spectral components. In Section B, the *spectral coherence function,* which is the spectral correlation coefficient obtained from the limit cross spectrum, is introduced and is shown to be a measure of the degree to which two time-series are related by a linear time-invariant transformation. The two extremes of completely coherent and completely incoherent time-series are described, and the role of coherence in the identification of dynamical systems is explained. Then the concept of coherence is generalized to that of *partial coherence* for a pair of time-series that are coherent with other time-series. In Section C, the central role of spectral correlation (coherence) in the spectral analysis of random data from periodic phenomena is explained, and the *spectral autocoherence function* for a single time-series is defined, and is illustrated for a class of amplitude- and phase-modulated sine waves.

In Section D, the various analog methods for statistical spectral analysis that are described in Chapter 4 are generalized for statistical cross-spectral analysis. This includes the methods of *temporal and spectral smoothing, Fourier transformation of tapered cross correlations, cross-wave analysis,* and *cross demodulation* (Figures 7-1, 7-2, 7-3). In Section E, the results on *resolution, leakage,* and *reliability* that are developed in Chapter 5 for spectral analysis are generalized for cross-spectral analysis, and it is explained that the variability and resolution of cross-spectrum estimates can be even more problematic than for (auto) spectrum estimates.

Appendix 7-1 describes the use of cross-spectral analysis for propagation path identification. Appendix 7-2 explains the use of cross-spectral analysis for distant-source detection. Appendix 7-3 reveals the connection between the cross-correlation function, the cross-spectral density function, and the *cross-ambiguity function* for the problems of time- and frequency-difference-of-arrival estimation.

EXERCISES

1. **(a)** Show that the periodogram of the sum of two time-series (1) is given by (2).
   (b) Show that the correlogram of the sum of two time-series (1) is given by (6).
2. Consider an LTI transformation with impulse-response function $h(t)$, (42). Show that

the limit cross correlation and limit cross spectrum for the excitation and response are given by (43)–(44).

3. **(a)** To prove that the Wiener filter (38a) minimizes the mean-squared error (29a), proceed as follows. First show that

$$e_{xy} = \hat{R}_x(0) - \hat{R}_{x\hat{x}}(0) - \hat{R}_{\hat{x}x}(0) + \hat{R}_{\hat{x}}(0). \tag{114}$$

Then use

$$\hat{R}_x(0) = \int_{-\infty}^{\infty} \hat{S}_x(f)\, df$$

together with (30a) and (44) to show that

$$e_{xy} = \int_{-\infty}^{\infty} [\hat{S}_x(f) - \hat{S}_{xy}(f)G^*(f) - \hat{S}_{yx}(f)G(f) + \hat{S}_y(f)|G(f)|^2]\, df. \tag{115}$$

Finally, let

$$G(f) = G_0(f) + G_*(f),$$

for which $G_0(f)$ is the Wiener filter (38a) and $G_*(f)$ is arbitrary, and show that

$$e_{xy} = e_{xy}^0 + e_{xy}^*,$$

where e_{xy}^0 is given by (31a) (which is independent of $G_*(f)$), that $e_{xy}^* \geqslant 0$, and that $e_{xy}^* = 0$ if and only if $G_*(f) \equiv 0$. Then conclude that e_{xy} is minimum if and only if $G(f) \equiv G_0(f)$. (A detailed treatment of Wiener filtering including geometrical interpretations is given in [Gardner 1985].)

(b) To verify that the integrand in (31a) is the spectral density of the error $e(t) \triangleq x(t) - \hat{x}(t)$, determine $\hat{R}_e(\tau)$ and then use (30a), (38a), and (44) (with H replaced by G_0).

4. Derive the formula

$$|\hat{C}_{12}(f)| = \{[1 + \rho_1(f)][1 + \rho_2(f)]\}^{-1/2}, \tag{116a}$$

in which

$$\rho_i(f) = \frac{\hat{S}_{n_i}(f)}{|H_i(f)|^2 \hat{S}_x(f)}, \qquad i = 1, 2, \tag{116b}$$

for the mutual coherence between the two time-series

$$\begin{aligned} y_1(t) &= h_1(t) \otimes x(t) + n_1(t) \\ y_2(t) &= h_2(t) \otimes x(t) + n_2(t), \end{aligned} \tag{117}$$

for which $n_1(t)$, $n_2(t)$, and $x(t)$ are mutually uncorrelated.

5. Consider two time-series $x(t)$ and $y(t)$ related by

$$y(t) = x(t) \otimes h(t) + n(t),$$

for which $x(t)$ and $n(t)$ are uncorrelated and have zero means ($\hat{R}_{xn} \equiv 0$), the Fourier transform of $h(t)$ is given by

$$H(f) = \begin{cases} 0, & |f| \leqslant B \\ 1, & |f| > B, \end{cases}$$

and the limit spectrum of $n(t)$ is given by

$$\hat{S}_n(f) = \begin{cases} N_0, & |f| \leqslant B \\ 0, & |f| > B. \end{cases}$$

Determine the mutual coherence of $x(t)$ and $y(t)$.

6. **(a)** Consider the problem of identifying a linear time-invariant system (estimating its

transfer function), using measurements $z(t) = x(t) + n(t)$ and $w(t) = y(t) + m(t)$ of the excitation $x(t)$ and the response $y(t)$ that are corrupted by measurement noise $n(t)$ and $m(t)$. Assume that $n(t)$ and $m(t)$ are uncorrelated with each other and with $x(t)$ and $y(t)$. Determine an explicit formula in terms of the ratio $\rho_1(f) = \hat{S}_n(f)/\hat{S}_x(f)$ for the error between the transfer function $H_0(f)$ and its estimate $H(f)$, normalized by $H_0(f)$, based on the estimation formula $H(f) = \hat{S}_{wz}(f)/\hat{S}_z(f)$. Then, if possible, simplify the general result for the two special cases of no input noise ($n(t) \equiv 0$) and no output noise ($m(t) \equiv 0$).

(b) Determine an explicit formula for the coherence magnitude $|\hat{C}_{zw}|$ in terms of the ratios $\rho_1(f) = \hat{S}_n(f)/\hat{S}_x(f)$ and $\rho_2(f) = \hat{S}_m(f)/|H_0(f)|^2\hat{S}_x(f)$ for the general case and two special cases described in (a).

7. (a) To verify the spectral density formulas (51) and (52), use (46) and determine $\hat{R}_{y\theta}(\tau)$, $\hat{R}_{x\theta}(\tau)$, and $\hat{R}_{x\theta y\theta}(\tau)$, and then evaluate the Fourier transforms using the convolution theorem.

(b) Generalize the procedure outlined in exercise 3, from scalar-filtering to vector-filtering, to derive the vector Wiener filter (48).

8. Consider three related processes, $x(t)$, $y(t)$, and $w(t)$, and show that the complex partial coherence function for $x(t)$ and $y(t)$ is given by (53).

9. Show that the generalized autocorrelation (58) yields the periodicity parameter (59).

10. The strength of an additive sine wave component in $w(t)$ at frequency α is given by

$$\hat{M}_w^\alpha \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} w(t)e^{-i2\pi\alpha t}\,dt. \tag{118a}$$

Show that, with $w(t)$ given by the QTI transformation (60), $\hat{M}_w^\alpha$ is given by

$$\hat{M}_w^\alpha = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(u, v)\hat{R}_x^\alpha(u - v)e^{-i\pi\alpha(u+v)}\,du\,dv. \tag{118b}$$

Therefore, it can be concluded that there exists a QTI transformation $k(u, v)$ such that $\hat{M}_w^\alpha$ is not zero (the QTI transformation generates a spectral line) if and only if $\hat{R}_x^\alpha(\tau)$ is not identically zero.

11. Verify the equivalence (79) between statistical cross spectra by using the following generalization of (8):

$$S_{xy_{\Delta t}}(t, f)_{\Delta f} \triangleq \int_{-\infty}^{\infty} S_{xy_{\Delta t}}(t, f - \nu)H_{\Delta f}(\nu)\,d\nu.$$

12. Generalize the approach outlined in exercise 4 in Chapter 4 to show that (93) is a close approximation to (91) for the demodulation cross-spectrum analyzer for real time-series. Do this by showing that the two terms in the real part of (91) are approximately equal and the two terms in the imaginary part are approximately equal, except for opposite signs, provided that $\Delta t\Delta f \gg 1$, and that

$$\lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right)y^*\left(t - \frac{\tau}{2}\right)e^{-i2\pi\alpha t}\,dt = 0 \qquad \text{for all } \alpha \neq 0,$$

which indicates the absence of second-order periodicity.

13. (a) Show that the mean of the cross periodogram is given by (94).

(b) Show that the covariance of the cross periodogram is given by (96) for zero-mean jointly Gaussian time-series $x(t)$ and $y(t)$. *Hint:* Follow a procedure analogous to that outlined in the hint for exercise 2(b) in Chapter 5 with the following Isserlis' formula:

$$\langle x(t+t_1)y^*(t+t_2)x^*(t+t_3)y(t+t_4)\rangle = \hat{R}_{xy}(t_1-t_2)\hat{R}_{yx}(t_4-t_3) + \hat{R}_x(t_1-t_3)\hat{R}_y(t_4-t_2) + \hat{R}_{xy^*}(t_1-t_4)\hat{R}_{y^*x}(t_2-t_3). \quad (119)$$

(c) Use (96) to obtain the formula (95) for the variance of the cross periodogram.

14. (a) Show that the mean of the statistical cross spectrum (100) is given by (101).

(b) Show that, for real-valued zero-mean jointly Gaussian time-series $x(t)$ and $y(t)$ and a real even spectral window $H_{\Delta f}$, the variance of the spectrally smoothed periodogram is given by

$$\text{var}\{S_{xy_{\Delta t}}(t,f)_{\Delta f}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_S(\nu_1,\nu_2)H_{\Delta f}(f-\nu_1)H_{\Delta f}(f-\nu_2)\,d\nu_1\,d\nu_2. \quad (120)$$

Then verify the approximation

$$\text{var}\{S_{xy_{\Delta t}}(t,f)_{\Delta f}\} \cong \frac{1}{\Delta t}\int_{-\infty}^{\infty}\hat{S}_x(\nu)\hat{S}_y(\nu)H_{\Delta f}^2(f-\nu)\,d\nu + \frac{1}{\Delta t}\int_{-\infty}^{\infty}[\hat{S}_{xy}(\nu)]^2H_{\Delta f}(f-\nu)H_{\Delta f}(f+\nu)\,d\nu, \qquad \Delta t\Delta f \gg 1. \quad (121)$$

Hint: Follow a procedure analogous to that outlined in the hint for exercise 13 in Chapter 5.

(c) Show that (121) corroborates the general formula (108), by use of approximation (105).

15. Derive formula (104) for the limit spectral density of any statistical cross spectrum of the form (100) for jointly Gaussian zero-mean time-series $x(t)$ and $y(t)$. *Hint:* See exercise 7 in Chapter 5.

16. (a) Determine $M(\nu,\mu)$ in (100) and (102) for each of the methods of cross-spectrum estimation corresponding to the auto-spectrum estimation methods enumerated in exercise 9 of Chapter 5.

(b) Use formula (101b) to determine the effective spectral smoothing window for each case in (a). Comment on the relationship between the effective spectral smoothing windows you obtain here and those obtained in exercise 9 in Chapter 5.

17. Use formula (101a) to show that the estimated coherence function, for example,

$$|C_{yx_{\Delta t}}(t,f)_{\Delta f}|^2 \triangleq \frac{|S_{yx_{\Delta t}}(t,f)_{\Delta f}|^2}{S_{y_{\Delta t}}(t,f)_{\Delta f}S_{x_{\Delta t}}(t,f)_{\Delta f}}, \quad (122)$$

for coherent time-series, $y(t) = h(t) \otimes x(t)$, can differ substantially from unity due to bias (leakage and limited resolution) only, regardless of how reliable these estimates are (i.e., regardless of how small their coefficients of variation are). Do this by substituting the means of the spectrum estimates in place of the estimates themselves in (122) and showing that even though

$$|\hat{C}_{yx}(f)|^2 = \frac{|\hat{S}_{yx}(f)|^2}{\hat{S}_y(f)\hat{S}_x(f)} \equiv 1,$$

$|C_{yx_{\Delta t}}(t,f)_{\Delta f}|$ can be much smaller than unity. Let $\hat{S}_x(f) \equiv 1$ for simplicity and assume that $E(f)$ has unity area to obtain

$$|C_{yx_{\Delta t}}(t,f)_{\Delta f}|^2 \cong \frac{|H(f) \otimes E(f)|^2}{|H(f)|^2 \otimes E(f)}.$$

Hint: To show the desired result by example, let $H(f)$ be flat except for a rectangle notch 50% deep and $2\Delta f$ wide, and let $E(f)$ be a rectangle of width Δf. Verify that this results in as much as 20% error, due entirely to limited resolution (no leakage).

18. Consider N propagation paths with impulse-response functions $\{h_i\}_1^N$ of the form

$$h_i(t) = g_i(t - T_i),$$

where

$$g_i(t) = 0, \qquad |t| > V_i$$

and

$$\min_{j \neq k} |T_j - T_k| > T_* + \max_i \{V_i\}.$$

This indicates that the differences between all pairs of propagation delays $\{T_i\}$ exceed the time-dispersive spreads $\{V_i\}$ plus a constant T_*. Let T_* exceed twice the correlation width τ_* of a common source of energy $x(t)$,

$$T_* > 2\tau_*,$$

where

$$\hat{R}_x(\tau) = 0, \qquad |\tau| > \tau_*.$$

Show that the cross correlation of the source and response

$$y(t) = \sum_{i=1}^{N} h_i(t) \otimes x(t) = \sum_{i=1}^{N} y_i(t) \tag{123}$$

decomposes into the sum of N nonoverlapping terms,

$$\hat{R}_{yx}(\tau) = \sum_{i=1}^{N} h_i(\tau) \otimes \hat{R}_x(\tau) = \sum_{i=1}^{N} \hat{R}_{y_i x}(\tau) \tag{124}$$

in which

$$R_{y_i x}(\tau) R_{y_j x}(\tau) \equiv 0, \qquad i \neq j. \tag{125}$$

19. Consider the model (1) for TDOA described in Appendix 7-3. Show that the limit cross-correlation function is given by (2) in Appendix 7-3.

20. The limit autocorrelation for a complex-valued time-series $x(t)$ is defined by

$$\hat{R}_x(\tau) \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) dt, \tag{126}$$

and the limit spectral density is given by the Wiener relation

$$\hat{S}_x(f) = \int_{-\infty}^{\infty} \hat{R}_x(\tau) e^{-i2\pi f\tau} \, d\tau.$$

Prove that $\hat{S}_x(f)$ is real-valued but is not necessarily an even function,

$$\hat{S}_x^*(f) = \hat{S}_x(f) \neq \hat{S}_x(-f).$$

21. **(a)** It is desired to estimate the cross spectrum for two real-valued time-series $x(t)$ and $y(t)$, which are modeled as

$$y(t) = x(t) \otimes h(t) + n(t),$$

where $\hat{S}_x(f) \equiv 1$ and $\hat{S}_n(f) \cong 10|H(f)|^2$, $n(t)$ and $x(t)$ are uncorrelated, and $H(f)$ has resolution width of $\Delta f^* = 100$ Hz. If it is desired to resolve $\hat{S}_{xy}(f)$ properly and to maintain a coefficient of variation of approximately $\frac{1}{10}$, then what is the minimum amount Δt of data needed?

(b) For (a) specify a particular cross-spectrum estimation method, and specify all required window shapes and widths.

22. Consider the problem of filtering the noisy signal

$$y(t) = x(t) + n(t)$$

in order to suppress the noise $n(t)$. Assume that the noise is white with intensity N_0 and the signal has limit autocorrelation function given by

$$\hat{R}_x(\tau) = S_0 e^{-f_0|\tau|}.$$

Determine an explicit formula in terms of only S_0, N_0, and f_0 for the transfer function of the Wiener filter that minimizes the mean-squared error between the filtered signal

$$\hat{x}(t) = y(t) \otimes g(t)$$

and the desired signal $x(t)$.

APPENDIX 7-1

Propagation-Path Identification

Consider a linear time-invariant propagation path for energy, such as acoustical waves traveling from one point to another, that is a parallel combination of separate paths such that the input-output relation is of the form

$$y(t) = \sum_{i=1}^{N} x(t) \otimes h_i(t) \triangleq \sum_{i=1}^{N} y_i(t), \tag{1}$$

where h_i is the impulse-response function for the ith path. In the event that the propagation delays for these N paths are quite distinct so that the differences between all delays exceed the time-dispersive spread for each path plus twice the correlation width of the source data x, then the transfer functions for each of the paths can be identified from the overall input-output measurements. Specifically, in this case it can be shown (exercise 18) that the cross correlation $\hat{R}_{yx}$ decomposes into the sum

$$\hat{R}_{yx}(\tau) = \sum_{i=1}^{N} h_i(\tau) \otimes \hat{R}_x(\tau) = \sum_{i=1}^{N} \hat{R}_{y_i x}(\tau) \tag{2}$$

of N nonoverlapping terms. Thus, once the individual terms in (2) have been separated, the transfer functions can be determined from the formula

$$H_i(f) = \frac{\hat{S}_{y_i x}(f)}{\hat{S}_x(f)}. \tag{3}$$

However, the accuracy of the LTI model for each path cannot be assessed by using the coherence functions

$$\hat{C}_{y_i x}(f) = \frac{\hat{S}_{y_i x}(f)}{\hat{S}_{y_i}(f)\hat{S}_x(f)}, \tag{4}$$

because the ith component of the response y_i is not accessible. All N components are superimposed, and therefore $\hat{S}_{y_i}$ cannot be measured. This and related approaches to propagation path identification are treated in [Bendat and Piersol 1980].

APPENDIX 7-2
Distant-Source Detection

We consider an important application of the coherence concept to the problem of detecting the presence of a distant source of energy propagating through a medium, based on noisy measurements of disturbance in the medium. In particular, we consider the ocean as a medium, and we consider detection of a distant underwater disturbance such as an earthquake or storm center [Munk et al. 1965].

Suppose there are two recording stations, separated by a considerable distance, at which measurements of ocean-wave amplitudes, $y_1(t)$ and $y_2(t)$, are taken. Suppose also that all contributions to $y_1(t)$ and $y_2(t)$ due to the distant source can be attributed to a single waveform $x(t)$ representing an idealized point source. Let $y_1(t)$ and $y_2(t)$ be modeled by

$$y_1(t) = h_1(t) \otimes x(t) + n_1(t)$$
$$y_2(t) = h_2(t) \otimes x(t) + n_2(t),$$

for which $n_1(t)$ and $n_2(t)$ are uncorrelated with $x(t)$ and are due primarily to local disturbance around each of the two recording stations, so that $n_1(t)$ and $n_2(t)$ are uncorrelated with each other. For this situation, it can be shown (exercise 4) that the coherence of $y_1(t)$ and $y_2(t)$ is given by

$$|\hat{C}_{12}(f)| = \{[1 + \rho_1(f)][1 + \rho_2(f)]\}^{-1/2},$$

where $\rho_1(f)$ is the ratio of noise and signal spectral densities

$$\rho_i(f) \triangleq \frac{\hat{S}_{n_i}(f)}{|H_i(f)|^2 \hat{S}_x(f)}.$$

Since the power in locally generated seawaves, such as $n_1(t)$ and $n_2(t)$, generally occupies a higher frequency range than that of waves that have traveled a long distance, such as $h_1(t) \otimes x(t)$ and $h_2(t) \otimes x(t)$ (because $h_1(t)$ and $h_2(t)$ represent low-pass filters), then we should have

$$\rho_1(f_l) \ll \rho_1(f_h)$$
$$\rho_2(f_l) \ll \rho_2(f_h)$$

for f_l in the low-frequency range and f_h in the high-frequency range. Consequently

$$|\hat{C}_{12}(f_l)| \gg |\hat{C}_{12}(f_h)|.$$

Moreover, if

$$\rho_1(f_l) \ll 1$$
$$\rho_2(f_l) \ll 1,$$

then

$$|\hat{C}_{12}(f_l)| \cong 1.$$

This would hold true if a distant source $x(t)$ were present. If not, then $\hat{S}_x(f) \to 0$ and, therefore,

$$|\hat{C}_{12}(f)| \longrightarrow \frac{|\hat{S}_{n_1 n_2}(f)|}{[\hat{S}_{n_1}(f)\hat{S}_{n_2}(f)]^{1/2}} = 0$$

since $n_1(t)$ and $n_2(t)$ are uncorrelated. Hence, a large value of coherence in the low frequency range indicates the presence of a distant source.

Of course, in practice our recordings are limited to finite time intervals, so that we must use statistical inference methods for estimation of the limit spectra and cross spectra of $y_1(t)$ and $y_2(t)$, as described in Section D.

Other energy source problems, such as source location, for which statistical spectral analysis is useful, are treated in [Bendat and Piersol 1980]. See also Appendix 7-3, and for improved methods that are tolerant to noise and interference see Chapter 14 in Part II.

APPENDIX 7-3

Time- and Frequency-Difference-of-Arrival Estimation

In many application areas such as radar and sonar, a fundamental problem is that of estimating the relative delay and/or relative frequency shift (doppler) between two corrupted versions, say $y(t)$ and $z(t)$, of a time-series, $x(t)$. For example, if $y(t)$ and $z(t)$ are the signals received by a pair of sensors on which a signal wavefront is impinging, then the planar direction of arrival of the wavefront can be determined from the relative delay of these two signals. If $y(t)$ and $z(t)$ are modeled by

$$y(t) = x(t - t_y) + n(t) \tag{1a}$$
$$z(t) = x(t - t_z) + m(t), \tag{1b}$$

where $n(t)$ and $m(t)$ are uncorrelated zero-mean measurement noises, then it follows (exercise 19) that the cross correlation of y and z is given by

$$\hat{R}_{yz}(\tau) = \hat{R}_x(\tau + t_z - t_y). \tag{2}$$

Consequently the *time-difference of arrival* (TDOA) of $y(t)$ and $z(t)$ can be estimated by measuring the cross correlogram

$$R_{yz_T}(\tau) = \frac{1}{T}\int_{-(T-|\tau|)/2}^{(T-|\tau|)/2} y\left(t + \frac{\tau}{2}\right) z\left(t - \frac{\tau}{2}\right) dt \tag{3}$$

and searching over the lag parameter τ for a peak. The accuracy of estimation is directly related to the width of the autocorrelation $\hat{R}_x$. The narrower (more peaked) $\hat{R}_x$ is, the more accurately the shift $t_z - t_y$ in the peak can be determined.

Similarly, if $y(t)$ and $z(t)$ represent noisy measurements of Doppler-shifted versions of $x(t)$, then the cross spectrum is given by

$$\hat{S}_{yz}(f) = \hat{S}_x(f + f_z - f_y)\delta_{f_z - f_y}. \tag{4}$$

This is an approximation that is accurate if the width of $\hat{S}_x$ is narrow relative to its center frequency (otherwise the doppler effect is not simply a frequency shift.) The Kronecker delta,

$$\delta_{f_z - f_y} = \begin{cases} 1, & f_z = f_y \\ 0, & f_z \neq f_y, \end{cases} \tag{5}$$

reflects the fact that frequency-shifted versions of a given stationary time-series $x(t)$ are uncorrelated with each other (see Chapter 5). It follows from (4) that the *frequency-difference of arrival* (FDOA) of $y(t)$ and $z(t)$ cannot be determined simply by measurement of the cross spectrum, since the limit cross spectrum vanishes for all nonzero FDOA. However, if $y(t)$ and $z(t)$ are frequency-shifted by $+\nu/2$ and $-\nu/2$, respectively, then when $\nu = f_z - f_y$ the cross spectrum will not vanish. Hence, the FDOA can be estimated by measuring the cross correlogram at zero lag as a function of the local frequency-shift, ν, and finding the peak of this function.

This FDOA estimation method can be combined with the TDOA estimation method. If the correlogram estimate is used, then the joint estimation procedure is described by

$$\max_{\nu,\tau} A(\tau, \nu), \tag{6a}$$

where

$$A(\tau, \nu) \triangleq \left| \frac{1}{T} \int_{-(T-|\tau|)/2}^{(T-|\tau|)/2} \gamma_y\left(t + \frac{\tau}{2}\right)\gamma_z^*\left(t - \frac{\tau}{2}\right)e^{i2\pi\nu t}\, dt \right|, \tag{6b}$$

which is the cross correlogram magnitude at zero lag for the two time- and frequency-shifted time-series

$$\tilde{y}(t) \triangleq \gamma_y\left(t + \frac{\tau}{2}\right)e^{i2\pi(\nu/2)(t+\tau/2)} \tag{7a}$$

$$\tilde{z}(t) \triangleq \gamma_z\left(t - \frac{\tau}{2}\right)e^{-i2\pi(\nu/2)(t-\tau/2)}, \tag{7b}$$

where γ_y and γ_z are the complex envelope representations (relative to some frequency f_0) of $y(t)$ and $z(t)$ (see Appendix 3-1),

$$A(\tau, \nu) = \left| \frac{1}{T} \int_{-(T-|\tau|)/2}^{(T-|\tau|)/2} \tilde{y}(t)\tilde{z}^*(t)\, dt \right| = \left| R_{\tilde{y}\tilde{z}_\tau}(0) \right|. \tag{8}$$

The function $A(\tau, \nu)$ is the *cross-ambiguity function* from radar theory [Rihaczek 1969]. The reason that the complex envelopes of the time-series are used in place of the time-series themselves is that the actual Doppler effect shifts the positive frequencies in one direction and the negative frequencies in the other.

In order to model this frequency-shift simply as a multiplication by a complex sine wave (which shifts *all* frequencies in the same direction), the negative frequencies must first be discarded (as done in the formation of the complex envelope (see Appendix 3-1)).

8

TIME-VARIANT SPECTRAL ANALYSIS

A. GENERAL VARIATION

As explained in Chapter 1, spectral analysis is especially appropriate in principle only for time-invariant linear transformations of data and for data from time-invariant phenomena. But, from a pragmatic point of view, we expect spectral analysis to be useful for time-variant linear transformations and for data from time-variant phenomena, provided that time variation is sufficiently slow. The purpose of this chapter is to clarify why time variation must be slow and how slow it must be. Only the situation in which the time-series data alone is to be used for spectral analysis is considered. That is, the model-fitting methods described in Chapter 9, which typically require additional information for the purpose of model-type selection, are omitted from consideration.

1. The Physical Spectrum

Regardless of the phenomenon or the data under study, the time-variant periodogram introduced in Chapter 2, Section A, is the normalized squared magnitude of a time-variant density of spectral components (namely, the time-variant finite-time complex spectrum; see (2b) in Chapter 2), and is therefore the most natural definition and the only generally appropriate definition of a nonstatistical time-variant spectrum. As explained in Chapter 5, Section C, if there is no ensemble of time-series available from the phenomenon of interest, then the degree of randomness of this spectrum can be reduced (reliability increased) only by temporal or spectral smoothing, with the result that temporal and spectral resolution are severely constrained by *Grenander's uncertainty condition,* (89) in Chapter 5,

$$\Delta t \Delta f \gg 1, \tag{1}$$

which is a necessary and sufficient condition (for a broad class of time-series data) for high reliability. Thus, the reliable resolvability of time variation in spectral characteristics of a phenomenon is limited by a temporal resolution width Δt that must greatly exceed the reciprocal of the spectral resolution width Δf. On the other hand, if an ensemble of replicas of a time-series (not time-translates of a single time-series but rather genuine random samples as would be obtained, for example, from replicating an experiment) from a given time-variant phenomenon is available, then as explained in Chapter 5, Section A, the degree of randomness can be reduced by ensemble averaging with no degradation in either temporal or spectral resolution. Thus, resolvability is limited by only *Gabor's uncertainty principle,* (88) in Chapter 5,

$$\Delta t \Delta f \cong 1, \tag{2}$$

which applies to the time-variant periodogram, whether or not it is ensemble-averaged. Furthermore, even when an ensemble is not physically available, the concept of an ensemble-averaged time-variant spectrum can be of value.[1] Specifically, the idealized limit of an ensemble-averaged time-variant periodogram, (4) in Chapter 5, which is the *expected time-variant periodogram*

$$\mathcal{S}_{x_T}(t, f) \triangleq E\{S_{x_T}(t, f)\}, \tag{3}$$

can be interpreted as a *completely reliable* (zero degree of randomness) measure of the time-variant spectral characteristics of a phenomenon. This most reliable measure exhibits the best possible temporal and spectral resolutions, that is, the resolutions satisfy (2). From this point of view, it apparently makes no sense to conceive of or seek a probabilistic measure of time-variant spectral characteristics that exhibits finer resolutions than those that satisfy (2).

Nevertheless, there is a probabilistic function of time and frequency that plays a fundamental role in the mathematical characterization of the expected periodogram $\mathcal{S}_{x_T}(t, f)$ and that is not subject to any counterpart of (2). Specifically, it can be shown (exercise 1) [Mark 1970], that the expected periodogram is a time- and frequency-smoothed version of a function $\mathcal{S}_x(t, f)$ called the *probabilistic instantaneous spectrum,*

$$\mathcal{S}_{x_T}(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{S}_x(t - \tau, f - \nu) w(\tau, \nu)\, d\tau\, d\nu, \tag{4}$$

in which the smoothing function w is defined by

$$w(\tau, \nu) \triangleq \frac{1}{T} \int_{-\infty}^{\infty} Tu_T\left(\tau + \frac{u}{2}\right) Tu_T\left(\tau - \frac{u}{2}\right) e^{-i2\pi\nu u}\, du \tag{5}$$

and the probabilistic instantaneous spectrum $\mathcal{S}_x$ is defined by

$$\mathcal{S}_x(t, \cdot) \triangleq F\{\mathcal{R}_x(t, \cdot)\} \tag{6}$$

$$\mathcal{R}_x(t, \tau) \triangleq E\left\{x\left(t + \frac{\tau}{2}\right) x\left(t - \frac{\tau}{2}\right)\right\}, \tag{7}$$

[1] *Wold's isomorphism* cannot be used to envision an appropriate ensemble in this case, as done in Chapter 5, Section B for time-invariant phenomena, because the equivalence between ensemble- and time-averaging would result in all time variation being averaged away by the ensemble-averaging operation.

where $\mathcal{R}_x(t, \tau)$ is called the *probabilistic instantaneous autocorrelation function*. Although the probabilistic instantaneous spectrum can in principle be measured as accurately as desired when a sufficiently large ensemble is physically available, in general it does not represent a time-variant spectrum of the data in any physical sense.[2] It can only approximate a time-variant spectrum, and this approximation can be close only when the time-variation is sufficiently slow that the fine structure of the function $\mathcal{S}_x(t, f)$ is accurately resolved with resolution widths satisfying

$$\Delta t \Delta f \geq 1 \tag{8}$$

(see [Mark 1970; Donati 1971]). But it can be shown by using (4)–(5) that when (8) is satisfied, then (using $T = \Delta t$)

$$\mathcal{S}_{x_T}(t, f) \cong \mathcal{S}_x(t, f). \tag{9}$$

As a reflection of the facts stated in the preceding discussion, the expected time-variant spectrum is called the *physical spectrum* to distinguish it from the generally nonphysical instantaneous probabilistic spectrum [Mark 1970; Eberly and Wodkiewicz 1977].

2. Linear Time-Variant Systems

To illustrate the ideas in the preceding subsection, let us consider a generalization of the motivating example discussed in Chapter 3, Section A. Specifically, we consider the problem of determining the time-variant spectral characteristics of a *linear time-variant* (LTV) system excited by white noise by measuring the time-variant spectrum of the system response. We begin with a definition of the time-variant spectral characteristics of an LTV system. The response of an LTV system can be expressed as a time-variant continuous linear combination of the excitation values,

$$y(t) = \int_{-\infty}^{\infty} h(t, u)x(u)\, du \tag{10}$$

for which $h(t, u)$ is the response at time t to an impulse excitation applied at time u. The *system function* for such a system is defined to be the Fourier transform [D'Angelo 1970]

$$G(t, \cdot) = F\{g(t, \cdot)\} \tag{11}$$

for which

$$g(t, \tau) \triangleq h(t, t - \tau). \tag{12}$$

The system function $G(t, f)$ is a generalization of the *transfer function* for a linear time-invariant (LTI) system for which $h(t, u) = h(t - u)$, because in this case we have

$$G(t, f) = H(f). \tag{13}$$

Moreover, for a sine wave excitation,

$$x(t) = e^{i2\pi ft}, \tag{14a}$$

[2] However, it is explained in Part II that for a cyclostationary or almost cyclostationary process, the Fourier coefficients of the periodic or almost periodic function $\mathcal{S}_x(t, f)$ are spectral correlation functions, which do indeed have a concrete physical interpretation.

the response is obtained simply by multiplication with the system function,

$$y(t) = G(t,f)e^{i2\pi ft}, \tag{14b}$$

analogous to the principal-component property of the transfer function of an LTI system, as described in Part 2 of Section B, Chapter 1. Consequently, *the system function magnitude* $|G(t,f)|$ *is defined* (by analogy with $|H(f)|$) *to be a representation of the time-variant spectral characteristics of an LTV system.*

Now, let us determine the relationship between the probabilistic instantaneous spectrum $\mathcal{S}_y(t,f)$ and the system function magnitude $|G(t,f)|$. For an LTI system with transfer function H, excited by white noise with probabilistic spectrum

$$\mathcal{S}_x(f) = \int_{-\infty}^{\infty} \mathcal{R}_x(\tau)e^{-i2\pi f\tau}\,d\tau = N_0, \tag{15a}$$

where

$$\mathcal{R}_x(\tau) \triangleq E\left\{x\left(t+\frac{\tau}{2}\right)x\left(t-\frac{\tau}{2}\right)\right\} = N_0\delta(\tau), \tag{15b}$$

we know that (see (28) in Chapter 3)

$$\mathcal{S}_y(f) = |H(f)|^2N_0. \tag{16}$$

Also, for a time-invariant phenomenon it follows from (15) in Chapter 5 and (6) and (7) that

$$\mathcal{S}_y(t,f) = \mathcal{S}_y(f). \tag{17}$$

However, even though $|G(t,f)|$ and $\mathcal{S}_y(t,f)$ are natural generalizations of $|H(f)|$ and $\mathcal{S}_y(f)$, it does not follow that (16) generalizes to

$$\mathcal{S}_y(t,f) = |G(t,f)|^2N_0. \tag{18}$$

But, it can be shown (exercise 2) that the approximation

$$\mathcal{S}_y(t,f) \cong |G(t,f)|^2N_0 \tag{19}$$

is close in general if (and only if) the system time variation is sufficiently slow that the fine structure of $G(t,f)$ can be accurately resolved with resolution widths satisfying (8). Moreover, in this case (9) applies so that we have

$$\mathcal{S}_{y_T}(t,f) \cong |G(t,f)|^2N_0. \tag{20}$$

Thus, only the physical spectrum $\mathcal{S}_{y_T}(t,f)$, not the probabilistic instantaneous spectrum $\mathcal{S}_y(t,f)$, is needed in general for determining the time-variant spectral characteristics of an LTV system. Moreover, when no ensemble is available, these characteristics as reflected in $|G(t,f)|$ can be reliably measured only for a resolution product satisfying (1).

Example: ***Locally Stationary Process***

Consider an LTV system with time-varying attenuation and negligible temporal dispersion. The input-output model for such a system is simply a product

$$y(t) = a(t)x(t). \tag{21}$$

This product can be put into the form of the superposition integral (10) by use of the impulse-response function

$$h(t,\tau) = a(t)\delta(t-\tau). \tag{22}$$

Formulas (11) and (12) applied to (22) yield the system function

$$G(t, f) = a(t), \tag{23}$$

which is independent of frequency f—a result of the dispersionless character of the system. Substitution of (21) into definition (7) (with x there replaced by y) yields the probabilistic instantaneous autocorrelation function

$$\begin{aligned} \mathscr{R}_y(t, \tau) &= E\left\{a\left(t + \frac{\tau}{2}\right)x\left(t + \frac{\tau}{2}\right)a\left(t - \frac{\tau}{2}\right)x\left(t - \frac{\tau}{2}\right)\right\} \\ &= a\left(t + \frac{\tau}{2}\right)a\left(t - \frac{\tau}{2}\right)\mathscr{R}_x(t, \tau). \end{aligned} \tag{24}$$

For white noise excitation we have

$$\mathscr{R}_x(t, \tau) = N_0\delta(\tau),$$

and therefore (24) reduces to

$$\mathscr{R}_y(t, \tau) = N_0 a^2(t)\delta(\tau). \tag{25}$$

Application of definition (6) to (25) yields the probabilistic instantaneous spectrum

$$\mathscr{S}_y(t, f) = N_0 a^2(t), \tag{26}$$

which is independent of frequency f. Consequently, no resolution in f is needed, and we can let $\Delta f = \infty$. As a result (8) is satisfied for arbitrarily small Δt, and therefore approximation (19) is exact. This is directly corroborated by substitution of (23) and (26) into (19). This example is the time-frequency dual of an LTI system that depends on f but is independent of t, in which case (19) is again exact, because we can let $\Delta t = \infty$ and satisfy (8) with arbitrarily small Δf. The stochastic process $y(t)$ in this example is sometimes called *locally stationary*. More generally, a probabilistic model for which (19) is a close approximation is called locally stationary (in the wide sense). The conditions under which a model consisting of an LTV system with white noise excitation is locally stationary are developed in exercise 2(a).

For a time-series $y(t)$ that can be modeled by (10) and (15) (with $N_0 = 1$), the squared magnitude of the system function $|G(t, f)|^2$ is called the *evolutionary spectrum* of $y(t)$ and has been suggested as an alternative to the probabilistic instantaneous spectrum that can be interpreted as an idealized probabilistic spectrum with no constraints on resolution [Priestley 1965, 1981; Tjøstheim 1976]. However as illustrated here, this function can be accurately determined in general with the use of ensemble averaging if and only if (8) is satisfied for $|G(t, f)|^2$, in which case both (19) and (20) are close approximations. A popular approach to time-variant spectral analysis based on the evolutionary spectrum employs the model-fitting methods described in Chapter 9. For example, the AR model fitting methods can be adapted to provide a time-variant AR model of the form

$$y(n) + a_M(n)y(n - M) + a_{M-1}(n)y(n - M + 1) + \cdots + a_1(n)y(n - 1) = x(n). \tag{27}$$

After having fit such a model to a given time-series $y(t)$, the function

$$|G_*(n, f)|^2 \triangleq \frac{N_0}{\left|1 - \sum_{p=1}^{M} a_p(n)(e^{-i2\pi f})^p\right|^2} \tag{28}$$

can be used as a time-variant spectrum estimate. This approach is reasonable for a sufficiently slowly time-variant phenomenon, in which case the function $G_*(n, f)$ can be a useful approximation to the system function for the model (27) (exercise 6), and the discrete-time counterparts to (19) and (20) can be useful approximations. However for time-variant phenomena with rapid time-variations, this approach is of questionable value; see exercise 6. (See [Loynes 1968] and [Priestley and Tong 1973] for interesting discussions of time-variant spectral analysis and [Gardner 1987c] for a general treatment of the related problem of measurement of time-variant correlation.)

3. Local Ergodicity

It is often thought that a probabilistic model that is locally stationary is also *locally ergodic,* which means that, for example, the probabilistic parameter $\mathcal{S}_y(t, f)$ can be accurately estimated using only a time average instead of an ensemble average. This thought is based on the reasoning that since the time variations are slow, much time averaging can be done while the physical spectrum $\mathcal{S}_{y_T}(t, f)$ changes very little, and therefore the ensemble average (expected value) in (3) can be closely approximated by a time average. But this is not true in general even if the probabilistic model is such that it becomes ergodic in the limit as the rate of time variations approaches zero so that the locally stationary process becomes stationary.[3] To illustrate why this is so, we consider an example.

Example: ***Photodetector Output***

As an example of a physical phenomenon that can be modeled probabilistically as time-variant, consider the voltage output $x(t)$ of a photodetector with light excitation whose intensity varies with time,

$$x(t) = \sum_i a_i g(t - \tau_i). \tag{29}$$

The ith photon incident on the detector gives rise to a voltage pulse at time τ_i with random amplitude a_i. The amplitudes $\{a_i\}$ can be modeled as a sequence of independent variables that are also independent of the times of arrival of the photons $\{\tau_i\}$. This sequence of arrival times can be modeled as a Poisson point process, with expected rate $\lambda(t)$, in which case the probabilistic instantaneous autocorrelation can be shown [Snyder 1975] to be given by

$$\begin{aligned} \mathcal{R}_x(t, \tau) &= \bar{\mathcal{R}}_a(0) \int_{-\infty}^{\infty} \lambda(u) g\left(t - u + \frac{\tau}{2}\right) g\left(t - u - \frac{\tau}{2}\right) du \\ &= \bar{\mathcal{R}}_a(0)\lambda(t) \otimes \left[g\left(t + \frac{\tau}{2}\right) g\left(t - \frac{\tau}{2}\right) \right], \end{aligned} \tag{30}$$

where $\bar{\mathcal{R}}_a(0)$ is the mean-squared value of $\{a_i\}$, and the convolution is with respect to the variable t. The probabilistic instantaneous spectrum is obtained by Fourier transformation of (30) and is given by

$$\mathcal{S}_x(t, f) = \bar{\mathcal{R}}_a(0)\lambda(t) \otimes W_g(t, f), \tag{31a}$$

where the function

$$W_g(t, f) \triangleq \int_{-\infty}^{\infty} g\left(t + \frac{\tau}{2}\right) g\left(t - \frac{\tau}{2}\right) e^{-i2\pi f\tau} \, d\tau \tag{31b}$$

[3] A detailed treatment of ergodicity is given in [Gardner 1985].

is called the *Wigner-Ville time-frequency energy density* of the pulse $g(t)$ [Wigner 1932; Ville 1948] (see [Claasen and Mecklenbräuker 1980]). Because of the smoothing effect of convolution, the temporal resolution of $\mathscr{S}_x(t, f)$ is limited by the smoothest of the two time functions $\lambda(t)$ and $W_g(t, f)$. If the pulse $g(t)$ is narrow enough to be treated as an impulse, then $W_g(t, f) = \delta(t)$, and therefore

$$\mathscr{S}_x(t, f) = \tilde{\mathscr{R}}_a(0)\lambda(t), \tag{32}$$

which indicates that this is a locally stationary process when $g(t)$ is sufficiently narrow. Does this mean that this probabilistic model is also locally ergodic? No it does not. This model is locally ergodic only if the expected rate $\lambda(t)$ of photon arrivals (which is proportional to the light intensity) is sufficiently high. If the expected rate is very low, then there will be very few photons and, therefore, very few voltage pulses to time average during the interval over which $\lambda(t)$ is nearly constant, and consequently the expected rate factor in (32) cannot be accurately estimated with a time average. This same conclusion applies to the example of a locally stationary probabilistic model in the previous subsection if the white noise excitation $x(t)$ in (21) is low-rate Poisson impulse noise. On the other hand, if the white noise $x(t)$ is Gaussian, then the locally stationary model (21) is indeed locally ergodic.[4] This example illustrates the subtleties that must be dealt with when probabilistic models are used.

B. PERIODIC VARIATION

In contrast with the general conclusion drawn in the preceding section, a *periodically time-variant linear system*[5] is a special case for which it can be shown that the probabilistic instantaneous spectrum, (6)–(7), for the system response (10) can be measured as accurately as desired *without ensemble averaging*. In fact, this is true for any periodically time-variant phenomenon with an appropriate (cycloergodic) probabilistic model [Boyles and Gardner 1983]. Therefore, (4) can be used to obtain the completely reliable measure of periodically time-variant spectral characteristics (3), which is subject only to the resolution-limiting condition (8) rather than the severely limiting condition (1). Moreover, it is shown in Part II that $|G(t, f)|^2$ can be measured as accurately as desired for a periodically time-variant system, with *no* limitation on resolution (not even (8)) and without ensemble averaging! In addition, the model-fitting approach has proven to be particularly useful for time-variant spectral analysis for periodic phenomena (see [Jones and Brelsford 1967; Parzen and Pagano 1979; Hasselmann and Barnett 1981]), and does not require that the periodic time variations be slow.

Example: *Swept-Frequency Analyzer*

As explained in Chapter 4, a waveform $x(t)$ can be analyzed over a broad range of frequencies by using a single low-pass filter preceded by a swept-frequency modulator. The output $y(t)$ of this modulator-filter combination is given by

$$y(t) = \int_{-\infty}^{\infty} h(t - u)c(u)x(u)\, du, \tag{33}$$

[4] A detailed treatment of the different white noise models is given in [Gardner 1985].

[5] The claims made here for a system with a single periodicity apply equally well to a system with multiple incommensurate periodicities.

where $h(t)$ is the impulse-response function for the filter and $c(u)$ is a sine wave with linearly swept frequency that is periodically repeated with period T_0,

$$c(t) = \cos(2\pi f_0 t + \pi\beta t^2), \qquad \frac{-T_0}{2} \leqslant t < \frac{T_0}{2}. \tag{34}$$

The probabilistic instantaneous autocorrelation of $y(t)$ is easily shown to be

$$\mathcal{R}_y(t, \tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h\left(t + \frac{\tau}{2} - u\right)h\left(t - \frac{\tau}{2} - v\right)c(u)c(v)\mathcal{R}_x\left(\frac{u + v}{2}, u - v\right) du\, dv. \tag{35}$$

If the waveform $x(t)$ is simply white noise, then

$$\mathcal{R}_x(t, \tau) = N_0\delta(\tau)$$

and therefore (35) reduces to

$$\mathcal{R}_y(t, \tau) = N_0 c^2(t) \otimes \left[h\left(t + \frac{\tau}{2}\right)h\left(t - \frac{\tau}{2}\right)\right] \tag{36}$$

and the probabilistic instantaneous spectrum is given by

$$\mathcal{S}_y(t, f) = N_0 c^2(t) \otimes W_h(t, f), \tag{37}$$

where W_h is the Wigner-Ville density for $h(t)$ (see (31b)). In practice the width of the impulse-response function $h(t)$ is typically chosen to be considerably larger than the varying period of the swept sine wave. Consequently, this is definitely not a locally stationary process, regardless of the sweep rate β. Nevertheless, it is a cycloergodic process and therefore the techniques described in Part II can be used to estimate $\mathcal{S}_y(t, f)$ as accurately as desired without ensemble averaging. And this would be true even if $c(u)$ were the sum of several periodic functions with incommensurate periods.

C. SUMMARY

In Section A the fundamental limitations on the simultaneous temporal and spectral resolution capabilities of statistical spectra are described, and the advantage of ensemble averaging (when possible) is explained. The *instantaneous probabilistic spectrum* is introduced and its relation to the *physical spectrum* is explained and illustrated with an application to the problem of identification of a time-variant linear system. An alternative to the instantaneous probabilistic spectrum as an idealized probabilistic time-variant spectrum, called the *evolutionary spectrum,* is introduced and its direct relationship with the system function is explained. The limitations on measurement of these idealized probabilistic spectra, when ensemble averaging cannot be used, are explained and the concepts of locally stationary and locally ergodic processes are introduced.

In Section B the fact that a periodically time-variant phenomenon is an exceptional case is briefly discussed. It is explained that for this special type of time-variant phenomenon, the fundamental resolution limitations do not apply. Another exception also should be mentioned: those situations in which the spectral analyst has more information about the phenomenon under study than can be obtained from the time-series alone. For example, if it is known that the time-series can be accurately modeled as an autoregression, then—as explained in

Chapter 9—spectral resolution performance that exceeds that of the periodogram is possible (but is not guaranteed).

A general treatment of the related problem of reliable measurement of time-variant auto- and cross-correlation functions is given in [Gardner 1987c].

EXERCISES

1. **(a)** Show that the expected periodogram is given by (4)–(7). *Hint:* Use the periodogram-correlogram relation, and the formula

$$R_{x_T}(t, \tau) = \frac{1}{T}\int_{-\infty}^{\infty} Tu_T\left(v + \frac{\tau}{2}\right)Tu_T\left(v - \frac{\tau}{2}\right)x\left(t + v + \frac{\tau}{2}\right)x\left(t + v - \frac{\tau}{2}\right) dv. \quad (38)$$

(b) Let $x(t)$ arise from a time-invariant phenomenon so that $\mathcal{S}_x(t, f)$ is independent of t, and show that (4)–(7) reduce to formula (34) in Chapter 5.

2. **(a)** Let $x(t)$ be white noise so that $\mathcal{R}_x(t, \tau) = N_0\delta(\tau)$, and show that the instantaneous autocorrelation for the response of the LTV system (10) is given by

$$\mathcal{R}_y(t, \tau) = N_0\int_{-\infty}^{\infty} g\left(t + \frac{\tau}{2}, \tau + v\right)g\left(t - \frac{\tau}{2}, v\right) dv. \quad (39)$$

If the resolution width Δt^* of time variation of the system (minimum over τ of the resolution width in t of $g(t, \tau)$) is greater than the memory length $\Delta\tau^*$ of the system (maximum-over-t width in τ of $g(t, \tau)$), then (39) is closely approximated by

$$\mathcal{R}_y(t, \tau) \cong N_0\int_{-\infty}^{\infty} g(t, \tau + v)g(t, v) dv. \quad (40)$$

Show that (40) yields (19) for $\Delta t^*\Delta f^* \geq 1$, where $\Delta f^* = 1/\Delta\tau^*$ is the minimum over t of the resolution width in f of $G(t, f)$.

(b) Verify that (40) and (19) are exact for the particular LTV system considered in the example in Part 2 of Section A.

3. Determine the system function for the swept-frequency system described in the example in Section B.

4. Consider a time-variant resistive-capacitive low-pass filter, which is governed by the differential equation

$$a(t)\frac{dy(t)}{dt} + y(t) = x(t), \quad (41)$$

where $a(t)$ represents the product of time-variant resistance and capacitance.

(a) Show that the system function is given by

$$G(t, f) = \int_{-\infty}^{\infty} a^{-1}(v)\exp\left\{-\int_v^t a^{-1}(u)\,du\right\}\exp\{-i2\pi f(t - v)\}\,dv. \quad (42)$$

Hint: Substitute $x(t) = e^{i2\pi ft}$ and $y(t) = G(t, f)e^{i2\pi ft}$ into (41), and then use the fact that the steady-state solution to the differential equation

$$\frac{dG(t)}{dt} = \alpha(t)G(t) + \beta(t) \quad (43a)$$

is given by

$$G(t) = \int_{-\infty}^{t}\exp\left\{-\int_v^t \alpha(u)\,du\right\}\beta(v)\,dv. \quad (43b)$$

Note that the analysis of the time-variant circuit in this exercise is deceptively simple because the time-variant parameters do not get differentiated in the differential equation. This is unfortunately not the case for most time-variant circuits of interest.

(b) Verify that if the product $a(t)$ of resistance and capacitance is time-invariant, then

$$G(t, f) = \frac{1}{1 + i2\pi fa}. \tag{44}$$

5. Consider the problem of time-variant spectral analysis of speech. In order to track certain fluctuations in the character of speech, a temporal resolution of 30 ms is desired.

(a) If a coefficient of variation of $\frac{1}{10}$ is desired, what is the minimum attainable spectral resolution width?

(b) How can the requirements on temporal resolution and/or coefficient of variation be modified in order to obtain a spectral resolution of 100 Hz?

(c) Assume that a temporal resolution of 30 ms and a spectral resolution of 200 Hz are required, and that the speech waveform to be analyzed is bandlimited to 5 KHz. If the Bartlett-Welch method with 50% overlap and no data-tapering is to be used, then determine how many time-samples are to be processed for each spectrum estimate (30-ms segment), how many zeros are to be added, what DFT size (integer power of 2 only) is needed, and how many DFT blocks are to be averaged for each spectrum estimate.

(d) What properties of the spectrum estimate will change and how will they change if data tapering is used?

6. (a) Consider the LTV model

$$y(t) - a_1(t)\frac{dy(t)}{dt} - a_2(t)\frac{d^2y(t)}{dt^2} - \cdots - a_M(t)\frac{d^My(t)}{dt^M} = x(t), \tag{45}$$

which is the time-variant counterpart of the all-pole LTI model described in exercise 9 of Chapter 3. Let $x(t) = e^{i2\pi ft}$ and let $G(t, f)$ denote the system function for (45). Then $y(t) = G(t, f)e^{i2\pi ft}$. Use this to verify that

$$G(t, f) \neq G_*(t, f) \triangleq \frac{1}{1 - \sum_{m=1}^{M} a_m(t)(i2\pi f)^m} \tag{46}$$

except in the special case for which $\{a_m(t)\}_1^M$ are all time-invariant, so that $G(t, f) = H(f)$. Also show that if $\{a_m(t)\}_1^M$ and $G(t, f)$ vary sufficiently slowly with time, then by treating derivatives of $G(t, f)$ with respect to t as negligible, (46) becomes a useful approximation.

(b) Do the discrete-time counterpart of (a), in which model (45) is replaced with the AR model (27).

9

PARAMETRIC METHODS

A. INTRODUCTION

All the methods of spectral analysis described in previous chapters are based on a direct decomposition of the data to be analyzed into spectral components using Fourier transformation or filtering and are therefore called *direct methods*. In contrast, the methods studied in this chapter are based on an entirely different philosophy. Specifically, each of these methods fits a particular form of model to the data by adjusting the values of parameters in the model. Once the model fitting is complete, the parameter values can be substituted into a formula to obtain an estimate of the limit spectrum for that model. There are many variations on this prototypical method. All such methods are referred to as *parametric methods,* whereas the direct methods described in previous chapters based on direct decomposition into spectral components are referred to as *nonparametric* methods.

Parametric methods of spectral analysis can yield better resolution of multiple spectral lines or other narrow features when the amount of data is severely limited, especially if the data-segment length is smaller than the reciprocal of the desired spectral resolution, and the experiment is repeatable so that an ensemble of data segments is available. (This occurs, for example, in some sensor-array signal-processing problems, where the spatial data-segment length is the number of sensors and ensemble averaging is performed by time-averaging.) But more generally the relative advantages depend on the appropriateness of the form of the model chosen. In some cases, nonparametric methods are helpful in selecting a form of model to be used as the basis for a parametric method.

In fact, in some applications the main objective is to fit a model to the data, and nonparametric spectral analysis methods are often used as a first step. But, even when an appropriate form of model for a signal is known, if the signal is masked by noise, then nonparametric methods can be superior for spectral analysis. However, parametric methods can be particularly useful for identification (detection and estimation) of additive sine wave components to be removed (to minimize spectral leakage) before application of a nonparametric method. Generally speaking, parametric methods are more computationally burdensome than nonparametric methods, but some methods based on autoregressive models are computationally competitive with direct methods based on the FFT. The practicality of autoregressive model-fitting methods of spectral analysis is attested to by the wide range of problems to which these methods have been applied. These include radar, sonar, image processing, radio astronomy, biomedicine, speech analysis and synthesis, geophysics, seismology, and oceanography.

Since parametric methods of spectral analysis are diverse and often tailored to special types of data, it is not possible to present a comprehensive and unified treatment paralleling that presented in previous chapters for the direct methods. However, the fundamental concepts and mathematical theory underlying the particularly important class of autoregressive methods can be and therefore are presented in a unified manner. In keeping with the philosophy of this book, the unified treatment is nonprobabilistic, whereas other treatments of the same material are typically couched within the conventional probabilistic framework.

Although the results of the fraction-of-time probabilistic analysis, such as bias and variance formulas derived in Chapter 5, apply to all direct methods, the parametric methods are not amenable to such straightforward probabilistic analysis. Useful results on bias and variance are typically obtainable only asymptotically as the data segment length used for spectral analysis approaches infinity (see [Kay 1987]).

In Section B, the theoretical background for autoregressive and related ARMA methods is presented. Then, in Sections C and D some of the methods that have proven to be of practical value are described. Finally, in Section E, an extensive experimental study and comparison of these methods is presented.

B. AUTOREGRESSIVE MODELING THEORY

There is a variety of different parametric methods of statistical spectral analysis that are based on the autoregressive (AR) model for time-series introduced in Chapter 3. In order to understand the motivation behind these various methods, we must first develop some background on the theory of AR modeling, which is accomplished in this section. The notation used in Section B of Chapter 6 is adopted here since it is conventional for this subject. Extension of the theory in this section and most of the methods in the following sections from real data to complex data can be found in [Kay 1987; Marple 1987].

The Mth order AR model for a time-series x_n is the difference equation

$$x_n + a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_M x_{n-M} = b z_n, \tag{1}$$

where z_n is a zero-mean sequence of uncorrelated variables,

$$\tilde{R}_z(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0. \end{cases} \tag{2}$$

This model enables us to interpret x_n as the response of a linear time-invariant system to white noise excitation z_n. The transfer function $\tilde{G}(f)$ of this system is obtained by substitution of the excitation $z_n = e^{i2\pi nf}$ and response $x_n = \tilde{G}(f)e^{i2\pi nf}$ into (1) to obtain

$$\tilde{G}(f) = \frac{b}{1 + \sum_{p=1}^{M} a_p(e^{-i2\pi f})^p}. \tag{3}$$

Since the power spectral density for z_n is unity (from (2)), then it follows from the limit-spectrum relation for filters and (3) that the power spectral density for x_n is given by

$$\tilde{S}_x(f) = \frac{b^2}{\left|1 + \sum_{p=1}^{M} a_p(e^{-i2\pi f})^p\right|^2}. \tag{4}$$

1. Yule-Walker Equations

The autocorrelation sequence $\tilde{R}_x(k)$ for the AR model satisfies a set of linear equations that can be derived as follows. Multiplication of both sides of (1), with n replaced by $n + k$, by x_n and evaluation of the limit time-average value of this product yields the equation (exercise 1)

$$\tilde{R}_x(k) = -a_1\tilde{R}_x(k-1) - a_2\tilde{R}_x(k-2) - \cdots - a_M\tilde{R}_x(k-M) + b\tilde{R}_{zx}(k). \tag{5}$$

Since z_n has zero mean value and x_n depends on only z_n, z_{n-1}, z_{n-2}, . . . (see (1)), then

$$\tilde{R}_{zx}(n) = 0, \qquad n \geq 1. \tag{6a}$$

Multiplication of both sides of (1) by z_n and evaluation of the limit time-average value of this product yields (exercise 1)

$$\tilde{R}_{zx}(0) = -a_1\tilde{R}_{zx}(1) - a_2\tilde{R}_{zx}(2) - \cdots - a_M\tilde{R}_{zx}(M) + b, \tag{7}$$

from which (6a) yields

$$\tilde{R}_{zx}(0) = b. \tag{6b}$$

It follows from (5)–(7) that

$$\tilde{R}_x(k) = -\sum_{p=1}^{M} a_p\tilde{R}_x(k-p), \qquad k \geq 1 \tag{8a}$$

$$\tilde{R}_x(0) = b^2 - \sum_{p=1}^{M} a_p\tilde{R}_x(-p). \tag{8b}$$

If the $M + 1$ values $\{\tilde{R}_x(k) = \tilde{R}_x(-k) : k = 0, 1, 2, 3, \ldots, M\}$ of the autocorrelation of x_n are known, then these $M + 1$ linear equations (for $k = 0, 1, 2, 3, \ldots, M$)

can be solved for the $M + 1$ model parameters $\{a_p : p = 1, 2, 3, \ldots, M\}$ and b^2. Thus, the AR model is fully specified (except for the sign of b, which is irrelevant) by the first $M + 1$ values of $\tilde{R}_x(k)$. Once the model is determined, $\tilde{R}_x(k)$ is fully specified for *all* k. Thus, there must be a way to determine $\{\tilde{R}_x(k) : |k| > M\}$ from $\{\tilde{R}_x(k) : |k| \leq M\}$. In fact (8a) for $k > M$ is a linear recursion that enables $\tilde{R}_x(k)$ to be determined from $\{\tilde{R}_x(j) : |j| < k\}$ for every $k > M$. Of course, this *autocorrelation extrapolation* procedure is valid (i.e., equations (8a)–(8b) are valid) only if $\tilde{R}_x(k)$ truly is the autocorrelation sequence for some Mth-order AR model. Equations (8a)–(8b) are known as the *Yule-Walker equations,* in honor of George Udny Yule's and Gilbert Walker's pioneering work on AR models [Yule 1927; Walker 1931]. They can be expressed jointly as a single matrix equation

$$\begin{bmatrix} \tilde{R}_x(0) & \tilde{R}_x(-1) & \tilde{R}_x(-2) & \cdots & \tilde{R}_x(-M) \\ \tilde{R}_x(1) & \tilde{R}_x(0) & \tilde{R}_x(-1) & \cdots & \tilde{R}_x(1-M) \\ \tilde{R}_x(2) & \tilde{R}_x(1) & \tilde{R}_x(0) & \cdots & \tilde{R}_x(2-M) \\ \vdots & \vdots & \vdots & & \vdots \\ \tilde{R}_x(M) & \tilde{R}_x(M-1) & \tilde{R}_x(M-2) & \cdots & \tilde{R}_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} b^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{8c}$$

or alternatively as

$$\begin{bmatrix} \tilde{R}_x(-1) & \tilde{R}_x(-2) & \tilde{R}_x(-3) & \cdots & \tilde{R}_x(-M) \\ \tilde{R}_x(0) & \tilde{R}_x(-1) & \tilde{R}_x(-2) & \cdots & \tilde{R}_x(1-M) \\ \tilde{R}_x(1) & \tilde{R}_x(0) & \tilde{R}_x(-1) & \cdots & \tilde{R}_x(2-M) \\ \vdots & \vdots & \vdots & & \vdots \\ \tilde{R}_x(M-1) & \tilde{R}_x(M-2) & \tilde{R}_x(M-3) & \cdots & \tilde{R}_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} b^2 - \tilde{R}_x(0) \\ -\tilde{R}_x(1) \\ -\tilde{R}_x(2) \\ \vdots \\ -\tilde{R}_x(M) \end{bmatrix} \tag{8d}$$

2. Levinson-Durbin Algorithm

The Yule-Walker equations can be solved using a particularly efficient recursive algorithm called the *Levinson-Durbin algorithm* [Levinson 1947; Durbin 1960; Wiggins and Robinson 1965]. Because of the fact that an AR model of order M_1 is identical to an AR model of order $M_2 > M_1$ if $a_{M_1+1} = a_{M_1+2} = \cdots = a_{M_2} = 0$, then one need not know the model order M in order to use the Yule-Walker equations to determine the model parameters. One can simply solve (8a) for $M = 1, 2, 3, \ldots$ until $a_p = 0$ for all $p > M$. Evidently, one would in principle have to solve for a_p for $p \to \infty$ to be sure that there are no nonzero a_p for $p > M$, but this problem can often be circumvented in practice. The following Levinson-Durbin algorithm is a computationally efficient recursion for carrying out this solution procedure. Let $\{a_p(M)\}$ and $b(M)$ denote the parameters for an Mth-order AR model. Then the algorithm is initialized by

$$a_1(1) = \frac{-\tilde{R}_x(1)}{\tilde{R}_x(0)} \tag{9a}$$

$$b^2(1) = [1 - a_1^2(1)]\tilde{R}_x(0) \tag{9b}$$

and the Mth step of the recursion is specified by

$$a_M(M) = \frac{-1}{b^2(M-1)}\left[\tilde{R}_x(M) + \sum_{q=1}^{M-1} a_q(M-1)\tilde{R}_x(M-q)\right] \tag{9c}$$

$$b^2(M) = [1 - a_M^2(M)]b^2(M-1) \tag{9d}$$

$$a_p(M) = a_p(M-1) + a_M(M)a_{M-p}(M-1), \qquad p = 1, 2, 3, \ldots, M-1. \tag{9e}$$

For any value of M, the solution $\{a_p(M) : p = 1, 2, 3, \ldots, M\}$ and $b^2(M)$ provided by (9) is identical to the solution provided by (8) for $k = 1, 2, 3, \ldots, M$. Thus at each step, say M, the Levinson-Durbin algorithm provides the solution to the Yule-Walker equations for an AR model of order M (see exercise 1).

3. Linear Prediction

The problem of fitting an AR model to a given autocorrelation sequence $\tilde{R}_x(k)$ is intimately related to the problem of predicting the value x_n of the time-series using the previous M values $\{x_{n-1}, x_{n-2}, \ldots, x_{n-M}\}$. This relationship is explained as follows. Since b^2 must be nonnegative, (9d) indicates that

$$|a_M(M)| \leq 1 \tag{10a}$$

and also that

$$b^2(M) \leq b^2(M-1). \tag{10b}$$

Furthermore, it follows from the AR model (1) that the temporal variance of the model error is equal to b^2,

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \left[x_n - \sum_{p=1}^{M} -a_p(M)x_{n-p}\right]^2 = b^2(M). \tag{11}$$

Hence, the sequence of model-error variances indexed by M is nonincreasing. The term *model-error variance* is used because the quantity

$$bz_n = x_n - \sum_{p=1}^{M} -a_p x_{n-p} \tag{12}$$

from (1) is the error in fitting the time-series x_n to an Mth-order linear regression on its past. The *Mth-order regressor* is denoted by

$$\hat{x}_n \triangleq \sum_{p=1}^{M} -a_p x_{n-p}. \tag{13}$$

It can be shown that the solution to the Yule-Walker equations of order M minimizes this error variance for any time-series x_n. This follows from the fact that the temporal variance (11) is minimum if and only if the *orthogonality condition,*

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} (x_n - \hat{x}_n)x_{n-p} = 0, \qquad p = 1, 2, 3, \ldots, M, \tag{14}$$

is satisfied. This condition can be derived simply by equating to zero the M partial derivatives of the variance (11) with respect to the M parameters $\{a_p\}$. Substitution of (13) into (14) yields the Yule-Walker equations (8a) (exercise 1).

Because the Yule-Walker equations are equivalent to the orthogonality condition and the terms *orthogonal* and *normal* are synonyms, they are also called the *normal equations*. Condition (14) is called an *orthogonality condition* because the limiting time average of the product of time-series can be interpreted as an inner product of vectors [Gardner 1985]. Furthermore, since the regressor $\hat{x}_n$ depends only on values of x_n at times prior to n, then it can be interpreted as a *predictor* of the value x_n. Thus, the solution to the Yule-Walker equations of order M yields the minimum-variance Mth-order linear predictor of the time-series and the minimum value, $b^2(M)$, of the prediction-error variance.

It can be shown (exercise 2) that, for any time-series x_n from a constant phenomenon whose limit spectrum contains no spectral lines, the model-error time-series, bz_n, becomes white, (2), in the limit $M \to \infty$. This reveals that *any* such time-series can be modeled exactly by an AR model, although the order M can be infinite. Nevertheless, since the model-error variance is a nonincreasing function of the model order M, then for any desired degree of accuracy in the fit of the model to the time-series (arbitrarily small difference between the finite-order model-error variance and the minimum error variance $b^2(\infty)$), there is a sufficiently large but finite value for the order M.

4. Wold-Cramér Decomposition

Any zero-mean time-series x_n from a constant phenomenon can be decomposed into two zero-mean time-series that are uncorrelated with each other,

$$x_n = x_n(r) + x_n(s) \tag{15}$$

$$\tilde{R}_{x(r)x(s)}(k) \equiv 0, \tag{16}$$

for which the component $x_n(s)$ is perfectly predictable (zero prediction-error variance) and is mean-square equivalent to a sum of sine waves and for which $x_n(r)$ is mean-square equivalent to the response of a linear, time-invariant, causal, stable system, whose inverse is also causal and stable, to a unity-variance white excitation, say z_n. The component $x_n(s)$ is called *singular,* and the component $x_n(r)$ is called *regular.*

The regular component admits a stable AR model (i.e., (1) with $x_n = x_n(r)$) with transfer function $\tilde{G}(f)$ given by (3) and with order M possibly infinite, and the inverse model is also stable and causal, with transfer function denoted by

$$\tilde{G}^{-1}(f) = \frac{1}{\tilde{G}(f)}$$

and corresponding impulse-response sequence denoted by g_n^{-1}. The white excitation of the inverse model

$$z_n = g_n^{-1} \otimes x_n(r) \tag{17a}$$

is called the *innovations representation* for $x_n(r)$,

$$x_n(r) = g_n \otimes z_n, \tag{17b}$$

because each new value z_n of the time-series is uncorrelated with all prior values $\{z_j : j < n\}$ of the time-series and therefore provides completely new information—an *innovation*. Since g_n is a causal stable sequence, then (17a) reveals that $x_n(r)$

also admits a stable MA model (see (87) in Chapter 3) with transfer function given by

$$\tilde{G}^{-1}(f) = \sum_{q=0}^{L-1} b_q (e^{-i2\pi f})^q \tag{18}$$

in which the order L is possibly infinite and $b_q = g_q^{-1}$. It is easy to show that if L is finite, then M in (3) is infinite, and if M is finite, then L in (18) is infinite. Also, either model can be obtained from the other by polynomial division of its transfer function.

The singular component $x_n(s)$ also admits an AR model, but with $b = 0$. That is, the system is marginally unstable and produces a response consisting of a sum of sine waves, without any excitation (see exercise 9). If there are K sine waves, then the AR model order is $M = 2K$, and the Levinson-Durbin algorithm terminates (with $b(M) = 0$) at this value of M. Unlike the regular component, $x_n(s)$ cannot admit a useful finite-order MA model because the sequence of coefficients g_n consists of a sum of sine waves that do not approach zero as $n \to \infty$.

The spectral density for the singular component consists of Dirac deltas only (a *pure line spectrum*), and the spectral density for the regular component is continuous and is equal to $|\tilde{G}(f)|^2$, which can be expressed as (4), with M possibly infinite. This representation of $\tilde{S}_{x(r)}(f)$ in terms of the product of conjugate factors,

$$\tilde{S}_{x(r)}(f) = \tilde{G}(f)\tilde{G}^*(f), \tag{19}$$

is called a *spectral factorization*. The particular factorization described here, in which both $\tilde{G}(f)$ and $1/\tilde{G}(f)$ are *minimum-phase functions*[1] (that is, they represent causal stable systems) is called the *canonical spectral factorization*. It follows from (17a) that this factorization identifies the *whitening filter* that produces the innovations representation z_n from $x_n(r)$.

The probabilistic counterpart of this decomposition[2] of x_n and its spectral density $\tilde{S}_x(f)$ is called the *Wold-Cramér decomposition* (see [Grenander and Rosenblatt 1984]). However, the Wold-Cramér decomposition is more general than that described here because it applies to nonstationary processes as well. Moreover the Wold-Cramér decomposition is a wide-sense version of a strict sense decomposition known as the *Doob decomposition,* in which the predictor (regressor) is not constrained to be linear (see [Larson and Shubert 1979; Doob 1953]). There is a partially analogous decomposition for continuous time-series (see [Larson and Schubert 1979; Doob 1953]). That is, such time-series can be decomposed into a sum of singular and regular components that are uncorrelated with each other. However, the singular component need not have a pure line spectrum, and therefore the component with continuous spectral density need

[1] If $\tilde{G}(f) = \mathcal{G}(e^{i2\pi f})$, then $\tilde{G}(f)$ is a *minimum-phase function* if $\mathcal{G}(z)$ and $1/\mathcal{G}(z)$ are analytic functions of the complex variable z for $|z| \geq 1$.

[2] Wold presented both the probabilistic and nonprobabilistic versions of this decomposition, which he referred to as the *stochastic* and *functional* decompositions, respectively [Wold 1948].

not be the regular component. Nevertheless, there is a condition on continuous spectral densities that guarantees that a time-series with a spectral density satisfying the condition is regular, and this is called the *Paley-Wiener condition* [Wiener and Paley 1934]. If the Paley-Wiener condition is satisfied for the component with the continuous spectral density, then it is the regular component, and its residual is the singular component and has a pure line spectrum.

5. Maximum-Entropy Model

An interesting question that arises in the practical application of AR models is what interpretation can be given to the Mth-order AR model obtained from the Yule-Walker equations when the only values of the autocorrelation sequence that are known are the $M + 1$ values used in these equations. That is, if the values $\{\tilde{R}_x(k) : k > M\}$ are unknown (or perhaps cannot be reliably estimated), then it cannot be known whether or not x_n is truly an Mth-order AR time-series. There is one particularly intriguing answer to this question. Specifically, it is shown in this section that among all possible time-series that have the same $M + 1$ autocorrelation values $\{\tilde{R}_x(k) : |k| \leq M\}$, the Mth-order AR time-series specified by the Yule-Walker equations is the *most random* in the sense of having maximum relative entropy rate for a given relative entropy rate of its innovations representation.

The *relative entropy rate,* denoted by $\overline{H}_x$, for a time-series x_n is defined by analogy with the probabilistic definition [Shannon and Weaver 1962] to be

$$\overline{H}_x \triangleq \lim_{L\to\infty} \frac{1}{L} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \ln\left\{\frac{1}{f_{x(L)}[\boldsymbol{x}_n(L)]}\right\}, \tag{20}$$

where $f_{x(L)}[\boldsymbol{x}_n(L)]$ is the Lth-order joint fraction-of-time probability density for the vector of L variables $\{x_{k+1}, x_{k+2}, \ldots, x_{k+L}\}$,

$$f_{x(L)}(\boldsymbol{z}) \triangleq \frac{\partial^L}{\partial z_1 \partial z_2 \cdots \partial z_L} \lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} U[z_1 - x_{k+1}]\, U[z_2 - x_{k+2}] \cdots U[z_K - x_{k+K}], \tag{21}$$

evaluated at $\boldsymbol{z} = \boldsymbol{x}_n(L) \triangleq [x_{n+1}, x_{n+2}, \ldots, x_{n+L}]'$. In (21), U is the unit step function, and therefore the summand is the indicator of the joint event

$$x_{k+l} < z_l, \qquad l = 1, 2, 3, \ldots, L.$$

The one-sided (positive time) averages in (20) and (21) are used instead of two-sided averages only because it is more common to consider one-sided time-series in subjects dealing with entropy. The relative entropy rate defined by (20) is a relative measure of the *average uncertainty per time-sample* of the time-series. It can be shown (exercise 3) that the difference between the *entropy rate* $\overline{H}_z$ of the input z_n and the entropy rate $\overline{H}_x$ of the output x_n of a minimum-phase linear time-invariant system with transfer function $\tilde{G}(f)$ is given by

$$\overline{H}_x - \overline{H}_z = \frac{1}{2} \int_{-1/2}^{1/2} \ln[|\tilde{G}(f)|^2]\, df \tag{22}$$

(which is completely analogous to the probabilistic result [Shannon and Weaver

1962]). It can also be shown that the model-error variance for $M \to \infty$ is given by the *Szegö-Kolmogorov formula* [Doob 1953; Grenander and Rosenblatt 1984],

$$b^2(\infty) = \exp\left\{\int_{-1/2}^{1/2} \ln[S_{x(r)}(f)]\, df\right\}. \tag{23}$$

Observe that the singular component has no effect on prediction-error variance (since it is perfectly predictable). It is assumed in the following that the singular component is zero so that the time-series is regular, $x_n = x_n(r)$. By letting z_n be the innovations representation for x_n, it follows from (19), (22), and (23) that the model-error variance for $M \to \infty$ can be expressed as

$$b^2(\infty) = \exp\{2[\overline{H}_x - \overline{H}_z]\}, \tag{24}$$

and it follows from (9d) that for finite-order models the model-error variance is nonincreasing,

$$b^2(M) \geq b^2(M+1) \geq b^2(M+2) \geq \cdots \geq b^2(\infty). \tag{25}$$

Consequently, the time-series with a given Mth-order model-error variance $b^2(M)$ will have maximum relative entropy rate $\overline{H}_x$, for a given relative entropy rate $\overline{H}_z$ of its innovations representation, if and only if

$$b^2(\infty) = b^2(M). \tag{26}$$

It follows from (26) that for a given model-error variance, $b^2(M)$, the time-series x_n exhibits maximum relative entropy rate if and only if the parameters in (9) satisfy

$$a_m(m) = 0, \qquad m > M, \tag{27}$$

in which case x_n is exactly modeled by an Mth-order AR time-series. Furthermore, since $b^2(M)$ is given if $\{\tilde{R}_x(k) : |k| \leq M\}$ is given (see (9)), then the maximum-entropy model for a zero-mean time-series with given $\{\tilde{R}_x(k) : |k| \leq M\}$ is the Mth-order AR model specified by the Levinson-Durbin algorithm, or equivalently the Yule-Walker equations.

Another approach to proving that the AR model specified by the Yule-Walker equations is a maximum-entropy model is to use (22) with the arbitrary transfer function $\tilde{G}(f)$ chosen to be that of the inverse whitening filter for x_n, in which case

$$\tilde{S}_x(f) = |\tilde{G}(f)|^2. \tag{28}$$

Then for a given relative entropy rate of the innovations z_n, the relative entropy rate of x_n will be maximum if and only if the integral

$$I \triangleq \int_{-1/2}^{1/2} \ln[\tilde{S}_x(f)]\, df \tag{29}$$

is maximum. Thus, the power spectral density of the maximum-entropy model is given by the function $\tilde{S}_x(f)$ that maximizes I subject to the constraints

$$\int_{-1/2}^{1/2} \tilde{S}_x(f) e^{i2\pi kf}\, df = \tilde{R}_x(k), \qquad k = 0, 1, 2, \ldots, M. \tag{30}$$

It is shown in exercise 4 that the solution to this constrained optimization problem is of the form (4), which corresponds to an AR model of order M. Thus, the

Yule-Walker equations specified by the $M + 1$ constraint values $\{\tilde{R}_x(k) : k = 0, 1, 2, \ldots, M\}$ yields the parameters $\{a_p(M)\}$ and $b(M)$ that specify the maximum-entropy model.

As an interesting aside, it is mentioned that although the AR model yields the maximum relative entropy rate for x_n only for a given relative entropy rate of the innovations z_n, it can be shown that a white (innovations) time-series has maximum relative entropy rate, for a given average power, if and only if it has a Gaussian fraction-of-time distribution.[3] Since the Gaussian property is preserved by linear transformations, then it follows that the time-series x_n that has maximum relative entropy rate, for a given $\{\tilde{R}_x(k) : |k| \leqslant M\}$, is specified by a Gaussian AR model.

6. Lattice Filter

The Mth-order minimum-variance linear predictor for a time-series x_n can be implemented in the form of a *lattice filter,* which has important properties from an implementation standpoint, especially if the predictor is to be made adaptive (see [Friedlander 1982a; Honig and Messerschmitt 1984]). This lattice-filter implementation can be derived from the Levinson-Durbin algorithm as follows. The Mth-order linear-prediction-error time-series denoted by $e_n(M)$ is given by

$$e_n(M) \triangleq x_n - \hat{x}_n(M), \tag{31}$$

where $\hat{x}_n(M)$ denotes the Mth-order linear predictor, which can be expressed (using (13)) as

$$\hat{x}_n(M) = -a_n(M) \otimes x_n, \tag{32}$$

where $\{-a_n(M) : n = 1, 2, 3, \ldots, M\}$ is interpreted as a finite-impulse-response sequence. Consequently, $e_n(M)$ can be expressed as

$$e_n(M) = c_n(M) \otimes x_n, \tag{33a}$$

where

$$c_n(M) = \delta_n + a_n(M). \tag{33b}$$

Alternatively, $\hat{x}_n(M)$ can be expressed as

$$\hat{x}_n(M) = [\delta_n - c_n(M)] \otimes x_n. \tag{34}$$

Thus, a realization for the prediction filter $\{-a_n(M)\}$ can always be obtained from a realization of the prediction-error filter $\{c_n(M)\}$, and vice versa. A realization of $\{c_n(M)\}$ can be obtained from the Levinson-Durbin algorithm as follows

$$e_n(M) = x_n + \sum_{p=1}^{M} a_p(M)x_{n-p} \tag{35a}$$

$$= x_n + \sum_{p=1}^{M-1} [a_p(M-1) + a_M(M)a_{M-p}(M-1)]x_{n-p} + a_M(M)x_{n-M} \tag{35b}$$

$$= e_n(M-1) + a_M(M)\check{e}_{n-1}(M-1), \tag{35c}$$

[3] This follows directly from the fact that the maximum-entropy distribution for a random variable with a given variance is the Gaussian distribution [Shannon and Weaver 1962].

where

$$\check{e}_n(M) \triangleq x_{n-M} + \sum_{p=1}^{M} a_p(M)x_{n-M+p}. \tag{36}$$

The quantity $\check{e}_n(M)$ is a *backward prediction-error* time-series, that is, it is the error in prediction of x_{n-M}, using $\{x_{n-M+p} : p = 1, 2, 3, \ldots, M\}$. It can be shown (using only the fact that $\bar{R}_x(k)$ is an even sequence) that the minimum-variance backward-prediction filter-coefficients are identical to those for minimum-variance forward prediction. Thus, (36) is the *minimum-variance* backward-prediction-error time-series. Consequently, the Levinson-Durbin algorithm can again be used to show (exercise 5) that

$$\check{e}_n(M) = \check{e}_{n-1}(M - 1) + a_M(M)e_n(M - 1). \tag{37}$$

The pair of joint recursions (35c) and (37) can be implemented directly with the lattice structure shown in Figure 9-1.

One of the greatest advantages of this lattice structure for implementation of a linear predictor is that as the order M is increased or decreased by the addition or deletion of final stages in the lattice, there is no effect on any of the previous stages. That is, their coefficients need not be readjusted. Moreover, it can be shown that the coefficients

$$\rho_p \triangleq -a_p(p) \tag{38}$$

that completely specify the minimum-variance prediction-error lattice filter are identical to the correlation coefficients for $e_n(p - 1)$ and $\check{e}_{n-1}(p - 1)$ [Kay 1987] and are called the PARCOR coefficients for x_n and x_{n-p} because

$$e_n(p - 1) = x_n - \hat{x}_n(p - 1) \tag{39a}$$

$$\check{e}_{n-1}(p - 1) = x_{n-p} - \check{x}_{n-p}(p - 1), \tag{39b}$$

where $\check{x}_{n-p}(p - 1)$ is the $(p - 1)$th-order backward linear predictor of x_{n-p}. Thus, ρ_p is the correlation coefficient for x_n and x_{n-p}, with the correlation of $\{x_{n-1}, x_{n-2}, \ldots, x_{n-p+1}\}$ first removed (see Chapter 7 on partial correlation). This explains why $|a_p(p)| \leq 1$, as concluded earlier from the Levinson-Durbin algorithm. Notice also that it follows directly from (9a), (9b), and (9d) in the Levinson-Durbin algorithm and (38) that the minimum prediction-error variance is easily determined from the PARCOR coefficients,

$$b^2(M) = \left[\prod_{p=1}^{M} (1 - \rho_p^2)\right]\bar{R}_x(0). \tag{40}$$

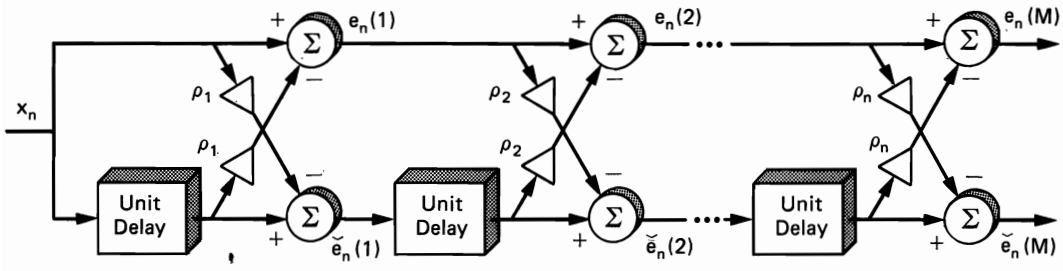


Figure 9-1 Lattice implementation of prediction-error filter.

It should be clarified that the AR model with input e_n (or z_n) and output x_n is an all-pole model, whereas the MA lattice model, with input x_n and output e_n, is an all-zero model, and these models are inverses of each other. Furthermore, it can be shown that the AR model is stable if and only if all coefficients $\{\rho_p\}$ in the lattice model for its inverse have magnitudes less than unity, and it is marginally stable if and only if some coefficients have unity magnitude. The preceding properties and various other properties of lattice models and recursive algorithms for linear prediction all can be given elegant interpretations as well as derivations in terms of the geometrical concept of orthogonal projection (see [Honig and Messerschmitt 1984; Kay 1987], and for the foundations of the geometric theory of random processes and time-series, see [Gardner 1985]).

7. Cholesky Factorization and Correlation Matrix Inversion

Since the Levinson-Durbin algorithm described in Part 2 of this section solves the Yule-Walker equations (8c), then it must in effect invert the correlation matrix

$$\tilde{\boldsymbol{R}}_x \triangleq \begin{bmatrix} \tilde{R}_x(0) & \tilde{R}_x(-1) & \tilde{R}_x(-2) & \dots & \tilde{R}_x(-M) \\ \tilde{R}_x(1) & \tilde{R}_x(0) & \tilde{R}_x(-1) & \dots & \tilde{R}_x(1-M) \\ \tilde{R}_x(2) & \tilde{R}_x(1) & \tilde{R}_x(0) & \dots & \tilde{R}_x(2-M) \\ \vdots & \vdots & \vdots & & \vdots \\ \tilde{R}_x(M) & \tilde{R}_x(M-1) & \tilde{R}_x(M-2) & \dots & \tilde{R}_x(0) \end{bmatrix} \tag{41}$$

As a matter of fact, the inverse of this matrix can easily be obtained from the results of the Levinson-Durbin algorithm. Specifically, let $\{a_p(m) : p = 1, 2, \dots, m\}$, denote the model parameters for the mth-order AR model specified by the Yule-Walker equations for $m = 1, 2, 3, \dots, M$, and define a lower-triangular matrix by

$$\boldsymbol{A} \triangleq \begin{bmatrix} 1 & & & & \\ a_1(M) & 1 & & & \\ a_2(M) & a_1(M-1) & 1 & & \\ a_3(M) & a_2(M-1) & a_1(M-2) & & \\ \vdots & \vdots & \vdots & & \\ a_M(M) & a_{M-1}(M-1) & a_{M-2}(M-2) & \dots & 1 \end{bmatrix} \tag{42}$$

Also, let $\{b^2(m) : m = 0, 1, 2, \dots, M\}$ denote the other AR model parameter (model-error variance), and define a diagonal matrix by

$$\boldsymbol{B} \triangleq \begin{bmatrix} b(M) & & & & \\ & b(M-1) & & & \\ & & b(M-2) & & \\ & & & \ddots & \\ & & & & b(0) \end{bmatrix} \tag{43}$$

With these definitions, the Yule-Walker equations can be used to show (exercise 6) that

$$\tilde{\boldsymbol{R}}_x^{-1} = \boldsymbol{A}\,\boldsymbol{B}^{-2}\boldsymbol{A}' \tag{44}$$

and, therefore,

$$\tilde{\boldsymbol{R}}_x^{-1} = \boldsymbol{L}\,\boldsymbol{L}', \tag{45}$$

where $\boldsymbol{L}$ is the lower-triangular matrix

$$\boldsymbol{L} = \boldsymbol{A}\,\boldsymbol{B}^{-1} = \begin{bmatrix} \dfrac{1}{b(1)} & & & & \\ \dfrac{a_1(M)}{b(1)} & \dfrac{1}{b(2)} & & & \\ \dfrac{a_2(M)}{b(1)} & \dfrac{a_1(M-1)}{b(2)} & \dfrac{1}{b(3)} & & \\ \vdots & \vdots & & \ddots & \\ \dfrac{a_M(M)}{b(1)} & \dfrac{a_{M-1}(M-1)}{b(2)} & \cdots & & \dfrac{1}{b(M)} \end{bmatrix} \tag{46}$$

and $\boldsymbol{L}'$ is its transpose, which is an upper-triangular matrix. It follows from (44) that the correlation matrix itself is given by

$$\tilde{\boldsymbol{R}}_x = \boldsymbol{L}^{-\prime}\boldsymbol{L}^{-1}, \tag{47}$$

where $\boldsymbol{L}^{-\prime}$ denotes the inverse of the transpose of $\boldsymbol{L}$ and is an upper-triangular matrix and $\boldsymbol{L}^{-1}$ is a lower triangular matrix. The decomposition (45) (as well as (47)) of a matrix into the product of upper- and lower-triangular matrices is called a *Cholesky factorization* [Lawson and Hanson 1974]. Thus, the Levinson-Durbin recursion not only provides the inverse of the correlation matrix, but it also provides its Cholesky factorization (see [Friedlander 1982a]). Formula (45)–(46) is equivalent to the original Gohberg-Semencul formula (see [Kailath et al. 1978]).

C. AUTOREGRESSIVE METHODS

1. Introduction

It follows from the Wold-Cramér decomposition described in Part 4 of Section B that a time-series that contains no finite additive sine waves can be exactly represented by either a stable moving average (MA) model

$$x_n = b_0 w_n + b_1 w_{n-1} + b_2 w_{n-2} + \cdots + b_L w_{n-L} \tag{48}$$

or a stable autoregressive (AR) model

$$x_n + a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_M x_{n-M} = w_n, \tag{49}$$

where w_n is white noise with variance σ_w^2. The limit spectrum for the MA model is

$$\tilde{S}_x(f) = \sigma_w^2 \left| \sum_{q=0}^{L} b_q (e^{-i2\pi f})^q \right|^2 \tag{50}$$

and for the AR model it is

$$\tilde{S}_x(f) = \frac{\sigma_w^2}{\left| 1 + \sum_{p=1}^{M} a_p (e^{-i2\pi f})^p \right|^2}. \tag{51}$$

For a given time-series, either of the model orders L or M or both can be infinite. If one is finite, the other must be infinite. If a close approximation (or exact representation) of the MA type with low order L exists, then a close approximation (or exact representation) of the AR type must necessarily be of high order, $M >> L$, and vice versa. Also, a time-series that cannot be closely approximated by either a low-order MA model or a low-order AR model can in some instances be closely approximated (or exactly represented) by a low-order ARMA model

$$\begin{aligned} x_n + a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_M x_{n-M} \\ = b_0 w_n + b_1 w_{n-1} + b_2 w_{n-2} + \cdots + b_L w_{n-L}, \end{aligned} \tag{52}$$

which has limit spectrum

$$\bar{S}_x(f) = \frac{\sigma_w^2 \left| \sum_{q=0}^{L} b_q (e^{-i2\pi f})^q \right|^2}{\left| 1 + \sum_{p=1}^{M} a_p (e^{-i2\pi f})^p \right|^2}. \tag{53}$$

(Observe that for a given time-series, L in (52) and (53) cannot be the same as L in (48) and (50) unless $M = 0$, and M in (52) and (53) cannot be the same as M in (49) and (51) unless $L = 0$.) These general observations reveal the practical importance of the choice of model type when it is desired to fit a model to data—the primary reason for this being that model-fitting methods typically work best when the number of parameters in the model is minimal, and this is especially crucial when the amount of data available for model-fitting is limited. Generally speaking, reliable model fitting requires a number of data points N that greatly exceeds the number of parameters required for a good fit. Therefore, small N requires small L for an MA model, small M for an AR model, and small L and M for an ARMA model.

In order to determine what types of limit spectra are most easily modeled by the three model types, we observe that the MA model can produce $L/2$ nulls or very low valleys for positive frequencies at the roots on or very near the unit circle in the complex plane of its Lth-order polynomial transfer function

$$\mathcal{H}(z) = \sum_{q=0}^{L} b_q z^{-q}. \tag{54}$$

The AR model can produce $M/2$ infinite spikes or very high peaks for positive frequencies at the roots on or very near the unit circle of its Mth-order polynomial reciprocal-transfer function,

$$\frac{1}{\mathcal{H}(z)} = 1 + \sum_{p=1}^{M} a_p z^{-p}. \tag{55}$$

The ARMA model can produce both $L/2$ valleys and $M/2$ peaks (exercise 8).

Model fitting can be used for statistical spectral analysis in a straightforward way. Once a model has been fit to a time-series of data, the model-parameter estimates ($\{\hat{b}_q\}$ and $\hat{\sigma}_w$ for MA, $\{\hat{a}_p\}$ and $\hat{\sigma}_w$ for AR, or $\{\hat{b}_q\}$, $\{\hat{a}_p\}$, and $\hat{\sigma}_w$ for ARMA) are simply substituted into the limit spectrum formula for the model ((50) for MA, (51) for AR, or (53) for ARMA), and the resultant function is taken

as the spectrum estimate. The difficulty thus lies only in the model-fitting procedures. Unfortunately, many MA and ARMA model-fitting procedures require highly nonlinear iterative algorithms with various undesirable characteristics, such as potential convergence to local rather than global maxima and potential divergence of maximum likelihood methods or potentially poor performance of more approximate maximum likelihood and other methods of suboptimum model fitting. These problems often discourage the user from adopting MA and ARMA model-fitting approaches in favor of other approaches, such as direct methods and AR model-fitting methods. As a matter of fact, the direct methods all produce spectrum estimates of the MA type (with the exception of the minimum leakage method, which produces spectrum estimates of the AR and ARMA types). It is shown in exercise 10 that for the temporally smoothed periodogram (Bartlett-Welch method), the model order is $L = T \cong 1/\Delta f$, and for the spectrally smoothed periodogram (Wiener-Daniell method) the model order is exactly $L = \Delta t$ but is effectively only $L \cong 1/\Delta f$ because the coefficients $\{b_q : |q| > 1/\Delta f\}$ are relatively small in magnitude.

As a consequence of the foregoing facts, MA model-fitting methods for spectral analysis are primarily of interest in two types of situations: 1) those where highly accurate determination of spectral nulls is important enough to merit relatively complicated algorithms and 2) those where highly accurate determination of spectral nulls as well as spikes is important, in which case AR modeling methods can be (but are not necessarily) inadequate and ARMA methods (which typically involve both AR and MA methods) can therefore be desirable. There are also some other situations where ARMA model-fitting methods can be of interest. For example, time-series consisting of either the sum of two independent AR time-series or the sum of an AR signal time-series and a broadband (white) noise time-series result in ARMA time-series with equal AR and MA orders, $L = M$. Some suboptimum but computationally manageable methods for ARMA model fitting have recently been proposed for such time-series.

In contrast to the relatively limited attractiveness of MA and ARMA parametric methods of spectral analysis, AR methods are very attractive in applications where low-order (and even some relatively high-order) AR models are appropriate. The primary reason is twofold: 1) many AR methods require only linear algorithms that are not computationally intensive and 2) AR methods can produce spectral resolution performance that is superior to that of direct methods, especially for spectra with multiple spikes, or at least dominant peaks, and short data segments. Motivated by the relative attractiveness of AR methods, the treatment of model-fitting methods presented here is limited primarily to those based on the AR model. However, a brief discussion of the ARMA model-fitting problem, which involves both MA and AR techniques, is given in Section D.

The fundamental difference between the resolution capabilities of direct methods and AR model-fitting methods can be seen by considering the fact that the narrowest peak resolvable with a direct method (without sequential sine wave removal) using N data points is on the order of $1/N$, and this can be attained only when no averaging (to reduce random effects) is done, whereas there is no fundamental limit on the resolvable width of a peak with an AR method since

the resolution width for a given spectral peak is simply the reciprocal of the distance between the location of the corresponding root of the polynomial (55) for the model and the unit circle in the complex plane. This difference in resolution capabilities is illustrated in Figures 9-2 and 9-3, where the *ideal spectra obtained from exact values of the limit autocorrelation* $\tilde{R}_x$ *for a limited number of lag values* are shown for the Blackman-Tukey (BT) direct method with triangle window ((143) in Chapter 6), the minimum leakage (ML) methods ((52) and (56) in Chapter 6), and the Yule-Walker (YW) autoregressive method ((51) and (8)). In Figure 9-2, the autocorrelation

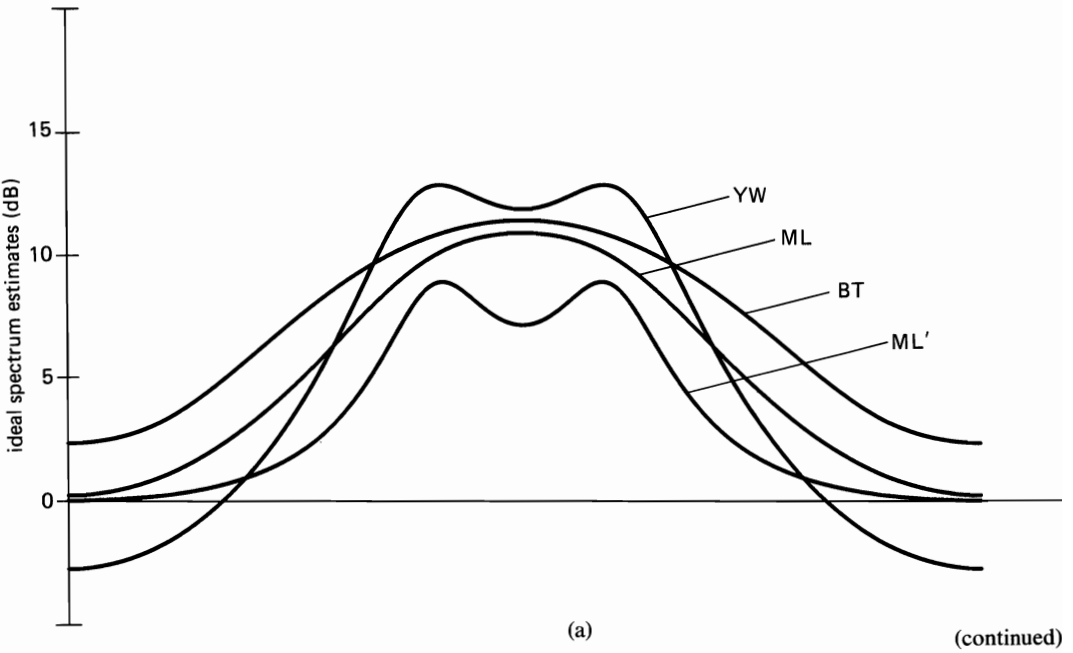
$$\tilde{R}_x(k) = \delta(k) + A\cos(\omega_1 k) + B\cos(\omega_2 k) \tag{56}$$

for the sum of two sine waves in additive white noise is used, and in Figure 9-3 the autocorrelation

$$\begin{aligned}\tilde{R}_x(k) = {} & Ae^{-\alpha|k|}\left[\cos(\omega_1 k) + \left(\frac{\alpha}{\omega_1}\right)\sin(\omega_1|k|)\right] \\ & + Be^{-\beta|k|}\left[\cos(\omega_2 k) + \left(\frac{\beta}{\omega_2}\right)\sin(\omega_2|k|)\right]\end{aligned} \tag{57}$$

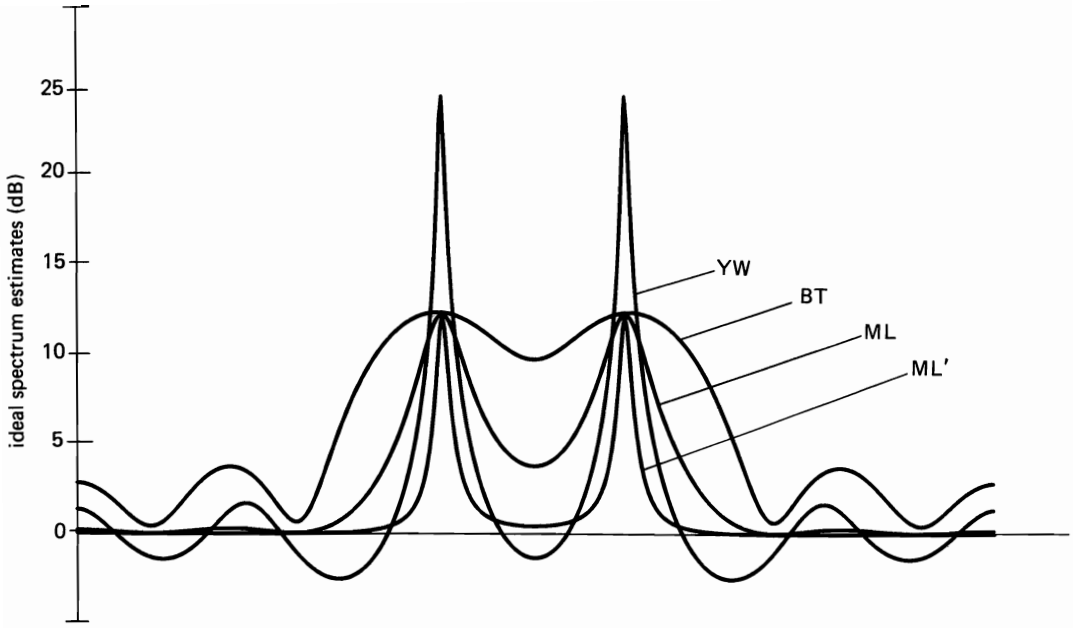
for a fourth-order ARMA time-series, obtained from adding two uncorrelated second-order AR time-series together, is used. The parameter values used are

$$\omega_1 = \frac{2\pi}{5},\ \omega_2 = \frac{3\pi}{5},\ A = B = 2\sqrt{2},\ \alpha = \frac{1}{50},\ \beta = \frac{1}{25}.$$

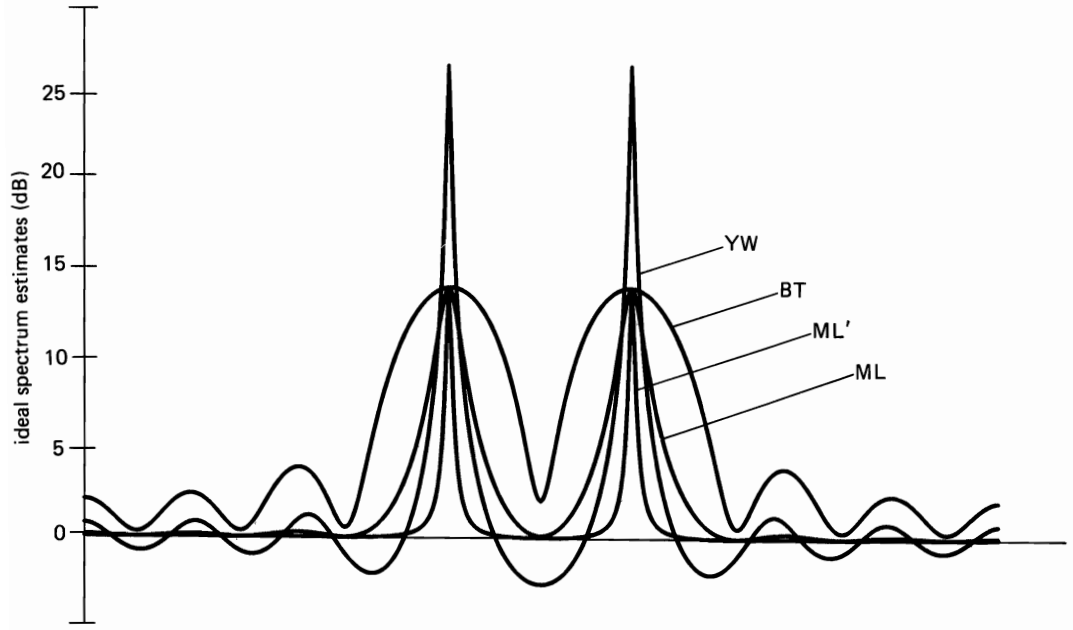


(continued)

Figure 9-2 Different spectrum estimates for $0 \leq f \leq 1/2$ obtained using exact values of the limit autocorrelation $\tilde{R}_x(k) = \delta(k) + 2\sqrt{2}[\cos(\omega_1 k) + \cos(\omega_2 k)]$: (a) $|k| \leq 6$. (b) $|k| \leq 12$. (c) $|k| \leq 18$.

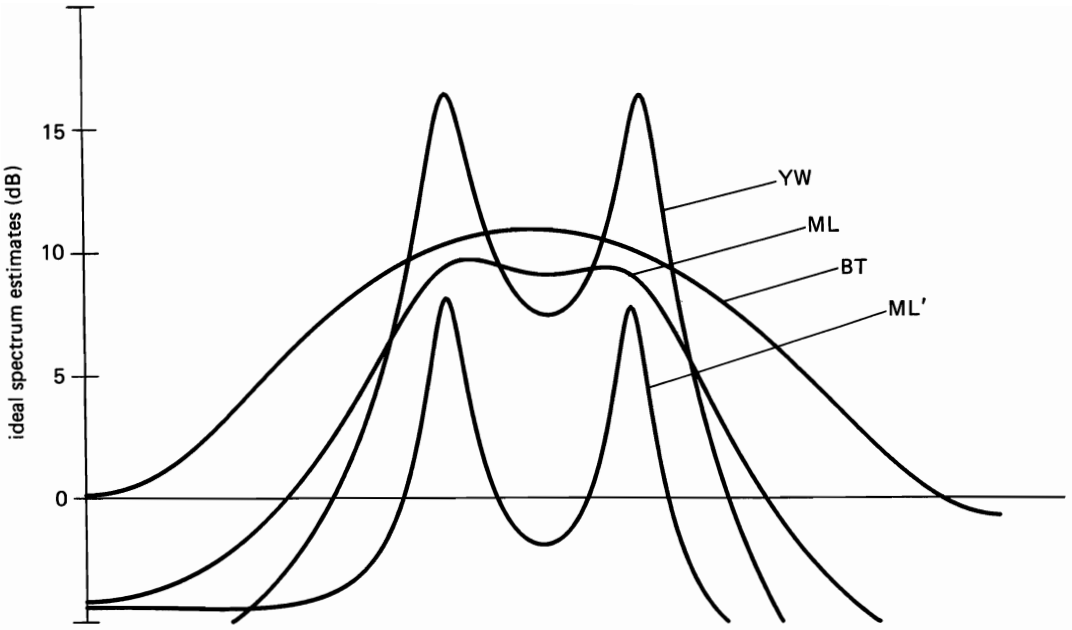


(b)

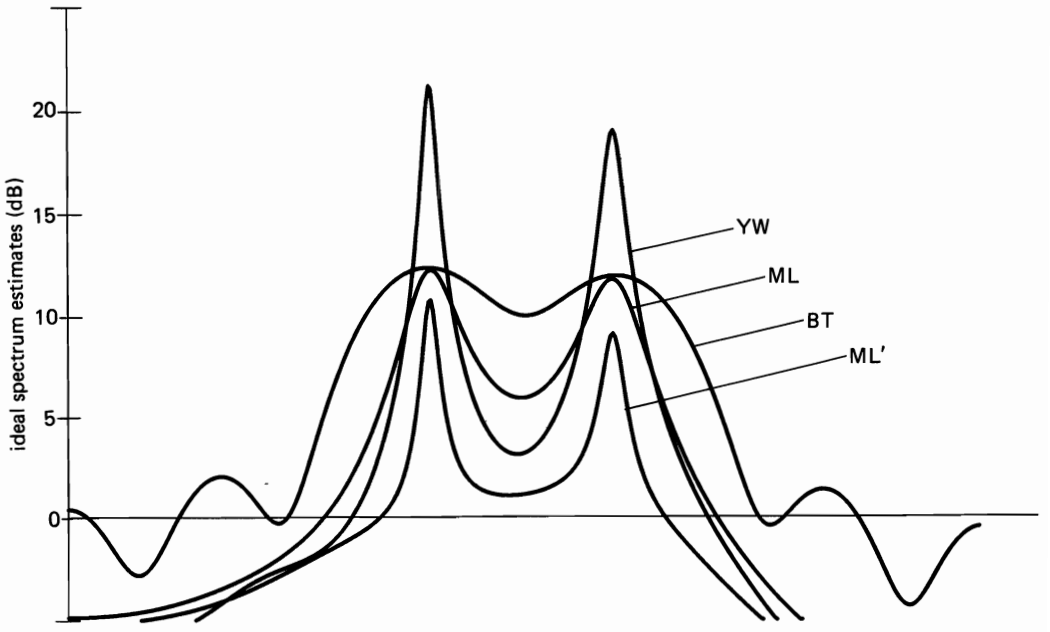


(c)

Figure 9-2 (continued)



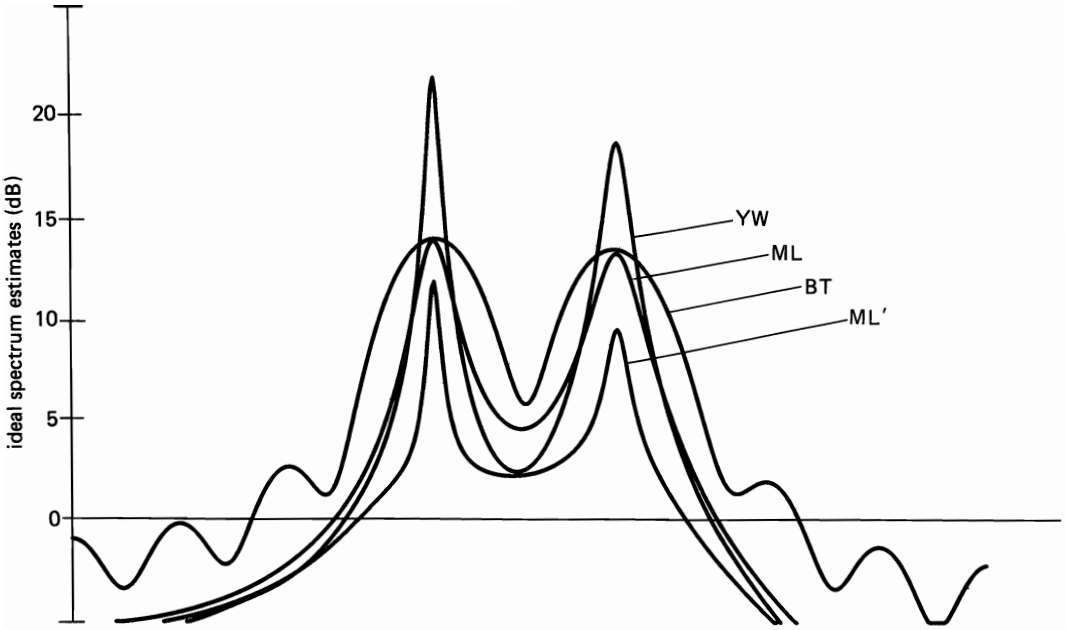
(a)



(b)

(continued)

Figure 9-3 Different spectrum estimates for $0 \leq f \leq 1/2$ obtained using exact values of the limit autocorrelation $\tilde{R}_x(k) = 2\sqrt{2}\{e^{-\alpha|k|}[\cos(\omega_1 k) + (\alpha/\omega_1)\sin(\omega_1 k)] + e^{-\beta|k|}[\cos(\omega_2 k) + (\beta/\omega_2)\sin(\omega_2 k)]\}$: (a) $|k| \leq 6$. (b) $|k| \leq 12$. (c) $|k| \leq 18$.



(c)

Figure 9-3 (continued)

In spite of the theoretically unlimited capability for resolving $M/2$ peaks with an Mth-order AR spectrum estimate, there is still a fundamental trade-off between resolution and reliability. Specifically, it can be shown [Berk 1974] that for zero-mean Gaussian data, the least squares procedures for AR model-fitting lead to an asymptotic ($N \to \infty$, $M \to \infty$) variance given by

$$\text{var}\{S_x(f)_{\text{AR}}\} \cong \frac{2M}{N}[\tilde{S}_x(f)]^2. \tag{58}$$

Thus, like the direct methods, the variance is inversely proportional to $\Delta t \cong N$; but unlike the direct methods, the variance is proportional to the number $M/2$ of resolvable peaks, rather than the reciprocal of the resolution width Δf. Also, in spite of the theoretical resolution limitation of the direct methods that simply time-average or frequency-smooth the periodogram, more sophisticated use of the DFT (e.g., sequential removal of sine wave component estimates; see exercise 22) can in fact produce much higher resolution of spectral peaks as illustrated in Section E. These observations reveal that even when high resolution is a major objective, the choice between methods based on the DFT and those based on AR model fitting is not clear-cut. It is also important to recognize that comparison of *ideal* spectra as in Figures 9-2 and 9-3, although of some utility, can be a misleading indication of comparative performance of corresponding spectrum estimators that use random data. For example, it is shown in Section E that the resolution capability of the ML method is not necessarily inferior to that of the YW method when variability effects are taken into account.

2. Least Squares Procedures

We know from the theory of AR modeling (Section B) that any time-series can be arbitrarily closely modeled by an AR model of sufficiently large but finite order M and that for an Mth-order AR time-series (49), the limit spectrum is given by (51), where the parameters $\{a_p\}_1^M$ minimize the prediction-error variance

$$\min_{\{a_p\}} \langle [x_n - \hat{x}_n]^2 \rangle = b^2, \tag{59a}$$

where

$$\hat{x}_n = -\sum_{p=1}^{M} a_p x_{n-p} \tag{59b}$$

$$b^2 = \sigma_w^2. \tag{59c}$$

This suggests the approach to spectrum estimation of using a finite number, say K, of time-samples of the prediction error to estimate the prediction-error variance and then minimizing this estimated variance with respect to the model parameters:

$$\min_{\{\hat{a}_p\}} \frac{1}{K} \sum_{n=0}^{K-1} w_n^2 = \hat{b}^2, \tag{60a}$$

where

$$w_n \triangleq x_n - \hat{x}_n \tag{60b}$$

$$\hat{x}_n \triangleq -\sum_{p=1}^{M} \hat{a}_p x_{n-p}. \tag{60c}$$

The solution $\{\hat{a}_p\}_1^M$ yields what is called the Mth-order *least squares* (LS) *linear predictor* (60c) and the corresponding LS AR spectrum estimate,

$$S_x(f)_{\text{AR}} \triangleq \frac{\hat{b}^2}{\left|1 + \sum_{p=1}^{M} \hat{a}_p (e^{-i2\pi f})^p\right|^2}. \tag{61}$$

The solution can be obtained from the necessary and sufficient *orthogonality condition*

$$\sum_{n=0}^{K-1} w_n x_{n-q} = 0, \qquad q = 1, 2, 3, \ldots, M, \tag{62}$$

which results from equating to zero the derivatives of the sum of squared errors in (60) with respect to the M parameters $\{\hat{a}_p\}_1^M$ (exercise 11). Substitution of (60b) for w_n into (62) yields

$$\sum_{n=0}^{K-1} \left(x_n + \sum_{p=1}^{M} \hat{a}_p x_{n-p}\right) x_{n-q} = 0, \qquad q = 1, 2, 3, \ldots, M, \tag{63}$$

which can be reexpressed as

$$\sum_{p=1}^{M} R_x(q, p)_K \hat{a}_p = -R_x(q, 0)_K, \qquad q = 1, 2, 3, \ldots, M, \tag{64}$$

where

$$R_x(q, p)_K \triangleq \frac{1}{K} \sum_{n=0}^{K-1} x_{n-p} x_{n-q} = R_x(p, q)_K. \tag{65}$$

The M equations (64) are the finite-data counterparts of the normal equations discussed in Part 3 of Section B (and are also called *normal equations*). In practice, only a finite segment $\{x_0, x_1, x_2, \ldots, x_{N-1}\}$ of data is available, and there are several ways to deal with this in (65), each of which gives rise to a distinct LS AR spectrum estimate. These are described in the following. For all methods it is required that $M < N$, at least, but the methods are often most useful for $M \ll N$.

The linear prediction equation (60c) can be expressed in matrix form as follows (for $K \geq N + M$):

$$\begin{bmatrix} 0 & & & \\ x_0 & & & \\ x_1 & x_0 & & \\ x_2 & x_1 & x_0 & \\ \vdots & & \ddots & \\ \hline x_{M-1} & & \cdots & x_0 \\ \vdots & & & \vdots \\ x_{N-2} & & \cdots & x_{N-M-1} \\ \hline x_{N-1} & & \cdots & x_{N-M} \\ & & \ddots & \vdots \\ & & & x_{N-1} \end{bmatrix} \begin{bmatrix} -\hat{a}_1 \\ -\hat{a}_2 \\ -\hat{a}_3 \\ \vdots \\ -\hat{a}_M \end{bmatrix} = \begin{bmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_{M-1} \\ \hline \hat{x}_M \\ \vdots \\ \hat{x}_{N-1} \\ \hline \hat{x}_N \\ \vdots \\ \hat{x}_{N+M-1} \end{bmatrix} \tag{66}$$

Only the predictors for the time points $n = 0, 1, 2, 3, \ldots, N + M - 1$ are included in (66) because these are the only times at which data is available. In fact the predictions $\hat{x}_n$ for $n = 0, 1, 2, 3, \ldots, M - 1$ are not truly of order M due to lack of data at the beginning; the same is true for $n = N + 1, N + 2, N + 3, \ldots, N + M - 1$ at the end. It can be argued that the resultant prediction filter specified by the solution $\{\hat{a}_p\}_1^M$ to (66) might be more appropriate if these $2M$ lower-order predictors were not included in the estimate of error variance to be minimized, (60a). This can be accomplished by changing the range of summation in (60a) and therefore in (65) from $[0, K - 1]$ to $[M, K - 1]$ and choosing $K = N + 1$. Also, since the Nth predictor $\hat{x}_N$ has no corresponding datum value x_N with which to determine the error w_N, then it can be argued that $K = N$ is a better choice. This corresponds to retaining only the portion of the matrices within the dashed lines in (66). Because the resultant matrix $R_x(q, p)_K$ is symmetric and nonnegative definite, it has all the properties of a covariance matrix, and this LS method is therefore often called the *covariance method*. In this case, $M < N/2$ is required for the existence of the inverse of the covariance matrix.

As an alternative to the covariance method, if all predictors in (66) are retained, then it can be shown (exercise 12) that (65) reduces (for $K \geq N + M$)

to

$$R_x(q, p)_K = \frac{1}{K} \sum_{n=\max\{p,q\}}^{N-1+\min\{p,q\}} x_{n-p} x_{n-q}$$
$$= \frac{1}{K} \sum_{n=0}^{N-1-|p-q|} x_{n+|p-q|} x_n \triangleq R_x(q - p)_K. \tag{67}$$

Because this matrix is symmetric, Toeplitz, and nonnegative definite, it has all the properties of a matrix obtained from an autocorrelation sequence (e.g., $\bar{R}_x(q, p) \triangleq \bar{R}_x(q - p)$), and this LS method is therefore often called the *autocorrelation method*. With (67) substituted into (64), we obtain the particular form of normal equations called the Yule-Walker equations (for finite data) as described in Part 1 of Section B, although only the first part, (8a), of the Yule-Walker equations is obtained here; nevertheless the second part, (8b), also arises, as explained later. Consequently, the autocorrelation method is also called the *Yule-Walker* (YW) *method*.

Now that we have methods for obtaining the parameter estimates $\{\hat{a}_p\}_1^M$ in the spectrum estimate (61), we must determine methods for obtaining the additional parameter estimate $\hat{b}^2$ in (61), which represents the intensity of the spectrum estimate. This parameter estimate is simply the value of the minimized sum of squared prediction errors (60a), and it can be shown (exercise 11) by using the orthogonality condition (62) that

$$\hat{b}^2 \triangleq \frac{1}{K} \sum_{n=0}^{K-1} w_n^2 = \frac{1}{K} \sum_{n=0}^{K-1} w_n x_n. \tag{68}$$

Substitution of (60b) and (60c) into (68) yields the desired result,

$$\hat{b}^2 = R_x(0, 0)_K + \sum_{p=1}^{M} \hat{a}_p R_x(0, p)_K. \tag{69}$$

For the autocorrelation method, this reduces to the other part (8b) of the (finite-data) Yule-Walker equations. Consequently, the parameter estimates $\{\hat{a}_p\}_1^M$ and $\hat{b}$ in the LS spectrum estimate (61) using the autocorrelation method can be obtained using the efficient Levinson-Durbin algorithm described in Part 2 of Section B. As a matter of fact, there is an efficient generalized Levinson recursion [Morf et al. 1977; Honig and Messerschmitt 1984; Marple 1987] that applies to the covariance method so that both methods possess efficient implementations.

Before proceeding, it is mentioned that an interesting interpretation of the AR spectrum estimate provided by the YW method can be obtained from the fact that the AR model specified by the Yule-Walker equations is a maximum-entropy model, as explained in Part 5 of Section B. Consequently, if the finite set of correlogram values (67) used in the finite-data Yule-Walker equations are interpreted as values of the limit autocorrelation, then the AR spectrum estimate can be interpreted as a maximum-entropy (ME) spectrum estimate. The maximum-entropy interpretation of AR spectrum estimation was first promoted by John P. Burg, who also exploited the connection with least squares prediction [Burg 1967]. Another interesting interpretation of the AR spectrum estimate provided

by the YW method can be obtained from the fact that the Yule-Walker equations extrapolate the limit autocorrelation, as explained in Part 1 of Section B. Consequently, the AR spectrum estimate can be interpreted as the FST of an extrapolated version of the correlogram. This interpretation reveals the fact that AR methods of spectral analysis avoid the leakage effects resulting from the actual or effective autocorrelation windowing operation that occurs in the direct methods. This advantage was apparently first explained in [Burg 1967].

It is also interesting that in the limit as the amount of data N approaches infinity, both the autocovariance and autocorrelation matrices equal the limit autocorrelation matrix $\tilde{R}_x(p - q)$, and therefore both methods yield the same spectrum estimate, (61), in the limit. Also, as the order M approaches infinity ($M < N$), this spectrum estimate becomes the exact limit spectrum $\tilde{S}_x(f)$. (If the time-series is truly of the AR type with order M_0, then only $M = M_0$ is needed.) Although these asymptotic properties are comforting, they are not particularly helpful in the case of primary concern in practice, namely, that of small N and, therefore, small M.

A primary source of poor performance of LS AR spectrum estimates in practice is an insufficient amount of data (small N). But some improvement can be made by using the available data more efficiently. One method for accomplishing this is based on the fact that for data from a time-invariant phenomenon, the ideal (infinite data) forward and backward prediction filters are identical. Consequently, we can double the number of prediction errors used to estimate the error variance to be minimized. That is, we can minimize the following sum of squared errors:

$$\hat{b}^2 \triangleq \frac{1}{K} \sum_{n=0}^{K-1} (w_n^2 + \breve{w}_n^2), \tag{70a}$$

where

$$\breve{w}_n \triangleq x_{n-M} - \breve{x}_{n-M} \tag{70b}$$

is the backward prediction error and

$$\breve{x}_{n-M} \triangleq -\sum_{p=1}^{M} \hat{a}_p x_{n-M+p} \tag{70c}$$

is the backward predictor. It can be shown (exercise 13) that the LS prediction parameters are specified by the linear equations

$$\sum_{p=1}^{M} [R_x(q, p)_K + \breve{R}_x(q, p)_K]\hat{a}_p = -[R_x(q, 0)_K + \breve{R}_x(q, 0)_K], \qquad q = 1, 2, 3, \ldots, M, \tag{71}$$

where $R_x(q, p)_K$ is given by (65) and

$$\breve{R}_x(q, p) \triangleq \frac{1}{K} \sum_{n=0}^{K-1} x_{n-M+p} x_{n-M+q} = \breve{R}_x(p, q)_K. \tag{72}$$

This method of spectrum estimation is called the *forward-backward* (FB) LS

method. In the autocorrelation version of this method, $R_x(p, q)_K$ is given by (67) and (exercise 12)

$$\begin{aligned}\check{R}_x(p, q)_K &= \frac{1}{K} \sum_{n=\max\{M-p,M-q\}}^{N-1+\min\{M-p,M-q\}} x_{n-M+p}x_{n-M+q} \\ &= \frac{1}{K} \sum_{n=0}^{N-1-|p-q|} x_{n+|p-q|}x_n \equiv R_x(p, q)_K,\end{aligned} \tag{73}$$

and therefore (71) is identical to (64); this method is thus identical to the forward-only autocorrelation method. Consequently, it is only the covariance version of this method that is of interest. In this case it can be shown that

$$\begin{aligned}R_x(p, q)_K + \check{R}_x(p, q)_K &= \frac{1}{K} \sum_{n=M}^{N-1} x_{n-p}x_{n-q} + \frac{1}{K} \sum_{n=M}^{N-1} x_{n-M+p}x_{n-M+q} \\ &= \frac{1}{K} \sum_{n=0}^{N-M-1} x_{n+M-p}x_{n+M-q} + \frac{1}{K} \sum_{n=0}^{N-1-M} x_{n+p}x_{n+q} \\ &\triangleq \mathring{R}_x(p, q)_N = \mathring{R}_x(q, p)_N.\end{aligned} \tag{74}$$

Also, it can be shown (exercise 13) that the minimized sum of squared errors is given by

$$\hat{b}^2 = \frac{1}{K} \sum_{n=0}^{K-1} [w_n x_n + \check{w}_n x_{n-M}]. \tag{75}$$

Substitution of (60b), (60c), (70b), and (70c) into (75) yields the explicit result

$$\hat{b}^2 = R_x(0, 0)_K + \check{R}_x(0, 0)_K + \sum_{p=1}^{M} \hat{a}_p[R_x(0, p)_K + \check{R}_x(0, p)_K]. \tag{76}$$

Equations (71) and (76) together with (74) can be combined into a single matrix equation

$$\begin{bmatrix} \mathring{R}_x(0, 0)_N & \mathring{R}_x(0,1)_N & \mathring{R}_x(0, 2)_N & \dots & \mathring{R}_x(0, M)_N \\ \mathring{R}_x(1, 0)_N & \mathring{R}_x(1, 1)_N & \mathring{R}_x(1, 2)_N & \dots & \mathring{R}_x(1, M)_N \\ \mathring{R}_x(2, 0)_N & \mathring{R}_x(2, 1)_N & \mathring{R}_x(2, 2)_N & \dots & \mathring{R}_x(2, M)_N \\ \vdots & \vdots & \vdots & & \vdots \\ \mathring{R}_x(M, 0)_N & \mathring{R}_x(M, 1)_N & \mathring{R}_x(M, 2)_N & \dots & \mathring{R}_x(M, M)_N \end{bmatrix} \begin{bmatrix} 1 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_M \end{bmatrix} = \begin{bmatrix} \hat{b}^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{77}$$

where the definition (74) of $\mathring{R}_x(p, q)_N$ is extended from $p, q = 1, 2, 3, \ldots, M$ to $p, q = 0, 1, 2, \ldots, M$. For the existence of the inverse of the matrix $\mathring{R}_x$, it is required that $M < N/3$. Efficient algorithms for solution of these linear equations are available [Dickinson and Turner 1979] (for $M < 50$), [Marple 1980] (for $M > 50$). The FB LS method was originally proposed independently in [Ulrych and Clayton 1976] and [Nuttall 1976].

In summary, for the autocorrelation (or YW) LS method, the parameter estimates $\hat{b}^2$ and $\{\hat{a}_p\}_1^M$ in the spectrum estimate (61) can be obtained by solving the normal equations (64) and (69) using the autocorrelation sequence estimate (67), and this can be accomplished with the Levinson-Durbin algorithm (9) using

the estimate (67) in place of the limit autocorrelation sequence $\tilde{R}_x$. For the covariance LS method, the parameter estimates in (61) can be obtained by solving the normal equations (64) and (69) using the covariance matrix estimate (65) with $[0, K-1]$ replaced with $[M, K-1]$ and $K = N$. For the FB LS method, the parameter estimates in (61) can be obtained by solving the normal equations (64) and (69) using the matrix (74) in place of that specified by (65). This is summarized in (77).

A variation that predates the FB LS method and is perhaps the most popular of the AR methods is called the *Burg method* in honor of its originator John P. Burg [Burg 1968]. The Burg method is based on constrained least squares forward-backward prediction. That is, the AR parameters are obtained by minimization of the sum of squared errors

$$\hat{b}^2(M) = \frac{1}{2(N-M)} \sum_{n=M}^{N-1} (w_n^2 + \breve{w}_n^2) \tag{78}$$

as in the FB LS method but subject to the constraint that the parameter estimates satisfy the recursion

$$\hat{a}_p(M) = \hat{a}_p(M-1) + \hat{a}_M(M)\hat{a}_{M-p}(M-1), \qquad p = 1, 2, 3, \ldots, M-1. \tag{79}$$

The notation $\hat{a}_p(M)$ denotes the estimate of the pth parameter in the Mth-order model. The recursion (79) arises in the Levinson-Durbin algorithm described in Part 2 of Section B and is automatically satisfied by the AR parameters obtained using the YW (autocorrelation) method. It is introduced here as a constraint in order to ensure that the AR parameter estimates will yield a stable filter (roots of (55) inside the unit circle). The limits on the sum in (78) are those corresponding to the covariance method and are used here to avoid the edge effects that were discussed in arriving at the covariance method. Thus, the Burg method combines features of the three preceding AR methods: the autocorrelation, covariance and forward-backward least squares methods.

It is easily shown (exercise 5) by substitution of the constraint (79) into the expressions for w_n and $\breve{w}_n$, (60) and (70), that these error sequences obey the lattice recursions (see (35) and (37))

$$w_n(M) = w_n(M-1) + \hat{a}_M(M)\breve{w}_{n-1}(M-1) \tag{80a}$$

$$\breve{w}_n(M) = \breve{w}_{n-1}(M-1) + \hat{a}_M(M)w_n(M-1). \tag{80b}$$

It follows immediately by substitution of these recursions into (78) that the constrained sum of squared errors to be minimized can be expressed as

$$\hat{b}^2(M) = \frac{1}{2(N-M)} \sum_{n=M}^{N-1} \{[w_n(M-1) + \hat{a}_M(M)\breve{w}_{n-1}(M-1)]^2 + [\breve{w}_{n-1}(M-1) + \hat{a}_M(M)w_n(M-1)]^2\}. \tag{81}$$

Since $w_n(M-1)$ and $\breve{w}_{n-1}(M-1)$ depend only on the data $\{x_0, x_1, x_2, \ldots, x_{N-1}\}$ and the parameters $\{a_1(M-1), a_2(M-1), a_3(M-1), \ldots, a_{M-1}(M-1)\}$ of the $(M-1)$st-order model, then $\hat{b}^2(M)$ can be minimized recursively. Given the parameters of the $(M-1)$st-order model, $\hat{b}^2(M)$ can be minimized by equating to zero its derivative with respect to $\hat{a}_M(M)$. This yields (exercise 14)

the solution

$$\hat{a}_M(M) = \frac{-2\sum_{n=M}^{N-1} w_n(M-1)\breve{w}_{n-1}(M-1)}{\sum_{n=M}^{N-1} [w_n^2(M-1) + \breve{w}_{n-1}^2(M-1)]} \tag{82}$$

in which the denominator, denoted by $d(M)$, can be reexpressed as

$$d(M) = 2(N-M+1)\hat{b}^2(M-1) - w_{M-1}^2(M-1) - \breve{w}_{N-1}^2(M-1). \tag{83}$$

Furthermore, (81)–(83) can be used to show (exercise 14) that the denominator obeys the recursion

$$d(M) = [1 - \hat{a}_{M-1}^2(M-1)]\, d(M-1) - w_{M-1}^2(M-1) - \breve{w}_{N-1}^2(M-1) \tag{84}$$

and also that

$$\hat{b}^2(M) = [1 - \hat{a}_M^2(M)]\frac{d(M)}{2(N-M)}. \tag{85}$$

In summary, the *Burg algorithm* consists of the following three steps:

1. Initialize the parameters with $M = 0$, $\hat{a}_0(0) = 0$, $w_n(0) = \breve{w}_n(0) = x_n$,
$$\hat{b}^2(0) = \frac{1}{N}\sum_{n=0}^{N-1} x_n^2, \qquad \text{and} \qquad d(1) = 2N\hat{b}^2(0) - x_0^2 - x_{N-1}^2.$$
2. Increase M by 1 and compute $\hat{a}_M(M)$ using (82); then (for $M \geq 2$) compute $\{\hat{a}_p(M) : p = 1, 2, 3, \ldots, M-1\}$ using recursion (79) and compute $\hat{b}^2(M)$ using (85).
3. If M is not large enough, update $w_n(M-1)$ and $\breve{w}_n(M-1)$ for $n = M-1, M, M+1, \ldots, N-1$ using recursions (80a) and (80b), update $d(M)$ using recursion (84), and return to Step 2.

Observe that unlike the three preceding AR methods, the Burg method computes the AR parameter estimates directly from the data without first computing an autocorrelation sequence or covariance matrix from the data.

For applications involving sine waves in noise, a recently proposed modification of the Burg algorithm can provide improved performance. The modified algorithm results from weighting the prediction errors before carrying out the constrained minimization (see [Kaveh and Lippert 1983; Paliwal 1985; Helme and Nikias 1985]). Another modification that can be applied not only to the Burg method but to the other AR methods as well and that can provide improved resolution is based on channelization and decimation, or subsampling (see Part 4 of Section C in Chapter 6), prior to AR parameter estimation [Quirk and Liu 1983].

A primary source of limitation on the performance of the preceding AR spectrum estimation methods is the limited accuracy of the autocorrelation estimates used (or implied in the case of the Burg method). The errors in these estimates

limit the accuracy of the AR parameter estimates. An interesting alternative approach that can reduce sensitivity to errors in the autocorrelation estimates introduces another stage of least squares approximation [Mehra 1971]. This approach is motivated by the Yule-Walker equations (8a), which can be expressed in matrix form (similar to (8d)) as

$$\begin{bmatrix} \tilde{R}_x(0) & \tilde{R}_x(-1) & \tilde{R}_x(-2) & \dots & \tilde{R}_x(1-M) \\ \tilde{R}_x(1) & \tilde{R}_x(0) & \tilde{R}_x(-1) & \dots & \tilde{R}_x(2-M) \\ \tilde{R}_x(2) & \tilde{R}_x(1) & \tilde{R}_x(0) & \dots & \tilde{R}_x(3-M) \\ \vdots & \vdots & \vdots & & \vdots \\ \tilde{R}_x(Q-1) & \tilde{R}_x(Q-2) & \tilde{R}_x(Q-3) & \dots & \tilde{R}_x(Q-M) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} -\tilde{R}_x(1) \\ -\tilde{R}_x(2) \\ -\tilde{R}_x(3) \\ \vdots \\ -\tilde{R}_x(Q) \end{bmatrix} \tag{86}$$

where $Q > M$. Although this means that there are more equations than unknowns, these are consistent equations, which possess a solution $\{a_p\}_1^M$ (assuming that $\tilde{R}_x$ is indeed the autocorrelation sequence of an Mth-order AR time-series). This is a result of the autocorrelation extrapolation property of AR models (see Part 1 of Section B), which reveals that there is linear dependence among the rows of the matrix in (86). However, when the entries of the matrix are replaced with the autocorrelation estimates $R_x(p-q)_K$ specified by (67), the errors in these estimates typically remove all linear dependence among the rows, yielding an inconsistent set of $Q > M$ equations, expressed in matrix form as

$$\boldsymbol{R}_x\hat{\boldsymbol{a}} \cong -\boldsymbol{r}_x, \tag{87}$$

where $\boldsymbol{R}_x$ and $\boldsymbol{r}_x$ are defined by

$$R_x(p, q) \triangleq R_x(p-q)_K \tag{88a}$$

$$r_x(p) \triangleq R_x(p)_K \tag{88b}$$

and the elements of $\hat{\boldsymbol{a}}$ are the desired AR parameter estimates. The approximation in (87) reflects the inconsistency of these Q equations. To accommodate this situation, let us obtain the AR parameter estimates by minimizing the sum of squared errors in (87). This sum of Q squared errors can be expressed as the squared Euclidean norm of the vector of errors, and therefore the parameter estimates are given by the solution to the norm-minimization problem

$$\min_{\hat{\boldsymbol{a}}} \|\boldsymbol{R}_x\hat{\boldsymbol{a}} + \boldsymbol{r}_x\|. \tag{89}$$

Thus, $\hat{\boldsymbol{a}}$ is the least squares solution to the set of inconsistent estimated Yule-Walker equations.[4] This solution is given by the pseudo-inverse of the nonsquare matrix $\boldsymbol{R}_x$ (exercise 16),

$$\hat{\boldsymbol{a}} = -[\boldsymbol{R}_x'\boldsymbol{R}_x]^{-1}\boldsymbol{R}_x'\boldsymbol{r}_x. \tag{90}$$

Various algorithms for this matrix inversion are given in [Lawson and Hanson 1974]. In spite of the fact that the $Q \times M$ matrix includes entries whose estimation errors are typically greater than those in the $M \times M$ matrix used in the YW

[4] By substitution of (88a) and (88b) into (89), it can be seen that the solution of (89) is the least squares extrapolator of the sequence $R_x(n)_K$.

method (because their lag values, $|p - q|$, are greater, and therefore the numbers of terms averaged to obtain the estimates (67) are smaller), this least squares procedure reportedly can outperform the other AR methods, especially for time-series with sharply peaked spectra, (e.g., for sine waves in additive noise) but not necessarily for smoother spectra (see [Kaveh and Bruzzone 1981]). This method is called the *overdetermined normal equations* (ODNE) *method* of spectrum estimation. Some users of this method recommend using the unbiased correlation matrix estimate $[K/(N - |p - q|)]\boldsymbol{R}_x$ in place of the biased estimate $\boldsymbol{R}_x$ (exercise 12), especially for time-series with sharply peaked spectra.

If $Q << N$ in the ODNE method, then the accuracy of the autocorrelation estimates that are added to the $M \times M$ matrix to obtain the $Q \times M$ matrix is essentially as good as the accuracy of the autocorrelation estimates in the $M \times M$ matrix, because the number of terms $N - |p - q|$ averaged to obtain the estimate $R_x(p - q)_K$ in (67) is nearly equal to N for all values of p and q. Thus, in this case the ODNE method is justified by the argument that the M AR parameters should satisfy every set of M Yule-Walker equations, not just the set with the smallest lag values $|p - q|$. Although more than one set cannot simultaneously be satisfied, the closest simultaneous fit to a number $(Q - M + 1)$ of sets is more appropriate than a perfect fit to one set, which need not be any more appropriate than the other sets. In fact, data that exhibit sharply peaked spectra have autocorrelations that oscillate for relatively large values of lag $|p - q|$ (see Figure 1-2). This oscillatory behavior is better reflected in the $Q >> M$ normal equations, which involve larger lag values, than in the M normal equations corresponding to the smallest lag values. (It is also better reflected in the unbiased correlation matrix estimate than in the biased estimate that is effectively attenuated at larger lag values.) Although this argument is not as strong when Q is not much smaller than N, in practice good results have been obtained with relatively large values of Q (see [Tufts and Kumaresan 1982; Cadzow 1982; Cadzow et al. 1983]).

Once the AR parameter estimates have been obtained using (90), the parameter b^2 in (4) can then be estimated by one of several possible methods. For example, the Yule-Walker equation (8b) can be used as a basis, together with the estimates of $\tilde{R}_x(k)$ and a_p. Or, as a potentially superior approach, the estimate of b^2 can simply be taken to be the time-averaged, squared prediction error obtained using the AR parameter estimates as the predictor coefficients.

3. Model-Order Determination

When any of the preceding AR methods of spectral analysis are used in practice, they require a technique for determining an appropriate model order M. For example, this task must be performed in order to carry out the third step in the Burg algorithm. In view of the preliminary discussion of AR, MA, and ARMA modeling in Part 1 of Section B, it would appear that the largest possible order should be used. However, the effects of a limited data segment intervene, and we find that there is a trade-off to be dealt with. Specifically, although the bias of the spectrum estimate can in principle be reduced by increasing M, the

variance of the spectrum estimate typically increases as M is increased. In particular, in order to maintain a small variance for the AR parameter estimates (and therefore for the AR spectrum estimate) it is generally required that $M \ll N$. Thus, the smaller the data-segment length N is, the smaller the order M must be. But if M is too small, then the resolution capability of the spectrum estimate is inadequate, and the bias is therefore large. For example, an Mth-order AR spectrum can contain no more than $M/2$ peaks (at positive frequencies). Thus, if M is too small, the spectrum estimate will be too smooth. But if M is too large, the large variance is typically exhibited in the form of spurious peaks. Thus, for a given length N of the data segment, there is in principle an optimal order M. However, this optimum depends on the underlying phenomenon that gives rise to the data as well as on the particular data segment and its length. Consequently, the order of the AR model to be used for spectral analysis must be determined from the data itself, except in those fortunate situations for which knowledge of the phenomenon dictates a specific order (that is manageable with the given data, e.g., $M < N/2$). As a general rule, the best order is typically within the range from $N/20$ to $N/5$, but for particularly short data segments the best order can be as large as $N/3$ or even $N/2$.

There are several techniques for automatically estimating an order according to specific optimization criteria. An ideal order-determining technique would use a concave-upward cost function, so that a unique minimum would exist and it would therefore not be necessary to try all orders $M < N$ to find the optimum. Rather, one would simply try $M = 1, 2, 3, \ldots$ in sequence. The corresponding values of the cost function would decrease until the optimum order is reached. The next larger order would produce an increase in the value of the cost function. Some techniques used in practice behave this way (see [Beamish and Priestley 1981], but not all do (see [Ulrych and Bishop 1975]).

There are three particular order-determining criteria that have proven to be of some use in practice. Furthermore, all three techniques typically produce comparable results and require essentially the same amount of computation because they all employ the same set of statistics, namely, the sum of squared prediction errors, $\hat{b}^2(M)$, for each model order, $M = 1, 2, 3, \ldots$. The cost functions used in these three methods are

$$\text{FPE}(M) \triangleq \left[\frac{N+M}{N-M}\right]\hat{b}^2(M) \tag{91}$$

$$\text{AIC}(M) \triangleq \frac{N}{2}\ln[\hat{b}^2(M)] + M \tag{92}$$

$$\text{CAT}(M) \triangleq \frac{1}{N}\sum_{m=1}^{M}\left[\frac{N-m}{N\hat{b}^2(m)}\right] - \frac{N-M}{N\hat{b}^2(M)}. \tag{93}$$

(Observe that minimization of AIC(M) is equivalent to minimization of $\hat{b}^2(M)e^{2M/N}$, which can be easier to compute.) Minimization of the cost function FPE is called the *final prediction error criterion*. It results from the objective of minimizing the sum of the variance of the prediction error and a quantity representing the error in estimating the AR parameters [Akaike 1969a, 1969b, 1970]. (If the sample

mean of the data segment is subtracted from the data before the AR model-fitting procedure is performed, then M should be replaced by $M + 1$ in the first factor in (91).) Minimization of the cost function AIC is called the *type A information criterion*. It results from the objective of minimizing the Kullback-Leibler distance (mean-information) between the true probability density of the data and the probability density of the assumed model, which is approximately the same as maximizing the difference between the likelihood function for the model-error variance and the model order [Akaike 1974]. Minimization of the cost function CAT is called the *autoregressive transfer function criterion*. It results from the objective of minimizing the integrated mean-squared relative error between the Mth-order fitted-model transfer function and the transfer function of the exact AR model (possibly of infinite order) [Parzen 1974, 1977].

Although all three techniques perform comparably in many situations, it has been observed that performance is often especially poor for short data segments, and even for relatively long data segments there is often much room for improvement.[5] Unfortunately, these techniques do not always exhibit an unambiguous minimum for time-series with sharp spectral peaks [Burshtein and Weinstein 1985]. If, for a given model-fitting method, $\hat{b}^2(M)$ is not monotonically decreasing in M, then none of these criteria can be expected to result in a unique local minimum that is the global minimum. Some interesting relationships among these three criteria are derived in [Burshtein and Weinstein 1985].

4. Singular-Value Decomposition

A recently proposed alternative to the methods of model-order determination based on the model-error variance $\hat{b}^2(M)$ has grown out of the ODNE approach to estimation of the AR parameters [Cadzow 1982] and earlier approaches to reduction of sensitivity of least squares solutions to perturbations of the matrix to be inverted (see [Tufts and Kumareson 1982; Cadzow et al. 1983; Owsley 1985]). This alternative approach is quite distinct in that it effectively determines a model order before any model fitting is done. The approach is motivated by the fact that the $Q \times M$ matrix in the Yule-Walker equations (86) has rank of only $M_0 < M$ if the true order of the AR time-series is M_0. (Also, for data consisting of M_0 sine waves in white noise, the approach is motivated by the fact that for $M_0 \ll M$ and not too low a signal-to-noise ratio, the *approximate* rank of the rank-M correlation matrix is only M_0.) The method is to replace M by Q to obtain a $Q \times Q$ matrix of correlation estimates $\boldsymbol{R}_x$ with elements defined by (88a) and to determine the *approximate rank*, say M_0, of $\boldsymbol{R}_x$, and then to use a rank-M_0 approximant to $\boldsymbol{R}_x$ in the Yule-Walker equations. Thus, the model order remains equal to the initially chosen large value $M = Q$ (e.g., $Q = N/2$ or even $Q = N$), but the solution tends to behave like a model of order M_0.

The approximate rank, say M_0, of a matrix and the best rank-M_0 approximating

[5] An alternative method, which uses a cost function that differs from AIC by the inclusion of the factor $\ln[N]$ multiplying the term that is linear in M in (92), produces lower order estimates that are apparently more appropriate for model fitting, but not necessarily for spectrum estimation [Hannan and Quinn 1979].

matrix (in the least squares sense) are determined using the *singular value decomposition* (SVD) method. Specifically, let $\boldsymbol{W}$ be the $Q \times Q$ matrix whose columns are the orthonormal eigenvectors of the $Q \times Q$ matrix $\boldsymbol{R}_x$, and let $\boldsymbol{\Lambda}$ be the diagonal $Q \times Q$ matrix whose elements are the corresponding eigenvalues, $\Lambda_{qq} = \lambda_q$. That is, the qth column of $\boldsymbol{W}$, denoted by $\boldsymbol{w}_q$, is specified by the eigenequation

$$\boldsymbol{R}_x \boldsymbol{w}_q = \lambda_q \boldsymbol{w}_q. \tag{94}$$

For convenience let the eigenvalues be ordered so that $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \cdots, \geqslant \lambda_Q$. (The λ_p are all real and nonnegative because $\boldsymbol{R}_x$ is symmetric and nonnegative definite.) $\boldsymbol{R}_x$ can be represented by $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$ as

$$\boldsymbol{R}_x = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}', \tag{95}$$

which is simply a reexpression of (94) for $q = 1, 2, 3, \ldots, Q$ (since $\boldsymbol{W}' = \boldsymbol{W}^{-1}$). The approximate rank of $\boldsymbol{R}_x$ is taken to be the smallest index, say $q = M_0$, such that

$$\rho \triangleq \frac{\sum_{q=1}^{M_0} \lambda_q}{\sum_{q=1}^{Q} \lambda_q} \cong 1. \tag{96}$$

Thus, when (96) is satisfied, $\lambda_{M_0+1}, \lambda_{M_0+2}, \lambda_{M_0+3}, \ldots, \lambda_Q$ are all negligibly small. The closest rank-M_0 matrix to $\boldsymbol{R}_x$, denoted by $\boldsymbol{R}_{(M_0)}$, which minimizes the sum of squared errors between all Q^2 elements of these two matrices, is simply (see [Eckart and Young 1936; Householder and Young 1950; Rao 1964])

$$\boldsymbol{R}_{(M_0)} = \boldsymbol{W}\boldsymbol{\Lambda}_{(M_0)}\boldsymbol{W}', \tag{97}$$

where

$$\boldsymbol{\Lambda}_{(M_0)} \triangleq \begin{bmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & & \\ & & \lambda_3 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_{M_0} & & \\ & & & & & 0 & \\ & & & & & & \ddots \\ & & & & & & & 0 \end{bmatrix} \tag{98}$$

Fortunately, this SVD ODNE method for AR model fitting can be carried out in a recursive fashion, which requires calculation of only M_0 eigenvalues and eigenvectors. Specifically, using the identity

$$\sum_{q=1}^{Q} \lambda_q = \operatorname{tr}\{\boldsymbol{R}_x\} \triangleq \sum_{p=1}^{Q} R_x(p, p)_K, \tag{99}$$

one can calculate λ_1 and check (96) with $M_0 = 1$, then calculate λ_2 and check (96) with $M_0 = 2$, and so on, until (96) is satisfied. Then the desired AR parameter estimates are obtained by solving the Q normal equations ((86) with $M = Q$) modified by replacement of $\boldsymbol{R}_x$ with its best rank-M_0 approximant $\boldsymbol{R}_{(M_0)}$. Since

$\boldsymbol{R}_{(M_0)}$ is not of full rank (assuming $M_0 < Q$, as will usually be the case) then the solution $\hat{\boldsymbol{a}}$ to the modified normal equations is not unique. If the particular solution that has minimum norm is used, then it can be expressed in terms of only the first M_0 eigenvectors and eigenvalues. Specifically, the minimum-norm solution is (exercise 16)

$$\hat{\boldsymbol{a}} = -\boldsymbol{W}\boldsymbol{\Lambda}_{(M_0)}^{-1}\boldsymbol{W}'\boldsymbol{r}_x, \tag{100}$$

where

$$\boldsymbol{\Lambda}_{(M_0)}^{-1} \triangleq \begin{bmatrix} \frac{1}{\lambda_1} & & & & & & \\ & \frac{1}{\lambda_2} & & & & & \\ & & \frac{1}{\lambda_3} & & & & \\ & & & \ddots & & & \\ & & & & \frac{1}{\lambda_{M_0}} & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \tag{101}$$

Substitution of (101) into (100) yields

$$\hat{\boldsymbol{a}} = -\sum_{q=1}^{M_0} \frac{1}{\lambda_q}(\boldsymbol{w}_q'\boldsymbol{r}_x)\boldsymbol{w}_q. \tag{102}$$

Finally, calculation of $\boldsymbol{w}_q$ as well as λ_q can be carried out recursively by calculating λ_1 and $\boldsymbol{w}_1$, then subtracting $\lambda_1\boldsymbol{w}_1\boldsymbol{w}_1'$ from $\boldsymbol{R}_x$ and calculating λ_2 and $\boldsymbol{w}_2$ from the resultant matrix, and so on. At the qth stage of this process, the largest eigenvalue is λ_q (assuming $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_Q$). Thus, a standard method for determining the largest eigenvalue and corresponding eigenvector can be used recursively (see [Golub and Kahan 1965; Lawson and Hanson 1974; Klema and Laub 1980]).

The result in (102) provides a means for interpreting how this rank-reduction method improves the parameter estimate $\hat{\boldsymbol{a}}$. When no rank reduction is used, the parameter estimate provided by the ODNE method is given simply by (102) with $M_0 = M = Q$. Thus, the rank-reduction method discards the $Q - M_0$ terms corresponding to the smallest eigenvalues λ_q, and it is these smallest eigenvalues and their corresponding eigenvectors that are typically the most sensitive to perturbations in the elements of the correlation matrix estimate $\boldsymbol{R}_x$ (see [Wilkinson 1965; Lawson and Hanson 1974]). Thus, the random errors in this matrix estimate typically have their greatest influence on the terms that are discarded in (102).

Once the AR parameter estimates have been obtained using (102), the parameter b^2 in (4) can be estimated by evaluating the average squared prediction error obtained using the AR parameter estimates as the predictor coefficients, or by using (69).

For sine waves in additive broadband noise, improved performance can reportedly be obtained by replacing the autocorrelation matrix $\boldsymbol{R}_x$ used in (95) with the forward-backward covariance matrix defined by (74). In this case

$Q = 3N/4$ often yields the best results. However, both types of autocorrelation matrix estimates used with this SVD ODNE method can result in poor spectrum estimates when there is no clear-cut best reduced-rank estimate. That is, when the eigenvalues roll off slowly rather than dropping abruptly, rank reduction, depending on its extent, can either cause important data components to be discarded or result in negligible improvement.

Although this method has worked well in practice on appropriate types of data, there is a modification that yields a model order of $M = M_0$ rather than $M = Q$, and the modified method can reportedly work even better in some applications [Cadzow 1982]. The method begins with the matrix in the estimated Yule-Walker equations (similar to (8c)),

$$\boldsymbol{R}_x \triangleq \begin{bmatrix} R_x(1)_K & R_x(0)_K & R_x(-1)_K & \cdots & R_x(1-Q)_K \\ R_x(2)_K & R_x(1)_K & R_x(0)_K & \cdots & R_x(2-Q)_K \\ R_x(3)_K & R_x(2)_K & R_x(1)_K & \cdots & R_x(3-Q)_K \\ \vdots & \vdots & \vdots & & \vdots \\ R_x(Q+1)_K & R_x(Q)_K & R_x(Q-1)_K & \cdots & R_x(1)_K \end{bmatrix} \tag{103}$$

This matrix is replaced by its best rank-M_0 approximant, $\boldsymbol{R}_{(M_0)}$, as before. The approximant is then used to construct $Q + 1 - M_0$ matrices of the form

$$\boldsymbol{R}_{(M_0)}(p) \triangleq \begin{bmatrix} R_{(M_0)}(1,p) & R_{(M_0)}(1,p+1) & \cdots & R_{(M_0)}(1,p+M_0) \\ R_{(M_0)}(2,p) & R_{(M_0)}(2,p+1) & \cdots & R_{(M_0)}(2,p+M_0) \\ R_{(M_0)}(3,p) & R_{(M_0)}(3,p+1) & \cdots & R_{(M_0)}(3,p+M_0) \\ \vdots & \vdots & & \vdots \\ R_{(M_0)}(Q+1,p) & R_{(M_0)}(Q+1,p+1) & \cdots & R_{(M_0)}(Q+1,p+M_0) \end{bmatrix} \tag{104}$$

for $p = 1, 2, 3, \ldots, Q + 1 - M_0$, where $R_{(M_0)}(p, q)$ is the pqth element of $\boldsymbol{R}_{(M_0)}$. (Observe that even though $\boldsymbol{R}_x$ is Toeplitz, $\boldsymbol{R}_{(M_0)}$ will in general not be Toeplitz.) Then $Q + 1 - M_0$ sets of normal equations (similar to (8c)) are formed,

$$\boldsymbol{R}_{(M_0)}(p)\hat{\boldsymbol{\alpha}} \cong \mathbf{0}, \qquad p = 1, 2, 3, \ldots, Q + 1 - M_0, \tag{105}$$

where

$$\hat{\boldsymbol{\alpha}} \triangleq [1, a_1, a_2, a_3, \ldots, a_{M_0}]'. \tag{106}$$

The approximation in (105) reflects the inconsistency of these equations: There are $Q + 1$ equations but only M_0 unknowns in each set. (Strictly speaking, nothing but zero itself can be approximately equal to zero.) To accommodate this situation, let us obtain the AR parameter estimates by minimizing the sum of weighted squared errors in (105),

$$\min_{\hat{\boldsymbol{\alpha}}} \sum_{p=1}^{Q+1-M_0} c_p \|\boldsymbol{R}_{(M_0)}(p)\hat{\boldsymbol{\alpha}}\|^2, \tag{107}$$

where $\{c_p\}$ is a set of positive weights. For example, since the equations in the midrange, $p \cong (Q - M_0)/2$, use the autocorrelation estimates that are expected to be the most accurate (since $|n - p|$ in (67) takes on the smallest values), then

the weight sequence could be selected to decay from a maximum value as p departs from the midpoint $(Q - M_0)/2$. Since the first element in the vector of unknowns $\hat{\boldsymbol{\alpha}}$ in (106) is fixed, then this is actually a constrained minimization problem. The solution is given (exercise 17) by the matrix equation

$$\hat{\boldsymbol{\alpha}} = \beta \boldsymbol{S}^{-1}\boldsymbol{e}, \tag{108}$$

where $\boldsymbol{S}^{-1}$ is the inverse of the $(M_0 + 1) \times (M_0 + 1)$ matrix

$$\boldsymbol{S} \triangleq \sum_{p=1}^{Q+1-M_0} c_p \boldsymbol{R}'_{(M_0)}(p)\boldsymbol{R}_{(M_0)}(p), \tag{109a}$$

$\boldsymbol{e}$ is the unit vector

$$\boldsymbol{e} = [1\ 0\ 0\ 0\ \ldots\ 0]', \tag{109b}$$

and β is a constant, which ensures that the first element of $\hat{\boldsymbol{\alpha}}$ is unity. The fact that $\boldsymbol{R}_{(M_0)}(p + 1)$ is obtained from $\boldsymbol{R}_{(M_0)}(p)$ simply by deleting the leftmost column and adding a new rightmost column can be exploited to obtain an efficient algorithm for computation of the matrix $\boldsymbol{S}$ [Cadzow 1982].

5. Maximum Likelihood Approach

An entirely different approach to AR model-fitting that predates all but the original YW method is based on the principle of maximum likelihood. Let $f_x(z \mid \boldsymbol{a}, \sigma_w)$ denote the fraction-of-time joint probability density for a segment of length N of the time-series x_n, which is assumed to be an Mth-order AR time-series, (49), with Gaussian innovations w_n having variance σ_w^2 and with AR parameter vector $\boldsymbol{a} \triangleq \{a_p\}_1^M$. This fraction-of-time density is given by

$$f_x(z \mid \boldsymbol{a}, \sigma_w) = [(2\pi)^N |\tilde{\boldsymbol{R}}_x|]^{-1/2} \exp\left\{ -\tfrac{1}{2} z' \tilde{\boldsymbol{R}}_x^{-1} z \right\}, \tag{110}$$

where $|\tilde{\boldsymbol{R}}_x|$ is the determinant of the $N \times N$ matrix $\tilde{\boldsymbol{R}}_x$, which has jkth element given by the limit autocorrelation sequence

$$\tilde{R}_x(j, k) = \tilde{R}_x(j - k), \tag{111}$$

and $\tilde{R}_x(k)$ is given by the inverse FST of the limit spectrum (51),

$$\tilde{R}_x(k) = \int_{-1/2}^{1/2} \sigma_w^2 \left| 1 + \sum_{p=1}^{M} a_p e^{-i2\pi pf} \right|^{-2} e^{i2\pi kf}\, df. \tag{112}$$

In the *maximum-likelihood* method, the fraction-of-time density (110) is evaluated at $z = \{x_0, x_1, x_2, \ldots, x_{N-1}\} \triangleq \boldsymbol{x}$, the available data segment, and then maximized with respect to the unknown parameters $\boldsymbol{a}$ and σ_w^2. Since σ_w^2 factors out of the matrix $\tilde{\boldsymbol{R}}_x$, we can easily maximize with respect to σ_w^2 first, substitute the maximum value into (110), and then maximize with respect to $\boldsymbol{a}$. The result of the first step is (exercise 18)

$$\hat{\sigma}_w^2 = \frac{1}{N} \boldsymbol{x}' \tilde{\boldsymbol{R}}^{-1} \boldsymbol{x} \tag{113}$$

and

$$\max_{\sigma_w} f_x(\boldsymbol{x} \mid \boldsymbol{a}, \sigma_w) = \left[(2\pi)^N \left(\frac{1}{N} \boldsymbol{x}' \tilde{\boldsymbol{R}}^{-1} \boldsymbol{x} \right)^N |\tilde{\boldsymbol{R}}| \right]^{-1/2} \exp\{-N/2\}, \tag{114}$$

where

$$\tilde{\boldsymbol{R}} \triangleq \frac{1}{\sigma_w^2}\tilde{\boldsymbol{R}}_x. \tag{115}$$

Therefore, the remaining maximization problem is (using $|\boldsymbol{R}^{-1}| = 1/|\boldsymbol{R}|$)

$$\max_{\boldsymbol{a}} \frac{|\tilde{\boldsymbol{R}}^{-1}|^{1/N}}{\boldsymbol{x}'\tilde{\boldsymbol{R}}^{-1}\boldsymbol{x}}, \tag{116}$$

where the matrix $\tilde{\boldsymbol{R}}$ has elements given by

$$\tilde{R}(j, k) = \int_{-1/2}^{1/2} \left|1 + \sum_{p=1}^{M} a_p e^{-i2\pi pf}\right|^{-2} e^{i2\pi(j-k)f}\, df. \tag{117}$$

This is, in general, a difficult maximization problem. One simplification is to use the approximation

$$|\tilde{\boldsymbol{R}}^{-1}|^{1/N} \cong 1 \tag{118}$$

for large N; however, unless N is very large, this is not a close approximation if the roots of the polynomial (55) are close to the unit circle in the complex plane, which is precisely the situation where the limit spectrum exhibits dominant peaks (exercise 8) and AR spectrum estimates are especially attractive. Because of the computational complexity of the maximization problem (116)–(117), the maximum-likelihood method of AR spectrum estimation has not been considered to be competitive with the preceding LS and SVD methods for many applications. In fact, if the available data-segment length N is very large, $N >> M$, then it can be shown (exercise 18) that the AR parameters obtained from the maximum-likelihood method are closely approximated by those obtained from the Yule-Walker least squares method.

Nevertheless, the maximum-likelihood approach continues to be a topic of research interest, and new and improved methods for obtaining approximate solutions to the optimization problem (116)–(117) are still being proposed (see [Kay 1983]).

6. Discussion

All the direct methods of spectral analysis described in preceding chapters yield spectrum estimates of the MA type except for the minimum-leakage method, which yields spectrum estimates of either the AR or ARMA types and generally provides better resolution of sharp spectral peaks. It has been reported that of the four AR LS methods, the best resolution of sharp peaks using short data-segment lengths is provided by the FB method, the poorest (relatively speaking) is provided by the YW method, and the Burg method does somewhat better than the covariance method. But there are many other performance criteria to be considered when selecting a method in practice. For example, both the YW and Burg methods guarantee a stable AR model, whereas neither the covariance nor FB LS methods guarantee stability, although in practice they typically do yield stable models, and even an unstable model can produce a useful spectrum estimate. The variance of the minimum-leakage AR method is reportedly smaller

than that of the YW method [Baggeroer 1976], but the YW method has smaller variance than both the covariance method and the Burg method for short data segments; the FB LS method has been said to have the smallest variance (as well as the best resolution) of all the LS AR methods [Swingler 1974; Ulrych and Bishop 1975; Nuttal 1976; Ulrych and Clayton 1976] but the minimum-leakage method can be superior as illustrated in Section E. Also, for sine waves in additive broadband noise, the YW, covariance, and Burg methods all occasionally produce two close spectral peaks around the frequency of a single sine wave (the spectral line-splitting phenomenon), but the FB LS method [Kay and Marple 1981] and the modified (weighted) Burg method [Helme and Nikias 1985], [Paliwal 1985] have not been observed to do this. Moreover, the location of the peaks in these latter two methods are generally closer to the correct frequencies than in the other three LS AR methods [Kay 1987]. It is also of interest that in all four LS AR methods, the amplitude of a spectral peak is proportional to the square of the power in the underlying spectral line associated with a sine wave for high signal-to-noise ratio, but the area under the peak is proportional to the power (because the width is inversely proportional to the power) [Lacoss 1971]. On the other hand, the amplitude of a spectral peak produced by the minimum-leakage AR method is indeed proportional to the power [Lacoss 1971] (assuming that the width of the spectral line is narrower than the effective resolution width for this method) as it is for all the direct methods. In summary, it appears that the best performing LS AR methods for sharply peaked spectra are the FB LS method and the modified Burg method. It should also be noted that additive white noise in the data generally has a smoothing effect on these AR spectrum estimates. Consequently, for sufficiently low signal-to-noise ratio sine waves in noise, the resolution performance of all these LS AR methods becomes inferior to that of the direct methods [Kay 1987]. (Also, for MA-type time-series, AR methods can be substantially inferior to direct methods [Beamish and Priestley 1981].) Nevertheless, other AR methods, such as the ODNE and SVD methods can reportedly provide improved performance for some low signal-to-noise ratio (e.g., as low as 0 dB) applications [Cadzow 1982; Tufts and Kumaresan 1982; Cadzow et al. 1983]. These SVD methods also considerably improve on other related but earlier parametric methods designed specifically for sine waves in noise, such as the *Hildebrand-Prony method* and the *Pisarenko method* (see [Tufts and Kumaresan 1982; Kay 1987]).

Another performance criterion to consider in selecting a method for spectral analysis is the ease with which spectrum estimates can be updated as time passes in applications where it is desired to do time-variant spectral analysis. Updating is relatively straightforward for direct methods, as explained in Chapter 6, and is also possible for many of the AR methods but is not quite as simple [Friedlander 1982a, 1982b, 1983a, 1983b; Cioffi and Kailath 1984; Honig and Messerschmitt 1984; Marple 1987]. It should be emphasized that the focus on computationally efficient methods of spectral analysis is appropriate in situations where large amounts of data are continually produced, such as in radar, sonar, and seismology signal-processing applications. However, in situations where the cost or other limitations on collecting data are dominant, this focus is not necessarily appropriate,

and the more computationally burdensome but potentially higher-performing exact or near-exact maximum-likelihood methods can be attractive. This is especially true for ARMA methods, which are discussed in the next section.

Finally, it should be emphasized that the appropriateness of parametric methods for sharply peaked spectra applies only when there are multiple peaks that are closely spaced. If the separation between peaks exceeds the resolution width $\Delta f = 1/N$ determined by the amount of data available, then direct nonparametric methods typically perform just as well, if not better. In fact it is shown in exercise 22 that for a spectrum with a single sharp peak, modeled by a time-series consisting of a single sine wave in additive white Gaussian noise, the periodogram provides optimum (maximum-likelihood) estimates of the amplitude and frequency of the sine wave. Furthermore, for multiple sine waves with frequency separation sufficiently in excess of $1/N$, the periodogram is still nearly optimum. However, when there are multiple spectral peaks spaced more closely than $1/N$, the parametric methods can perform better than the direct methods. If the complex spectrum obtained from the DFT with substantial zero-padding is used to estimate (phase as well as frequency and amplitude) and subtract sequentially each of a multiplicity of additive sine wave components, and the spectrum of the residual is then estimated by direct methods, then performance can be considerably improved, but is still not comparable with parametric methods for data consisting of multiple closely spaced sine waves in additive noise. Furthermore, parametric methods typically provide substantial improvements in performance relative to direct methods for sensor-array signal-processing problems that arise in radar, sonar, and seismology applications. The reason for this is that in these problems, the number of spatial samples N is very small, but the $N \times N$ correlation matrix for these N samples can be relatively accurately estimated (since the number of products averaged is not restricted by N but rather is determined by the number of time samples taken at each of the N sensors). Thus, the problems of energy-source detection and direction-of-arrival estimation, which are the spatial analogs of the temporal counterparts of spectral-line detection and spectral-line-frequency estimation, can benefit from exploitation of special structure in the correlation matrix, including AR structure and special eigenstructure that can be revealed by singular value decomposition.

D. ARMA METHODS

Although high-order AR methods of spectral analysis are being developed to perform well for some ARMA time-series as well as AR time-series, ARMA model fitting continues to be of interest because of its inherent appropriateness for ARMA and related time-series. For example a time-series consisting of M sine waves in additive white noise behaves much like an ARMA time-series with order parameters M and $L = M$ (exercise 9), and therefore an ARMA spectrum estimate would appear to be the most appropriate of the three parametric types. Similarly the sum of two independent AR time-series of orders M_1 and M_2 is an ARMA time-series with order parameters of $M \leq M_1 + M_2$ and $L \leq M_1 + M_2$. However, because of the inherent nonlinearity of the ARMA

model-fitting problem, currently available ARMA methods are not as well proven as the AR methods described in the previous section.

ARMA methods of model fitting can be conveniently classified into three groups. The first group consists of those methods that attempt to maximize the exact likelihood function for the $L + M + 2$ ARMA parameters $\{a_p\}_1^M$, $\{b_q\}_0^L$, and σ_w in (52) and (53) and those that attempt to maximize a close approximation to the likelihood function, which includes some nonlinear least squares methods and other interative procedures.[6] These methods typically involve the most computationally intensive and potentially problematic algorithms for optimization but also are potentially the best performing. The second group consists of those methods that use estimates of the autocorrelation sequence to jointly estimate both the AR parameters $\{a_p\}_1^M$ and the MA parameters $\{b_q\}_0^L$. These first two groups of methods are beyond the scope of this introductory treatment. The third group consists of those methods that use estimates of the autocorrelation sequence first to estimate the AR parameters and then use these estimates together with the autocorrelation estimates or the data to estimate the MA parameters.[7] These methods are typically the most computationally attractive but often at the cost of poorer performance relative to methods in the first two groups (see [Kaveh and Bruzzone 1981; Kaveh and Bruzzone 1983; Bruzzone and Kaveh 1984]). Nevertheless, their performance can be superior to AR methods for some time-series. It is this third approach to ARMA spectrum estimation that is briefly described here.

1. Modified Yule-Walker Equations

Part of the Yule-Walker equations for AR time-series, which are derived in Part 1 of Section B, apply equally well to ARMA time-series. Specifically, it is shown in exercise 19 that (8a) applies for $k \geq L + 1$,

$$\tilde{R}_x(k) = -\sum_{p=1}^{M} a_p \tilde{R}_x(k-p), \qquad k \geq L+1, \tag{119a}$$

where L is the MA order and M is the AR order of the ARMA time-series (52). It is also shown in exercise 19 that another part of the Yule-Walker equations, (8b), applies equally well:

$$\tilde{R}_x(0) = b_0\sigma_w^2 - \sum_{p=1}^{M} a_p \tilde{R}_x(-p). \tag{119b}$$

However, the remaining part of the Yule-Walker equations, (8a) for $1 \leq k \leq L$, must be modified in order to apply to an ARMA time-series. Specifically, it is

[6] See [Dugré et al. 1981; Kaveh and Bruzzone 1983; Friedlander 1983a; Kay 1987] for a multitude of references.

[7] Most of these methods are reported on in [Walker 1962; Hsia and Landgrebe 1967; Gersh 1970; Graupe et al. 1975; Sakai and Arase 1979; Kaveh 1979; Kay 1980; Bruzzone and Kaveh 1980; Cadzow 1980, 1982, 1983; Friedlander 1983a, 1983b; Kaveh and Bruzzone 1983; Friedlander and Porat 1984a, 1984b; Moses et al. 1985]. See also references in [Kay 1987].

shown in exercise 19 that these L equations contain an extra sum of terms so that the modified equations are given by

$$\tilde{R}_x(k) = \sigma_w^2 \sum_{q=k}^{L} b_q h_{q-k} - \sum_{p=1}^{M} a_p \tilde{R}_x(k-p), \qquad k = 1, 2, 3, \ldots, L, \tag{119c}$$

where $\{h_k : k = 0, 1, 2, 3, \ldots\}$ is the impulse-response sequence of the ARMA model (52) and is given by the inverse FST of the transfer function

$$\tilde{H}(f) = \frac{\sum_{q=0}^{L} b_q e^{-i2\pi fq}}{1 + \sum_{p=1}^{M} a_p e^{-i2\pi fp}}, \tag{120}$$

$$h_k = \int_{-1/2}^{1/2} \tilde{H}(f) e^{i2\pi fk}\, df. \tag{121}$$

It can be shown (exercise 19) that

$$h_k = 0, \qquad k < 0, \tag{122a}$$

$$h_0 = b_0, \tag{122b}$$

and if $M = 0$ (in which case the ARMA model degenerates to an MA model), then

$$h_k = \begin{cases} b_k, & k = 1, 2, 3, \ldots, L \\ 0, & k > L. \end{cases} \tag{122c}$$

Otherwise, h_k is a complicated nonlinear function, (120)–(121), of both $\{a_p\}_1^M$ and $\{b_q\}_0^L$.

Because of the highly nonlinear character of the *modified Yule-Walker equations* (119c), they will not be used here directly as a basis for estimating the ARMA parameters $\{a_p\}_1^M$ and $\{b_q\}_0^L$. However, the linear equations (119a) will indeed be used. The general approach to be described in the next two parts of this section is to use (119a) as a basis for estimating the AR parameters $\{a_p\}_1^M$ from autocorrelation estimates and then to use these estimated parameters together with the data and/or autocorrelation estimates to estimate the MA parameters $\{b_q\}_0^L$. Since equations (119a) are valid only for the lag parameter k extended beyond L, they are typically referred to as the *extended Yule-Walker equations*.

2. Estimation of the AR Parameters

Since the extended Yule-Walker equations (119a) are valid for ARMA time-series as well as AR time-series, then all methods for estimating the AR parameters $\{a_p\}_1^M$ in an AR model that are based on the Yule-Walker equations (8a) can be used to estimate the AR parameters $\{a_p\}_1^M$ in an ARMA model provided only that the lag values, k, used are appropriately extended so that (8a) is used as a basis only for $k \geq L + 1$ as in (119a). For example, the YW and ODNE methods can be used if the equations (86) used as a basis are replaced by the equations

$$\begin{bmatrix} \tilde{R}_x(L) & \tilde{R}_x(L-1) & \dots & \tilde{R}_x(L-M+1) \\ \tilde{R}_x(L+1) & \tilde{R}_x(L) & \dots & \tilde{R}_x(L-M+2) \\ \tilde{R}_x(L+2) & \tilde{R}_x(L+1) & \dots & \tilde{R}_x(L-M+3) \\ \vdots & \vdots & & \vdots \\ \tilde{R}_x(L+Q-1) & \tilde{R}_x(L+Q-2) & \dots & \tilde{R}_x(L+Q-M) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} -\tilde{R}_x(L+1) \\ -\tilde{R}_x(L+2) \\ -\tilde{R}_x(L+3) \\ \vdots \\ -\tilde{R}_x(L+Q) \end{bmatrix} \tag{123}$$

in which the lag values have been extended by L. For $Q = M$, with $\tilde{R}_x(k)$ replaced by its estimate $R_x(k)_K$, (123) is an adaptation of the Yule-Walker autocorrelation method from AR modeling to ARMA modeling. Similarly, for $Q > M$, (123) can be used as the basis for an adaptation of the ODNE method. The estimates for this adapted method are given by (90), with $\boldsymbol{R}_x$ there given by the $Q \times M$ matrix of autocorrelation estimates corresponding to the ideal matrix in (123). Similarly, the SVD method for AR model-order determination and parameter estimation can be adapted simply by replacement of the matrix (103) of autocorrelation estimates with the matrix

$$\boldsymbol{R}_x \triangleq \begin{bmatrix} R_x(L+1)_K & R_x(L)_K & \dots & R_x(L-Q+1)_K \\ R_x(L+2)_K & R_x(L+1)_K & \dots & R_x(L-Q+2)_K \\ R_x(L+3)_K & R_x(L+2)_K & \dots & R_x(L-Q+3)_K \\ \vdots & \vdots & & \vdots \\ R_x(L+Q+1)_K & R_x(L+Q)_K & \dots & R_x(L+1)_K \end{bmatrix} \tag{124}$$

in which the lag values have been extended by L. The rest of the procedure is the same: $\boldsymbol{R}_x$ is approximated by a rank-M_0 matrix, $\boldsymbol{R}_{(M_0)}$, which is then used in the formula (108)–(109). It should be mentioned that the approach of finding the best rank M_0 with which to approximate the rank of the $(Q+1) \times (Q+1)$ matrix (124) is motivated by the fact that the extended Yule-Walker equations (119a) reveal that the rank of the ideal correlation matrix corresponding to (124) is precisely M for an ARMA time-series of AR order M and MA order less than or equal to L (just as for an AR time-series). However, for ARMA model fitting, there is the practical problem of selecting an appropriate value for the MA order L (the lag-extension parameter) in (124) or in (123). As long as L exceeds the true MA order of an ARMA time-series, the preceding reasoning applies and the method is therefore appropriate. Thus, in practice, one should attempt to select a sufficiently large value for L. An approach to determining if the value selected is large enough is described in the next part of this section.

3. Estimation of the MA Parameters

In some applications of parametric spectral analysis, the primary objective is to detect and/or locate any sharp peaks (e.g., due to additive sine wave components in the data). Although an ARMA model might be required for a good fit to the data, only the AR part of the model is especially relevant to spectral-peak analysis. In such cases, the only reason for estimating the MA parameters after having estimated the AR parameters is to obtain a means for checking on the

value of the MA order (lag extension) parameter L used to estimate the AR parameters. However, in other applications where all features of the spectrum (e.g., valleys, notches, and plateaus as well as peaks) are of interest, the effects of the MA part of the model on the spectrum also must be determined. Nevertheless, this does not necessarily require obtaining explicit estimates of the MA parameters $\{b_q\}_0^L$. This can be seen as follows. The limit spectrum for an ARMA time-series is given by (53), which can be expressed as

$$\bar{S}_x(f) = \frac{|B(f)|^2}{|A(f)|^2}, \tag{125a}$$

where

$$B(f) \triangleq \sigma_w \sum_{q=0}^{L} b_q e^{-i2\pi fq} \tag{125b}$$

$$A(f) \triangleq 1 + \sum_{p=1}^{M} a_p e^{-i2\pi fp}. \tag{125c}$$

If the time-series is filtered using the transfer function

$$\bar{G}(f) = A(f), \tag{126}$$

then the limit spectrum of the filtered time-series

$$y_n = x_n \otimes g_n \tag{127}$$

is given by

$$\bar{S}_y(f) = |\bar{G}(f)|^2 \bar{S}_x(f) = |B(f)|^2. \tag{128}$$

Consequently, the MA part of the spectrum for x_n is simply the spectrum of the filtered time-series y_n. Consequently, if the estimate of the AR parameters is sufficiently accurate, then one can estimate the MA part of the spectrum simply by estimating the spectrum of the time-series $\hat{y}_n$ obtained by filtering x_n using the transfer function

$$\hat{G}(f) = \hat{A}(f) \tag{129}$$

obtained from (125c) with $\{a_p\}$ replaced by their estimates $\{\hat{a}_p\}$. Thus, if the estimate of $\bar{S}_y(f)$ is denoted by $|\hat{B}(f)|^2$, then the composite ARMA spectrum estimate of $\bar{S}_x(f)$ is given by

$$S_x(f)_{\text{ARMA}} = \frac{|\hat{B}(f)|^2}{|\hat{A}(f)|^2}. \tag{130}$$

Furthermore, if the estimate $|\hat{A}(f)|^2$ is sufficiently accurate, then the filtered time-series $\hat{y}_n$ is closely approximated by an MA model with spectrum (128). Thus, direct spectral analysis methods can be appropriate for this second stage of estimation. For example, since the limit autocorrelation of an MA time-series of order L is zero for lags exceeding L (exercise 20),

$$\bar{R}_y(k) = \begin{cases} \sigma_w^2 \sum_{q=0}^{L-|k|} b_q b_{q+|k|}, & |k| = 0, 1, 2, \ldots, L \\ 0, & |k| > L, \end{cases} \tag{131}$$

then a Blackman-Tukey type of spectrum estimate with correlogram window width equal to L should be appropriate. This follows from the fact that the above autocorrelation is invariant to multiplication by a rectangle window of width L,

$$\tilde{R}_y(k)w(k) = \tilde{R}_y(k), \qquad -\infty < k < \infty \tag{132a}$$

$$w_k \triangleq \begin{cases} 1, & |k| \leq L \\ 0, & |k| > L, \end{cases} \tag{132b}$$

and therefore the spectrum is invariant to convolution (smoothing) with a sinc window

$$\tilde{S}_y(f) \otimes W(f) = \tilde{S}_y(f), \qquad -\infty < f < \infty \tag{133a}$$

$$W(f) \triangleq \frac{\sin(2\pi Lf)}{\sin \pi f}. \tag{133b}$$

Thus, the usual leakage effects expected from the Blackman-Tukey method with a rectangle correlogram-tapering window need not be as problematic here.

It follows from the finiteness of the extent of the autocorrelation of an MA time-series that the appropriateness of the value for L used in estimating the AR part of the spectrum can be assessed by inspecting the correlogram of the filtered time-series $\hat{y}_n$. If this correlogram does not become particularly small for lags exceeding L, then the value used for L is too small (or the estimate of the AR part of the spectrum is poor.) On the other hand, if the correlogram becomes particularly small for lags exceeding $L_0 < L$, then the first stage of spectrum estimation can be repeated with L replaced by L_0. This could improve the AR spectrum estimate.

It is especially interesting that this approach to ARMA spectral analysis can be interpreted as nothing more than the prewhitening approach described in Chapter 4, Section C, which was proposed as a means for minimizing the effects of spectral leakage.

There is an approach to estimating the MA part of the spectrum that is based on the preceding filtering concept but does not actually require filtering the time-series x_n to obtain $\hat{y}_n$ and then estimating the autocorrelation for $\hat{y}_n$. Specifically, it follows from (128) that the limit autocorrelation for the filtered time-series y_n is given by the convolution

$$\tilde{R}_y(k) = r_g(k) \otimes \tilde{R}_x(k), \tag{134a}$$

where

$$r_g(k) \triangleq g_k \otimes g_{-k}, \tag{134b}$$

and it follows from (131) that

$$\tilde{R}_y(k) = 0, \qquad |k| > L.$$

Thus, the limit spectrum for y_n is given by

$$\tilde{S}_y(f) = \sum_{k=-L}^{L} [r_g(k) \otimes \tilde{R}_x(k)]e^{-i2\pi fk} = [|\tilde{G}(f)|^2 \tilde{S}_x(f)] \otimes W(f), \tag{135}$$

where $W(f)$ is given by (133b). Consequently, an estimate of the MA part of the spectrum (128) can be obtained by replacing $\bar{R}_x(k)$ in (135) with the autocorrelation estimates $R_x(k)_K$ and using the estimated filter $\hat{g}_k$ from (129) in place of the ideal filter g_k in (134b) to obtain $\hat{r}_g(k)$. This requires only two convolutions to obtain $\hat{r}_g(k)$ and then $\hat{r}_g(k) \otimes R_x(k)_K$, and an FST (or DFT) corresponding to (135)

$$S_y(f)_{\text{MA}} = 2\,\text{Re}\left\{\sum_{k=0}^{L} [\hat{r}_g(k) \otimes R_x(k)_K] e^{-i2\pi fk}\right\} - [\hat{r}_g(k) \otimes R_x(k)_K]_{k=0}$$

$$\triangleq |\hat{B}(f)|^2. \tag{136}$$

As an alternative, an estimate of the MA part of the spectrum (128) can be obtained by replacing $\bar{S}_x(f)$ in (135) with the FST of $R_x(k)_K$, and using the estimated filter $\hat{G}(f)$ in place of the ideal filter $\bar{G}(f)$. For example, if $R_x(k)_K$ is taken to be the correlogram, $R_{x_N}(k)$, then its FST is the periodogram, $S_{x_N}(f)$, and the estimate is given by (using (129))

$$S_y(f)_{\text{MA}} = [|\hat{A}(f)|^2 S_{x_N}(f)] \otimes W(f) = |\hat{B}(f)|^2. \tag{137}$$

Thus, the composite ARMA spectrum estimate is given by (using (130))

$$S_x(f)_{\text{ARMA}} = \frac{[|\hat{A}(f)|^2 S_{x_N}(f)] \otimes W(f)}{|\hat{A}(f)|^2}. \tag{138}$$

A variation on this method that is intended to improve the estimates $\hat{r}_g(k) \otimes R_x(k)_K$ of the autocorrelation $\bar{R}_y(k)$ is based on the fact that this limit autocorrelation is the same whether x_n is filtered in the forward direction,

$$\hat{y}_n = \sum_{p=0}^{M} \hat{a}_p x_{n-p}\,, \tag{139a}$$

or in the backward direction,

$$\check{y}_n = \sum_{p=0}^{M} \hat{a}_p x_{n+p}\,. \tag{139b}$$

Thus, we can obtain additional distinct lag products to average in forming the autocorrelation estimate

$$\mathring{R}_y(k)_K \triangleq \frac{1}{2K} \sum_{n=0}^{N-M-|k|} [\hat{y}_{n+|k|+M}\hat{y}_{n+M} + \check{y}_{n+|k|}\check{y}_n]. \tag{140}$$

This estimate can then be used as in (136),

$$S_y(f)_{\text{MA}} = 2\,\text{Re}\left\{\sum_{k=0}^{L} \mathring{R}_y(k)_K e^{-i2\pi fk}\right\} - \mathring{R}_y(0)_K. \tag{141}$$

There is an alternative to the preceding general approach based on filtering that can reportedly yield better performance [Friedlander 1983a]. This alternative approach is based on a decomposition of the spectrum $\bar{S}_x(f)$ into its causal part,

$$S_+(f) \triangleq \sum_{k=1}^{\infty} \bar{R}_x(k) e^{-i2\pi fk}, \tag{142}$$

and its anticausal part, $S_+(-f)$, and is known as Shanks' method [Shanks 1967]. It is easily verified (exercise 21) that

$$\tilde{S}_x(f) = \tilde{R}_x(0) + S_+(f) + S_+(-f), \tag{143}$$

and that for a limit spectrum $\tilde{S}_x(f)$ of the ARMA form (125), $S_+(f)$ is of the form

$$S_+(f) = \frac{C(f)}{A(f)}, \tag{144a}$$

where $A(f)$ is given by (125c) and $C(f)$ is given by

$$C(f) \triangleq \sum_{q=1}^{M} c_q \, e^{-i2\pi fq}, \tag{144b}$$

in which the parameters $\{c_q\}_1^M$ are to be determined. Substitution of (144a) and (125a) into (143) yields (exercise 21) the representation

$$|B(f)|^2 = \tilde{R}_x(0)|A(f)|^2 + A(-f)C(f) + A(f)C(-f) \tag{145}$$

for the MA part of the spectrum. The parameters $\{c_q\}_1^M$ can be obtained by inverse FST using (144a)

$$c_q = \int_{-1/2}^{1/2} S_+(f)A(f)e^{i2\pi fq} \, df. \tag{146}$$

Application of the convolution theorem to (146) yields

$$c_q = \sum_{k=1}^{q} \tilde{R}_x(k)a_{q-k}, \qquad q = 1, 2, 3, \ldots, M. \tag{147}$$

Thus, given estimates of $\{a_p\}_1^M$ and $\{\tilde{R}_x(k)\}_1^M$, we can estimate $\{c_q\}_1^M$ using (147) as a basis. Then, using (144b) to obtain the estimate $\hat{C}(f)$ and using the estimates $\hat{A}(f)$ and $R_x(0)_K$, we can estimate $|B(f)|^2$ using (145) as a basis. However, this approach, like the preceding approach, is limited in performance by the accuracy of the $M + 1$ autocorrelation estimates $\{R_x(k)_K\}_0^M$ used in the counterparts of (147) and (145) (or by the $M + L$ autocorrelation estimates used in (136) in the preceding approach.) But this situation can be improved upon by the same technique used in the ODNE approach to estimation of the AR part of the spectrum. Specifically, rather than solving (144a) for $C(f)$ and $\{c_q\}_1^M$ as in (146)–(147), let us express $\{c_q\}_1^M$ in implicit form analogous to the Yule-Walker equations for $\{a_p\}_1^M$. This is accomplished by introducing the sequence

$$f_k \triangleq \int_{-1/2}^{1/2} \left[\frac{1}{A(f)}\right] e^{i2\pi fk}, \qquad k = 0, 1, 2, \ldots \tag{148}$$

(for which it can be shown that $f_n = 0$ for $n < 0$). It follows from (142), (148), and the convolution theorem applied to (144a) that $\tilde{R}_x(k)$ is given by

$$\tilde{R}_x(k) = f_k \otimes c_k = \sum_{q=1}^{\min\{k,M\}} f_{k-q}c_q, \qquad k = 1, 2, 3, \ldots, \tag{149a}$$

which can be expressed in matrix form as

$$\boldsymbol{F}\boldsymbol{c} = \boldsymbol{r}_x \tag{149b}$$

or (terminating k at $k = P$)

$$\begin{bmatrix} f_0 & & & & \\ f_1 & f_0 & & & \\ f_2 & f_1 & f_0 & & \\ \vdots & \vdots & & \ddots & \\ f_{M-1} & f_{M-2} & \cdots & & f_0 \\ \vdots & \vdots & & & \vdots \\ f_{P-1} & f_{P-2} & \cdots & & f_{P-M} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_M \end{bmatrix} = \begin{bmatrix} \bar{R}_x(1) \\ \bar{R}_x(2) \\ \bar{R}_x(3) \\ \vdots \\ \bar{R}_x(P) \end{bmatrix} \tag{149c}$$

Now, if $x(t)$ is truly an ARMA time-series of order M (or lower) and $L \leq M$, then this overdetermined (for $P > M$) set of equations (149c) will possess a solution that is identical to (147). However, if $\bar{R}_x(k)$ is replaced with estimated values directly in (149) and indirectly in (149) through estimates $\hat{f}_k$ obtained from the counterpart of (148) in which $\hat{A}(f)$ depends on estimates of $\bar{R}_x(k)$, then (149c) will in general be an overdetermined inconsistent set of equations for all $P > M$. Thus, rather than choose $P = M$ to obtain a perfect fit to the minimal subset of equations relating $\hat{f}_k$ to $R_x(k)_K$ (a set that need not be any more appropriate than other subsets of these equations), we can choose $P >> M$ and obtain the closest simultaneous fit to a much larger set of equations (all of which would indeed be satisfied if there were no estimation error). As usual, this is accomplished by the method of least squares, which yields the solution

$$\boldsymbol{c} = [\boldsymbol{F}'\boldsymbol{F}]^{-1}\boldsymbol{F}'\boldsymbol{r}_x. \tag{150}$$

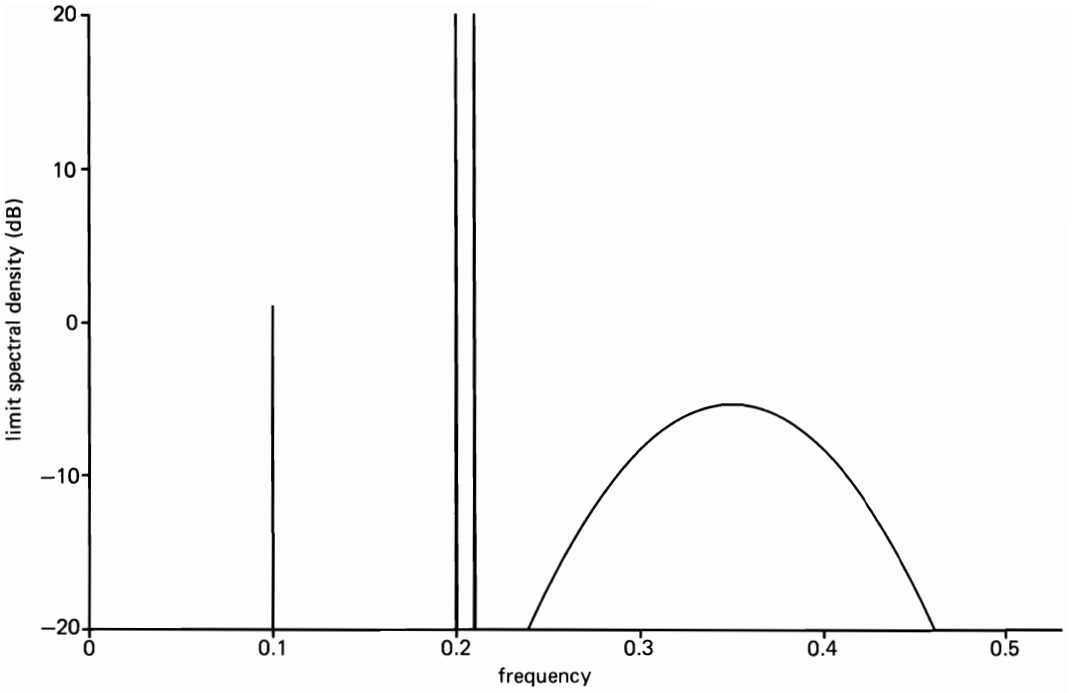
Because of the Toeplitz-like structure of (149c), the solution (150) can be computed efficiently (see [Friedlander, 1983a].)

E. EXPERIMENTAL STUDY

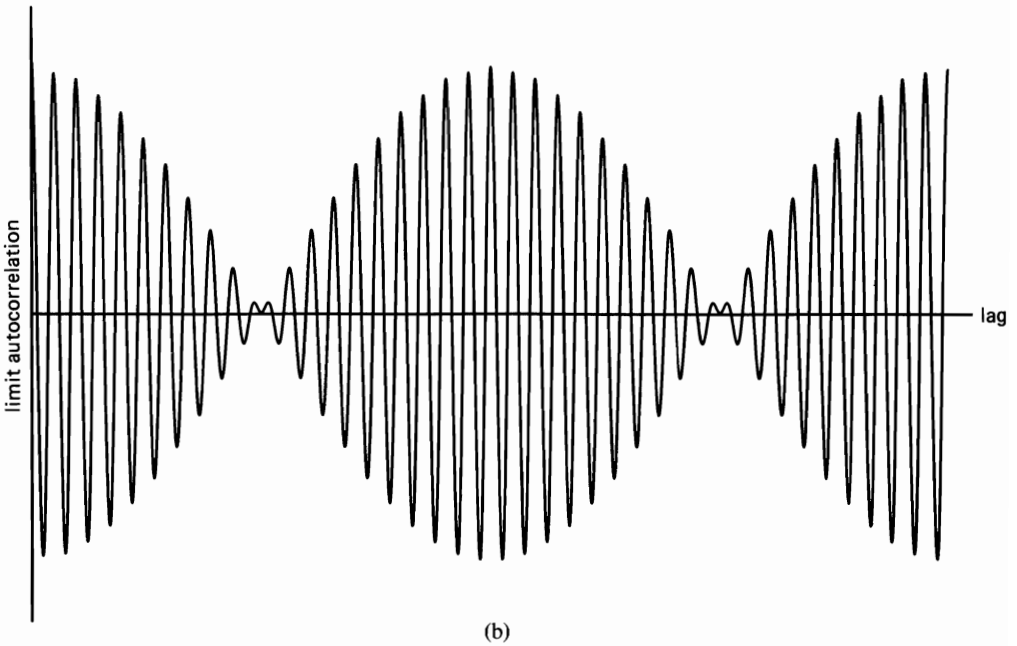
In order to illustrate some of the ways in which the performances of various methods of spectral analysis differ, a summary of the results of an extensive experimental study is presented in this concluding section.[8] The methods compared include the periodogram with zero-padding, with and without data tapering, with and without frequency-smoothing, and with and without sequential sine wave removal and spectral-line reinsertion; the minimum-leakage method with covariance-type correlation matrix; the Yule-Walker, Burg, and forward-backward least squares AR methods; the over-determined-normal-equations AR method with biased and unbiased correlation estimates; and a singular-value-decomposition AR method.

The data was generated using a model that has the limit spectrum and limit autocorrelation shown in Figure 9-4. Appendix 9-1 gives a table of 1024 time samples from this model, and graphs of data segments of lengths 64 and 256 from this data set are shown in Figure 9-5. The data consists of three sine waves in additive highly colored Gaussian noise, with a bandlimited Gaussian-shaped spectrum centered at $0.35/T_s$ Hz. The power of each of the two sine waves at frequencies $0.20/T_s$ Hz and $0.21/T_s$ Hz is -3 dB (relative to unity) and the

[8] The experimental study reported in this section was carried out jointly by Messrs. B. G. Agee, W. A. Brown, C. K. Chen, J. H. Reed, and R. S. Roberts under the author's supervision.

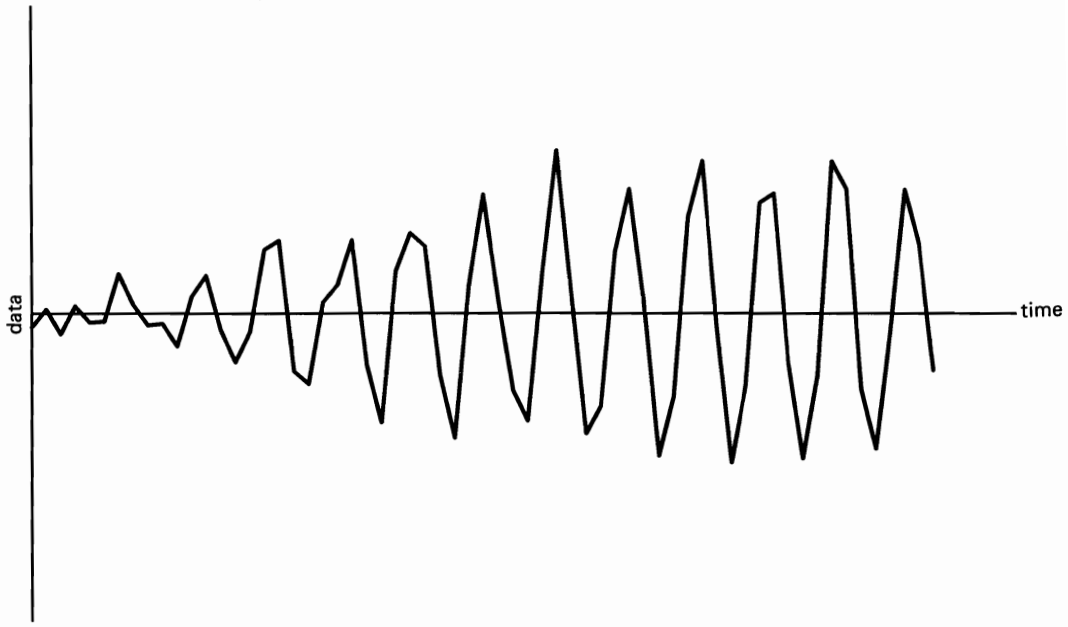


(a)

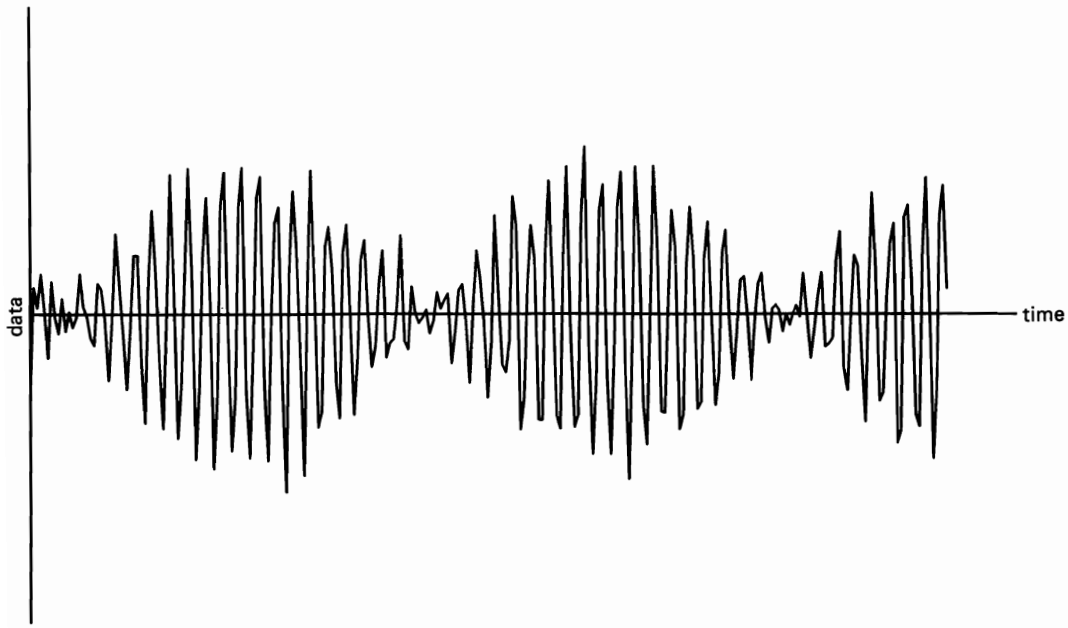


(b)

Figure 9-4 (a) Limit spectrum for data used in experimental study (spectrum smoothed with rectangle window of width 1/256; highest peak is 21 dB). (b) Limit autocorrelation for data used in experimental study (200 lag increments shown).



(a)



(b)

Figure 9-5 (a) Data segment of length 64 from Table 9-1 (starting point is 64). (b) Data segment of length 256 from Table 9-1 (starting point is 256).

phases are 106.2° and 41.5°, respectively. The power of the third sine wave at $0.10/T_s$ Hz is -23 dB and the phase is 32.6°. The power of the colored noise is -15 dB.

For each method and each set of parameter values considered, three samples of spectrum estimates are shown superimposed. Thus, resolution, leakage, and reliability properties are all reflected in these graphical results. All methods are applied to two lengths of data segments, $N = 64$ and $N = 256$. The three statistical samples of data are obtained from Appendix 9-1 using starting points of 64, 128, and 192 for $N = 64$, and 256, 512, and 768 for $N = 256$. For convenience, T_s is taken to be unity in the graphs.

For all the parametric methods considered that require use of a model-order-determining algorithm, the three methods FPE, AIC, and CAT were studied. However, the performances of these methods were usually unacceptable. Trial-and-error experimentation revealed that the order $M = 16$ was typically the best for $N = 64$, and $16 \leq M \leq 32$ (usually $M \cong 32$) was typically the best for $N = 256$. Consequently, results are presented for primarily $M = 16$ and $M = 32$. (Note that M is used to denote both the model order for model-fitting methods and the smoothing parameter for the periodogram.)

1. Periodogram Methods

The first spectrum estimate considered is the periodogram without data tapering, time-averaging, or frequency-smoothing. This raw periodogram was obtained using a DFT algorithm with $K = 8$ zero-padding factor (total points transformed is KN) and is shown in Figure 9-6(a) and Figure 9-7(a) for the two data-segment lengths of $N = 64$ and $N = 256$. It can be seen that the two closely spaced spectral lines are not reliably resolved for $N = 64$ but are for $N = 256$. These periodograms were then modified by use of a raised cosine data-tapering window (with height $a_T(0) = 2$), and the results are shown in Figure 9-6(b) and Figure 9-7(b). All these periodograms were then frequency-smoothed using smoothing parameter $M = 2$ (MK DFT bins are averaged together), and the results are shown in Figure 9-6(c), (d) and Figure 9-7(c), (d). It can be seen that data tapering greatly reduces spectral leakage and that frequency-smoothing improves reliability (although the amount of improvement is small for $M = 2$).

Since Figure 9-6(a)–(b) suggested to us (pretending ignorance of the true model) that there were spectral lines in the vicinities of $f = 0.1/T_s$ and $f = 0.2/T_s$, the next method we studied removed sinewaves from the data as explained in the following and then the preceding methods, corresponding to Figures 9-6(a)–(d) and 9-7(a)–(d), were repeated, except $M = 4$ for $N = 64$ and $M = 8$ for $N = 256$ were used instead of $M = 2$. The sine wave removal was accomplished by using least-squares estimates of the frequency (in the vicinity of the peaks in the raw periodogram), amplitude, and phase of each sine wave, which are obtained directly from the DFT of the data (see exercise 22). Since a substantial spectral peak remained in the periodogram after subtraction of a sinewave with frequency near $0.2/T_s$, the estimation procedure was repeated in this vicinity and a second sine wave was removed. Then the procedure was repeated in the vicinity of $f = 0.1/T_s$ since a spectral peak was clearly evident

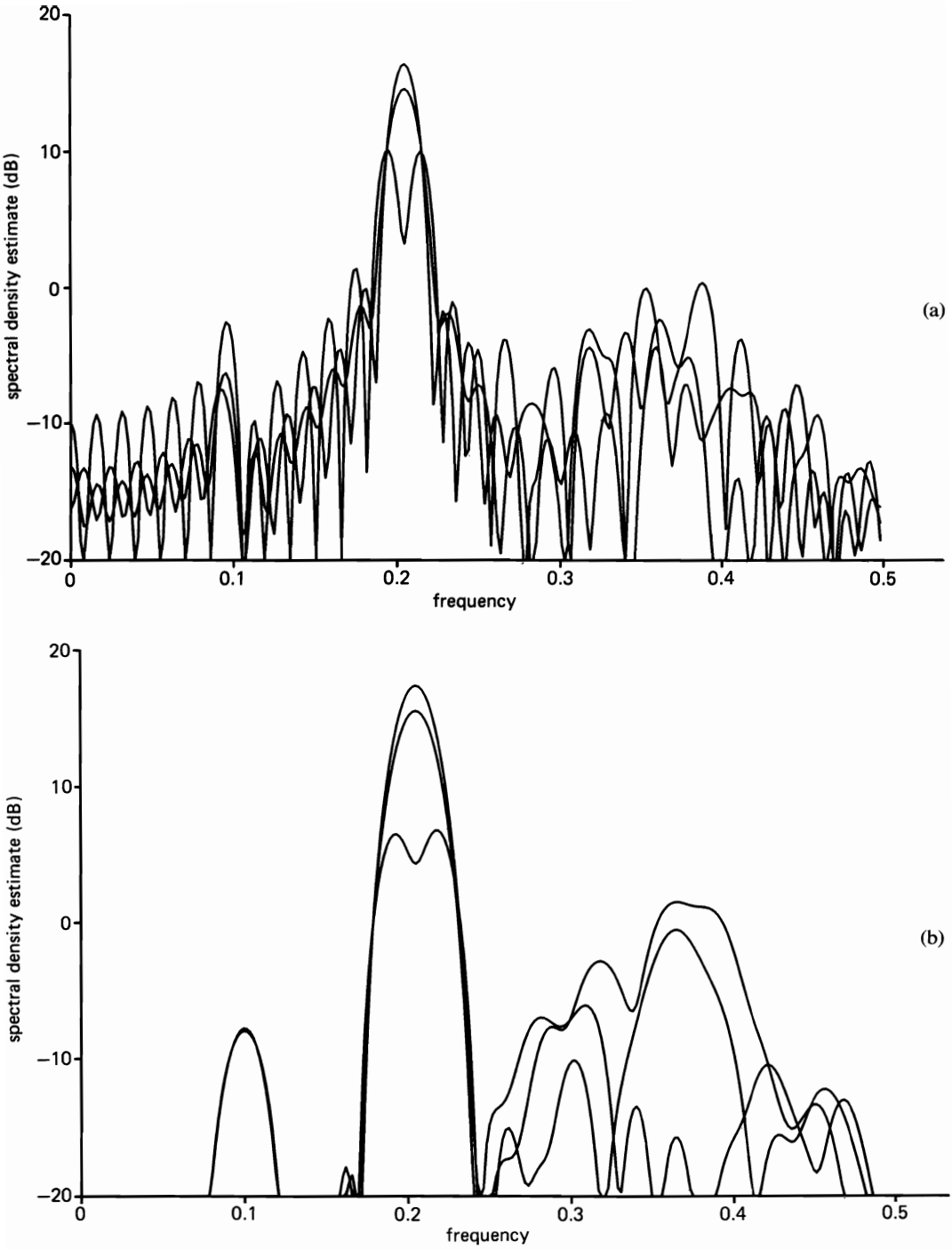
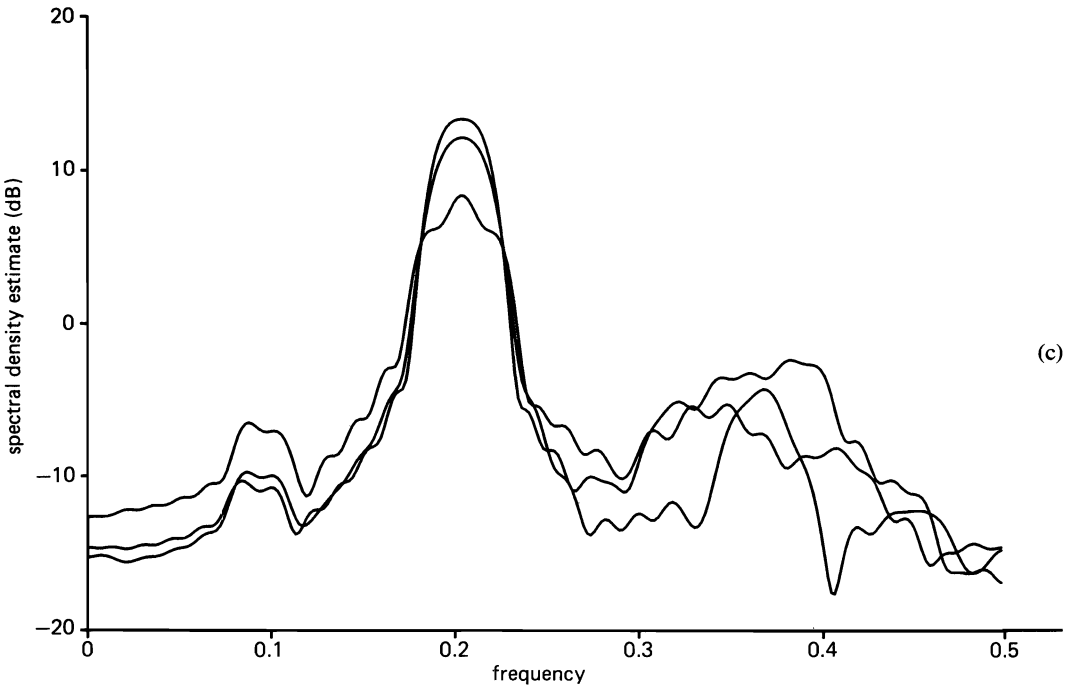
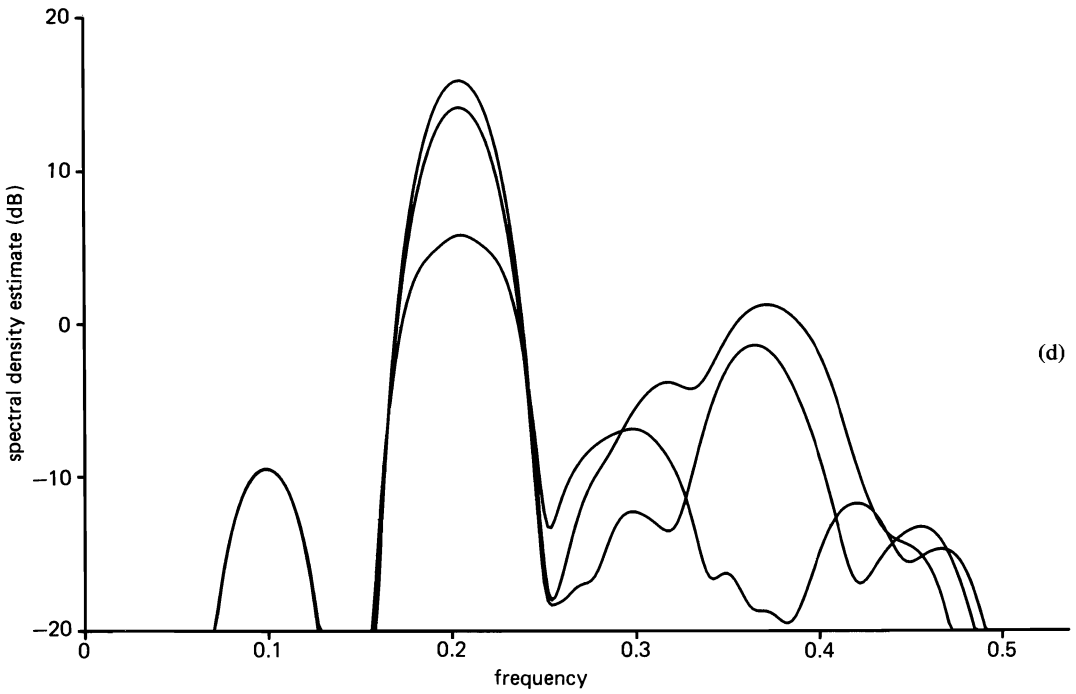


Figure 9-6 (a) Three periodograms for $N = 64$ data points each with $K = 8$ zero-padding factor and no data tapering. (b) Same as (a) but with raised-cosine data tapering.



(c)



(d)

Figure 9-6 (continued) (c) Same as (a) but with frequency smoothing ($M = 2$). (d) Same as (b) but with frequency smoothing ($M = 2$).

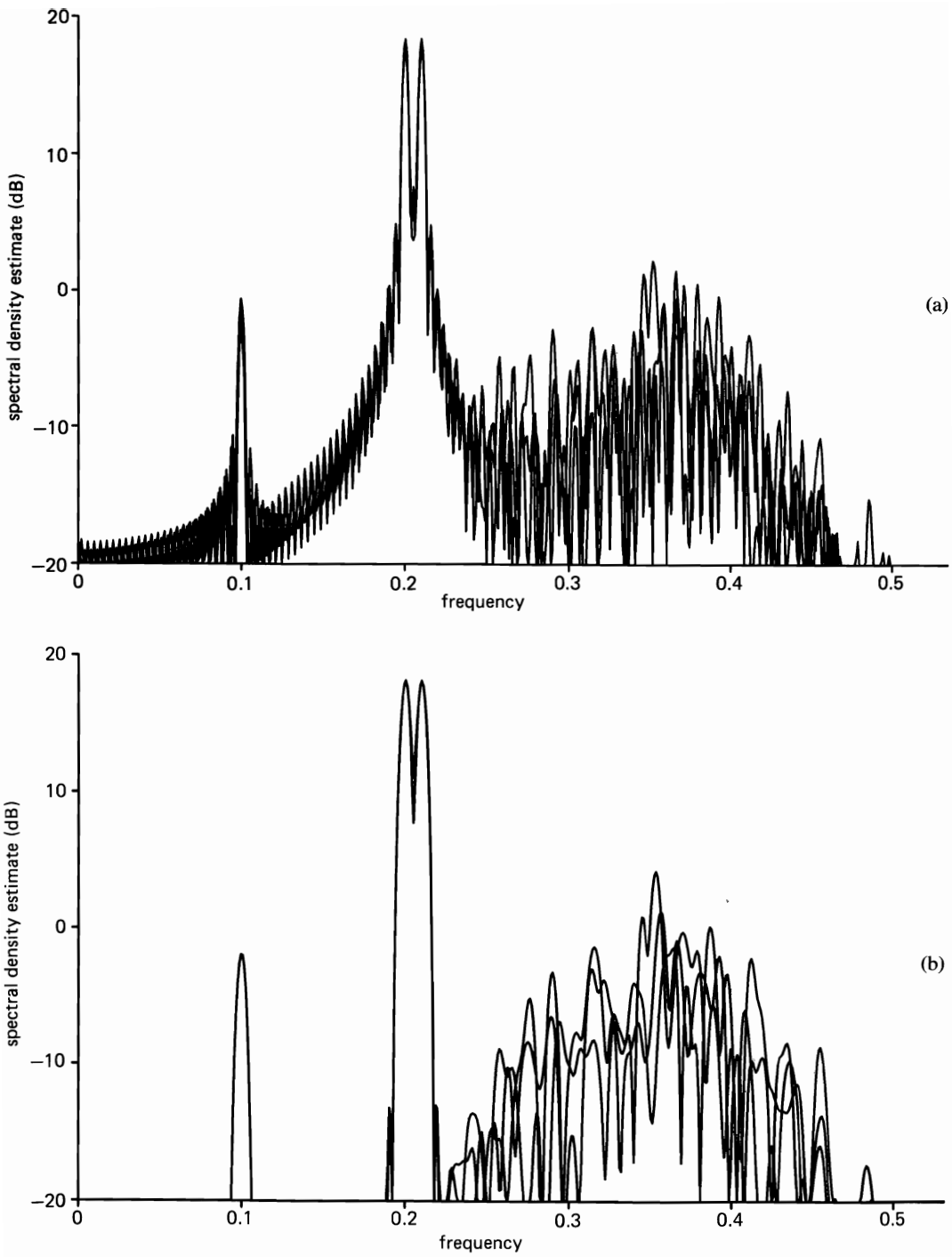


Figure 9-7 (a) Three periodograms for $N = 256$ data points each with $K = 8$ zero-padding factor and no data tapering. (b) Same as (a) but with raised-cosine data tapering.

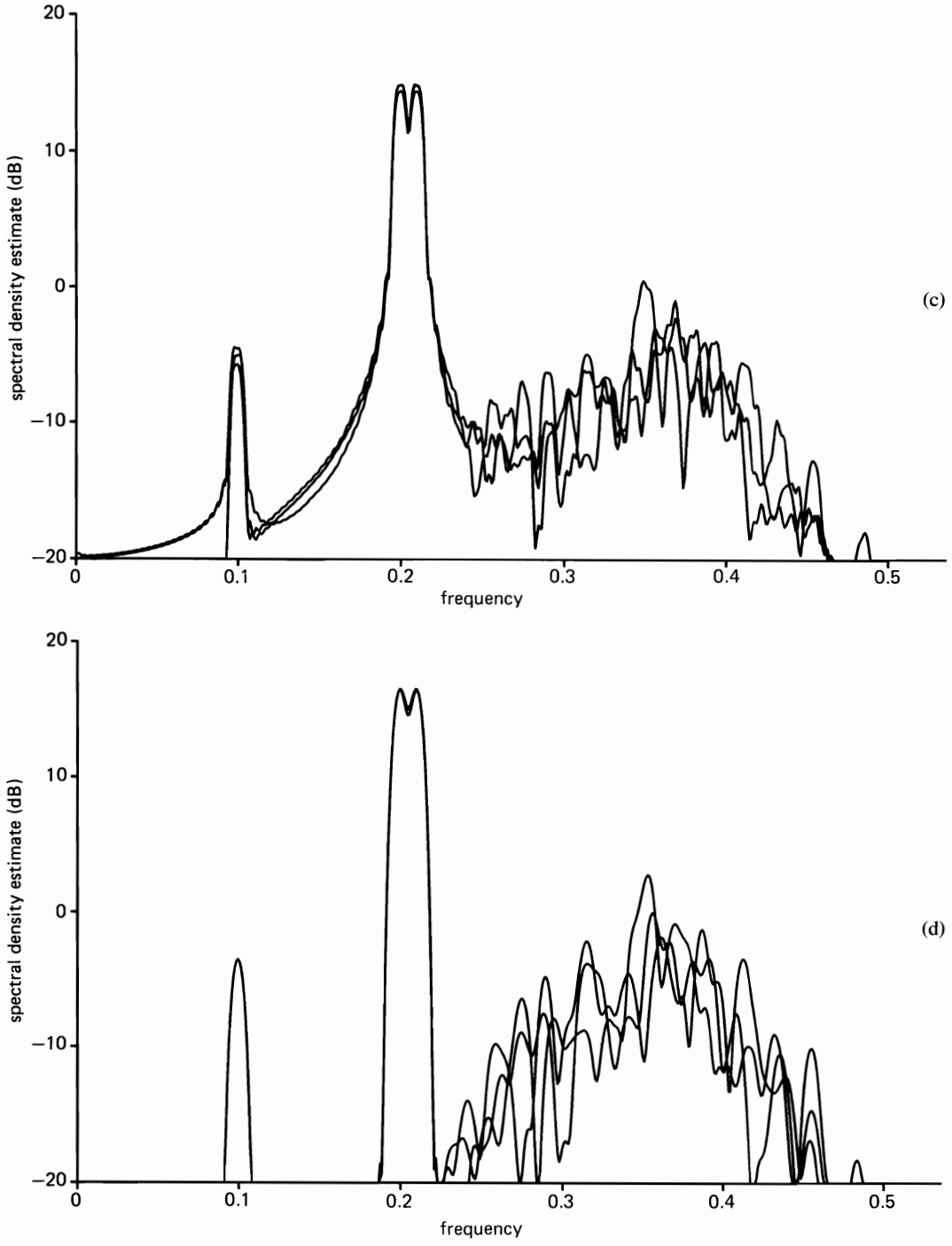


Figure 9-7 (continued) (c) Same as (a) but with frequency smoothing ($M = 2$). (d) Same as (b) but with frequency smoothing ($M = 2$).

with the leakage effects of the two peaks near $f = 0.2/T_s$ largely removed. With the three estimated sine waves subtracted from the data, the periodogram was calculated for the residual time-series. Finally, spectral lines with one DFT bin-width and with magnitude determined by the least squares estimates of the sine wave amplitudes were added to the periodogram. The resultant spectrum estimates, which are shown in Figure 9-8(a)–(d) for $N = 64$ and Figure 9-9(a)–(d) for $N = 256$, exhibit greatly reduced spectral leakage, as would be expected. Although the variability of the pair of spectral lines in the vicinity of $f = 0.2/T_s$ is substantial for $N = 64$, it is very small for $N = 256$. Also the variability of the single weak spectral line at $f = 0.1/T_s$ is quite small even for $N = 64$. The variability of the pair of spectral lines for $N = 64$, as well as the corresponding residual leakage, can be reduced by using a joint least squares estimate or other parametric method as demonstrated in Part 6 of this section.

2. Minimum-leakage Method

The next spectrum estimates considered are those provided by the minimum-leakage (ML) method. Both the modified ((56) in Chapter 6) and unmodified ((52) in Chapter 6) ML estimates were considered, and two versions of each of these were studied. The two versions correspond to use of either the covariance-type data-correlation matrix (which prevents the adaptive filter from running off the ends of the finite segment of data) or the autocorrelation-type matrix (which allows the filter to run completely off both ends of the data segment). The best results were obtained with the unmodified method using the covariance-type matrix and are shown in Figure 9-10(a)–(d). (However, when the autocorrelation-type matrix was used, the modified method gave the best results.) For $N = 64$ it can be seen that both resolution and reliability performances for $M = 24$ are definitely better than that provided by the periodogram (both with and without sine wave removal). However, this superior performance for $N = 64$ is obtained with the unfair advantage of knowing that the filter order $M = 24$ yields the best results. If the limit spectrum were not known in advance, then the choice $M = 24$ would not necessarily have been made. The ML spectrum estimates for $M < 20$ were too smooth and those for $M > 24$ exhibited spurious peaks. For $N = 256$, the best ML estimate, which is shown in Figure 9-10(c), does not approximate the limit spectrum as accurately as does the periodogram method (with sine wave removal) shown in Figure 9-9(c)–(d).

3. Yule-Walker, Burg, and Forward-Backward Least-Squares AR Methods

The least-squares autoregressive spectrum estimates provided by the methods of Yule-Walker (YW), or maximum-entropy, Burg, and forward-backward (FB) linear-prediction are presented in Figures 9-11(a)–(c), 9-12(a)–(c), and 9-13(a)–(c) for $N = 64$ with $M = 16$ and for $N = 256$ with $M = 16$ and $M = 32$. It can be seen that the YW method is consistently inferior to the ML method shown in Figure 9-10(a)–(d) except in the accuracy of the overall level of the spectrum

estimate. For $N = 64$, it barely resolves the peak at $f = 0.1/T_s$ and does not resolve the pair of peaks near $f = 0.2/T_s$, nor does it produce as accurate an estimate of the continuous part of the spectrum centered at $f = 0.35/T_s$. For $N = 256$, both resolution and variability are poorer than they are for both the ML method (Figure 9-10d) and the periodogram method (Figure 9-9(c)–(d)). The Burg method is also consistently inferior to the ML method (except for the accuracy of the overall level and the relative heights of the weak and strong spectral lines) because of its higher variability. However, the Burg method is superior to the periodogram method for $N = 64$ (Figure 9-9(c)–(d)). The FB method clearly does the best job of resolving the three spectral lines for $N = 64$, and it is clearly the best in terms of low variability in the locations of the spectral lines. However, the Burg method is superior in terms of the accuracy of estimating the relative heights of the weak and strong spectral lines. For $N = 256$, the same conclusions apply in comparing the Burg and FB methods. However these two methods are inferior to the periodogram methods (and the ML method in Figure 9-10(d) except for the accuracy of the overall level) for $N = 256$. Comparison of Figures 9-12(b)–(c) and 9-13(b)–(c) with Figure 9-9(c)–(d) reveals that the periodogram method provides considerably more accurate estimates of the limit spectrum.

The superiority of the Burg and FB methods relative to the periodogram method for $N = 64$ must be tempered by the fact that these two parametric methods were given the unfair advantage of having the best model orders specified. When this advantage is removed by using the estimated orders produced by the model-order-determining methods of Part 3 in Section C, both parametric methods become inferior to the periodogram method because the estimated order is too high for $N = 64$, resulting in spurious peaks and high variability.

Although all three order-determining-methods, FPE, AIC, and CAT, produced comparable order estimates for each spectrum estimation method, these estimates were too low for the YW method ($5 \leqslant M \leqslant 11$ for $N = 64$ and $12 \leqslant M \leqslant 22$ for $N = 256$) and too high for the Burg method ($M \geqslant 48$ for $N = 64$ and $M \geqslant 84$ for $N = 256$) and for the FB method, except for $N = 256$ ($M \cong 24$ for $N = 64$, $M \cong 32$ for $N = 256$). Unfortunately, in practice where there is only one data segment to analyze, there is apparently no way to determine from these parametric methods which peaks are correct and which are spurious. Sample spectrum estimates with estimated orders are shown in Figures 9-14, 9-15, and 9-16.

4. Overdetermined-Normal-Equations AR Method

The spectrum estimates provided by the overdetermined-normal-equations (ODNE) AR method specified by (90) and (69) are shown in Figures 9-17(a)–(c) and 9-18(a)–(c) for the parameter Q, which specifies the number of normal equations, given by $Q = 48$. The cases included are $N = 64$ with $M = 16$ and $N = 256$ with $M = 16$ and $M = 32$. The results shown in Figure 9-17(a)–(c) were obtained using the biased correlation estimates from the autocorrelation method of least squares, whereas the results shown in Figure 18(a)–(c) were

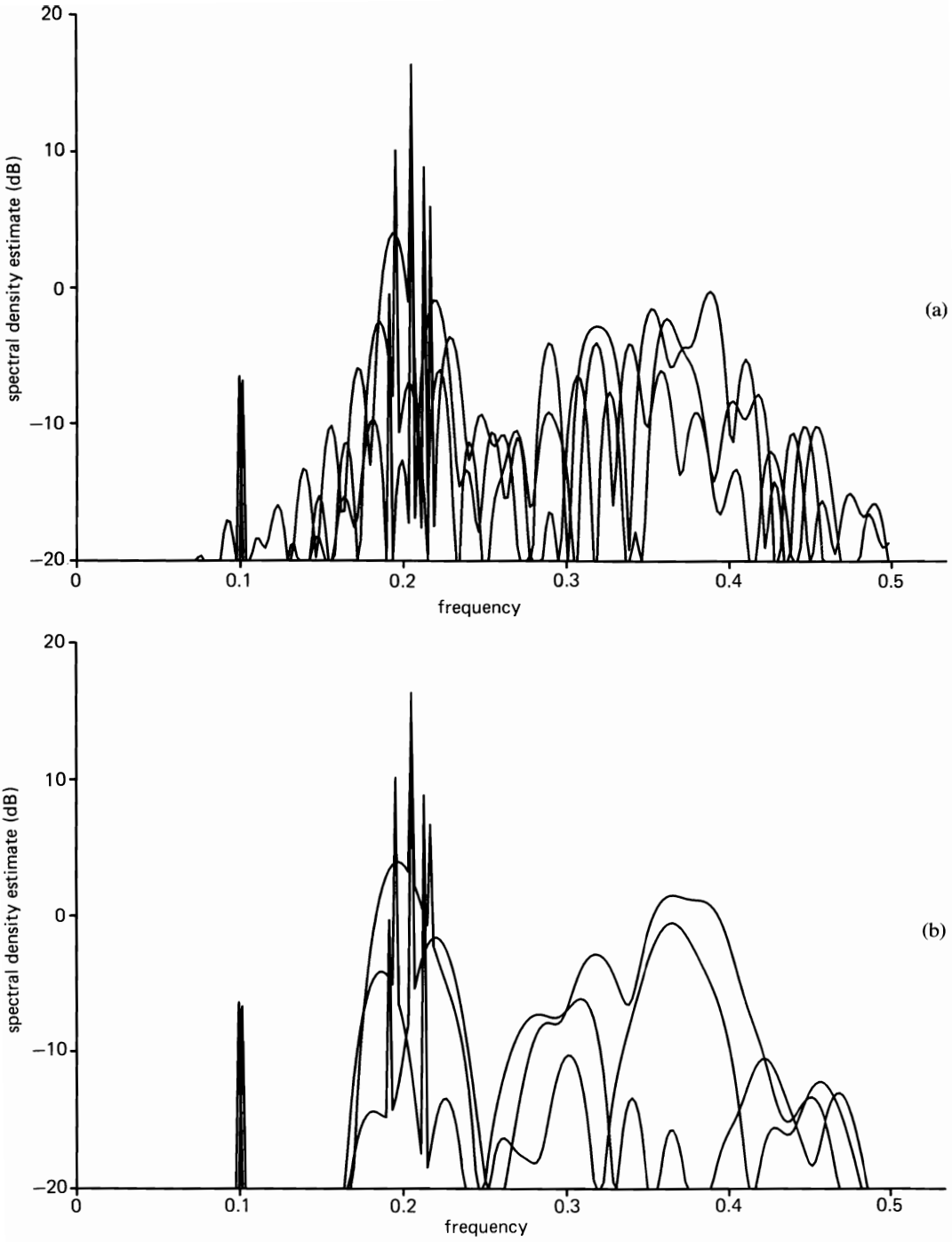
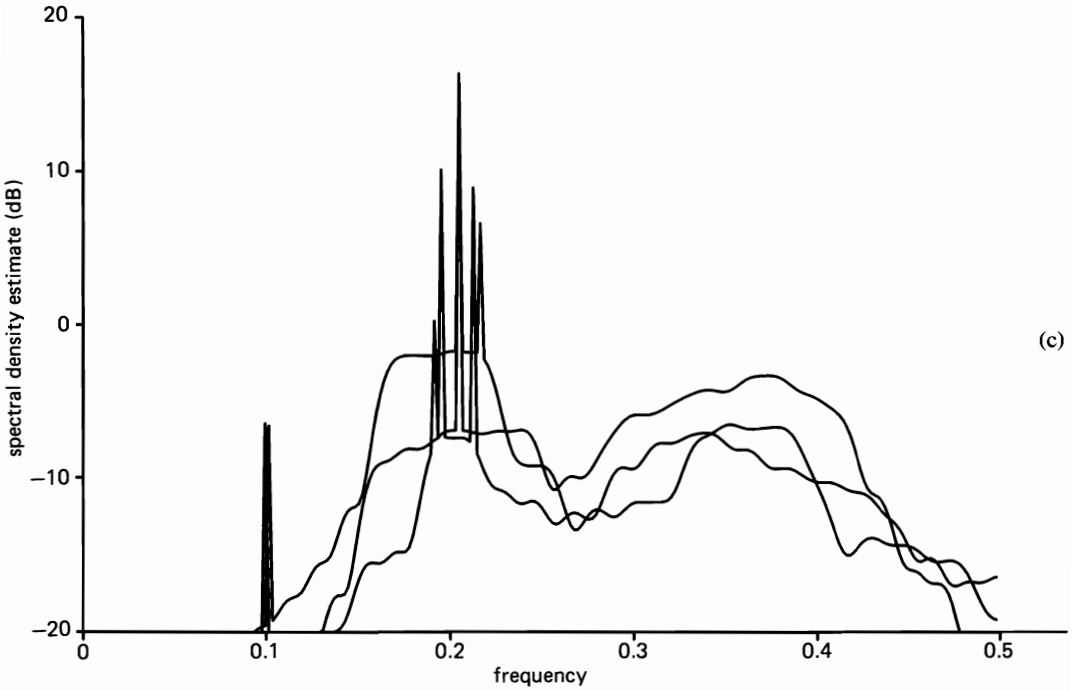
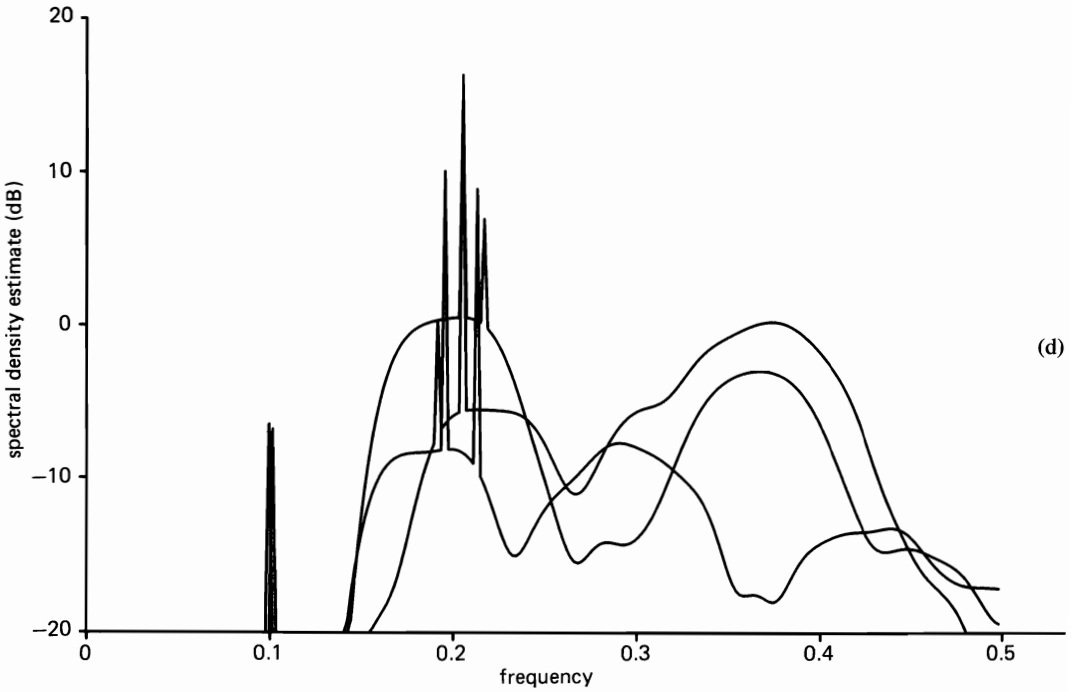


Figure 9-8 (a) Three periodograms with sine-wave removal and spectral line reinsertion for $N = 64$ data points each, and with $K = 8$ zero-padding factor and no data tapering. (b) Same as (a) but with raised-cosine data tapering.



(c)



(d)

Figure 9-8 (continued) (c) Same as (a) but with frequency smoothing ($M = 4$). (d) Same as (b) but with frequency smoothing ($M = 4$).

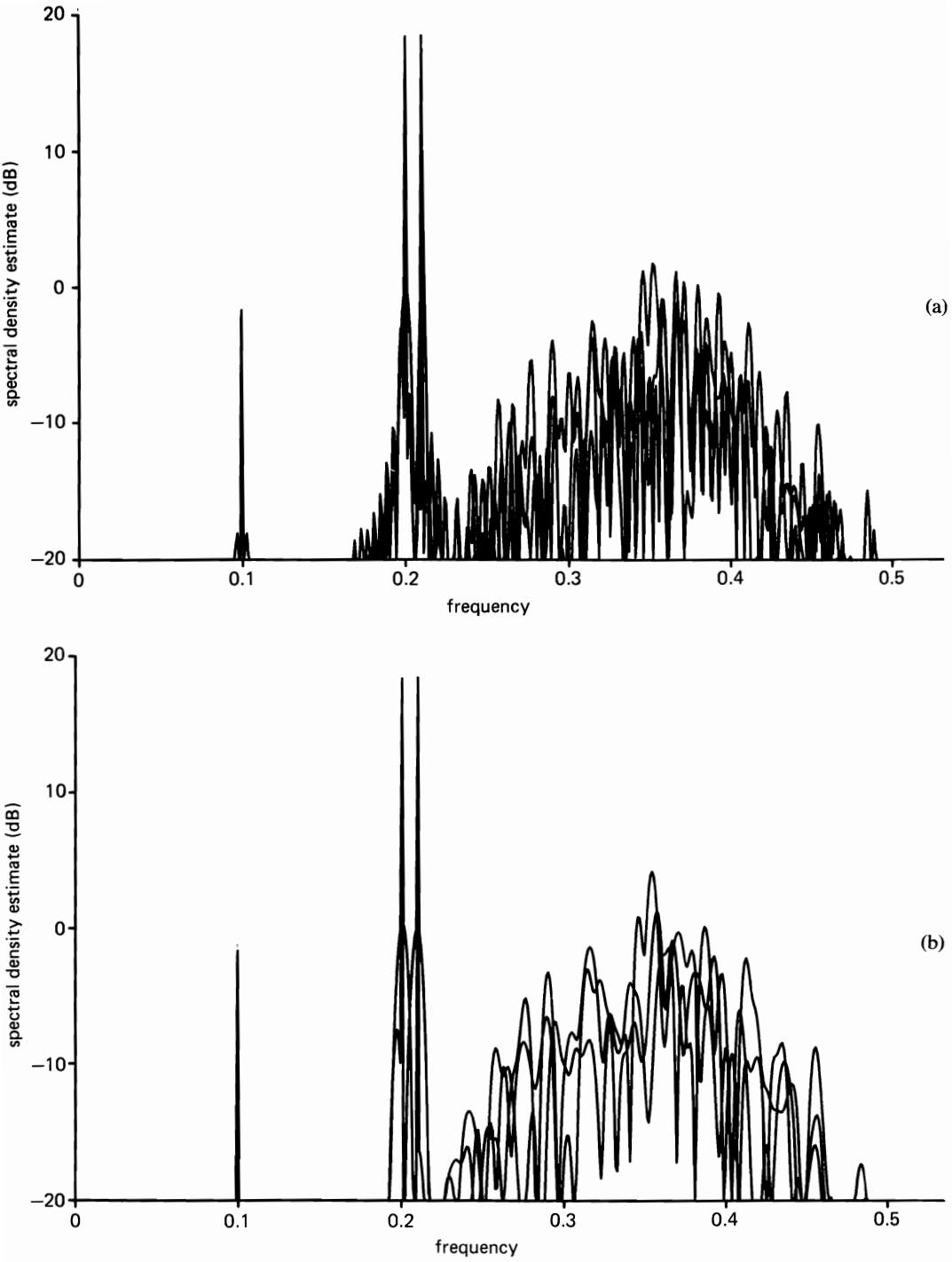


Figure 9-9 (a) Three periodograms with sine-wave removal and spectral line reinsertion for $N = 256$ data points each, and with $K = 8$ zero-padding factor and no data tapering. (b) Same as (a) but with raised-cosine data tapering.

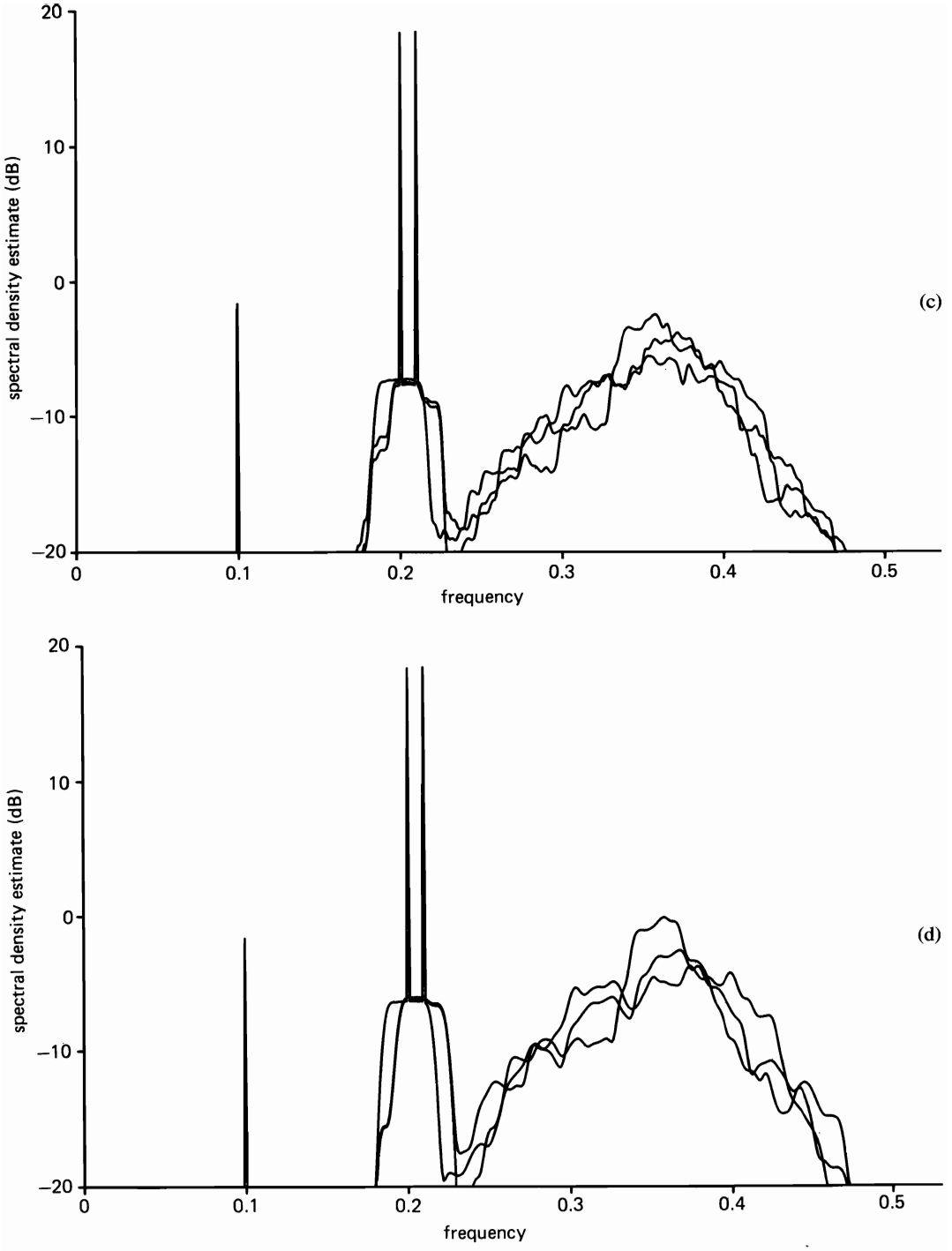


Figure 9-9 (continued) (c) Same as (a) but with frequency smoothing ($M = 8$). (d) Same as (b) but with frequency smoothing ($M = 8$).

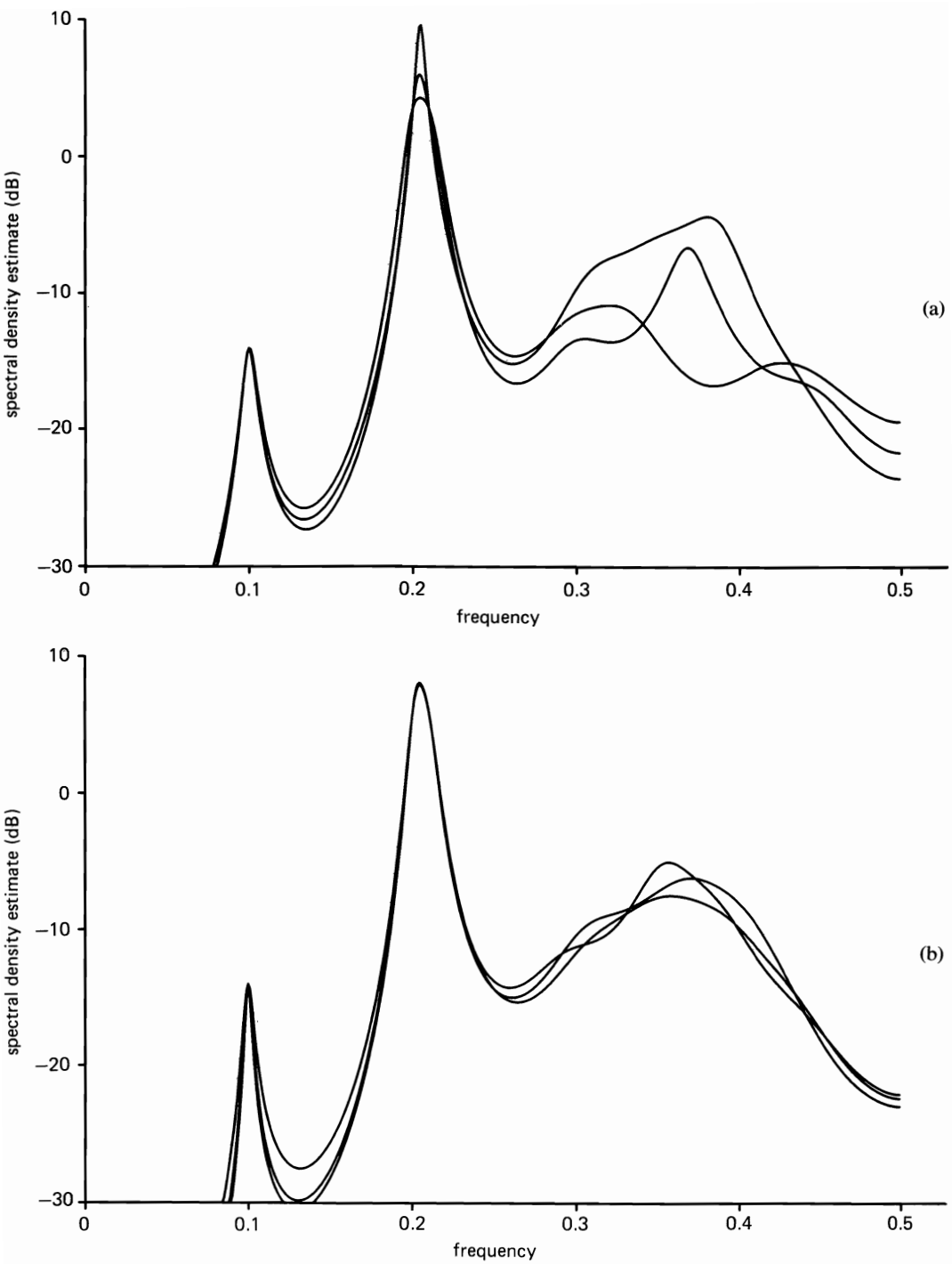


Figure 9-10 (a) Three minimum-leakage spectrum estimates with $N = 64$, $M = 16$. (b) Same as (a) but with $N = 256$, $M = 16$.

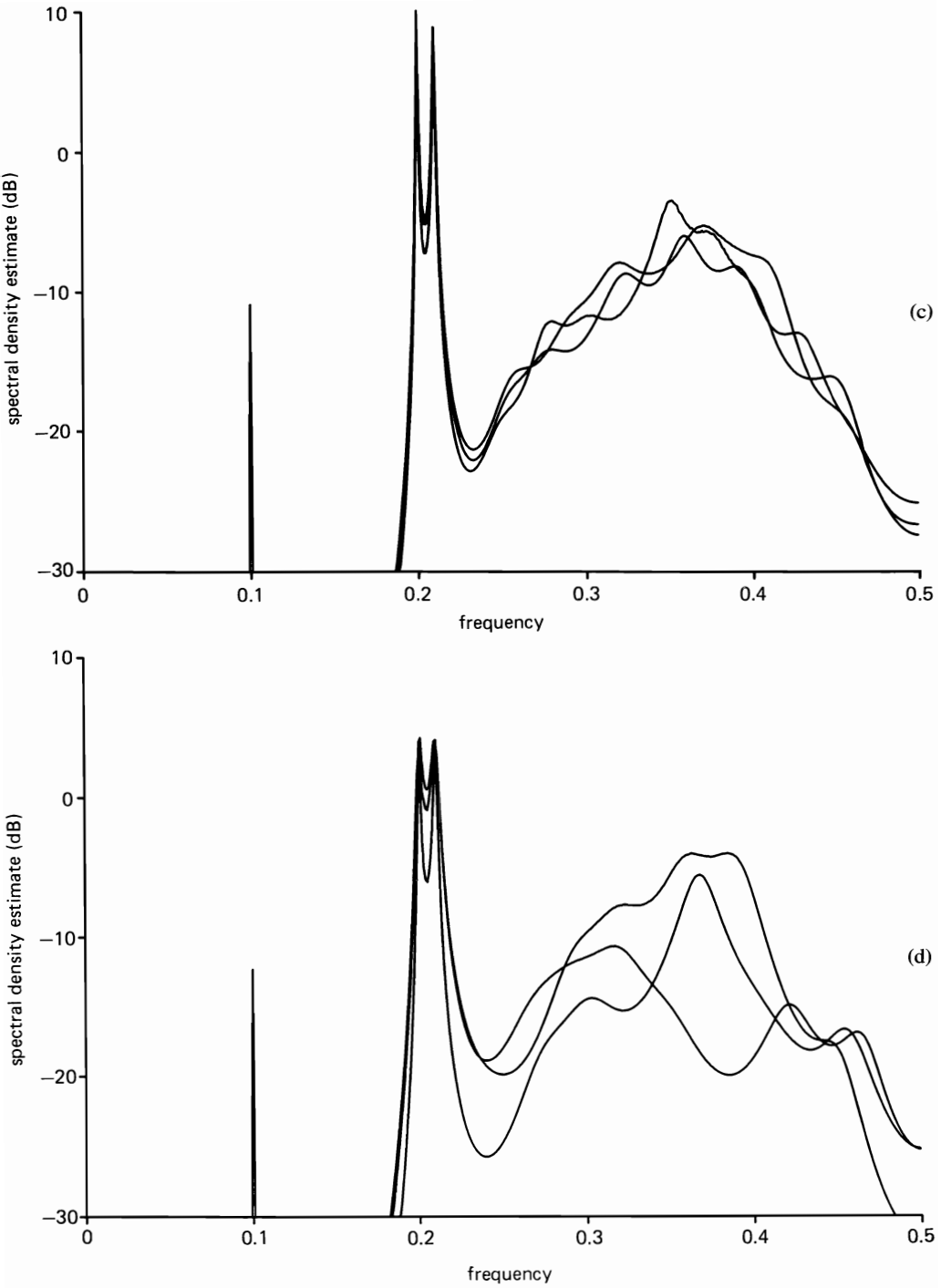


Figure 9-10 (continued) (c) Same as (a) but with $N = 256$, $M = 32$. (d) Same as (a) but with $M = 24$.

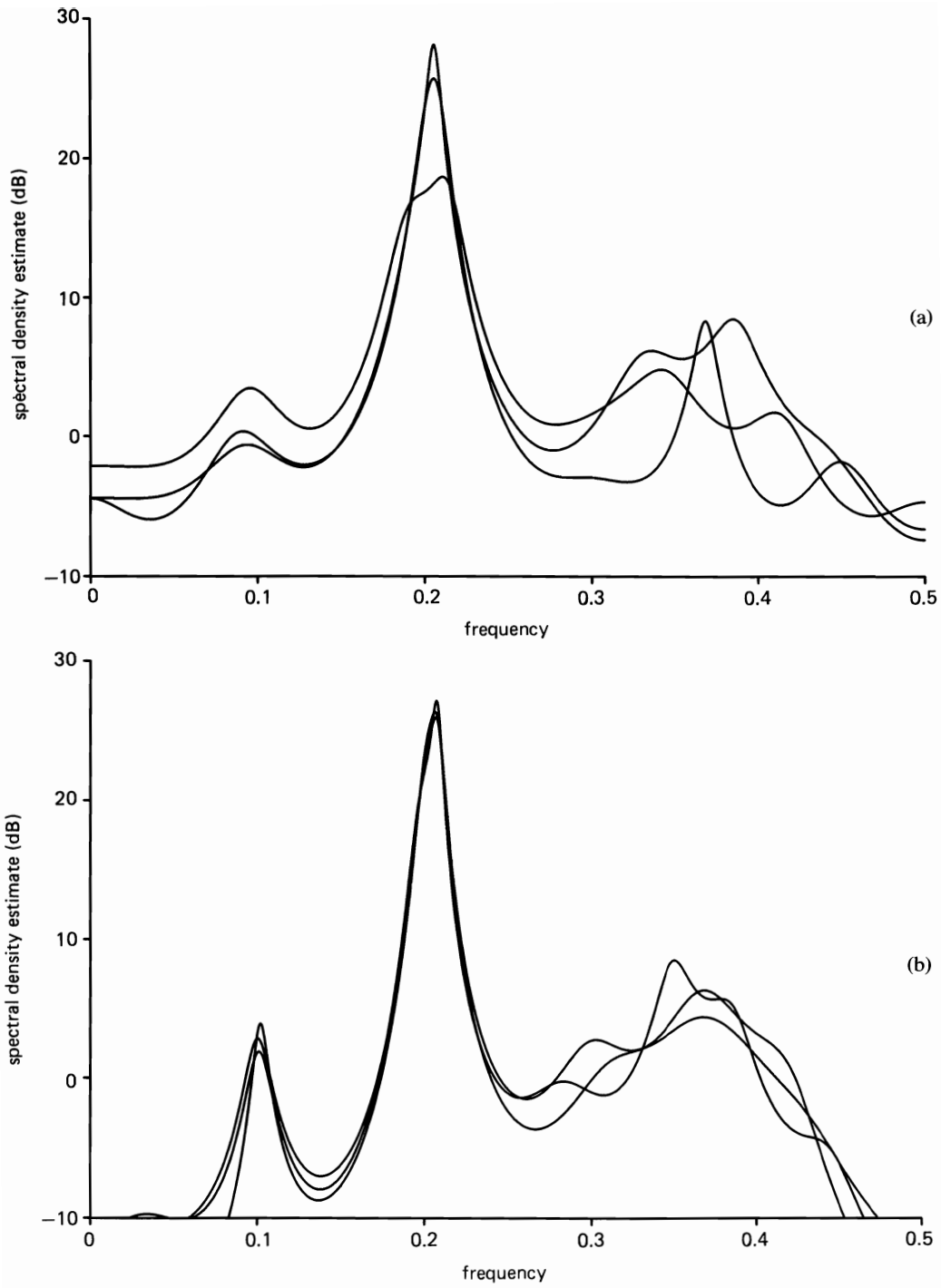


Figure 9-11 (a) Three Yule-Walker AR spectrum estimates with $N = 64$, $M = 16$. (b) Same as (a) but with $N = 256$, $M = 16$.

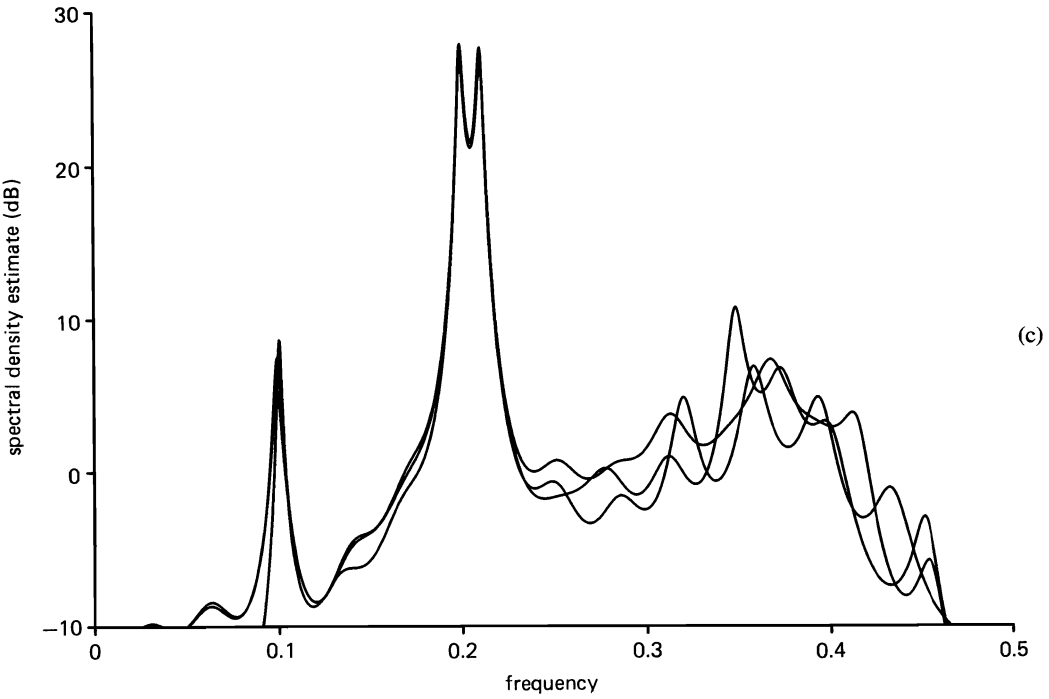


Figure 9-11 (continued) (c) Same as (a) but with $N = 256$, $M = 32$.

obtained using the corresponding unbiased estimates (see exercise 12). As expected, the method that uses the unbiased estimates provides better resolution but the increase in variability is surprisingly small. For $N = 64$, the overall performance of the unbiased ODNE method is slightly inferior to that of the Burg method and is strongly inferior to that of the FB method. For $N = 256$, the performance of the unbiased ODNE method is comparable to that of the Burg and FB methods and is, therefore, inferior to that of both the ML method (Figure 9-10(d)) and the periodogram method (Figure 9-9(c)–(d)). Also, the accuracy of the overall level of the spectrum estimate provided by both ODNE methods is poor.

5. Singular-Value-Decomposition AR Method

The spectrum estimates provided by the singular-value-decomposition (SVD) method specified by (102) and (69) are shown in Figure 9-19(a)–(d). Since the data is not from an AR model and does not consist of simply strong sinewaves in a white noise background, the eigenvalues do not partition into one set of relatively large values corresponding to spectral features of interest and a remaining set of negligible values or a set of small values corresponding to a flat spectrum. As a consequence, the procedure of using the test (96) to determine a best rank M_0 cannot be expected to perform well, particularly since there will be no obvious way to set the threshold ratio ρ in (96). Therefore, M_0 was just chosen to be

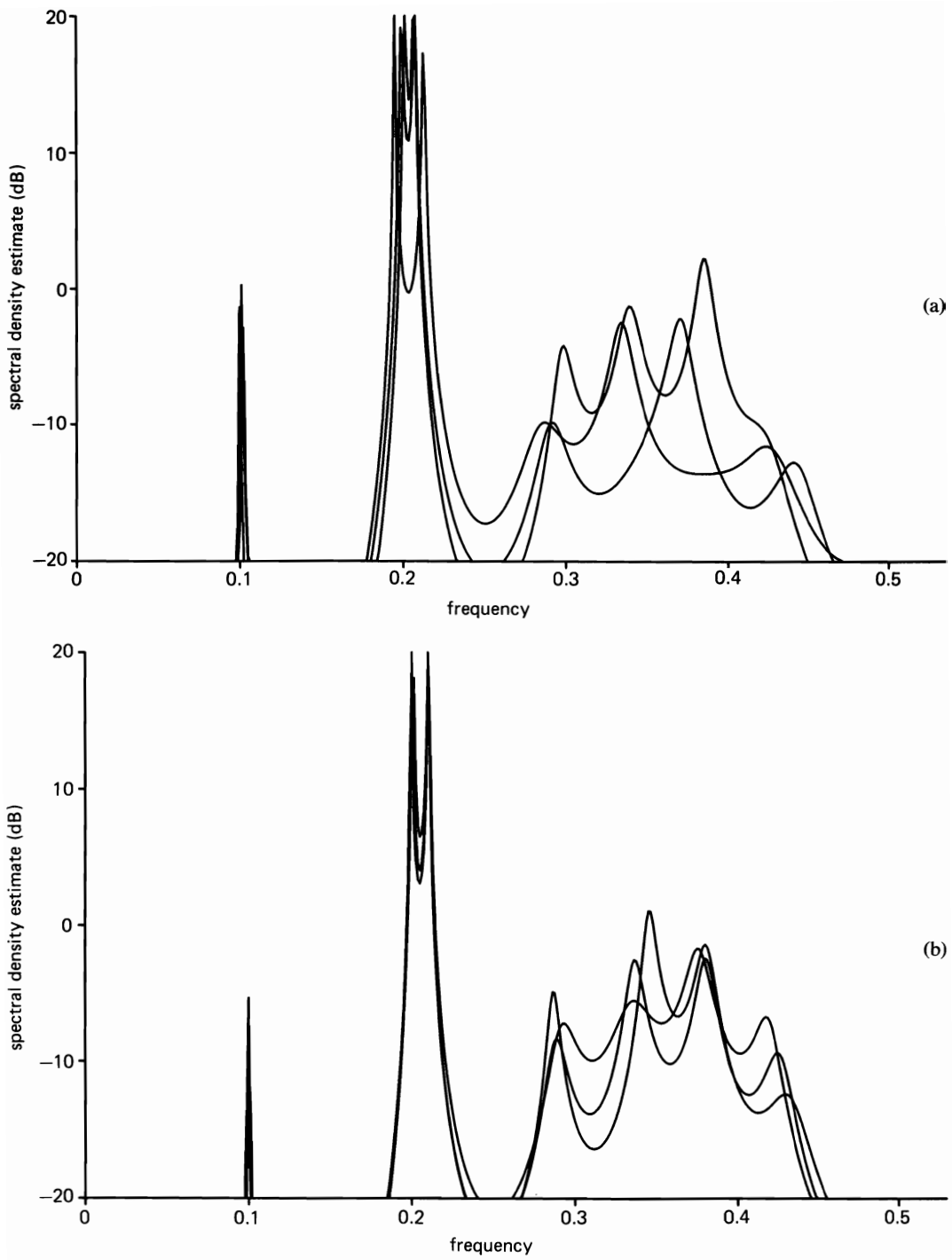


Figure 9-12 (a) Three Burg AR spectrum estimates with $N = 64$, $M = 16$ (highest peak is 33 dB). (b) Same as (a) but with $N = 256$, $M = 16$ (highest peak = 34 dB).

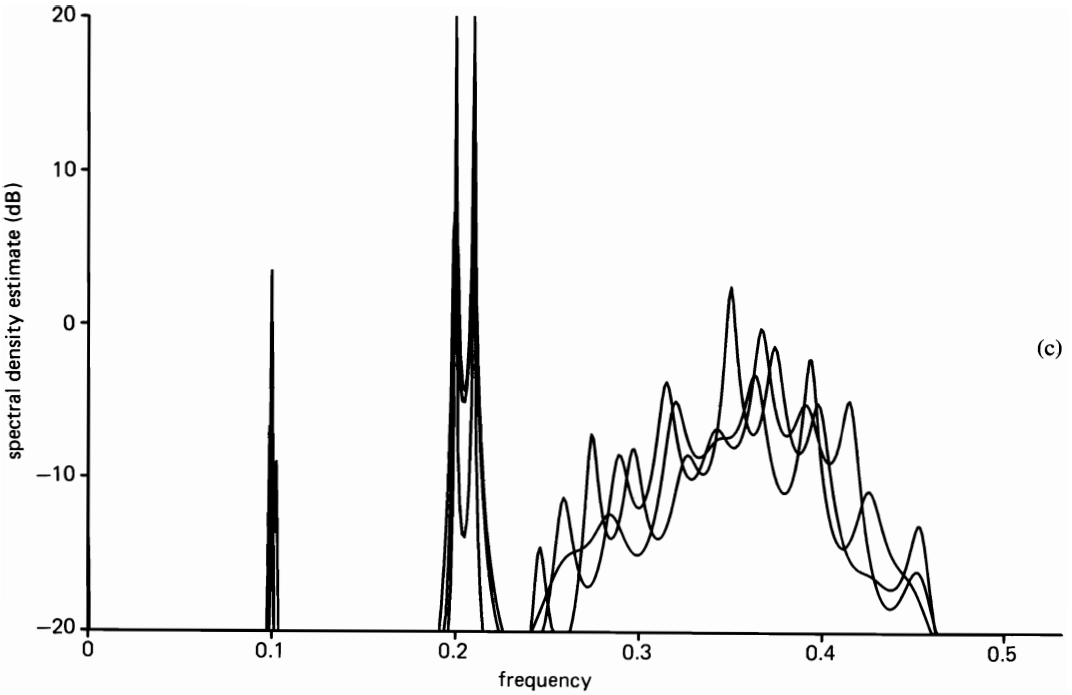


Figure 9-12 (continued) (c) Same as (a) but with $N = 256$, $M = 32$ (highest peak = 35 dB).

$M_0 = 16$ and $M_0 = 32$, and the corresponding ratio ρ was calculated (as an aside). For $M_0 = 16$, $\rho \cong 0.96$ for $N = 64$ and $\rho \cong 0.97$ for $N = 256$. For $M_0 = 32$, $\rho \cong 0.99$ for $N = 64$ and $N = 256$. It can be seen from Figure 9-19(a)–(d), that for both $N = 64$ and $N = 256$ the performance is inferior to that of all other parametric methods as well as the ML method and the periodogram method (Figures 9-8(c)–(d) and 9-9(c)–(d)). As with all parametric methods, experimentation with the order M_0 showed that the best results were obtained with $M_0 \cong 16$ for $N = 64$ and with $M_0 \cong 32$ for $N = 256$.

6. Hybrid Method

It can be seen from the preceding results that the best possible spectrum estimate can be obtained by exploiting the strong points of each of several methods. To illustrate the quality of spectrum estimates obtainable with such hybrid methods, the FB method was used to detect and estimate the frequencies of spectral lines, and the DFT (with $K = 8$ zero-padding factor) was then used to estimate the amplitudes and phases of the three corresponding sinewaves; after subtraction of these estimated sinewaves from the data, the 32-point smoothed periodogram ($M = 4$ and $K = 8$) was computed. The results are shown in Figure 9-20. This is clearly the most accurate of all the spectrum estimates considered for short data segments ($N = 64$).

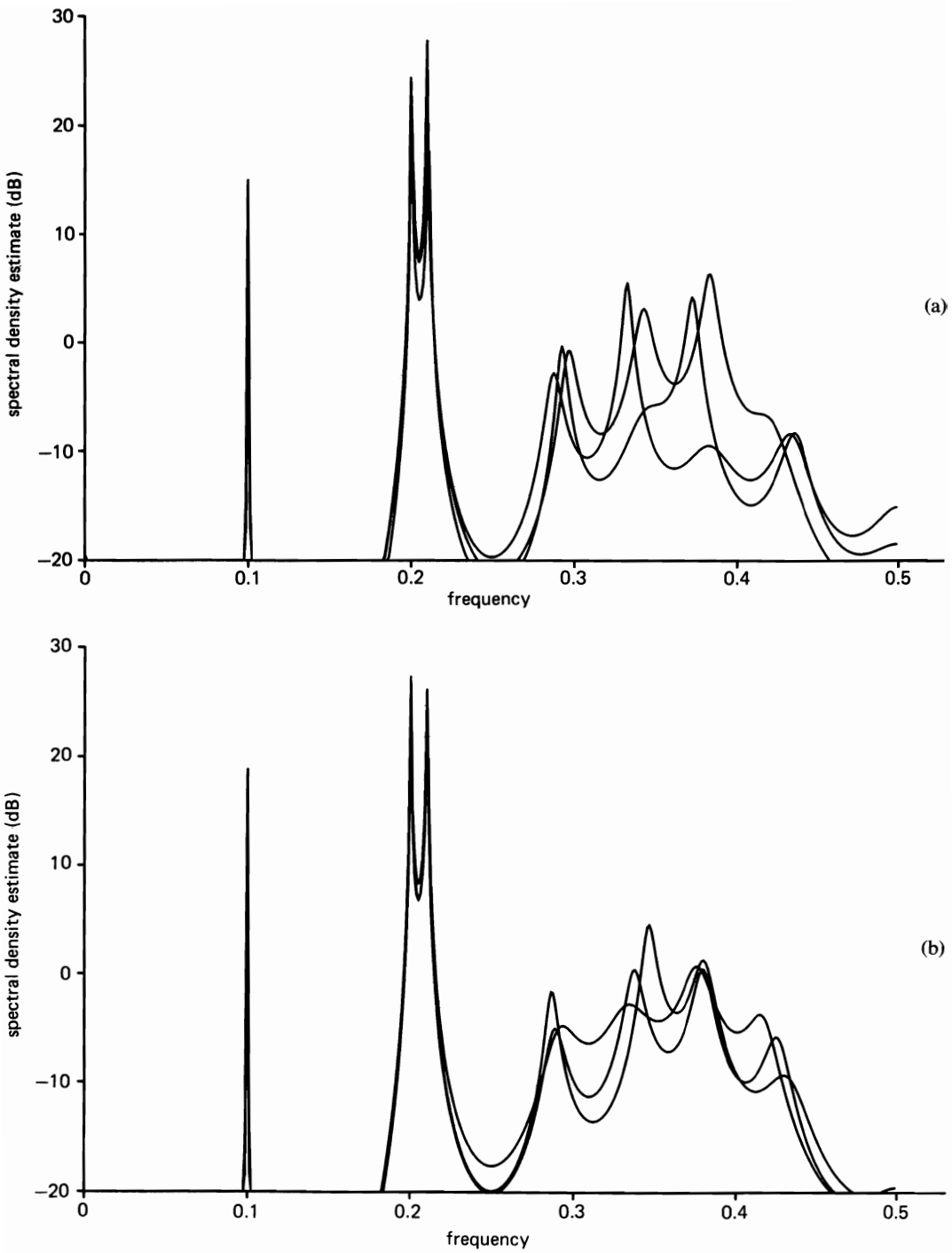


Figure 9-13 (a) Three forward-backward linear-prediction AR spectrum estimates with $N = 64$, $M = 16$. (b) Same as (a) but with $N = 256$, $M = 16$.

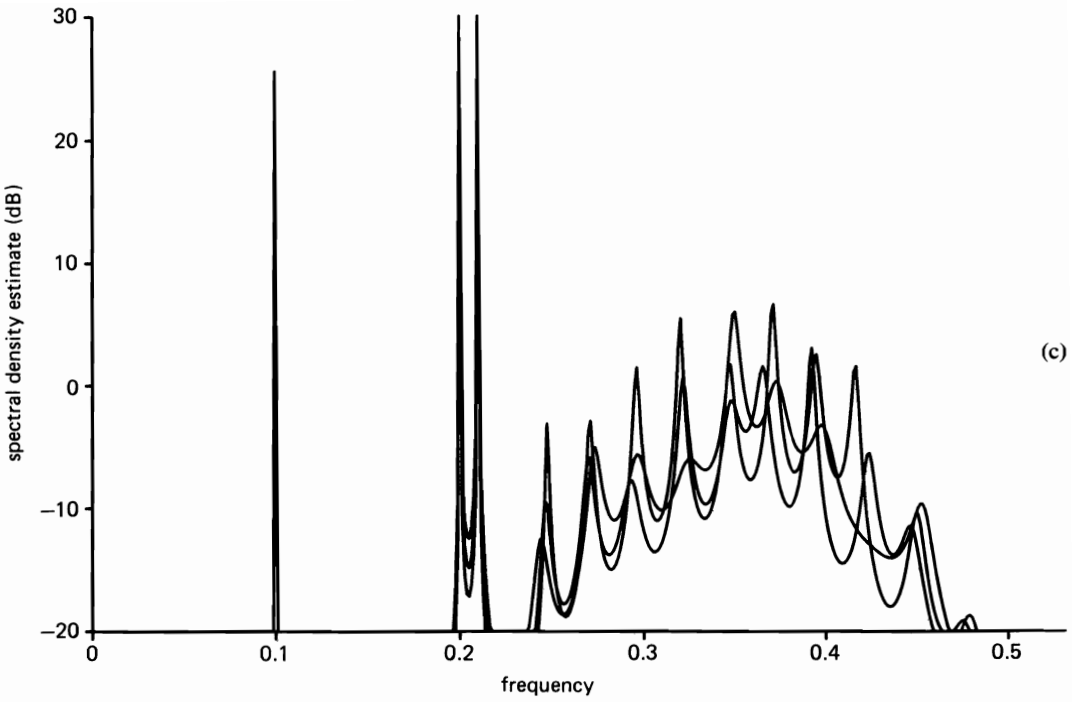


Figure 9-13 (continued) (c) Same as (a) but with $N = 256$, $M = 32$ (highest peak is 39 dB).

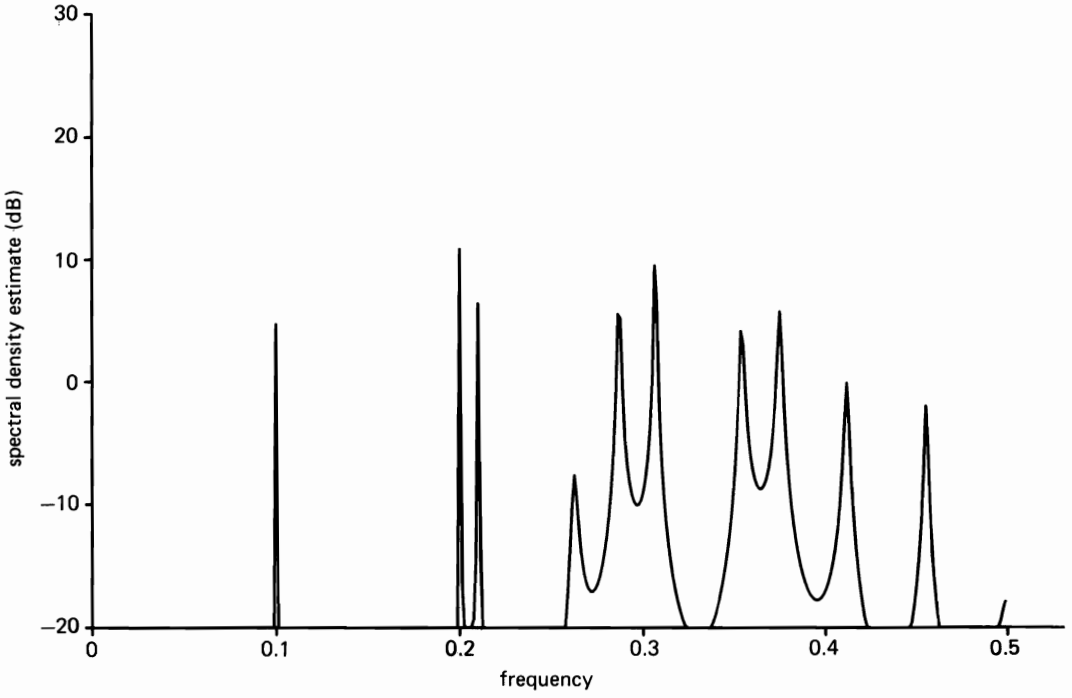


Figure 9-14 Sample of a poor spectrum estimate obtained with the forward-backward linear-prediction method with too high an order ($M = 24$) for $N = 64$.

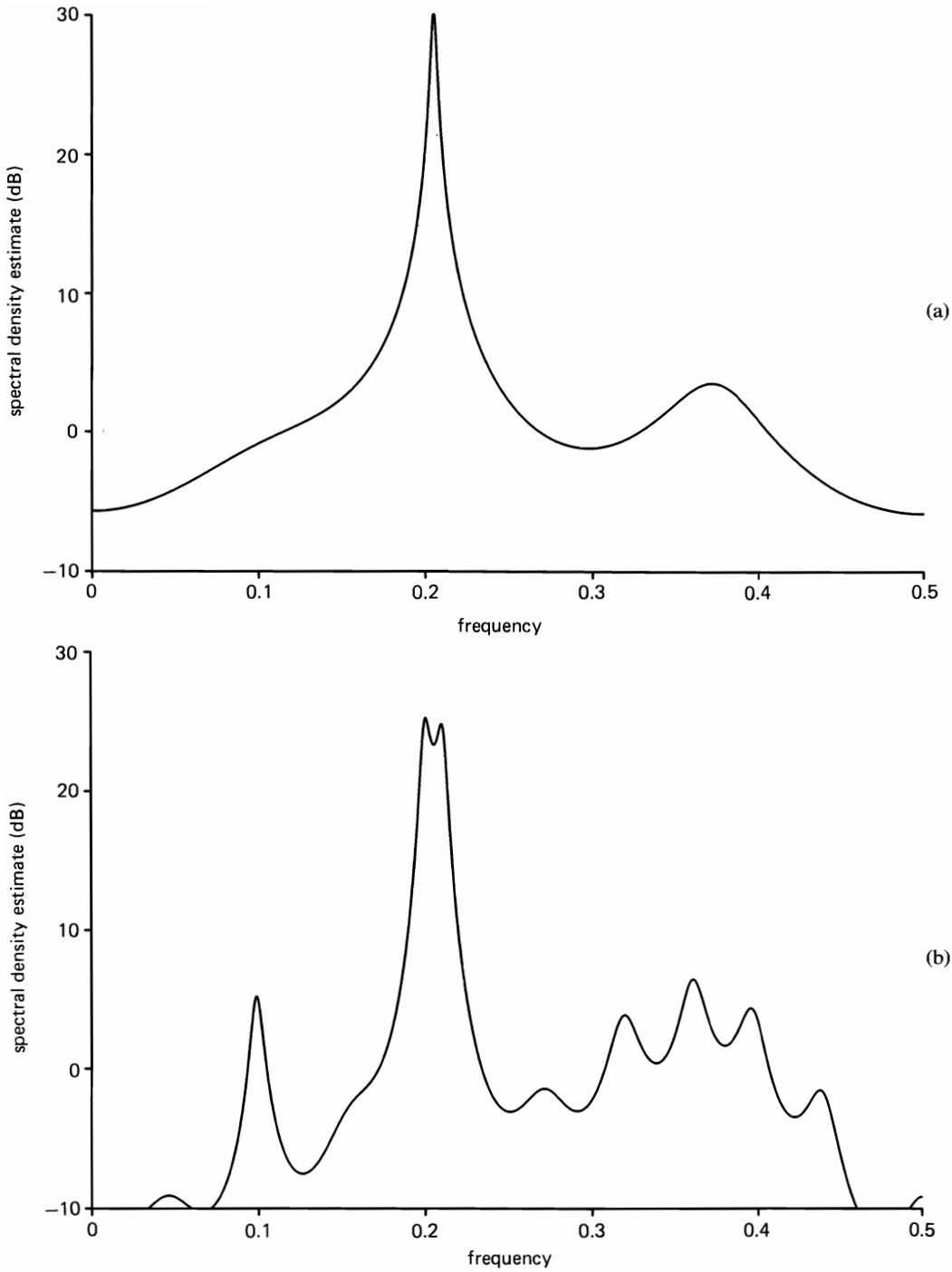


Figure 9-15 (a) Sample of poor spectrum estimates obtained with the Yule-Walker AR method with too low an order ($M = 7$) for $N = 64$. (b) Same as (a) but with $M = 22$, $N = 256$.

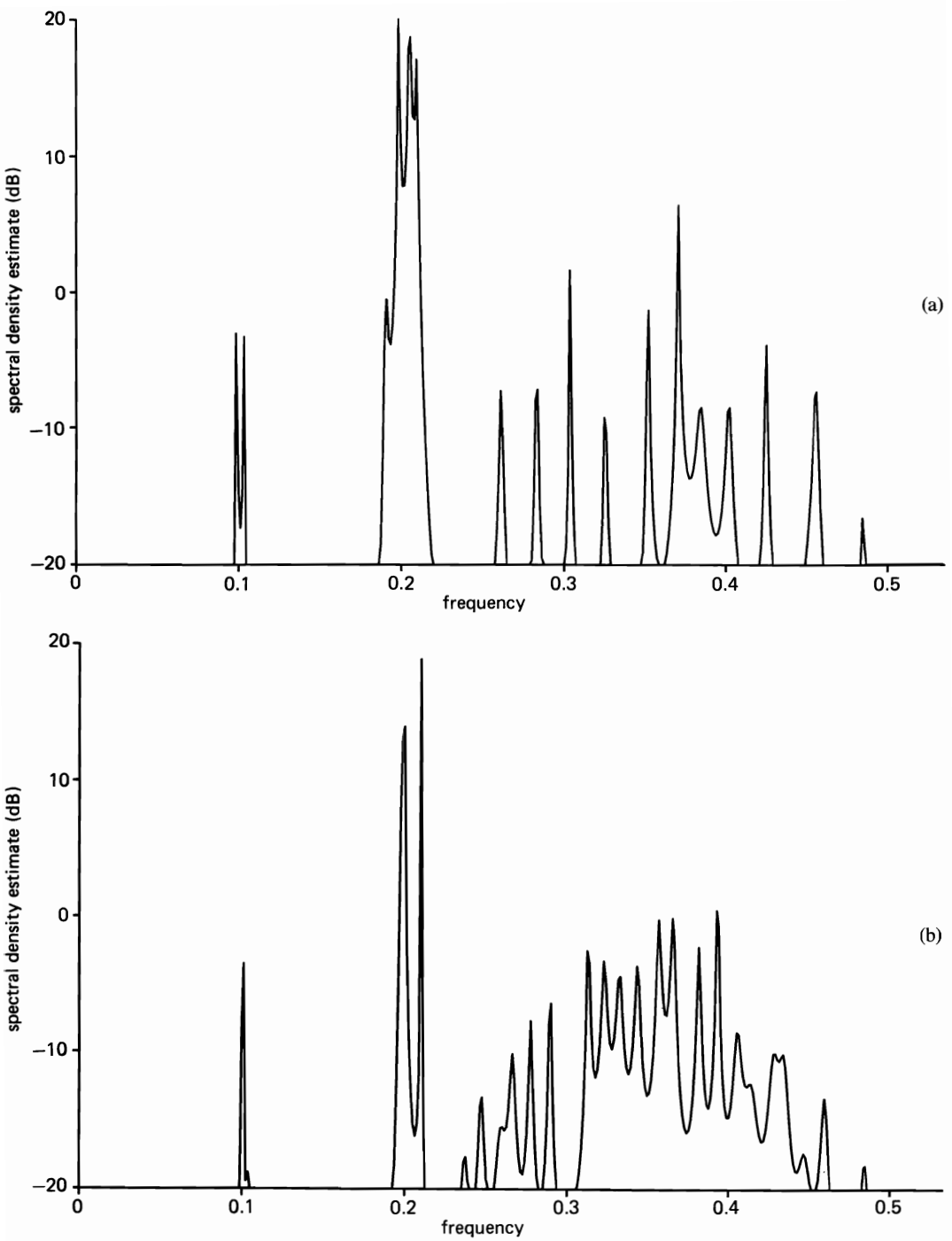


Figure 9-16 (a) Sample of poor spectrum estimates obtained with the Burg AR method with too high an order ($M = 48$) for $N = 64$. (b) Same as (a) but with $M = 90$, $N = 256$.

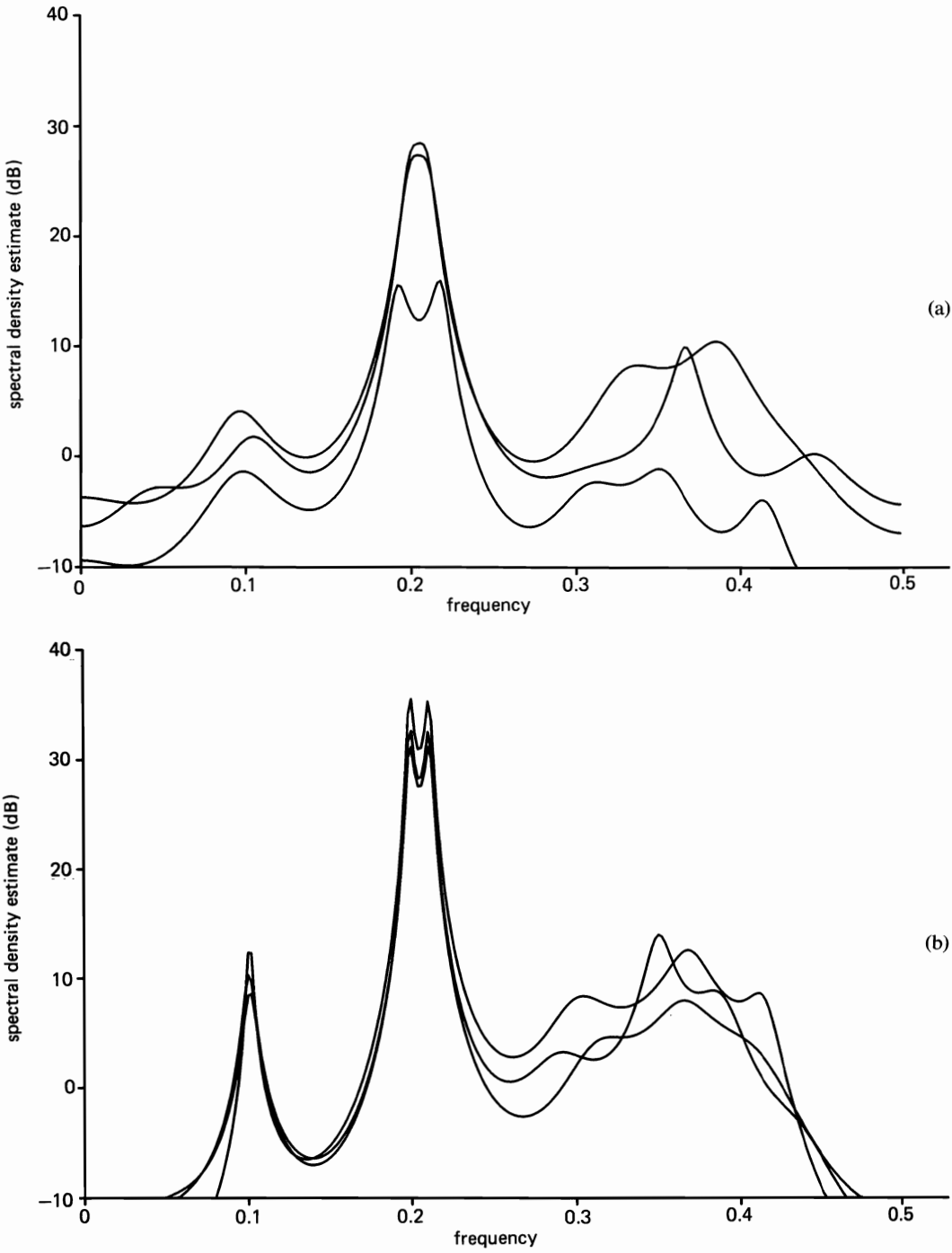


Figure 9-17 (a) Three overdetermined-normal-equations AR spectrum estimates obtained using biased autocorrelation estimates and $Q = 48$ normal equations with $N = 64$, $M = 16$. (b) Same as (a) but with $N = 256$, $M = 16$.

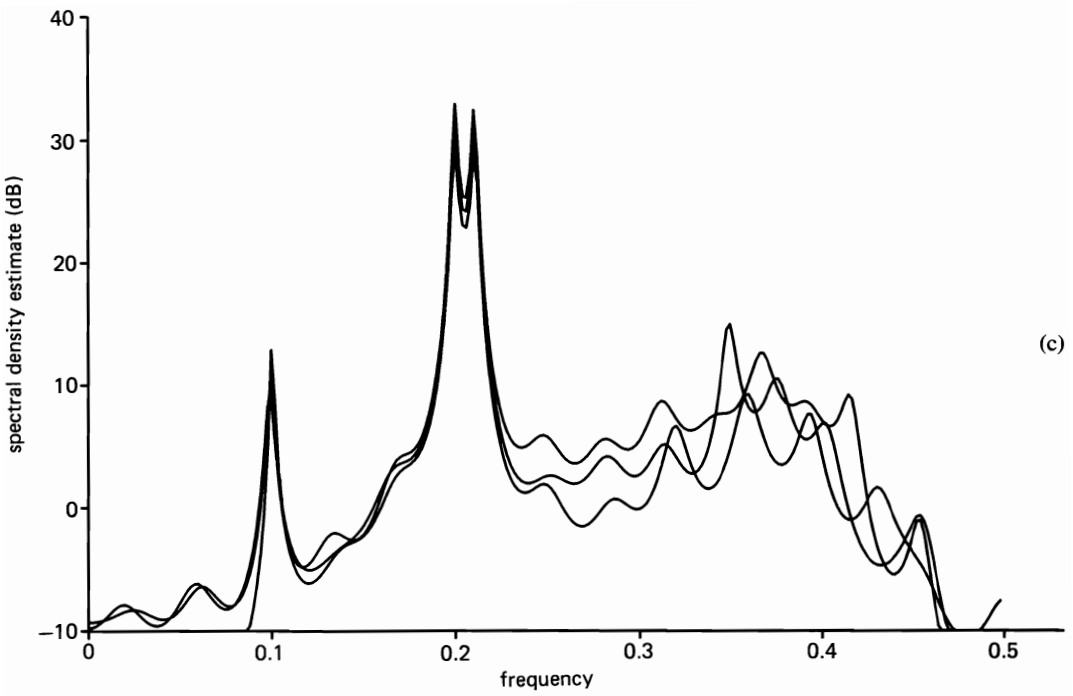


Figure 9-17 (continued) (c) Same as (a) but with $N = 256$, $M = 32$.

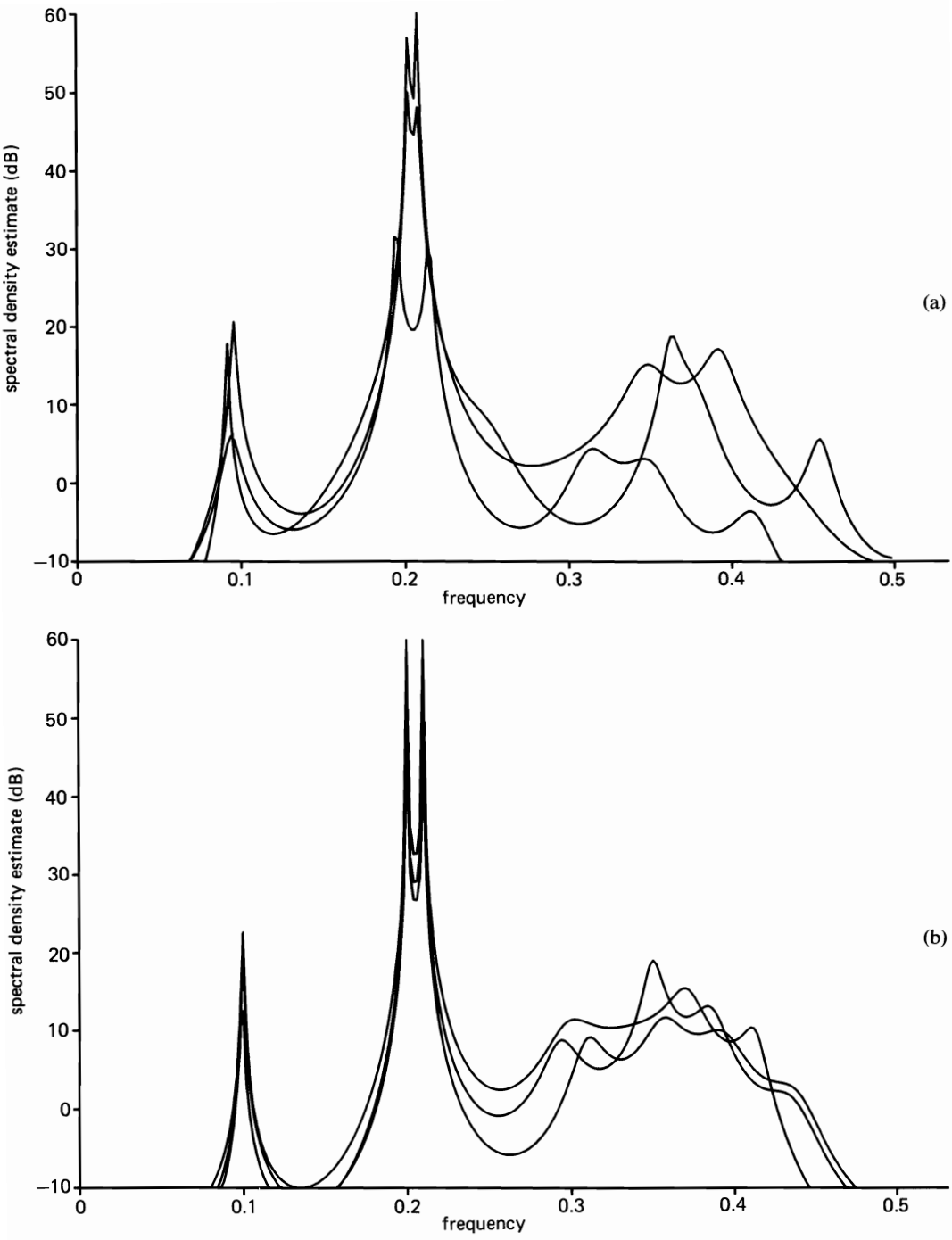


Figure 9-18 (a) Three overdetermined-normal-equations AR spectrum estimates obtained using unbiased autocorrelation estimates and $Q = 48$ normal equations with $N = 64$, $M = 16$. (b) Same as (a) but with $N = 256$, $M = 16$ (highest peak is 64 dB).

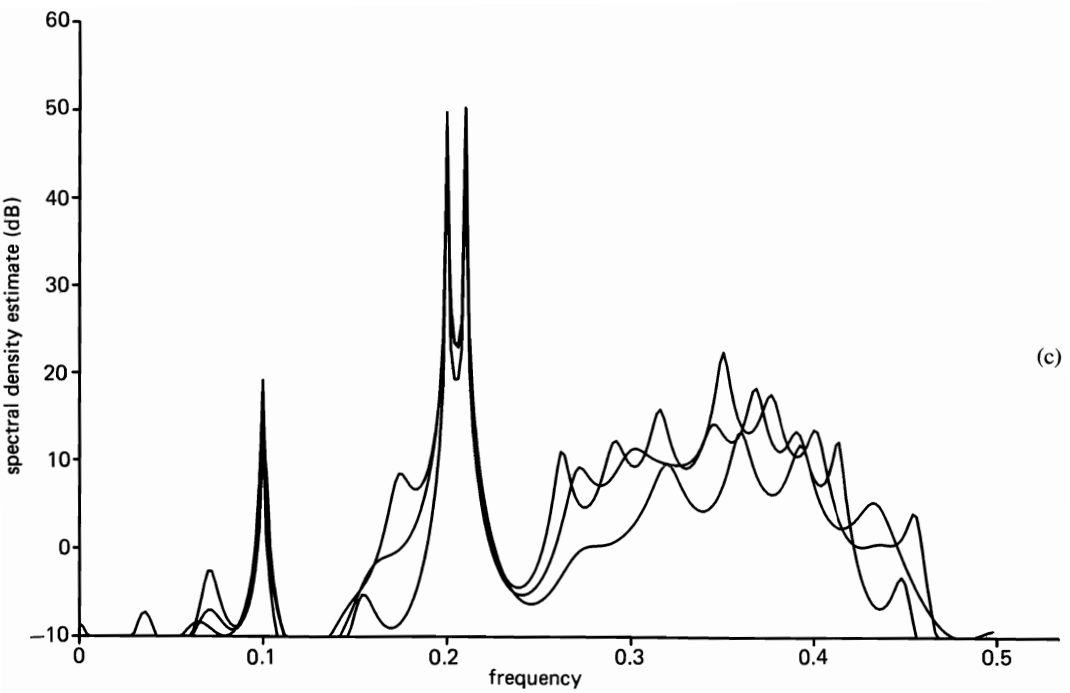


Figure 9-18 (continued) (c) Same as (a) but with $N = 256$, $M = 32$.

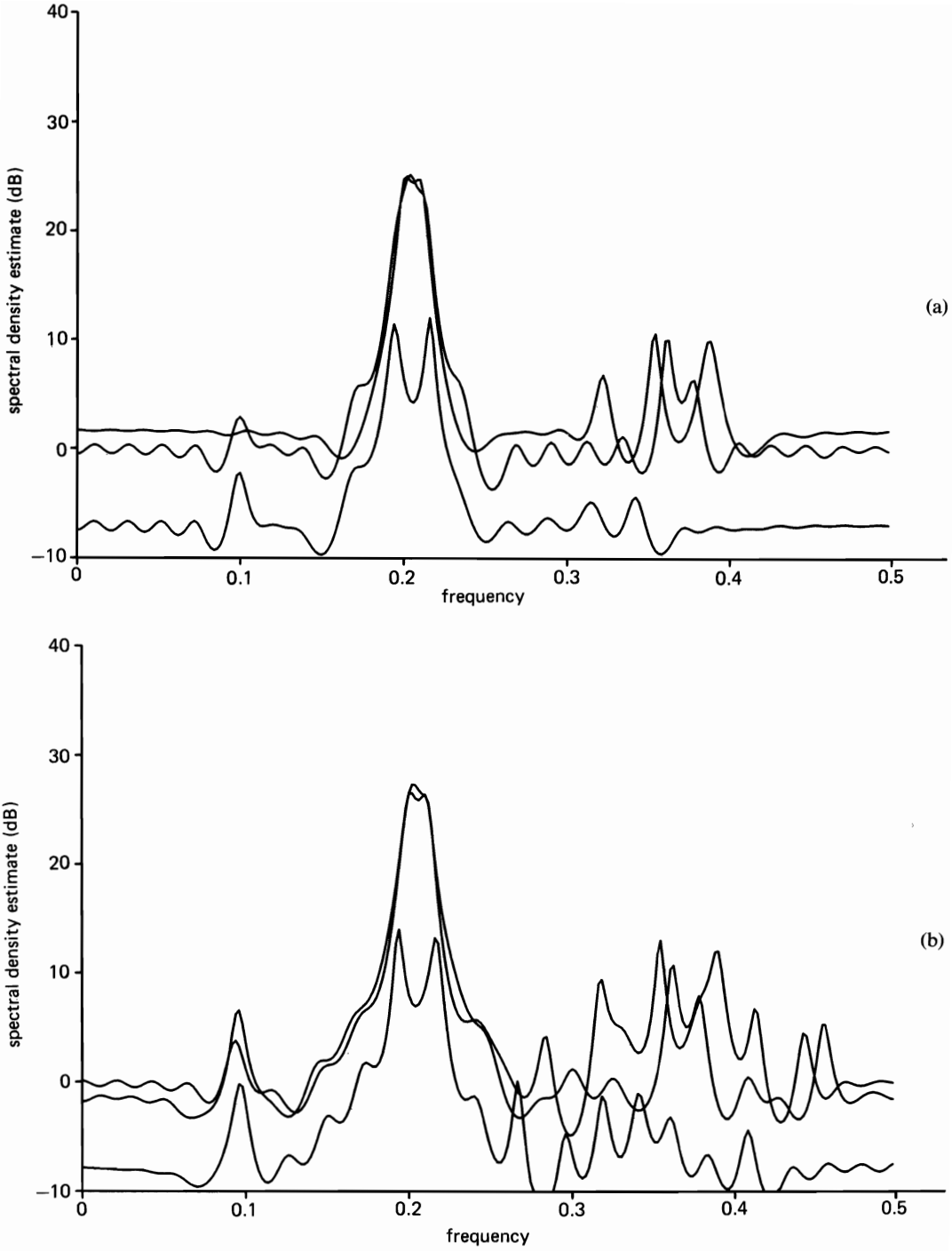
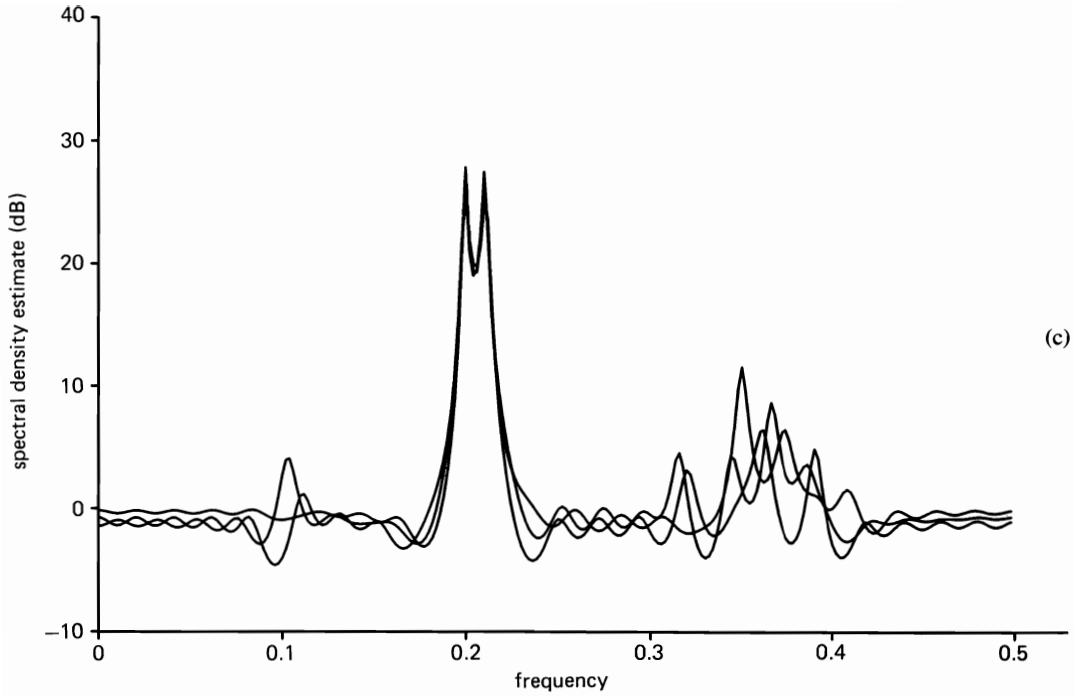


Figure 9-19 (a) Three singular-value-decomposition spectrum estimates obtained using $Q = 48$, $N = 64$, $M_0 = 16$. (b) Same as (a) but with $M_0 = 32$.



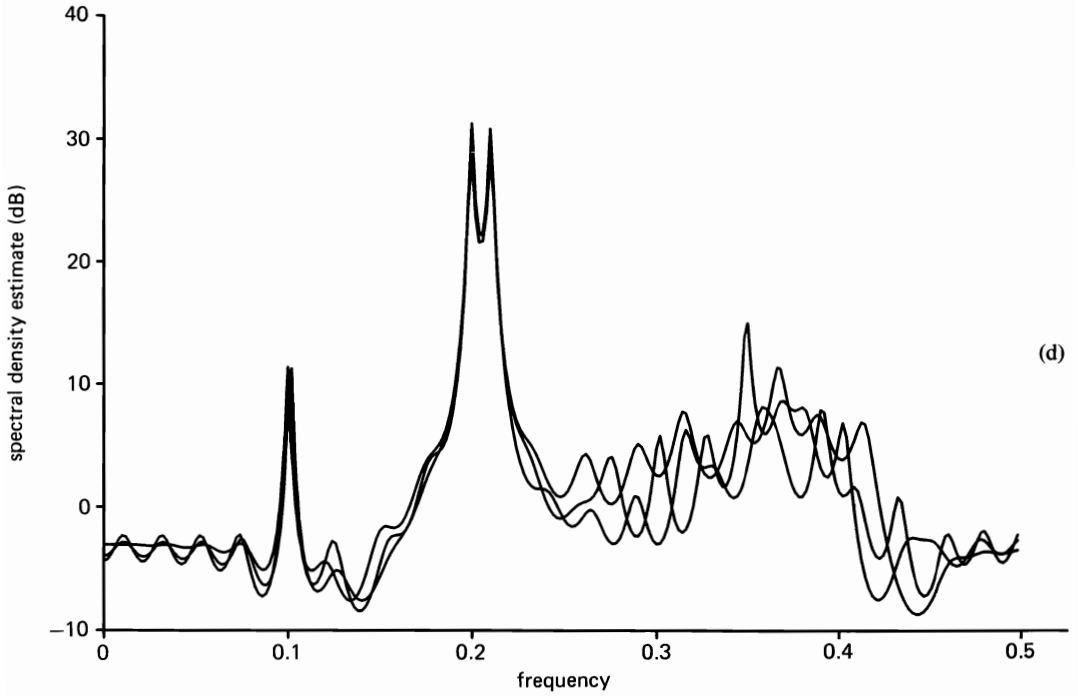


Figure 9-19 (continued) (c) Same as (a) but with $N = 256$, $M_0 = 16$. (d) Same as (a) but with $N = 256$, $M_0 = 32$.

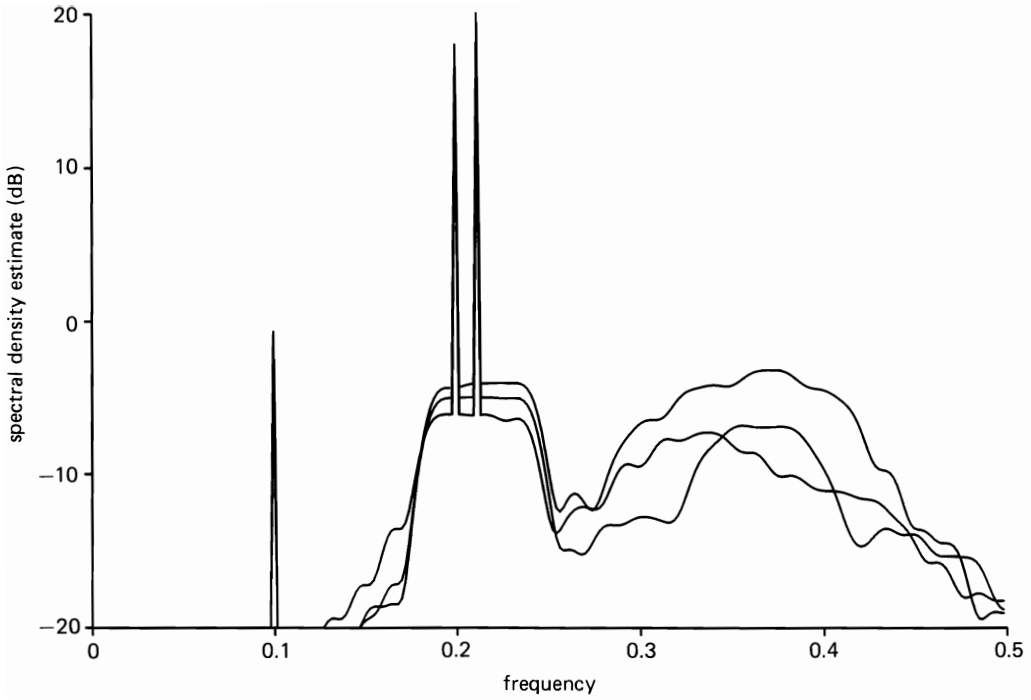


Figure 9-20 Three spectrum estimates obtained using the hybrid method described in Part 6 of Section E with $N = 64$.

F. SUMMARY

In Section B, the theory of autoregressive modeling of time-series, which underlies many of the parametric methods of spectral analysis, is presented in a concise but thorough form. This includes the topics of Yule-Walker equations, Levinson-Durbin algorithm, linear prediction, Wold-Cramér decomposition, maximum-entropy model, Lattice filter, and Cholesky factorization. Then in Section C, after a general discussion of the relative merits of AR, MA, and ARMA modeling, the theory of autoregressive modeling is exploited in a step-by-step development of the most popular AR model-fitting methods for spectral analysis. These include 1) the Yule-Walker (or autocorrelation type of least squares linear predictive) method and its interpretations in terms of maximum-entropy and autocorrelation extrapolation; 2) the covariance type of least squares method and its improved forward-backward linear predictive version, and 3) the lattice-constrained variant of this, known as the Burg method; 4) the overdetermined-normal-equations variation on the least squares linear predictive methods (autocorrelation and autocovariance types), and 5) two modifications based on singular value decomposition of the data-correlation matrix; and 6) the maximum-likelihood approach. Also, four model-order-determining methods, known as final prediction error, information criterion, autoregressive transfer-function criterion, and singular-

value-decomposition, are described. The subsection on AR model fitting concludes with a comparative discussion of the relative merits of these various methods. In Section D, the many methods for ARMA model fitting are classified into three primary groups. The group that includes the most computationally attractive methods is then focused on, and the extended and modified Yule-Walker equations, which form the basis for these methods, are derived. It is explained that most of the AR methods can be simply adapted to the task of estimating the AR parameters in the ARMA model, and an adaptation of the ODNE-SVD method is described. Then a variety of methods for utilizing the AR estimates, together with the autocorrelation estimates or the data, to estimate the MA parameters are described. These include (1) direct methods preceded by an inverse AR filtering operation on the data, (2) a variation on the Blackman-Tukey and Wiener-Daniell versions of this approach that circumvents the data-filtering operation, (3) another variation that utilizes forward-backward filtering, (4) Shanks' method, which is based on a decomposition of the spectrum into causal and anticausal parts, and (5) a variation on this that uses the overdetermined-equations technique. The chapter concludes with an extensive experimental study that compares and contrasts the performances of many of the methods described herein.

EXERCISES

1. **(a)** Use the AR model (1) to verify (5) and (7).
   (b) Derive the Yule-Walker equations (8a) from (13) and the orthogonality condition (14).
   (c) Verify that for $M = 2$ the Levinson-Durbin algorithm yields the solution to the Yule-Walker equations.
2. In order to show that the model-error time-series bz_n becomes white in the limit as the model order approaches infinity, $M \to \infty$, proceed as follows. Substitute (12) and (13) into the necessary and sufficient orthogonality condition (14) to show that

$$\langle z_n x_{n-p} \rangle = 0, \qquad p = 1, 2, 3, \ldots, M, \tag{151}$$

and therefore

$$\langle z_n z_{n-p} \rangle = \frac{-1}{b(M)} \sum_{q=1}^{M} a_q(M) \langle z_n x_{n-p-q} \rangle, \qquad p = 1, 2, 3, \ldots, M. \tag{152}$$

Then use (151) in (152) to obtain

$$\bar{R}_z(p) = \frac{-1}{b(M)} \sum_{q=M-p+1}^{M} a_q(M) \bar{R}_{zx}(p + q), \qquad p = 1, 2, 3, \ldots, M. \tag{153}$$

Finally, since the sum in (153) contains only p terms, then as long as $\bar{R}_{zx}(q)$ is bounded, $1/b(M)$ is bounded, and $a_q(M) \to 0$ as $q \to \infty$ (and $M \to \infty$), this sum must converge to zero for fixed p as $M \to \infty$, and therefore $\bar{R}_z(p) = 0$ for all $p \neq 0$ in the limit $M \to \infty$. Since $\langle [x_n - \hat{x}_n]^2 \rangle \leq \langle x_n^2 \rangle$ and $\langle z_n x_n \rangle^2 \leq \langle z_n^2 \rangle \langle x_n^2 \rangle$, then

$$|\bar{R}_{zx}(q)|^2 \leq \bar{R}_z(0)\bar{R}_x(0) \leq \frac{1}{b^2(M)} \bar{R}_x^2(0), \tag{154}$$

and it is therefore sufficient for $\bar{R}_z(p) = 0$ if $1/b(M)$ is bounded and $a_q(M)$ converges to zero as $q \to \infty$ (and $M \to \infty$). As explained in Part 4 of Section B, this will be

so if x_n is a regular time-series, because then $b^2(M) > b^2(\infty) > 0$ and a stable model exists, and $a_q \to 0$ as $q \to \infty$ for a stable model. (More generally, it can be shown that z_n is white in the limit $M \to \infty$ as long as x_n is not a singular time-series.)

3. To derive formula (22) for the gain in entropy rate due to passage through a filter, proceed as follows:

(a) Consider the *relative entropy* of the n-vector

$$\boldsymbol{x}_n(L) \triangleq [x_{n+1}, x_{n+2}, \ldots, x_{n+L}]',$$

defined by

$$H[\boldsymbol{x}_n(L)] \triangleq \left\langle \ln\left\{\frac{1}{f_{x(L)}[\boldsymbol{x}_n(L)]}\right\}\right\rangle, \tag{155}$$

and show that the relative entropy of the linearly transformed vector

$$\boldsymbol{y}_n(L) \triangleq \boldsymbol{A}\boldsymbol{x}_n(L) \tag{156}$$

(where $\boldsymbol{A}$ is a time-invariant $L \times L$ matrix) is given by

$$H[\boldsymbol{y}_n] = H[\boldsymbol{x}_n] + \ln(|\boldsymbol{A}|), \tag{157}$$

in which $|\boldsymbol{A}|$ is the absolute value of the determinant of $\boldsymbol{A}$. *Hint:* Use the fact that

$$f_{y(L)}(\boldsymbol{z}) = \frac{f_{x(L)}(\boldsymbol{A}^{-1}\boldsymbol{z})}{|\boldsymbol{A}|}. \tag{158}$$

(b) Consider the lower triangular matrix

$$\boldsymbol{A} = \begin{bmatrix} g_0 & & & & \\ g_1 & g_0 & & & \\ g_2 & g_1 & g_0 & & \\ \vdots & \vdots & & \ddots & \\ g_{L-1} & g_{L-2} & & \cdots & g_0 \end{bmatrix}$$

corresponding to a causal time-invariant filter with discrete-impulse response sequence g_n and show that

$$\ln(|\boldsymbol{A}|) = (L)\ln[g_0].$$

(c) Use the results of (a) and (b) to show that the relative entropy rate (20) for the time-series defined by

$$\boldsymbol{y}_n \triangleq [y_0, y_1, y_2, \ldots, y_{n-1}]'$$

with $n \to \infty$ is given by

$$\overline{H}_y = \overline{H}_x + \ln[g_0]. \tag{159}$$

(d) It can be shown [Doob 1953] that for a minimum-phase linear time-invariant transformation, with transfer function $\tilde{G}(f)$, we have

$$\ln[g_0] = \int_{-1/2}^{1/2} \ln|\tilde{G}(f)|\, df. \tag{160}$$

Use (159) and (160) to prove that

$$\overline{H}_y - \overline{H}_x = \tfrac{1}{2}\int_{-1/2}^{1/2} \ln|\tilde{G}(f)|^2\, df. \tag{161}$$

4. To show that the solution to the problem of maximizing (29) subject to the constraint (30) is of the form (4), proceed as follows. Substitute the FST formula

$$\tilde{S}_x(f) = \sum_{k=-\infty}^{\infty} \tilde{R}_x(k)e^{-i2\pi kf} \tag{162}$$

into (29), and then use the fact that the integral (29) is maximum, subject to the constraint (30), only if the partial derivatives are zero,

$$\frac{\partial I}{\partial \tilde{R}_x(k)} = 0, \qquad |k| > M,$$

to show that a necessary condition is

$$\int_{-1/2}^{1/2} \left[\frac{1}{\tilde{S}_x(f)}\right] e^{-i2\pi kf}\, df = 0, \qquad |k| > M. \tag{163}$$

This reveals that the Fourier series coefficients of the function $1/\tilde{S}_x(f)$ are zero for $|k| > M$, and therefore its Fourier series expansion contains only $2M + 1$ terms, say

$$\frac{1}{\tilde{S}_x(f)} = \sum_{k=-M}^{M} c_k e^{i2\pi kf}. \tag{164}$$

The canonical spectral factorization (see Part 4 of Section B) of $\tilde{S}_x(f)$, as specified by (164), yields (4).

5. Derive the backward-prediction-error recursion (37) by analogy with (35)–(36).

6. **(a)** To verify formula (44), first show that the Yule-Walker equations (8) of orders $m = 0, 1, 2, \ldots, M$ can be jointly expressed in the matrix form

$$\tilde{\boldsymbol{R}}_x \boldsymbol{A} = \boldsymbol{C}, \tag{165}$$

where $\boldsymbol{C}$ is an upper triangular $(M + 1) \times (M + 1)$ matrix with diagonal elements (from top to bottom) $b^2(M), b^2(M - 1), \ldots, b^2(0)$ (the other nonzero elements in $\boldsymbol{C}$ are unimportant here), $\boldsymbol{A}$ is the lower triangular matrix defined by (42), and $\tilde{\boldsymbol{R}}_x$ is the $(M + 1) \times (M + 1)$ limit-autocorrelation matrix with pqth element $\tilde{R}_x(p - q)$, and with $p, q = 0, 1, 2, \ldots, M$, (41). Use (165) to show that

$$\boldsymbol{A}' \tilde{\boldsymbol{R}}_x \boldsymbol{A} = \boldsymbol{B}^2, \tag{166}$$

where $\boldsymbol{B}^2$ is the diagonal matrix with elements $b^2(M), b^2(M - 1), \ldots, b^2(0)$. (Note that because $\tilde{\boldsymbol{R}}_x = \tilde{\boldsymbol{R}}_x'$ and $\boldsymbol{A}'$ and $\tilde{\boldsymbol{R}}_x \boldsymbol{A}$ are upper triangular, then $\boldsymbol{B}$ is upper triangular and $\boldsymbol{B} = \boldsymbol{B}'$, and therefore $\boldsymbol{B}$ is diagonal; thus only the diagonal elements in (166) need to be checked.) Taking the inverse of (166) yields the desired result (44). As an aside, it is mentioned that the matrix $\boldsymbol{A}' \tilde{\boldsymbol{R}}_x \boldsymbol{A}$ is the limit autocorrelation matrix of the prediction-error vector $[e_n(M), e_n(M - 1), \ldots, e_n(0)]' = \boldsymbol{A}'[x_n, x_{n-1}, \ldots, x_{n-M}]'$ (see (31)).

(b) Use (166) to show that the determinant of $\tilde{\boldsymbol{R}}_x$ is given by

$$|\tilde{\boldsymbol{R}}_x| = b^2(0)b^2(1)b^2(2) \cdots b^2(M).$$

7. There is an interesting relationship between the MLM spectrum of order M (from (52) in Chapter 6),

$$\tilde{S}^M_{MLM}(f) = \frac{1}{\boldsymbol{s}'(f)\tilde{\boldsymbol{R}}_x^{-1}\boldsymbol{s}^*(f)}, \tag{167}$$

where $\tilde{\boldsymbol{R}}_x$ is the $(M + 1) \times (M + 1)$ limit-autocorrelation matrix, and the maximum-entropy method (MEM) spectra of orders $m = 0, 1, 2, 3, \ldots, M$,

$$\tilde{S}^m_{MEM}(f) = \frac{-b^2(m)}{\left|1 + \sum_{p=1}^{m} a_p(m) e^{-i2\pi pf}\right|^2}. \tag{168}$$

Specifically, it is shown in this exercise that [Burg 1972]

$$\frac{1}{\tilde{S}^M_{MLM}(f)} = \sum_{m=0}^{M} \frac{1}{\tilde{S}^m_{MEM}(f)}. \tag{169}$$

To verify (169), simply substitute the formula (44) into (167) to obtain

$$\frac{1}{\tilde{S}^M_{MEM}(f)} = [\boldsymbol{A}'\boldsymbol{s}(f)]^T \boldsymbol{B}^{-1} [\boldsymbol{A}'\boldsymbol{s}(f)], \tag{170}$$

where the superscript T denotes transpose conjugate. Now use (42) and (170) to verify (169). (Another revealing relationship between MLM and MEM is given in [Musicus 1985].)

8. **(a)** Consider the MA spectrum (50),

$$\tilde{S}_x(f) = \sigma_w^2 |\mathcal{H}(e^{-i2\pi f})|^2, \tag{171}$$

in its factored form (using $z = e^{i2\pi f}$)

$$z^L \mathcal{H}(z) = (z - z_1)(z - z_2) \cdots (z - z_L). \tag{172}$$

If there are complex roots, then these exist in conjugate pairs $z_j = z_i^*$, where

$$z_i = |z_i| \exp(i \underline{/z_i}).$$

Show that as the magnitude of a root approaches unity, $|z_i| \to 1$, $\tilde{S}_x(f)$ approaches zero at $f = \dfrac{1}{2\pi} \underline{/z_i}$.

(b) Consider the AR spectrum (51),

$$\tilde{S}_x(f) = \sigma_w^2 |\mathcal{H}(e^{i2\pi f})|^2 \tag{173}$$

in its factored form

$$z^{-M} \mathcal{H}(z) = \frac{1}{(z - z_1)(z - z_2) \cdots (z - z_M)}. \tag{174}$$

Show that as the magnitude of a root approaches unity, $|z_i| \to 1$, $\tilde{S}_x(f)$ approaches infinity at $f = (1/2\pi) \underline{/z_i}$.

(c) Consider the partial fraction expansion of (174)

$$z^{-M} \mathcal{H}(z) = \frac{A_1}{z - z_1} + \frac{A_2}{z - z_2} + \cdots + \frac{A_M}{z - z_M}, \tag{175}$$

where it is assumed that no roots are repeated: $z_i \neq z_j$ for $i \neq j$. Show that when $z \to z_i$, then

$$z^{-M} \mathcal{H}(z) \to \frac{A_i}{z - z_i}, \tag{176}$$

and therefore

$$\tilde{S}_x(f) \cong \frac{\sigma_w^2 |A_i|^2}{|\exp(i2\pi f) - |z_i| \exp(\underline{/z_i})|^2} \tag{177}$$

for $f \cong (1/2\pi) \underline{/z_i}$ and $|z_i| \cong 1$. Thus, for a root near the unit circle, $z = e^{i2\pi f}$, the height of the corresponding peak in the spectrum is closely approximated by

$$\tilde{S}_x(f) \cong \frac{\sigma_w^2 |A_i|^2}{(1 - |z_i|)^2}, \qquad f = \frac{1}{2\pi} \underline{/z_i}. \tag{178}$$

(d) Use the result of (c) to show that for a root near the unit circle, the depth of the

corresponding notch in the MA spectrum is closely approximated by

$$\tilde{S}_x(f) \cong \frac{\sigma_w^2(1 - |z_i|^2)}{|A_i|^2}. \tag{179}$$

(e) Use (177) to show that the 3-dB bandwidth of the peak of $\tilde{S}_x(f)$ centered at $f = (1/2\pi)\underline{/z_i}$ is approximately $(1/\pi)(1 - |z_i|)$ for $|z_i| \cong 1$. That is, show that

$$\tilde{S}_x\left[\frac{1}{2\pi}\underline{/z_i} \pm \frac{1}{2\pi}(1 - |z_i|)\right] \cong \frac{1}{2}\tilde{S}_x\left(\frac{1}{2\pi}\underline{/z_i}\right). \tag{180}$$

Hint: Reduce the problem to that of solving for $\theta \triangleq \underline{/z_i} - 2\pi f$ in

$$|1 - \gamma e^{i\theta}|^2 = 2|1 - \gamma|^2,$$

where $\gamma \triangleq |z_i|$. Then show that this equation is equivalent to

$$\sin^2\left(\frac{\theta}{2}\right) = \frac{1}{4}\left(\gamma + \frac{1}{\gamma} - 2\right).$$

Finally, make the approximations

$$\frac{1}{1 - \epsilon} \cong 1 + \epsilon + \epsilon^2, \qquad \epsilon << 1,$$

for $\gamma = 1 - \epsilon$, and

$$\sin^2\left(\frac{\theta}{2}\right) \cong \left(\frac{\theta}{2}\right)^2, \qquad \theta << 1,$$

and show that the resultant approximate solution is

$$\theta \cong 1 - \gamma.$$

9. To see that additive sine waves in a time-series can be thought of as corresponding to roots on the unit circle in the z-plane for the AR part of an ARMA spectrum, consider the time-series

$$s_n = \sin(\omega_0 n). \tag{181}$$

(a) Use the standard trigonometric identity

$$\sin(\theta) = 2\cos(\phi)\sin(\theta - \phi) - \sin(\theta - 2\phi) \tag{182}$$

to show that s_n satisfies the homogeneous autoregression

$$s_n = 2\cos(\omega_0)s_{n-1} - s_{n-2}, \tag{183}$$

which reveals that s_n is perfectly predictable.

(b) The corresponding characteristic polynomial for (183) is

$$\mathcal{G}(z) = 1 + a_1 z^{-1} + a_2 z^{-2} = z^{-2}(z - z_1)(z - z_2), \tag{184a}$$

where

$$a_1 = -2\cos(\omega_0) \tag{184b}$$

$$a_2 = 1.$$

Show that the roots are given by (184c)

$$z_1 = z_2^* = e^{i\omega_0}.$$

(c) Now consider a sine wave in additive white noise y_n,

$$x_n = \sin(\omega_0 n) + y_n. \tag{185}$$

Use (183) to show that

$$x_n - 2\cos(\omega_0)x_{n-1} + x_{n-2} = y_n - 2\cos(\omega_0)y_{n-1} + y_{n-2}, \tag{186}$$

which reveals that x_n is *like* an ARMA time-series with its AR parameters identical to its MA parameters. However, the fact that the roots are on the unit circle means that this is an unstable model. Observe that the ARMA spectrum formula (53) for (186) is simply $\bar{S}_x(f) = \sigma_w^2$, which is certainly not correct for a sine wave in noise. Formula (53) for $\bar{S}_x(f)$ is valid only for a stable model.

(Observe that once the coefficients in (186)—the parameters a_1 and a_2 in (184)—have been estimated, an estimate of the frequency, $\omega_0/2\pi$, of the sine wave can be obtained by solving for the roots of the polynomial (184a). This is true also for multiple sine waves in noise and is the basis for a method of sine wave frequency estimation [Pisarenko 1973].)

10. **(a)** Show that a time-averaged periodogram yields an MA type of spectrum estimate with MA order T, where $T = N - 1$ and N is the length of data segment used to obtain the periodogram before time-averaging. Specifically, show that

$$S_{x_T}(t, f)_{\Delta t} = \sum_{n=-T}^{T} \beta_n(t) e^{-i2\pi nf}, \tag{187}$$

which is potentially factorable into the form

$$S_{x_T}(t, f)_{\Delta t} = \left| \sum_{q=0}^{T} b_q(t) e^{-i2\pi qf} \right|^2, \tag{188}$$

where $\{\beta_n(t)\}$ are given by quadratic transformations of the data $x(t)$

$$\beta_n(t) = \sum_u \sum_v k_n(u, v) x(t-u) x(t-v). \tag{189}$$

(b) Show that a frequency-smoothed periodogram yields an MA type of spectrum estimate with MA order closely approximated by $T = 1/\Delta f$, where Δf is the width of the spectral smoothing window. Specifically, show that

$$S_{x_{\Delta t}}(t, f)_{\Delta f} = \sum_{n=-\Delta t}^{\Delta t} \beta_n(t) e^{-i2\pi nf} \cong \sum_{n=-T}^{T} \beta_n(t) e^{-i2\pi nf}, \tag{190}$$

where

$$\beta_n(t) \triangleq \frac{1}{N} \sum_{v=0}^{\Delta t - |n|} x(t + v + |n|) x(t + v) \frac{T}{KN} \sum_{w=-(KN/T-1)/2}^{(KN/T-1)/2} e^{-i2\pi nw/KN}. \tag{191}$$

11. **(a)** Derive the orthogonality condition (62) by equating to zero the derivatives of (60a) with respect to each of the a_p.

(b) Show that (62) can be reexpressed as (64).

(c) Use (62) to show that the minimized sum of squared errors can be expressed as in (69). *Hint:* Use the form

$$\sum_{n=0}^{K-1} w_n \hat{x}_n = 0 \tag{192}$$

of the orthogonality condition (62) to verify (68). Then substitute (60c) into (60b) and the result into (68) to obtain the desired result, (69).

12. **(a)** Show that (65) with $K \geq N + M$ reduces to (67). *Hint:* Show that since $x_n = 0$ for $n < 0$ and for $n > N - 1$, then

$$x_{n-p} x_{n-q} = 0 \qquad \text{for } n < \max\{p, q\} \quad \text{and} \quad \text{for } n > N - 1 + \min\{p, q\}.$$

Use this in (65), together with a change of index of summation, to show that

$$R_x(p, q)_K = \frac{1}{K} \sum_{m=\max\{p,q\}-q}^{N-1+\min\{p,q\}-q} x_{m+q-p} x_m.$$

Now, consider the two cases $p \geq q$ and $p \leq q$ separately to obtain (67).

(b) Use a similar approach to show that (72) reduces to (73).

(c) Prove that the temporal mean of the correlation estimate (67) is given by

$$\langle R_x(p-q)_K \rangle = \frac{N - |p-q|}{K} \tilde{R}_x(p-q), \tag{193}$$

and therefore this estimate is biased. How can you modify it to make it unbiased? Does this modified estimate have smaller or larger variance than the unmodified estimate? (It is generally accepted that the biased estimate (67) has smaller mean-squared error than its unbiased counterpart for most time-series of interest; see [Jenkins and Watts 1968].)

13. (a) Derive the forward-backward least squares prediction equations (71) by equating to zero the partial derivatives of $\hat{b}^2$ in (70) with respect to each of the M parameters $\{a_p\}_1^M$.

(b) Use (71) to show that

$$\sum_{n=0}^{K-1} [w_n \hat{x}_n + \breve{w}_n \breve{x}_{n-M}] = 0, \tag{194}$$

which is a type of orthogonality condition. Then expand the squares in (70a) and use this orthogonality condition to verify (75).

14. (a) Equate to zero the derivative of $\hat{b}^2(M)$ in (81) with respect to $\hat{a}_M(M)$, and show that $\hat{a}_M(M)$ must satisfy

$$\hat{a}_M(M) \sum_{n=M}^{N-1} [w_n^2(M-1) + \breve{w}_{n-1}^2(M-1)] + 2 \sum_{n=M}^{N-1} w_n(M-1)\breve{w}_{n-1}(M-1) = 0, \tag{195}$$

which yields (82).

(b) Derive (83) from definition (78).

(c) Expand the squares in (81) and use (195) to show that

$$\hat{b}^2(M) = [1 - \hat{a}_M^2(M)] \frac{1}{2(N-M)} \sum_{n=M}^{N-1} [w_n^2(M-1) + \breve{w}_{n-1}^2(M-1)]. \tag{196}$$

Then use (196) to obtain (85) and use (196) together with (83) to obtain (84).

15. (a) Show that for the Burg method

$$\hat{a}_1(1) = \frac{-\sum_{n=1}^{N-1} x_n x_{n-1}}{\sum_{n=1}^{N-2} x_n^2 + \frac{1}{2}x_0^2 + \frac{1}{2}x_{N-1}^2}, \tag{197}$$

whereas for the Yule-Walker method

$$\hat{a}_1(1) = \frac{-\sum_{n=1}^{N-1} x_n x_{n-1}}{\sum_{n=0}^{N-1} x_n^2}. \tag{198}$$

Hint: For the Yule-Walker method, use the Levinson-Durbin algorithm (9) with $\tilde{R}_x(q)$ replaced by $R_x(q)_K$ from (67).

(b) As an extension of (a), compare the formulas for each of $\hat{a}_1(2)$ and $\hat{a}_2(2)$ obtained from the Burg and Yule-Walker methods.

16. In order to characterize the solution to the linear least squares approximation problem in terms of the pseudo-inverse matrix, we reexpress the sum of squared errors (60a) to be minimized as (using $K = N + M$)

$$\sum_{n=0}^{K-1} (x_n - \hat{x}_n)^2 = \|\boldsymbol{x} + \boldsymbol{X}\,\hat{\boldsymbol{a}}\|^2, \tag{199}$$

where $\boldsymbol{x}$ is the K-vector

$$[x_0, x_1, x_2, x_3, \ldots, x_{N-1}, 0, 0, 0, \ldots, 0]',$$

$\boldsymbol{X}$ is the $K \times M$ matrix shown in (66), and $\hat{\boldsymbol{a}}$ is the M-vector

$$\hat{\boldsymbol{a}} = [\hat{a}_1, \hat{a}_2, \hat{a}_3, \ldots, \hat{a}_M]'.$$

As indicated by (64), the solution is specified implicitly by the M linear equations

$$\sum_{p=1}^{M} \left[\sum_{n=0}^{K-1} x_{n-q}x_{n-p}\right]\hat{a}_p = -\sum_{n=0}^{K-1} x_{n-q}x_n, \qquad q = 1, 2, 3, \ldots, M. \tag{200}$$

(a) Show that these M equations can be expressed as

$$\boldsymbol{X}'\boldsymbol{X}\,\hat{\boldsymbol{a}} = -\boldsymbol{X}'\boldsymbol{x}. \tag{201}$$

Thus, in terms of the inverse of the $M \times M$ matrix $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{R}_x$, we have

$$\hat{\boldsymbol{a}} = -[\boldsymbol{X}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{x}. \tag{202}$$

The $M \times K$ matrix

$$\boldsymbol{X}^{(-1)} \triangleq [\boldsymbol{X}'\boldsymbol{X}]^{-1}\boldsymbol{X}' \tag{203}$$

is called the *pseudo-inverse* of the $K \times M$ matrix $\boldsymbol{X}$. In most cases of practical interest, $K \geq M$. If at least M of the K rows of $\boldsymbol{X}$ are linearly independent, then the rank of $\boldsymbol{X}'\boldsymbol{X}$ is full, and this matrix is indeed invertible. (Observe that by comparison of (89) with (199), the solution (202) yields the result (90) for the ODNE approach to AR parameter estimation.)

(b) If only fewer than M of the K equations are linearly independent, then the solution to (201) is nonunique. A particularly useful solution in many applications is the one that has the smallest norm. If the rank of $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{R}_x$ is $M_0 < M$, then the minimum-norm solution to (201) is a linear combination of the M_0 eigenvectors of $\boldsymbol{R}_x$ corresponding to the M_0 nonzero eigenvalues of $\boldsymbol{R}_x$. (This is a result of the fact that the remaining $M - M_0$ eigenvectors are in the null space of $\boldsymbol{R}_x$ and their unnecessary presence in $\hat{\boldsymbol{a}}$ can only increase the norm of $\hat{\boldsymbol{a}}$.) Show that the solution must then be given by

$$\hat{\boldsymbol{a}} = -[\boldsymbol{W}\,\boldsymbol{\Lambda}_{(M_0)}^{-1}\boldsymbol{W}']\boldsymbol{X}'\boldsymbol{x}, \tag{204}$$

where $\boldsymbol{\Lambda}_{(M_0)}^{-1}$ is a diagonal matrix with the first M_0 elements on the diagonal given by the reciprocals of the nonzero eigenvalues (assumed to be unique) of $\boldsymbol{R}_x$ and the remaining $M - M_0$ elements equal to zero, and the columns of $\boldsymbol{W}$ are the corresponding eigenvectors in the same order as the eigenvalues (except that the ordering of the last $M - M_0$ columns is irrelevant.) *Hint:* Substitute the singular value decomposition

$$\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{W}\,\boldsymbol{\Lambda}\,\boldsymbol{W}' \tag{205}$$

into (201).

17. Consider the problem of minimizing the sum of weighted squared norms

$$f(\hat{\boldsymbol{\alpha}}) \triangleq \sum_p c_p \|\boldsymbol{R}_{(M_0)}(p)\hat{\boldsymbol{\alpha}}\|^2 \tag{206}$$

subject to the constraint

$$g(\hat{\boldsymbol{a}}) \triangleq \sum_{m=0}^{M_0} e_m \hat{\alpha}_m = 1 \tag{207}$$

on the inner product of $\hat{\boldsymbol{a}}$ and the unit vector

$$\boldsymbol{e} \triangleq [1, 0, 0, 0, \ldots, 0]'. \tag{208}$$

The theory of Lagrange multipliers establishes that a necessary and (in this particular case) sufficient condition for $\hat{\boldsymbol{a}}$ to minimize $f(\hat{\boldsymbol{a}})$ subject to $g(\hat{\boldsymbol{a}}) = 1$ is that it satisfy the linear equations

$$\frac{\partial f(\hat{\boldsymbol{a}})}{\partial a_p} + \lambda \frac{\partial g(\hat{\boldsymbol{a}})}{\partial a_p} = 0, \qquad p = 0, 1, 2, \ldots, M_0, \tag{209}$$

where λ is chosen to yield a solution $\hat{\boldsymbol{a}}$ that satisfies the constraint (207). Show that these M_0 linear equations can be put into the explicit form (108).

18. **(a)** Maximize the likelihood function (110) for an AR time-series with respect to the innovations variance, σ_w^2, and show that the resultant maximum-likelihood estimate is given by (113) and the corresponding maximized likelihood function is given by (114). *Hint:* Substitute (115) into (110), and then equate to zero the derivative of the natural logarithm of the result. (Since the logarithm is monotonic, this is a necessary and, in this particular case, sufficient condition for maximization.)

(b) To show that for sufficiently large N, the maximum-likelihood estimates of the AR parameters $\boldsymbol{a}$ are approximated by the least squares prediction parameters specified by the Yule-Walker equations, proceed as follows. Let $\boldsymbol{w}_{N\setminus M} \triangleq [w_M, w_{M+1}, \ldots, w_{N-1}]'$ and define $\boldsymbol{x}_{N\setminus M}$ similarly; let $\boldsymbol{x}_M \triangleq [x_0, x_1, x_2, \ldots, x_{M-1}]'$. Then it follows from the AR time-series expression (49), together with (59b), that

$$\boldsymbol{w}_{N\setminus M} = \boldsymbol{C}\boldsymbol{x}_{N\setminus M} + \boldsymbol{D}\boldsymbol{x}_M = \boldsymbol{x}_{N\setminus M} - \hat{\boldsymbol{x}}_{N\setminus M},$$

where the $(N - M) \times (N - M)$ matrix $\boldsymbol{C}$ depends only on $\boldsymbol{a}$ and has determinant equal to unity, and the $(N - M) \times M$ matrix $\boldsymbol{D}$ depends only on $\boldsymbol{a}$. It follows that (see [Gardner 1985])

$$\begin{aligned} f_{\boldsymbol{x}_{N\setminus M}|\boldsymbol{x}_M}(\boldsymbol{x}_{N\setminus M} \mid \boldsymbol{x}_M, \boldsymbol{a}, \sigma_w) &= f_{\boldsymbol{w}_{N\setminus M}}(\boldsymbol{C}\boldsymbol{x}_{N\setminus M} + \boldsymbol{D}\boldsymbol{x}_M \mid \boldsymbol{a}, \sigma_w) \\ &= f_{\boldsymbol{w}_{N\setminus M}}(\boldsymbol{x}_{N\setminus M} - \hat{\boldsymbol{x}}_{N\setminus M} \mid \boldsymbol{a}, \sigma_w), \end{aligned}$$

and therefore Bayes' law of conditional probability,

$$f_{\boldsymbol{x}_{N\setminus M}|\boldsymbol{x}_M} f_{\boldsymbol{x}_M} = f_{\boldsymbol{x}_{N\setminus M}, \boldsymbol{x}_M} \equiv f_{\boldsymbol{x}_N},$$

yields the expression

$$f_{\boldsymbol{x}_N}(\boldsymbol{x}_N \mid \boldsymbol{a}, \sigma_w) = f_{\boldsymbol{w}_{N\setminus M}}(\boldsymbol{x}_{N\setminus M} - \hat{\boldsymbol{x}}_{N\setminus M} \mid \boldsymbol{a}, \sigma_w) f_{\boldsymbol{x}_M}(\boldsymbol{x}_M \mid \boldsymbol{a}, \sigma_w) \tag{210}$$

for the likelihood function (110). Now, it follows from (210) and the white noise property of w_n that

$$f_{\boldsymbol{x}_N}(\boldsymbol{x}_N \mid \boldsymbol{a}, \sigma_w) = [2\pi\sigma_w^2]^{-(N-M)/2} \exp\left\{ -\frac{1}{2\sigma_w^2} \sum_{n=M}^{N-1} (x_n - \hat{x}_n)^2 \right\} f_{\boldsymbol{x}_M}(\boldsymbol{x}_M \mid \boldsymbol{a}, \sigma_w), \tag{211}$$

where $\hat{x}_n$ is the predictor given by (59b). It is shown in (c) that for sufficiently large N the influence of the last factor in (211) on the maximum with respect to $\boldsymbol{a}$ is negligible. Show that in this case the maximum is given by the solution to

$$\min_{\boldsymbol{a}} \sum_{n=M}^{N-1} (x_n - \hat{x}_n)^2. \tag{212}$$

(c) Substitute (110) for both $z = x_N$ and $z = x_M$ into (211) and take the natural logarithm of the result to verify the identity

$$\left[\frac{1}{\sigma_w^2} S_0^N - x_N' \tilde{R}_{x_N}^{-1} x_N - \ln|\tilde{R}_{x_N}|\right] = \left[\frac{1}{\sigma_w^2} S_0^M - x_M' \tilde{R}_{x_M}^{-1} x_M - \ln|\tilde{R}_{x_M}|\right], \tag{213}$$

where

$$S_0^N \triangleq \sum_{n=0}^{N-1} (x_n - \hat{x}_n)^2. \tag{214}$$

It is shown in (d) that for sufficiently large N the terms in the right member of (213) are negligible compared with those in the left member and, therefore, (combining S_0^M with S_0^N)

$$x_N' \tilde{R}_{x_N}^{-1} x_N + \ln|\tilde{R}_{x_N}| \cong \frac{1}{\sigma_w^2} \sum_{n=M}^{N-1} (x_n - \hat{x}_n)^2. \tag{215}$$

Substitute (215) into the logarithm of (110) with $z = x_N$ to verify that

$$\ln\{f_x(x_N \mid a, \sigma_w)\} \cong c - \frac{1}{2\sigma_w^2} \sum_{n=M}^{N-1} (x_n - \hat{x}_n)^2, \tag{216}$$

where c is a constant, independent of a and σ_w. Thus, maximization of the likelihood function with respect to a is approximately equivalent to minimization of the sum of squared prediction errors, as in (212). (We cannot use (216) to maximize with respect to σ_w because large σ_w violates the approximation (215).)

(d) It can be shown that

$$\lim_{N\to\infty} \frac{2}{N} \ln|\tilde{R}_{x_N}| = \int_{-1/2}^{1/2} \ln[\tilde{S}_x(f)]\, df. \tag{217}$$

Thus, for large N, $\ln|\tilde{R}_{x_N}|$ is proportional to N. Show that the temporal mean of $x_N' \tilde{R}_{x_N}^{-1} x_N$ is equal to N for a in $\tilde{R}_{x_N}$ equal to the correct values for the time-series x_N, and that the temporal mean of S_0^N is $N\sigma_w^2$ for a in $\hat{x}_N$ equal to the correct values for the time-series x_N. Consequently, all terms in the left member of (213) are proportional to N, but none of those in the right member are. However, the term $\ln|\tilde{R}_{x_M}|$ can be quite large when the roots of (55) are close to the unit circle. This is suggested by (217), since $\hat{S}_x(f) \to \infty$ as a root approaches the unit circle (see (177)). Also, the term $x_M' \tilde{R}_{x_M}^{-1} x_M$ can be quite large if $\tilde{R}_{x_M}$ is ill-conditioned. But regardless of how large these terms are, if they are finite, then there is a value of N large enough to validate the approximation (215).

A heuristic approach to explaining (217) follows. Since $\tilde{R}_x$ is a Toeplitz matrix, $\tilde{R}_x(p, q) = \tilde{R}_x(p - q)$, the characteristic equation

$$\tilde{R}_{x_N} v = \lambda v \tag{218}$$

approaches the convolution

$$\tilde{R}_x(n) \otimes v_n \cong \lambda v_n \tag{219}$$

as N increases. Fourier-series transformation of (219) yields

$$\tilde{S}_x(f)\tilde{V}(f) \cong \lambda \tilde{V}(f). \tag{220}$$

Thus, for large N the eigenvalues are closely approximated by

$$\lambda_n \cong \tilde{S}_x(f_n), \tag{221}$$

and it can be shown that the values of f_n are approximately equally spaced throughout the interval $[0, \frac{1}{2}]$. Consequently, the determinant identity

$$\ln|\tilde{\boldsymbol{R}}_{x_N}| = \ln[\lambda_1\lambda_2\lambda_3 \cdots \lambda_N] \tag{222}$$

can be approximated by

$$\begin{aligned} \ln|\tilde{\boldsymbol{R}}_{x_N}| &\cong \ln\left[\tilde{S}_x(0)\tilde{S}_x\left(\frac{1}{2N}\right)\tilde{S}_x\left(\frac{2}{2N}\right) \cdots \tilde{S}_x\left(\frac{N-1}{2N}\right)\right] \\ &= \sum_{n=0}^{N-1} \ln\left[\tilde{S}_x\left(\frac{n}{2N}\right)\right] \cong N\int_0^{1/2} \ln[\tilde{S}_x(f)]\, df, \end{aligned} \tag{223}$$

from which (217) follows.

19. To derive the extended and modified Yule-Walker equations (119) for an ARMA time-series, proceed as follows.

(a) Multiply both sides of (52), with n replaced by $n + k$, by x_n, and show that the limit time-average of this product is given by

$$\begin{aligned} \tilde{R}_x(k) = &-a_1\tilde{R}_x(k-1) - a_2\tilde{R}_x(k-2) - \cdots - a_M\tilde{R}_x(k-M) \\ &+ b_0\tilde{R}_{wx}(k) + b_1\tilde{R}_{wx}(k-1) + \cdots + b_L\tilde{R}_{wx}(k-L). \end{aligned} \tag{224}$$

Then use (52) to argue that x_n depends only on w_n, w_{n-1}, w_{n-2}, . . ., and therefore

$$R_{wx}(n) = 0, \qquad n \geqslant 1,$$

from which (224) yields (119a).

(b) Let $\{h_n\}$ denote the discrete-impulse response sequence of the ARMA model (52),

$$x_n = \sum_{m=0}^{\infty} h_{n-m}w_m. \tag{225}$$

Use (225) to show that

$$R_{wx}(k) = \sigma_w^2 h_{-k}. \tag{226}$$

Then use (224) and (226) to derive (119b) and (119c).

(c) Use (52) to argue that (122a) must be valid. Then use the initial value theorem,

$$h_0 = \lim_{z\to\infty} \mathcal{H}(z), \tag{227}$$

where $\mathcal{H}(z)$ is the transfer function of the ARMA model (52) (which is the ratio of (54) to (55)), to verify (122b).

20. Derive formula (131) for the limit autocorrelation of an MA time-series y_n. *Hint:* First show that

$$\tilde{R}_y(k) = \sum_{\substack{p,q=0 \\ q-p=k}}^{L} b_p b_q. \tag{228}$$

Then consider $k > 0$ and $k < 0$ separately.

21. **(a)** Use definition (142) to verify (143).

(b) Show that (125a), (143), and (144a) yield (145).

22. The maximum-likelihood estimates of the amplitude, phase, and frequency parameters a, ϕ, and f of a sine wave in additive white Gaussian noise are obtained by maximizing the *likelihood function* for the time-series

$$x(t) = a\cos(2\pi ft + \phi) + n(t). \tag{229}$$

It can be shown using the periodically time-variant fraction-of-time distribution for

a Gaussian time-series (Part II) that this is equivalent to minimizing the function

$$\int_{t-T/2}^{t+T/2} [x(u) - a\cos(2\pi fu + \phi)]^2 \, du, \tag{230}$$

which is the continuous-time counterpart of a least squares fit of $x(t)$ to a sine wave. Use the approximation

$$\int_{t-T/2}^{t+T/2} [\cos(2\pi ft - \phi)]^2 \, dt \cong \frac{T}{2}, \qquad T >> \frac{1}{f} \tag{231}$$

to show that the maximum-likelihood estimate of a is given by

$$\hat{a} = \frac{2}{T} \int_{t-T/2}^{t+T/2} x(u)\cos(2\pi \hat{f} u - \hat{\theta}) \, du, \tag{232}$$

and then use (230) and (232) to show that the maximum-likelihood estimate of ϕ is given by

$$\hat{\phi} = -\tan^{-1}\left[\frac{\displaystyle\int_{t-T/2}^{t+T/2} x(u)\sin(2\pi \hat{f} u) \, du}{\displaystyle\int_{t-T/2}^{t+T/2} x(v)\cos(2\pi \hat{f} v) \, dv}\right]. \tag{233}$$

Then use (230) and (233) to show that the maximum-likelihood estimate of f is obtained from the periodogram by the maximization

$$\max_{\hat{f}} S_{x_T}(t, \hat{f}). \tag{234}$$

Finally, substitute (233) into (232) to show that the estimate $\hat{a}$ also is obtainable directly from the periodogram,

$$\hat{a} = 2\left[\frac{1}{T} S_{x_T}(t, \hat{f})\right]^{1/2}. \tag{235}$$

Unfortunately, the simplicity of this solution does not extend to time-series containing more than one sine wave, unless the number of such sine waves is known and the separation between their frequencies greatly exceeds $1/T$ [Walker 1971].

23. Consider a linear array of N equally spaced sensors (hydrophones, electromagnetic antennas, seismometers, etc.) in a homogeneous medium with a sinusoidal wavefront of energy impinging on the array at an angle of θ radians measured from the normal to the linear array, as shown in Figure 9-21. Show that the intensity at the nth sensor is given by

$$x_n(t) = \sin[\omega n + \alpha t + \phi],$$

where

$$\omega = \frac{2\pi d}{\lambda}\cos\theta$$

$$\alpha = \frac{2\pi c}{\lambda},$$

λ is the wavelength, c is the speed of the wavefront, and d is the distance between sensors (preferably, $d = \lambda/2$). Thus, if there are M sources of energy arriving at angles $\{\theta_i\}_1^M$, all with the same wavelength and speed, and there is additive measurement noise present, then

$$x_n(t) = \sum_{i=1}^{M} A_i \sin[\omega_i n + \alpha t + \phi_i] + z_n(t), \qquad n = 0, 1, 2, \ldots, N-1,$$

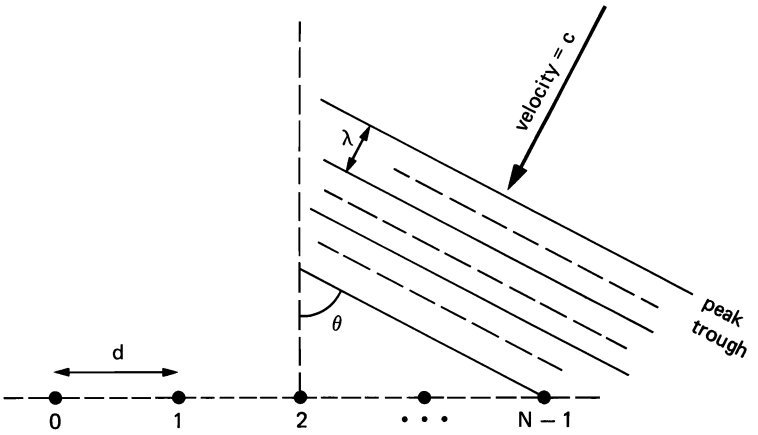


Figure 9-21 Linear array of sensors with impinging sine wave (e.g., acoustical wave or electromagnetic wave).

where A_i represents the strength of the ith energy source. Consequently, we have a set of discrete spatial series consisting of M sine waves in noise, and the set is indexed by time t. Therefore, t indexes an ensemble over which we can average in order to obtain a statistical spectrum of this spatial series. Hence, we can obtain relatively accurate estimates of the $N \times N$ correlation matrix, even for very small values of the spatial segment length N. Discuss the potential relative merits of direct methods versus parametric methods for detecting the presence of narrow-band energy sources and estimating their directions of arrival $\{\theta_i\}_1^M$ by detecting spectral peaks and estimating their locations in the spectrum.

24. (a) Consider the problem of estimating the range r of an energy-reflecting object by transmitting a pulse of energy and measuring the elapsed time $\tau = 2r/c$ until the reflected pulse is received (c is the speed of propagation). If the energy is electromagnetic (radio frequency), this is called radar range estimation. If the energy is acoustical (sound frequency), this is called sonar range estimation. In order to obtain a strong echo, the transmitted pulse energy must be large. However, because of limitations on dynamic range, sufficiently strong pulses often cannot be obtained simply by choosing a sufficiently large amplitude. Instead, the length of the pulse must be made long. However, in order to obtain good range resolution, the pulse cannot be a simple function; it must be a sophisticated function (Chapter 2, Section B) having bandwidth greatly exceeding the reciprocal of its duration, so that its autocorrelation width is much narrower than its duration. This requirement results from the technique of correlating the echo with the transmitted pulse and estimating the elapsed time to be the lag at which this cross correlation peaks. The width of this peak determines the resolution capability. One useful type of sophisticated function is a binary valued waveform obtained from a pseudonoise sequence. Another is the chirp pulse

$$x(t) = \sin(\beta t^2), \qquad 0 \leq t \leq T.$$

(The limit autocorrelation for an infinitely long chirp pulse has zero width; see (c).) A particularly attractive way to estimate the elapsed time from a chirp pulse is based on the fact that the product of a chirp and a delayed replica is given by

$$y(t) = x(t)x(t - \tau_0) = \tfrac{1}{2}\cos(\omega_0 t - \theta_0) - \tfrac{1}{2}\cos(\omega_0 t - \theta_0 - 2\beta t^2),$$

where

$$\omega_0 = 2\beta\tau_0$$

$$\theta_0 = \beta\tau_0^2.$$

(Verify this equation.) Thus, $y(t)$ consists of a sine wave of power $\frac{1}{4}$ at frequency $f_0 = \omega_0/2\pi$ plus a broadband chirp waveform of total power $\frac{1}{4}$ with bandwidth $2\beta T$ centered at f_0. Thus, the delay τ_0 can be estimated by estimating the frequency ω_0 of the sine wave. Show that if the transmitted waveform is band-pass with center frequency f_* rather than low-pass,

$$x(t) = \sin(\omega_* t + \beta t^2), \qquad \frac{-T}{2} \leq t \leq \frac{T}{2},$$

then the preceding result is essentially the same.

(b) Consider the situation where there are M closely spaced reflecting objects and generalize the preceding result for the product,

$$y(t) = x(t) \sum_{i=1}^{M} A_i x(t - \tau_i),$$

of the transmitted chirp and the multichirp echo. Based on the result obtained, discuss the conditions on the ranges $\{\tau_i\}_1^M$ under which parametric methods of spectral analysis for sine waves in noise would be potentially attractive alternatives to direct methods for determining the number of objects present and estimating their ranges.

(c) Prove that for the infinite-duration chirp pulse

$$x(t) = \sin(\beta t^2),$$

we have

$$\hat{R}_x(\tau) = \begin{cases} \frac{1}{2}, & \tau = 0 \\ 0, & \tau \neq 0. \end{cases}$$

Consequently, $\hat{S}_x(f) \equiv 0$ (see Part 1 of Section D in Chapter 3). Use this result to prove that the limit spectrum for the lag product

$$y(t) = x(t)x(t - \tau_0)$$

is

$$\hat{S}_y(f) = \frac{1}{16}\delta(f - f_0) + \frac{1}{16}\delta(f + f_0),$$

where

$$f_0 = \beta\tau_0/\pi.$$

Consequently, as the chirp pulse length T increases, the time-averaged spectral level of the broadband component in the lag product waveform decreases and approaches zero as $T \to \infty$.

25. Describe explicitly the dependence of the likelihood function for an ARMA time-series on the ARMA parameters $\{a_p\}_1^M$, $\{b_q\}_0^M$, σ_w^2. *Hint:* Start with the right member of (110), which applies to all Gaussian time-series.

REFERENCES FOR PART I

AKAIKE, H. 1969a. Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* 21:243–47.

AKAIKE, H. 1969b. Power spectrum estimation through autoregression model fitting. *Ann. Inst. Statist. Math.* 21:407–19.

AKAIKE, H. 1970. Statistical predictor identification. *Ann. Inst. Statist. Math.* 22:203–17.

AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control.* AC-19:716–23.

ALTER, D. 1927. A group or correlation periodogram, with application to the rainfall of the British Isles. *Monthly Weather Review.* 55:263–66.

BAGGEROER, A. B. 1976. Confidence intervals for regression (MEM) spectral estimates. *IEEE Trans. Information Th.* IT-22:534–45.

BARNETT, V. 1973. *Comparative Statistical Inference.* New York: John Wiley & Sons.

BARTLETT, M. S. 1948. Smoothing periodograms from time-series with continuous spectra. *Nature.* 161:686–87.

BARTLETT, M. S. 1950. Periodogram analysis and continuous spectra. *Biometrika.* 37:1–16.

BASS, J. 1962. *Les Fonctions Pseudo-Aleatoires.* Paris: Gauthier-Villars.

BEAMISH, N., and M. B. PRIESTLY. 1981. A study of autoregressive and window spectral estimation. *Appl. Statist.* 30:41–58.

BENDAT, J. S., and A. G. PIERSOL. 1980. *Engineering Applications of Correlation and Spectral Analysis.* New York: John Wiley & Sons.

BENNETT, W. R. 1953. The correlatograph. *Bell System Tech. J.* 32:1173–85.

BERK, K. N. 1974. Consistent autoregressive spectral estimates. *The Annals of Statistics.* 2:489–502.

BLACKMAN, R. B., and J. W. TUKEY. 1958. *The Measurement of Power Spectra.* New York: American Telephone and Telegraph Co. (also New York: Dover, 1959).

BLAHUT, R. E. 1985. *Fast Algorithms for Digital Signal Processing*. Reading, Mass.: Addison-Wesley.

BOYLES, R. A., and W. A. GARDNER. 1983. Cycloergodic properties of discrete-parameter nonstationary stochastic processes, *IEEE Trans. Information Th.* IT-29:105–14.

BRACEWELL, R. N. 1985. *The Hartley Transform*. New York: Oxford University Press.

BRACEWELL, R. N. 1986. *The Fourier Transform and Its Applications*. 2d ed., rev. New York: McGraw-Hill.

BRENNAN, D. G. 1961. Probability theory in communication system engineering. Chapter 2 in *Communication System Theory*. Ed. E. J. Baghdady. New York: McGraw-Hill.

BRILLINGER, D. R. 1975. *Time-Series*. New York: Holt, Rinehart and Winston.

BRILLINGER, D. R. 1976. Some history of the data analysis of time-series in the United States. In *History of Statistics in the United States*. Ed. D. B. Owen. New York: Marcel Dekker, pp. 267–80.

BROWN, W. A. 1987. "On the theory of cyclostationary signals," Ph.D. dissertation, Department of Electrical and Computer Engineering, University of California, Davis, Calif.

BRUZZONE, S. P., and M. KAVEH. 1980. On some suboptimum ARMA spectral estimators. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-28:753–55.

BRUZZONE, S. P., and M. KAVEH. 1984. Information tradeoffs in the sample autocorrelation function in ARMA parameter estimation. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-32:701–15.

BURG, J. P. 1967. Maximum entropy spectral analysis. *Proc. 37th Meeting Society of Exploration Geophysicists*. Oklahoma City, Okla., October 31.

BURG, J. P. 1968. A new analysis technique for time-series data. *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*. Enschede, The Netherlands, August 12–23.

BURG, J. P. 1972. The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics*. 37:375–76.

BURRUS, C. S., and T. W. PARKS. 1985. *DFT/FFT and Convolution Algorithms*. New York: John Wiley & Sons.

BURSHSTEIN, D., and E. WEINSTEIN. 1985. Some relations between the various criteria for autoregressive model order determination. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-33:1017–19.

BUYS-BALLOT, C. H. D. 1847. *Les Changements Périodiques de Température Dépendants de la Nature du Soleil et de la Lune* . . . . Utrecht: Kemink & Fils.

CADZOW, J. A. 1980. High performance spectral estimation—A new ARMA method. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-28:524–29.

CADZOW, J. A. 1982. Spectral estimation: An overdetermined rational model equation approach. *Proceedings IEEE*. 70:907–39.

CADZOW, J. A., B. BASEGHI, and T. HSU. 1983. Singular-value decomposition approach to time-series modelling. *IEE Proceedings, Part F*. 130:202–9.

CAPON, J. 1969. High-resolution frequency-wavenumber spectrum analysis. *Proceedings IEEE*. 57:1408–18.

CAPON, J., R. J. GREENFIELD, and R. J. KOLKER. 1967. Multidimensional maximum-likelihood processing of a large-aperture seismic array. *Proceedings IEEE*. 55:192–211.

CARSON, J. R. 1931. The statistical energy-frequency spectrum of random disturbances. *Bell System Tech. J.* 10:374–81.

CHAPMAN, S., and J. BARTELS. 1940. *Geomagnetism, Vol. II, Analysis of the Data and Physical Theories*. London: Oxford University Press (2d ed. 1951).

CIOFFI, J. M., and T. KAILATH. 1984. Fast, recursive-least-squares transversal filters for adaptive filtering. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-32:304–37.

CLAASEN, T. A. C. M., and W. F. G. MECKLENBRÄUKER. 1980. The Wigner distribution—a tool for time-frequency signal analysis, Parts I–III. *Phillips J. of Research*. 35:217–50, 276–300, 372–89.

CLAYTON, H. H. 1917. Effect of short period variation of solar radiation on the earth's atmosphere. *Smithsonian Miscellaneous Collections*. 68, no. 3, publ. 2446:19 pp.

CRAMÉR, H. 1940. On the theory of stationary random processes. *Ann. Math.* 41:215–30.

CRAMÉR, H. 1942. On harmonic analysis in certain functional spaces. *Arkiv Math. Astr. Fysik*. 28:1–7.

CROCHIERE, R. E., and L. R. RABINER. 1983. *Multirate Digital Signal Processing*. Englewood Cliffs, N.J.: Prentice-Hall.

D'ANGELO, H. 1970. *Linear Time-Varying Systems: Analysis and Synthesis*. Boston, Mass.: Allyn and Bacon.

DANIELL, P. J. 1946. Discussion of "On the theoretical specification and sampling properties of autocorrelated time-series". *J. Royal Statist. Soc.* 8B, no. 1:27–97 (especially p. 89).

DAVIS, H. T. 1941. *The Analysis of Economic Time Series*. Bloomington, Ind.: The Principia Press.

DICKINSON, B. W., and J. M. TURNER. 1979. Reflection coefficient estimation using Cholesky decomposition. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-27:146–49.

DOLPH, C. L. 1946. A current distribution for broadside arrays which optimizes the relationship between width and sidelobe level. *Proceedings IRE*. 34:335–48.

DONATI, F. 1971. Finite-time averaged power-spectra. *IEEE Trans. Information Th.* IT-17:7–16.

DOOB, J. L. 1953. *Stochastic Processes*. New York: John Wiley & Sons.

DUGRÉ, J-P, L. L. SCHARF, and C. GUEGUEN. 1981. Exact likelihood for stationary vector autoregressive moving-average processes. *Proc. of Workshop on Fast Algorithms For Linear Systems*. Aussois, France, September, pp. 9.1–9.24.

DURBIN, J. 1960. The fitting of time series models. *Revue L'Institut Internationale de Statistique*. 28:233–43.

EBERLY, J. H., and K. WODKIEWICZ. 1977. The time-dependent phsyical spectrum of light. *J. Optical Soc. Amer.* 67:1252–61.

ECKART, C., and G. YOUNG. 1936. The approximation of a matrix by another lower rank matrix. *Psychometrika*. 1:211–18.

EINSTEIN, A. 1906. On the theory of the Brownian movement. *Annalen der Physik*. 19:371–81.

EINSTEIN, A. 1914. Méthode pour la détermination de valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières. *Archives des Sciences Physiques et Naturelles*. 37:254–56.

FANO, R. M. 1950. Short-time autocorrelation functions and power spectra. *J. Acoustical Soc. Amer.* 22:546–50.

FINCH, P. D. 1969. Linear least squares prediction in non-stochastic time series. *Adv. Appl. Prob.* 1:111–22.

FRANKS, L. E. 1969. *Signal Theory*. Englewood Cliffs, N.J.: Prentice-Hall.

FRIEDLANDER, B. 1982a. Lattice filters for adaptive proceessing. *Proceedings IEEE*. 70:829–67.

FRIEDLANDER, B. 1982b. Recursive maximum likelihood algorithm for ARMA spectral estimation. *IEEE Trans. Information Th.* IT-28:639–46.

FRIEDLANDER, B. 1983a. Efficient algorithm for ARMA spectral estimation. *IEE Proceedings, Part F.* 130:195–201.

FRIEDLANDER, B. 1983b. Instrumental variable methods for ARMA spectral estimation. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-31: 404–15.

FRIEDLANDER, B., and B. PORAT. 1984a. A spectral matching technique for ARMA parameter estimation. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-32:338–43.

FRIEDLANDER, B., and B. PORAT. 1984b. The modified Yule-Walker method of ARMA spectral estimation. *IEEE Trans. Aerospace Electronic Systems*. AES-20:158–73.

GABOR, D. 1946. Theory of communication, part III. *Journal IEE*. 93:429–57.

GARDNER, W. A. 1985. *Introduction to Random Processes with Applications to Signals and Systems*. New York: Macmillan.

GARDNER, W. A. 1987a. Common pitfalls in the application of stationary process theory to time-sampled and modulated signals. *IEEE Trans. Communications*. COM-35:529–34.

GARDNER, W. A. 1987b. Rice's representation for cyclostationary processes. *IEEE Trans. Communications*. COM-35:74–8.

GARDNER, W. A. 1987c. Correlation estimation and time-series modeling for nonstationary processes. Signal and Image Processing Lab. Tech. Report No. SIPL-87-3, Department of Electrical and Computer Engineering, University of California, Davis, Calif.

GERSCH, W. 1970. Estimation of the autoregressive parameters of a mixed autoregressive moving-average time series. *IEEE Trans. Automatic Control*. AC-15:583–88.

GOLOMB, S. W. 1967. *Shift Register Sequences*. San Francisco: Holden-Day.

GOLUB, G., and W. KAHAN. 1965. Calculating the singular values and pseudo-inverse of a matrix. *J. SIAM Numerical Analysis* (Ser. B). 2:205–24.

GOUTEREAU, C. (1906) Sur la variabilité de la température. *Annuaire de la Soc. Mét. de France*. 54:122–27.

GRAUPE, D., D. J. KRAUSE, and J. B. MOORE. 1975. Identification of autoregressive and moving-average parameters of time-series. *IEEE Trans. Automatic Control*. AC-20:104–7.

GRENANDER, U. 1951. On empirical spectral analysis of stochastic processes. *Ark. Mat.* 1:503–31.

GRENANDER, U. 1958. Bandwidth and variance in estimation of the spectrum. *J. Royal Statist. Soc.* 20:152–57.

GRENANDER, U., and M. ROSENBLATT. 1953. Statistical spectral analysis of time-series arising from stationary stochastic processes. *Ann. Math. Stat.* 24:537–58.

GRENANDER, U., and M. ROSENBLATT. 1984. *Statistical Analysis of Stationary Time-Series*. 2d [corrected] ed. New York: Chelsea Publishing Co.

HAMMING, R. W. 1983. *Digital Filters*. 2d. ed. Englewood Cliffs, N.J.: Prentice-Hall.

HANNAN, E. J. 1970. *Multiple Time Series*. New York: John Wiley & Sons.

HANNAN, E. J., and B. G. QUINN. 1979. The determination of the order of an autoregression. *J. Royal Statist. Soc.* Ser. B. 41:190–95.

HARRIS, F. J. 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings IEEE.* 66:51–83.

HASSELMANN, K. and T. P. BARNETT. 1981. Techniques of linear prediction for systems with periodic statistics. *J. Atmospheric Sciences.* 38:2275–83.

HELME, B. I., and C. L. NIKIAS. 1985. Improved spectrum performance via a data-adaptive weighted Burg technique. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-33:903–10.

HOFSTETTER, E. M. 1964. Random processes. Chapter 3 in *The Mathematics of Physics and Chemistry, Vol. II.* Ed. H. Margenau and G. M. Murphy. Princeton, N.J.: D. Van Nostrand Co.

HONIG, M. L., and D. G. MESSERSCHMITT. 1984. *Adaptive Filters: Structures, Algorithms, and Applications.* Boston: Kluwer Academic Publishers.

HOOKER, R. H. 1901. Correlation of the marriage-rate with trade. *J. Royal Statist. Soc.* 64:485–92.

HOUSEHOLDER, A. S., and G. YOUNG. 1950. Matrix approximation and latent roots. *Amer. Math. Mon.* 45:165–67.

HSIA, T. C., and D. A. LANDGREBE. 1967. On a method for estimating power spectra. *IEEE Trans. Instrumentation and Measurement.* IM-16:255–57.

ISSERLIS, L. 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika.* 12:134–39.

JAMES, H. M., N. B. NICHOLS, and R. S. PHILLIPS. 1947. *Theory of Servomechanisms.* New York: McGraw-Hill.

JENKINS, G. M., and D. G. WATTS. 1968. *Spectral Analysis and Its Applications.* San Francisco: Holden-Day.

JONES, R. H., and W. M. BRELSFORD. 1967. Time-series with periodic structure. *Biometrika.* 54:403–7.

KAILATH, T., A. VIEIRA, and M. MORF. 1978. Inverses of Toeplitz operators, innovations, and orthonormal polynomials. *SIAM Rev.* 20:106–19.

KAISER, J. F. 1966. Digital filters. Chapter 7 in *System Analysis by Digital Computer.* Ed. F. F. Kuo and J. F. Kaiser. New York: John Wiley & Sons.

KAMPÉ DE FÉRIET, J. 1954. Introduction to the statistical theory of turbulence, I and II. *J. Soc. Indust. Appl. Math.* 2, nos. 1 and 3:1–9 and 143–74.

KAVEH, M. 1979. High resolution spectral estimation for noisy signals. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-27:286–87.

KAVEH, M., and S. P. BRUZZONE. 1981. A comparative overview of ARMA spectral estimation. *Proc. of First ASSP Workshop on Spectral Est.* Hamilton, Ontario, August 17–18, 2.4.1–2.4.8.

KAVEH, M., and S. P. BRUZZONE. 1983. Statistical efficiency of correlation-based methods for ARMA spectral estimation. *IEE Proceedings, Part F.* 130:211–17.

KAVEH, M., and G. A. LIPPERT. 1983. An optimum tapered Burg method for linear prediction and spectral analysis. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP 31:438–44.

KAY, S. M. 1980. Noise compensation for autoregressive spectral estimates. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-28:292–303.

KAY, S. M. 1983. Recursive maximum likelihood estimation of autoregressive processes. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-31:56–65.

KAY, S. M. 1987. *Modern Spectral Estimation.* Englewood Cliffs, N.J.: Prentice-Hall.

KAY, S. M., and S. L. MARPLE. 1981. Spectrum analysis—A modern perspective. *Proceedings IEEE.* 69:1380–1419.

KHARKEVICH, A. A. 1960. *Spectra and Analysis.* Trans. from Russian. New York: Consultants Bureau.

KHINCHIN, A. J. 1934. Korrelations theorie der stationären stochastischen Prozesse. *Mathematische Annalen.* 109:604–15.

KLEMA, V. C., and A. J. LAUB. 1980. The singular value decomposition: its computation and some applications. *IEEE Trans. Automatic Control.* AC-25:164–76.

KOENIG, W., H. K. DUNN, and L. Y. LACY. 1946. The sound spectrograph. *J. Acoustical Soc. Amer.* 18:19–49.

KOLMOGOROV, A. N. 1941a. Stationary sequences in Hilbert space. *Bull. Moscow State U. Math.* 2, no. 6. 40 pp.

KOLMOGOROV, A. N. 1941b. Interpolation and extrapolation von stationären zufälligen folgen. *Bull. Acad. Sci. de l'U.R.S.S.* 5:3–14.

KOOPMANS, L. H. 1974. *The Spectral Analysis of Time-Series.* New York: Academic Press.

LACOSS, R. T. 1971. Data adaptive spectral analysis methods. *Geophysics.* 36:661–75.

LAGRANGE, J. L. 1772. Recherches sur la manière de former des tables des planètes d'apres les seules observations. *Oeuvres de Lagrange.* 6:507–627, 1873. Sur les interpolations. *Oeuvres de Lagrange.* 7:535–53, in particular pp. 541 et. seq. 1877.

LAGUNAS-HERNÁNDEZ, M. A., and A. GASULL-LLAMPALLAS. 1984. An improved maximum likelihood method for power spectral density estimation. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-32:170–73.

LANNING, J. H., and R. H. BATTIN. 1956. *Random Processes in Automatic Control.* New York: McGraw-Hill.

LARSON, H. J., and B. O. SHUBERT. 1979. *Probabilistic Models in Engineering Sciences.* Vol. II. New York: John Wiley & Sons.

LAWSON, C. L. and R. J. HANSON. 1974. *Solving Least Squares Problems.* Englewood Cliffs, N.J.: Prentice-Hall.

LAWSON, J. L., and G. E. UHLENBECK. 1949. *Threshold Signals.* Massachusetts Inst. of Tech. Radiation Lab. Series. vol. 24, also Boston: Boston Tech. Publ., Inc., 1964.

LEVINSON, N. 1947. The Wiener RMS (root mean square) error criterion in filter design and prediction. *J. Math. Phys.* 25, no. 4:261–78.

LOYNES, R. M. 1968. On the concept of the spectrum for non-stationary processes, *J. Royal Statist. Soc.* Ser. B. 30:1–30.

MARK, W. D. 1970. Spectral analysis of the convolution and filtering of non-stationary stochastic processes. *J. Sound Vib.* 11:19–63.

MARPLE, S. L. 1980. A new autoregressive spectrum analysis algorithm. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-28:441–54.

MARPLE, S. L. 1987. *Digital Spectral Analysis with Applications.* Englewood Cliffs, N.J.: Prentice-Hall.

MASANI, P. R. 1979. "Commentary on the memoir on generalized harmonic analysis,"

pp. 333–79 in *Norbert Wiener: Collected Works, Volume II*. Cambridge, Mass.: Massachusetts Institute of Technology.

MEHRA, R. K. 1971. On-line identification of linear dynamical systems with applications to Kalman filtering. *IEEE Trans. Automatic Control*. AC-16:12–22.

MORF, M., B. DICKINSON, T. KAILATH, and A. VIEIRA. 1977. Efficient solution of covariance equations for linear prediction. *IEEE Trans. Acoustics, Speech, and Sig. Proc*. ASSP-25:429–33.

MOSES, R. L., J. A. CADZOW, and A. A. BEEX. 1985. A recursive proccedure for ARMA modeling. *IEEE Trans. Acoustics, Speech, and Sig. Proc*. ASSP 33:1188–96.

MUNK, W. H., F. E. SNODGRASS, and M. J. TUCKER. 1959. "Spectra of low-frequency ocean waves," *Bull. Scripps Inst. Oceanography*. 7:283–362.

MUSICUS, B. R. 1985. Fast MLM spectrum estimation from uniformly spaced correlations. *IEEE Trans. Acoustics, Speech. and Sig. Proc*. ASSP-33:1333–35.

NUSSBAUMER, H. J. 1982. *Fast Fourier Transform and Convolution Algorithms*. Berlin: Springer-Verlag.

NUTTALL, A. H. 1976. Spectral analysis of a univariate process with bad data points, via maximum entropy and linear predictive techniques. *Tech. Report* Tr-5303. Naval Underwater Systems Center, New London, Conn. March 26.

NUTTALL, A. H. 1981. Some windows with very good sidelobe behavior, *IEEE Trans. Acoustics, Speech, and Sig. Proc*. ASSP-29:84–91.

NYQUIST, H. 1928. Certain topics in telegraph transmission theory. *Trans. AIEE*. 47:617–44.

OWSLEY, N. L. 1985. High-resolution spectrum analysis by dominant-mode enhancement. Chapter 4 in *VLSI and Modern Signal Processing*. Ed. S. Y. Kung, H. J. Whitehouse, and T. Kailath. Englewood Cliffs, N.J.: Prentice-Hall.

PALIWAL, K. K. 1985. Further results on tapered and energy-weighted Burg methods. *IEEE Trans. Acoustics, Speech, and Sig. Proc*., ASSP-33:1624–26.

PAPOULIS, A. 1962. *The Fourier Integral and its Applications*. New York: McGraw-Hill.

PAPOULIS, A. 1977. *Signal Analysis*. New York: McGraw-Hill.

PAPOULIS, A. 1984. *Probability, Random Variables, and Stochastic Processes*. 2d ed., rev. New York: McGraw-Hill.

PARZEN, E. 1957a. On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat*. 28:329–48.

PARZEN, E. 1957b. On choosing an estimate of the spectral density function of a stationary time series. *Ann. Math. Stat*. 28:921–32.

PARZEN, E. 1974. Some recent advances in time-series modeling. *IEEE Trans. Automatic Control*. AC-19:723–30.

PARZEN, E. 1977. "Multiple time-series: Determining the order of approximating autoregressive schemes." In *Multivariate Analysis IV*. Ed. P. Krishnaiah, Amsterdam: North-Holland, pp. 283–95.

PARZEN, E., and M. PAGANO. 1979. An approach to modeling seasonally stationary time-series, *J. Econometrics*. 9:137–53.

PFAFFELHUBER, E. 1975. Generalized harmonic analysis for distributions. *IEEE Trans. Information Th*. IT-21:605–611.

PISARENKO, V. F. 1973. The retrieval of harmonics from a covariance function, *Geophysical J. Royal Astronomical Soc*. 33:347–66.

POYNTING, J. H. 1884. A comparison of the fluctuations in the price of wheat and in the cotton and silk imports into Great Britain. *J. Royal Statist. Soc.* 47:34–64.

PRIESTLEY, M. B. 1965. Evolutionary spectra and non-stationary processes. *J. Royal Statist. Soc.* Ser. B. 27:204–37.

PRIESTLEY, M. B. 1981. *Spectral Analysis and Time-Series. Vols. 1, 2.* London: Academic Press.

PRIESTLEY, M. B., and H. TONG. 1973. On the analysis of bivariate non-stationary processes. *J. Royal Statist. Soc.* Ser. B. 35:153–88.

QUIRK, M. P., and B. LIU. 1983. Improving resolution for autoregressive spectral estimation by decimation. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-31:630–37.

RAO, C. R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya Ind. J. Statist. A.* 26:329–58.

RAO, C. R. 1973. *Linear Statistical Inference and its Applications.* 2d ed. New York: John Wiley & Sons.

RICE, S. O. 1944. Mathematical analysis of random noise, parts I, II. *Bell System Tech. J.* 23:282–332.

RICE, S. O. 1945. Mathematical analysis of random noise, parts III and IV. *Bell System Tech. J.* 24:46–156.

RICE, S. O. 1948. Statistical properties of a sine wave plus random noise. *Bell System Tech. J.* 27:109–57.

RIHACZEK, A. W. 1969. *Principles of High-Resolution Radar.* New York: McGraw-Hill.

ROBINSON, E. A. 1982. A historical perspective of spectrum estimation. *Proceedings IEEE.* 70:885–907.

SAKAI, H., and M. ARASE. 1979. Recursive parameter estimation of an autoregressive process disturbed by white noise. *Internat. J. Control.* 30:949–66.

SCHROEDER, M. R., and B. S. ATAL. 1962. Generalized short-time power spectra and autocorrelation functions. *J. Acoustical Soc. Amer.*, 34:1679–83.

SCHUSTER, A. 1894. On interference phenomena. *Phil. Mag.* 37:509–45.

SCHUSTER, A. 1897. On lunar and solar periodicities of earthquakes, *Proc. Royal Soc. of London.* 61:455–65.

SCHUSTER, A. 1898. On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena. *Terr. Magnet.* 3:13–41.

SCHUSTER, A. 1900. The periodogram of magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871–1895. *Trans. Cambridge Phil. Soc.* 18:107–35.

SCHUSTER, A. 1904. *The Theory of Optics.* London: Cambridge University Press.

SCHUSTER, A. 1906a. The periodogram and its optical analogy. *Proc. Royal Soc. of London.* 77A:136–40.

SCHUSTER, A. 1906b. On the periodicity of sunspots. *Phil Trans. Royal Soc. of London.* 206A:69–100.

SCHUSTER, A. 1911. On the periodicity of sunspots. *Proc. Royal Soc. of London.* 85A:50–53.

SHANKS, J. L. 1967. Recursion filters for digital processing. *Geophysics.* 32:33–51.

SHANNON, C. E. and W. WEAVER. 1962. *The Mathematical Theory of Communication.* Urbana, Ill.: The University of Illinois Press.

SLUTSKY, E. E. 1927. The summation of random causes as the source of cyclic processes. *Problems of Economic Conditions*. Ed. Conjuncture Institute, Moscow. 3, no. 1 (reprinted in English in *Econometrica*, 5 (1937):267–98.

SLUTSKY, E. E. 1929. Sur l'extension de la theórie de periodogrammes aux suites des quantités dépendentes. *Comptes Rendues*. 189:722–33.

SLUTSKY, E. E. 1934. Alcuni applicazioni di coefficienti di Fourier al analizo di sequenze eventuali coherenti stazionarii. *Giorn. d. Instituto Italiano degli Atuari*. 5:435–82.

SNYDER, D. L. 1975. *Random Point Processes*. New York: John Wiley & Sons.

SORENSON, H. V., D. L. JONES, C. S. BURRUS, and M. T. HEIDEMAN. 1985. On computing the discrete Hartley transform. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-33:1231–38.

STOKES, G. G. 1879. Note on searching for periodicities. *Proc. Royal Soc. of London*. 29:122–23.

STUMPFF, K. 1927. *Analyse periodischer Vorgänge*. Berlin: Gebruder Borntraeger.

SWINGLER, D. N. 1974. A comparison between Burg's maximum entropy method and a nonrecursive technique for the spectral analysis of deterministic signals. *J. Geophysical Research*. 84:674–85.

TAYLOR, G. I. 1920. Diffusion by continuous movements. *Proc. London Math. Soc.* 20:196–212.

TAYLOR, G. I. 1938. The spectrum of turbulence. *Proc. Royal Soc. of London*. 164A:476–90.

THOMSON, D. J. 1982. Spectrum estimation and harmonic analysis. *Proceedings IEEE*. 70:1055–96.

TJØSTHEIM, D. 1976. Spectral generating operators for non-stationary processes. *Adv. Appl. Prob*. 8:831–46.

TUFTS, D. W., and R. KUMARESAN. 1982. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proceedings IEEE*. 70:975–89.

TUKEY, J. W. 1949. "The sampling theory of power spectrum estimates," Symposium on Applications of Autocorrelation Analysis to Physical Problems. Woods Hole, Mass. June 1949, pp. 47–67; also, *The Collected Works of John W. Tukey*, Belmont, Calif.: Wadsworth, 1984.

ULRYCH, T. J. and T. N. BISHOP. 1975. Maximum entropy spectral analysis and autoregressive decomposition. *Rev. Geophysics and Space Physics*. 13:183–200.

ULRYCH, T. J., and R. W. CLAYTON. 1976. Time-series modelling and maximum entropy. *Physics of the Earth and Planetary Interiors*. 12:188–200.

URKOWITZ, H. 1983. *Signal Theory and Random Processes*. Dedham, Mass.: Artech House.

VAKMAN, D. E. 1968. *Sophisticated Signals and the Uncertainty Principle in Radar*. Trans. from Russian. New York: Springer-Verlag.

VAN DER POL, B. 1946. Frequency modulation. *J. IEE*. London. 93, Part III:153–8.

VILLE, J. 1948. Theory and application of the notion of the complex signal. *Câbles et Transmission*. 2:67–74. (Trans. to English by I. Selin, Rand Corp. Report T-92, August 1, 1958).

VON SMOLUCHOWSKI, M. 1914. *The Kinetic Theory of Matter and Electricity*. Leipzig and Berlin, Germany: B. G. Teubner.

WALKER, A. M. 1962. Large sample estimation of parameters for autoregressive processes with moving-average residuals. *Biometrika*. 49:117–31.

WALKER, A. M. 1971. On the estimation of a harmonic component in a time series with independent residuals. *Biometrika*. 58:21–36.

WALKER, G. 1931. On periodicity in series of related terms. *Proc. Royal Soc*. 131:518–32.

WELCH, P. D. 1967. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio and Electroacoustics*. AU-15:70–73.

WIENER, N. 1923. Differential space. *J. of Math and Physics*. 2:131–74.

WIENER, N. 1930. Generalized harmonic analysis. *Acta Mathematika*. 55:117–258.

WIENER, N. 1938. The historical background of harmonic analysis. *Amer. Math. Soc. Semicentennial Publ*. 2:56–68.

WIENER, N. 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time-Series*. New York: The Technology Press of MIT and John Wiley & Sons. Originally issued in February, 1942, as a classified National Defense Res. Counsel Report.

WIENER, N., and R. E. A. C. PALEY. 1934. Fourier transforms in the complex domain. *Amer. Math. Soc. Colloquium Publication*. 19.

WIGGINS, R. A., and E. A. ROBINSON. 1965. Recursive solution to the multichannel filtering problem. *J. Geophysics Research*. 70:1885–91.

WIGNER, E. 1932. On the quantum correction for thermodynamic equilibrium. *Phys. Rev*. 40:749–59.

WILKINSON, J. H. 1965. *The Algebraic Eigenvalue Problem*. Oxford, England: Clarendon.

WOLD, H. O. A. 1938. *A Study in the Analysis of Stationary Time Series*. Uppsala: Almqvist and Wiksell. (2d ed., Stockholm, 1954).

WOLD, H. O. A. 1948. On prediction in stationary time series. *Ann. Math. Stat*. 19:558–67.

WOLD, H. O. A., ed. 1965. *Bibliography on Time Series and Stochastic Processes*. Cambridge, Mass.: The M.I.T. Press.

YULE, G. U. 1926. Why do we sometimes get nonsense correlations between time-series?—a study in sampling and the nature of time-series. *J. Royal Statist. Soc*. 89:1–69.

YULE, G. U. 1927. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans. Royal Soc. London*. A. 226:267–98.

REFERENCES FOR PART II

ACKROYD, M. H. 1984. Stationary and cyclostationary finite buffer behaviour computation via Levinson's Method. *AT&T Bell Laboratories Tech. J.* 63:2159–70.

AGEE, B. G., S. V. SCHELL, and W. A. GARDNER. 1987. Self-coherence restoral: A new approach to blind adaptation of antenna arrays. *Proc. of Twenty First Annual Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, Calif. Nov. 2–4, 1987.

ALBUQUERQUE, J. P. A., O. SHIMBO, and L. N. NGUGEN. 1984. Modulation transfer noise effects from a continuous digital carrier to FDM/FM carriers in memoryless nonlinear devices. *IEEE Trans. Communications*. COM-32:337–53.

BENNETT, W. R. 1958. Statistics of regenerative digital transmission. *Bell System Tech. J.* 37:1501–42.

BERG, N. J., and J. N. LEE. 1983. *Acousto-Optic Signal Processing*. New York: Marcel Dekker.

BLACKMAN, R. B., and J. W. TUKEY. 1958. *The Measurement of Power Spectra*. New York: American Telephone and Telegraph Co. (also New York: Dover, 1959).

BOYLES, R. A., and W. A. GARDNER. 1983. Cycloergodic properties of discrete-parameter nonstationary stochastic processes. *IEEE Trans. Information Th.* IT-29:105–14.

BRELSFORD, W. M. 1967. "Probability predictions and time series with periodic structure," Ph.D. dissertation, Johns Hopkins University, Baltimore, Maryland.

BROWN, W. A. 1987. "On the theory of cyclostationary signals," Ph.D. dissertation, Department of Electrical and Computer Engineering, University of California, Davis, Calif.

BUYS-BALLOT, C. H. D. 1847. *Les Changements Périodiques de Température Dépendants de la Nature du Soleil et de la Lune . . . .* Utrecht: Kemink & Fils.

CAMPBELL, J. C., A. J. GIBBS, and B. M. SMITH. 1983. The cyclostationary nature of crosstalk interference from digital signals in multipair cable—Part I: Fundamentals. *IEEE Trans. Communications*. COM-31:629–37.

CARIOLARO, G. L., G. L. PIEROBON, and G. P. TRONCA. 1983. Analysis of codes and spectra calculations. *Int. J. Electronics*. 55:35–79.

CHAPMAN, S., and J. BARTELS. 1940. *Geomagnetism, Vol. II, Analysis of the Data and Physical Theories*. London: Oxford University Press (2d ed. 1951).

CLAASEN, T. A. C. M., and W. F. G. MECKLENBRÄUKER. 1980. The Wigner distribution—A tool for time-frequency signal analysis, Parts I–III. *Phillips J. Research*. 35:217–50, 276–300, 372–89.

CORDUNEAUNU, C. 1961. *Almost Periodic Functions*. New York: John Wiley & Sons.

DAVENPORT, W. B., and W. L. ROOT. 1958. *An Introduction to Random Signals and Noise*. New York: McGraw-Hill.

EISENSTEIN, B. A., and L. R. CERRATO. 1978. Statistical deconvolution of electrocardiograms. *IEEE Trans. Biomed. Eng*. BME-25:96–99.

ERICSON, T. H. E. 1981. Modulation by means of linear periodic filtering. *IEEE Trans. Information Th*. IT-27:322–27.

FERRARA, E. 1985. Frequency-domain implementations of periodically time-varying adaptive filters. *IEEE Trans. Acoustics, Speech, and Sig. Proc*. ASSP-33:883–92.

FRANKS, L. E. 1980. Carrier and bit synchronization in data communication—A tutorial review. *IEEE Trans. Communications*. COM-28:1107–21.

FRANKS, L. E., and J. BUBROUSKI. 1974. Statistical properties of timing jitter in a PAM timing recovery scheme. *IEEE Trans. Communications*. COM-22:913–30.

FRENCH, C. A., and W. A. GARDNER. 1986. Despreading spread-spectrum signals without the code. *IEEE Trans. Communications*. COM-34:404–07.

FURRER, B., and W. GUGGENBUHL. 1981. Noise analysis of sampled-data circuits. *AEU Journal*. 35:426–30.

GARDNER, W. A. 1972. "Representation and estimation of cyclostationary processes," Ph.D. dissertation, Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, Mass. Reprinted as Signal and Image Processing Lab. Tech. Rept. No. SIPL-82-1, Department of Electrical and Computer Engineering, University of California, Davis, Calif., 1982.

GARDNER, W. A. 1973. The structure of linear least-mean-square estimators for synchronous M-ary signals. *IEEE Trans. Information Th*. IT-19:240–43.

GARDNER, W. A. 1978. Stationarizable random processes. *IEEE Trans. Information Th*. IT-24:8–22.

GARDNER, W. A. 1982. Structural characterization of locally optimum detectors in terms of locally optimum estimators and correlators. *IEEE Trans. Information Th*. IT-28:924–32.

GARDNER, W. A. 1984. Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Processing*. 6:113–33. Errata: *Signal Processing*, 12:211

GARDNER, W. A. 1985. *Introduction to Random Processes with Applications to Signals and Systems*. New York: Macmillan.

GARDNER, W. A. 1986. The role of spectral correlation in design and performance analysis of synchronizers. *IEEE Trans. Communications*. COM-34:1089–95.

GARDNER, W. A. 1987a. Common pitfalls in the application of stationary process theory to time-sampled and modulated signals. *IEEE Trans. Communications*. COM-35:529–34.

GARDNER, W. A. 1987b. Signal interception: A unifying theoretical framework for feature detection. *Proc. of Onzieme Colloque Sur Le Traitement Du Signal Et Des Images*. Nice, France. June 1–5. pp. 69–72.

GARDNER, W. A., and B. G. AGEE. 1980. Two-stage adaptive noise cancellation for intermittent-signal applications. *IEEE Trans. Information Th.* IT-26:746–50.

GARDNER, W. A., and C. K. CHEN. 1987. Interference-tolerant time-difference-of-arrival estimation for modulated signals. Signal and Image Processing Lab. Tech. Report No. SIPL-87-4. Department of Electrical and Computer Engineering, University of California, Davis, Calif.

GARDNER, W. A., and L. E. FRANKS. 1975. Characterization of cyclostationary random signal processes. *IEEE Trans. Information Th.* IT-21:4–14.

GARDNER, W. A., C. M. SPOONER, and L. PAURA. 1987. Performance evaluation of detectors for signal interception. Signal and Image Processing Lab. Tech. Report No. SIPL-87-8. Department of Electrical and Computer Engineering, University of California, Davis, Calif.

GOLOMB, S. W. 1967. *Shift Register Sequences*. San Francisco: Holden-Day.

GRAEF, F. K. 1983. Joint optimization of transmitter and receiver for cyclostationary random signal processes. *Proc. of the Nato Advanced Study Institute on Nonlinear Stochastic Problems*. Algarve, Portugal. May 16–28, 1982. Dordrecht, Netherlands: Reidel. pp. 581–92.

HASSELMAN, K., and T. P. BARNETT. 1981. Techniques of linear prediction for systems with periodic statistics. *J. Atmospheric Science*. 38:2275–83.

HAYKIN, S. 1986. *Adaptive Filter Theory*. Englewood Cliffs, N.J.: Prentice-Hall.

HINICH, M. J. 1982. Detecting a hidden periodic signal when its period is unknown. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-30:747–50.

HOLL, J. P., and W. A. GARDNER. 1985. Adaptive extraction of periodic signals from noise: Outperforming coherent averaging. Signal and Image Processing Lab. Tech. Rept. No. SIPL-85-14, Department of Electrical and Computer Engineering, University of California, Davis, Calif.

HURD, H. 1969. "An investigation of periodically correlated stochastic processes," Ph.D. dissertation, Duke University, Durham, No. Carolina.

ISSERLIS, L. 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*. 12:134–39.

JOHNSON, W. K. 1981. The dynamic pneumocardiogram: An application of coherent signal processing to cardiovascular measurement. *IEEE Trans. Biomed. Eng.* BME-28:471–75.

JONES, R. H., and W. M. BRELSFORD. 1967. Time series with periodic structure. *Biometrika*. 54:403–7.

KAPLAN, M. 1983. Single-server queue with cyclostationary arrivals and arithmetic service. *Operations Research*. 31:184–205.

MARTIN, W. 1982. Time-frequency analysis of random signals. *Proc. Intl. Conf. on Acoustics, Speech, and Sig. Proc.* Paris, France. pp. 1325–28.

MARTIN, W., and P. FLANDRIN. 1985. Wigner-Ville spectral analysis of nonstationary processes. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-33:1461–70.

MENGALI, U., and E. PEZZANI. 1978. Tracking properties of phase-locked loops in optical communication systems. *IEEE Trans. Communications*. COM-26:1811–18.

MESIYA, M. F., P. J. MCLANE, and L. L. CAMPBELL. 1978. Optimal receiver filters for BPSK transmission over a bandlimited nonlinear channel. *IEEE Trans. Communications*. COM-26:12–22.

MOENECLAEY, M. 1982a. Comment on "Tracking performance of the filter and square bit synchronizer". *IEEE Trans. Communications*. COM-30:407–10.

MOENECLAEY, M. 1982b. Linear phase-locked loop theory for cyclostationary input disturbances. *IEEE Trans. Communications*. COM-30:2253–59.

MOENECLAEY, M. 1983. The optimum closed-loop transfer function of a phase-locked loop used for sychronization purposes. *IEEE Trans. Communications*. COM-31:549–53.

MOENECLAEY, M. 1984. A fundamental lower bound on the performance of practical joint carrier and bit synchronizers. *IEEE Trans. Communications*. COM-32:1007–12.

MONIN, A. S. 1963. Stationary and periodic time series in the general circulation of the atmosphere. *Proc. Symp. Time Series Anal.* Ed. M. Rosenblatt. New York: John Wiley & Sons., pp. 144–51.

MONTI, C. M., and G. L. PIEROBON. 1985. Block codes for linear timing recovery in data transmission systems. *IEEE Trans. Communications*. COM-33:527–34.

O'REILLY, J. J. 1984. Timing extraction for baseband digital transmission. *Problems of Randomness in Communication Engineering*. Ed. K. W. Cattermole and J. J. O'Reilly. London: Plymouth.

ORTIZ, M. J., and A. RUIZ DE ELVIRA. 1985. A cyclo-stationary model of sea surface temperature in the Pacific Ocean. *Tellus*. 37A:14–23.

PAPOULIS, A. 1984. *Probability, Random Variables, and Stochastic Processes*. 2d ed., rev. New York: McGraw-Hill.

PARZEN, E., and M. PAGANO. 1979. An approach to modeling seasonally stationary time-series. *J. Econometrics*. 9:137–53.

PELKOWITZ, L. 1981. Frequency domain analysis of wraparound error in fast convolution algorithms. *IEEE Trans. Acoustics, Speech, and Sig. Proc.* ASSP-29:413–22.

PUPOLIN, S., and C. TOMASI. 1984. Spectral analysis of line regenerator time jitter. *IEEE Trans. Communications*. COM-32:561–66.

RICE, S. O. 1944. Mathematical analysis of random noise, Parts I and II. *Bell System Tech. J.* 23:282–332.

RICE, S. O. 1945. Mathematical analysis of random noise, Parts III and IV. *Bell System Tech. J.* 24:46–156.

RICE, S. O. 1948. Statistical properties of a sine wave plus random noise. *Bell System Tech. J.* 27:109–57.

RIHACZEK, A. W. 1968. Signal and energy distribution in time and frequency. *IEEE Trans. Information Th.* IT-14:369–74.

RIHACZEK, A. W. 1969. *Principles of High-Resolution Radar*. New York: McGraw-Hill.

RHYNE, V. E. 1969. A Comparison of coherent averaging techniques for repetitive biological signals. *Medical Research Engrg.* August–September:22–6.

STRATONOVICH, R. L. 1967. *Topics in the Theory of Random Noise*. Volumes I and II. Revised English edition trans. by R. A. Silverman. New York: Gordon and Breach.

STROM, T., and S. SIGNELL 1977. Analysis of periodically switched linear circuits. *IEEE Trans. Circuits and Systems*. CAS-24:531–41.

VECCHIA, A. V. 1985. Periodic autoregressive-moving average (parma) modeling with applications to water resources. *Water Resour. Bull.* 21:721–30.

VILLE, J. 1948. Theory and application of the notion of the complex signal. *Câbles et Transmission*. 2:67–74. Trans. to English by I. Selin, Rand Corp. Rept. T-92, August 1, 1958.

WIDROW, B., and S. D. STEARNS. 1985. *Adaptive Signal Processing*. Englewood Cliffs, N. J.: Prentice-Hall.

WIENER, N. 1930. Generalized harmonic analysis. *Acta Mathematika*. 55:117–258.

WIENER, N. 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: The Technology Press of MIT and John Wiley & Sons. Originally issued in February, 1942, as a classified National Defense Res. Counsel Report.

WIGNER, E. 1932. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.* 40:749–59.

WILLIS, D. M. 1964. The statistics of a particular non-homogeneous Poisson process. *Biometrika*. 51:399–404.

WOLD, H. O. A. 1938. *A Study in the Analysis of Stationary Time Series*. Uppsala: Almqvist and Wiksell. (2d ed., Stockholm, 1954).

WOODWARD, P. M. 1953. *Probability and Information Theory, with Applications to Radar*. New York: Pergamon Press.

YULE, G. U. 1927. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans. Royal Soc. London*. A, 226:267–98.

AUTHOR INDEX

SUBJECT INDEX

D

E

F

G

H

I

K

L

M

Q

R

S

T

U

V

W