TIME–FREQUENCY ANALYSIS

Leon Cohen

Hunter College and Graduate Center of
The City University of New York

Prentice Hall PTR, Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data
Cohen, Leon
Time-frequency analysis / Leon Cohen
p cm
Includes bibliographical references and index
ISBN 0-13-594532-1
1 Signal processing 2 Time-series analysis 3 Frequency spectra I Title
TK5102 9 C557 1995
621 382 23--dc20 94-39843
CIP

Editorial/production supervision *bookworks*
Manufacturing buyer Alexis R Heydt

A Pearson Education Company
Upper Saddle River, NJ 07458

The publisher offers discounts on this book when ordered in bulk quantities. For more information, contact: Corporate Sales Department, Prentice Hall PTR, 1 Lake Street, Upper Saddle River, New Jersey, 07694. Phone: 800-382-3419. Fax 201-236-7141 E-mail: corporatesales@prenhall.com

Printed in the United States of America

10 9 8 7 6 5 4

ISBN 0-13-594532-1

Prentice-Hall International (UK) Limited, London
Prentice-Hall of Australia Pty Limited, Sydney
Prentice-Hall Canada Inc., Toronto
Prentice-Hall Hispanoamericana, S.A., Mexico
Prentice-Hall of India Private Limited, New Delhi
Prentice-Hall of Japan, Inc., Tokyo
Pearson Education Asia Pte. Ltd., Singapore
Editoria Prentice-Hall do Brasil, Ltda., Rio De Janeiro

To Carol, Valerie, Ken,
Livia, and Douglas

Contents

Preface

Changing frequencies is one of the most primitive sensations since we are surrounded by light of changing color, by sounds of varying pitch, and by many other phenomena whose periodicities change in time. A sunset is dramatic because of the colors and the change in colors. The aim of time-frequency analysis is to describe how the frequency or spectral content of a signal evolves and to develop the physical and mathematical ideas needed to understand what a time-varying spectrum is. The attempt to represent a signal simultaneously in time and frequency is full of challenges, both physical and mathematical, and I hope this book conveys the fascination and elegance the subject holds for me.

My aim is to give a simple exposition of the physical ideas and origins, motivations, methods, underpinnings, and scaffolding of the field. I have attempted to be clear on questions of what is known, what is not known, what is speculation, and the limitations of our current knowledge.

I never understood or learned unless I saw simple examples illustrating the ideas introduced. Naturally, I assume everyone else is like me. So, whenever possible, I illustrate each point with examples.

The basic ideas and methods that have been developed are readily understood by the uninitiated; the book is self-contained. The mathematics is elementary, with the possible exception of the last few chapters. There is an attitude these days that one should use from the beginning the most "sophisticated" mathematics. The reason generally given is that the sophisticated mathematics has to be learned eventually. I have attempted to do everything with the simplest of mathematics and only use sophisticated methods when absolutely needed or when there is an overwhelming advantage, either from a manipulative point of view or a simplification of the physical ideas.

Time-frequency analysis spans many fields, including engineering, physics, astronomy, chemistry, geophysics, biology, medicine, and mathematics. I have strived for a minimum of jargon so that the book may be understandable to a wide audience.

I wish to express my appreciation to Carol Frishberg, Pat Loughlin, Jim Pitton, and Ted Posch for reading the manuscript and making many valuable suggestions.

Leon Cohen

New York

Notation in Brief

The main notational conventions we use are as follows:

1. Integrals. All integrals without limits imply integration from minus $-\infty$ to ∞,

$$\int \equiv \int_{-\infty}^{\infty} \quad .$$

2. Fourier Transform Pairs. We use $s(t)$ to denote a signal and $S(\omega)$ its Fourier transform and normalize symmetrically:

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int s(t)\, e^{-j\omega t} dt \qquad ; \qquad s(t) = \frac{1}{\sqrt{2\pi}} \int S(\omega)\, e^{j\omega t} d\omega$$

For other quantities there will be exceptions in regard to the factor of 2π due to historically accepted conventions. We consistently use angular frequency, ω.

3. Magnitude and Phase. It is often advantageous to express a signal and its Fourier transform in terms of their respective amplitudes and phases. The notation we use is

$$s(t) = A(t)e^{j\varphi(t)} \qquad ; \qquad S(\omega) = B(\omega)e^{j\psi(\omega)}$$

and we use "phase" and "spectral phase" to denote φ and $\psi(\omega)$, respectively, and "amplitude" and "spectral amplitude" to denote $A(t)$ and $B(\omega)$.

4. Functions. We often use the variable to denote a function. That is, $f(x)$ and $f(y)$ may not necessarily be the same function, the individuality of the functions being denoted by the variable, x or y. Where confusion may arise we will use different notation to emphasize the distinction.

5. Averages. Global and conditional averages are denoted by the following conventions:

$\langle w \rangle$	e.g., average weight
$\langle w \rangle_h$	e.g., average weight for a given height
$\sigma_w^2 = \langle w^2 \rangle - \langle w \rangle^2$	e.g., standard deviation of weight
$\sigma_{w\|h}^2 = \langle w^2 \rangle_h - \langle w \rangle_h^2$	e.g., standard deviation of weight for a given height

6. Operators. Symbols in calligraphic letters are operators. For example, the frequency operator, $\mathcal{W}$, and time operator, $\mathcal{T}$, are

$$\mathcal{W} = -j\frac{d}{dt} \qquad ; \qquad \mathcal{T} = j\frac{d}{d\omega}$$

Time–Frequency Analysis

Chapter 1

The Time and Frequency Description of Signals

1.1 INTRODUCTION

In this chapter we develop the basic ideas of time and frequency analysis from the perspective that the standard treatments already contain the seeds and motivations for the necessity of a combined time-frequency description. *Signal analysis* is the study and characterization of the basic properties of signals and was historically developed concurrently with the discovery of the fundamental signals in nature, such as the electric field, sound wave, and electric currents. A signal is generally a function of many variables. For example, the electric field varies in both space and time. Our main emphasis will be the time variation, although the ideas developed are easily extended to spatial and other variables, and we do so in the latter part of the book.

The time variation of a signal is fundamental because time is fundamental. However, if we want to gain more understanding, it is often advantageous to study the signal in a different representation. This is done by expanding the signal in a complete set of functions, and from a mathematical point of view there are an infinite number of ways this can be done. What makes a particular representation important is that the characteristics of the signal are understood better in that representation because the representation is characterized by a physical quantity that is important in nature or for the situation at hand. Besides time, the most important representation is frequency. The mathematics of the frequency representation was invented by Fourier, whose main motivation was to find the equation governing the behavior of heat. The contributions of Fourier were milestones because indeed he did find the fundamental equation governing heat, and, in addition, he invented the remarkable mathematics to handle discontinuities (1807). He had to be able to

handle discontinuities because in one of the most basic problems regarding heat, namely, when hot and cold objects are put in contact, a discontinuity in temperature arises. Fourier's idea, that a discontinuous function can be expressed as the sum of continuous functions – an absurd idea on the face of it, which the great scientists of that time, including Laplace and Lagrange, did not hesitate to call absurd in nondiplomatic language – turned out to be one of the great innovations of mathematics and science.[1] However, the reason spectral analysis is one of the most powerful scientific methods ever discovered is due to the contributions of Bunsen and Kirchhoff about sixty years after Fourier presented his ideas (1807) and about 35 years after his death in 1830. Spectral analysis turned out to be much more important than anyone in Fourier's time could have envisioned. This came about with the invention of the spectroscope[2] and with the discovery that by spectrally analyzing light we can determine the nature of matter; that atoms and molecules are fingerprinted by the frequency spectrum of the light they emit. This is the modern usage of *spectral analysis*. Its discoverers, Bunsen and Kirchhoff, observed (around 1865) that light spectra can be used for recognition, detection, and classification of substances because they are unique to each substance.

This idea, along with its extension to other waveforms and the invention of the tools needed to carry out spectral decomposition, certainly ranks as one of the most important discoveries in the history of mankind. It could certainly be argued that the spectroscope and its variations are the most important scientific tools ever devised. The analysis of spectra has led to the discovery of the basic laws of nature and has allowed us to understand the composition and nature of substances on earth and in stars millions of light years away. It would be appropriate to refer to spectral analysis as Bunsen-Kirchhoff analysis.

1.2 TIME DESCRIPTION OF SIGNALS

Fundamental physical quantities such as the electromagnetic field, pressure, and voltage change in time and are called time waveforms or signals. We shall denote a signal by $s(t)$. In principle, a signal can have any functional form and it is possible to produce signals, such as sound waves, with extraordinary richness and complexity. Fortunately, simple signals exist, hence the motivation to study and characterize the simple cases first in order to build up one's understanding before tackling the more complicated ones.

[1] Laplace and Lagrange weren't thrilled about Fourier's theory of heat either. However, his ideas were eventually widely accepted in his own lifetime and he succeeded to Lagrange's chair. Fourier was heavily involved in politics and had his ups and downs in that realm also. At one time he accompanied Napoleon to Egypt and had a major impact in establishing the field of Egyptology.

[2] The spectroscope was invented by Fraunhofer around 1815 for the measurement of the index of refraction of glasses. Fraunhofer was one of the great telescope makers and realized that the accurate determination of the index of refraction is essential for building optical instruments of high quality. In using the spectroscope for that purpose Fraunhofer discovered and catalogued spectral lines which have come to be known as the Fraunhofer lines. However, the full significance of spectral analysis as a finger print of elements and molecules was first understood by Bunsen and Kirchhoff some fifty years after the invention of the spectroscope.

The simplest time-varying signal is the sinusoid. It is a solution to many of the fundamental equations, such as Maxwell equations, and is common in nature. It is characterized by a constant amplitude, a, and constant frequency, ω_0,

$$s(t) = a\cos\omega_0 t \tag{1.1}$$

We say that such a signal is of constant amplitude. This does not mean that the signal is of constant value, but that the maxima and minima of the oscillations are constant. The frequency, ω_0, has a clear physical interpretation, namely the number of oscillations, or ups and downs, per unit time.

One attempts to generalize the simplicity of the sinusoid by hoping that a general signal can be written in the form

$$s(t) = a(t)\cos\vartheta(t) \tag{1.2}$$

where the amplitude, $a(t)$, and phase, $\vartheta(t)$, are now arbitrary functions of time. To emphasize that they generally change in time, the phrases amplitude modulation and phase modulation are often used, since the word modulation means change.

Difficulties arise immediately. Nature does not break up a signal for us in terms of amplitude and phase. Nature only gives us the left-hand side, $s(t)$. Even if the signal were generated by a human by way of Eq. (1.2) with specific amplitude and phase functions, that particular $a(t)$ and $\vartheta(t)$ would not be special since there are an infinite number of ways of choosing different pairs of amplitudes and phases that generate the same signal. Is there one pair that is special?

Also, it is often advantageous to write a signal in complex form

$$s(t) = A(t)\,e^{j\varphi(t)} = s_r + js_i \tag{1.3}$$

and we want to take the actual signal at hand to be the real part of the complex signal. How do we choose A and φ or, equivalently, how do we choose the imaginary part, s_i? It is important to realize that the phase and amplitude of the real signal are not generally the same as the phase and amplitude of the complex signal. We have emphasized this by using different symbols for the phases and amplitudes in Eqs. (1.2) and (1.3).

How to unambiguously define amplitude and phase and how to define a complex signal corresponding to a real signal will be the subject of the next chapter. From the ideas and mathematics developed in this chapter we will see why defining a complex signal is advantageous and we will lay the groundwork to see how to do it. In this chapter we consider complex signals but make no assumptions regarding the amplitude and phase.

Energy Density or Instantaneous Power. How much energy a signal has and specifically how much energy it takes to produce it is a central idea. In the case of electromagnetic theory, the electric energy density is the absolute square of the electric field and similarly for the magnetic field. This was derived by Poynting using Maxwell's equations and is known as Poynting's theorem. In circuits, the energy

density is proportional to the voltage squared. For a sound wave it is the pressure squared. Therefore, the energy or intensity of a signal is generally $|s(t)|^2$. That is, in a small interval of time, Δt, it takes $|s(t)|^2 \Delta t$ amount of energy to produce the signal at that time. Since $|s(t)|^2$ is the energy per unit time it may be appropriately called the energy density or the instantaneous power since power is the amount of work per unit time. Therefore

$$|s(t)|^2 = \text{energy or intensity per unit time at time } t$$

$$(\text{energy density or instantaneous power})$$

$$|s(t)|^2 \Delta t = \text{the fractional energy in the time interval } \Delta t \text{ at time } t$$

Signal analysis has been extended to many diverse types of data, including economical and sociological. It is certainly not obvious that in those cases we can meaningfully talk about the energy density per unit time and take $|s(t)|^2$ to be its value. However, that is what is done by "analogy", which is appropriate if the results are fruitful.

Total Energy. If $|s(t)|^2$ is the energy per unit time, then the total energy is obtained by summing or integrating over all time,

$$E = \int |s(t)|^2 dt \tag{1.4}$$

For signals with finite energy we can take, without loss of generality, the total energy to be equal to one. For many signals the total energy is infinite. For example, a pure sine wave has infinite total energy, which is reasonable since to keep on producing it, work must be expended continually. Such cases can usually be handled without much difficulty by a limiting process.

Characterization of Time Wave Forms: Averages, Mean Time, and Duration. If we consider $|s(t)|^2$ as a density in time, the average time can be defined in the usual way any average is defined:

$$\langle t \rangle = \int t\,|s(t)|^2 dt \tag{1.5}$$

The reasons for defining an average are that it may give a gross characterization of the density and it may give an indication of where the density is concentrated. Many measures can be used to ascertain whether the density is concentrated around the average, the most common being the standard deviation, σ_t, given by

$$T^2 = \sigma_t^2 = \int (t - \langle t \rangle)^2\,|s(t)|^2 dt \tag{1.6}$$

$$= \langle t^2 \rangle - \langle t \rangle^2 \tag{1.7}$$

where $\langle t^2 \rangle$ is defined similarly to $\langle t \rangle$. The standard deviation is an indication of the duration of the signal: In a time $2\sigma_t$ most of the signal will have gone by. If the standard deviation is small then most of the signal is concentrated around the mean time and it will go by quickly, which is an indication that we have a signal of short duration; similarly for long duration. It should be pointed out that there are signals for which the standard deviation is infinite, although they may be finite energy signals. That usually indicates that the signal is very long lasting.

The average of any function of time, $g(t)$, is obtained by

$$\langle g(t) \rangle = \int g(t) \mid s(t) \mid^2 dt \tag{1.8}$$

Note that for a complex signal, time averages depend only on the amplitude.

Example 1.1: Gaussian Envelope.

Consider the following signal where the phase is arbitrary

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha(t-t_0)^2/2+j\varphi(t)} \tag{1.9}$$

The mean time and duration are calculated to be

$$\langle t \rangle = \sqrt{\frac{\alpha}{\pi}} \int t\, e^{-\alpha(t-t_0)^2} dt = t_0 \tag{1.10}$$

$$\langle t^2 \rangle = \sqrt{\frac{\alpha}{\pi}} \int t^2\, e^{-\alpha(t-t_0)^2} dt = \frac{1}{2\alpha} + t_0^2 \tag{1.11}$$

Hence

$$\sigma_t^2 = \langle t^2 \rangle - \langle t \rangle^2 = \frac{1}{2\alpha} \tag{1.12}$$

Example 1.2: Rectangular Amplitude.

A signal with constant amplitude from time t_1 to t_2 and zero otherwise is

$$s(t) = \sqrt{\frac{1}{t_2 - t_1}}\, e^{j\varphi(t)} \qquad t_1 \le t \le t_2 \tag{1.13}$$

The mean time and duration are

$$\langle t \rangle = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} t\, dt = \tfrac{1}{2}(t_2 + t_1) \tag{1.14}$$

$$\langle t^2 \rangle = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} t^2\, dt = \tfrac{1}{3}\left(t_2^2 + t_2 t_1 + t_1^2\right) \tag{1.15}$$

which gives

$$\sigma_t = \frac{1}{2\sqrt{3}}(t_2 - t_1) \tag{1.16}$$

For this case the signal unambiguously lasts $(t_2 - t_1)$. However, $2\sigma_t$ is pretty close to the true duration and has the advantage that it applies to any signal.

1.3 FREQUENCY DESCRIPTION OF SIGNALS

There are four main reasons for frequency analysis or spectral analysis. First, by spectrally analyzing a waveform we learn something about the source. That is how we have learned about the composition of the stars, paper, blood, and almost everything else.

Second, the propagation of waves through a medium generally depends on frequency. That is why visible light goes through glass but not through aluminum, while X-rays go through aluminum but not so readily through glass. The propagation of a wave through a medium is quite complicated but the basic effect is that waves of different frequencies propagate with different velocities. This is called dispersion because the earliest discovered manifestation was that a prism can "disperse" white light into different colors. The other important effect in propagation is the attenuation, the dying out or absorption of a wave. The amount of attenuation depends on the medium and the frequency. In the case of sound in normal conditions there is almost no attenuation, which is why we are able to hear from far away. In contrast, high frequency electromagnetic waves are damped within a short distance of entering the surface of a conductor. To study the propagation through frequency dependent mediums, we decompose the signal into its different frequency components, do the analysis for each frequency component, and then reconstruct the signal to obtain the resulting wave form. Hence, the need to decompose a signal into individual frequencies, which is what Fourier analysis does.

The third reason for spectral decomposition is that it often simplifies our understanding of the waveform. Simple sinusoids are common in nature which is consistent with the fact that for some of the fundamental equations of motion sinusoids are possible solutions. So are sums of sinusoids if the equation governing the physical quantity is linear. In general, a signal is messy, but often the mess is really the simple superposition of sine waves, which is simpler to understand and characterize.

Finally, Fourier analysis is a powerful mathematical tool for the solution of ordinary and partial differential equations.

Fourier Expansion. The signal is expanded in terms of sinusoids of different frequencies

$$s(t) = \frac{1}{\sqrt{2\pi}} \int S(\omega)\, e^{j\omega t} d\omega \tag{1.17}$$

The waveform is made up of the addition (linear superposition) of the simple waveforms, $e^{j\omega t}$, each characterized by the frequency, ω, and contributing a relative amount indicated by the coefficient, $S(\omega)$. $S(\omega)$ is obtained from the signal by

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int s(t)\, e^{-j\omega t}\, dt \tag{1.18}$$

and is called the spectrum or the Fourier transform. Since $S(\omega)$ and $s(t)$ are uniquely

related we may think of the spectrum as the signal in the frequency domain or frequency space or frequency representation.

Spectral Amplitude and Phase. As with the signal, it is often advantageous to write the spectrum in terms of its amplitude and phase,

$$S(\omega) = B(\omega)\, e^{j\psi(\omega)} \tag{1.19}$$

We call $B(\omega)$ the *spectral* amplitude and $\psi(\omega)$ the *spectral* phase to differentiate them from the phase and amplitude of the signal.

Energy Density Spectrum. In analogy with the time waveform we can take $|\, S(\omega)\,|^2$ to be the energy density per unit frequency:

$$|\, S(\omega)\,|^2 = \text{energy or intensity per unit frequency at frequency } \omega$$
$$\text{(energy density spectrum)}$$

$$|\, S(\omega)\,|^2\, \Delta\omega = \text{the fractional energy in the frequency interval } \Delta\omega$$
$$\text{at frequency } \omega$$

That $|\, S(\omega)\,|^2$ is the energy density can be seen by considering the simple case of one component, $s(t) = S(\omega_0)\, e^{j\omega_0 t}$, characterized by the frequency, ω_0. Since the signal energy is $|\, s(t)\,|^2$, then for this case the energy density is $|\, S(\omega_0)\,|^2$. Since all the energy is in one frequency, $|\, S(\omega_0)\,|^2$ must then be the energy for that frequency. In Chapter 15 we consider arbitrary representations and discuss this issue in greater detail. Also, the fact that the total energy of the signal is given by the integration of $|\, S(\omega)\,|^2$ over all frequencies, as discussed below, is another indication that it is the density in frequency.

The total energy of the signal should be independent of the method used to calculate it. Hence, if the energy density per unit frequency is $|\, S(\omega)\,|^2$, the total energy should be the integral of $|\, S(\omega)\,|^2$ over all frequencies and should equal the total energy of the signal calculated directly from the time waveform

$$E = \int |\, s(t)\,|^2\, dt = \int |\, S(\omega)\,|^2\, d\omega \tag{1.20}$$

This identity is commonly called Parceval's or Rayleigh's theorem.[3] To prove it consider

[3]The concept of the expansion of a function in a set of orthogonal functions started around the time of Laplace, Legendre, and Fourier. However, the full importance and development of the theory of orthogonal functions is due to Rayleigh some one hundred years later, around 1890.

$$E = \int |s(t)|^2\,dt = \frac{1}{2\pi}\iiint S^*(\omega')\,S(\omega)\,e^{j(\omega-\omega')t}\,d\omega\,d\omega'\,dt \tag{1.21}$$

$$= \iint S^*(\omega')\,S(\omega)\,\delta(\omega-\omega')\,d\omega\,d\omega' \tag{1.22}$$

$$= \int |S(\omega)|^2\,d\omega \tag{1.23}$$

where in going from Eq. (1.21) to (1.22) we have used

$$\frac{1}{2\pi}\int e^{j(\omega-\omega')t}\,dt = \delta(\omega-\omega') \tag{1.24}$$

Mean Frequency, Bandwidth, and Frequency Averages. If $|S(\omega)|^2$ represents the density in frequency then we can use it to calculate averages, the motivation being the same as in the time domain, namely that it gives a rough idea of the main characteristics of the spectral density. The average frequency, $\langle\omega\rangle$, and its standard deviation, σ_ω (commonly called the root mean square bandwidth and signified by B), are given by

$$\langle\omega\rangle = \int \omega\,|S(\omega)|^2\,d\omega \tag{1.25}$$

$$B^2 = \sigma_\omega^2 = \int (\omega - \langle\omega\rangle)^2\,|S(\omega)|^2\,d\omega \tag{1.26}$$

$$= \langle\omega^2\rangle - \langle\omega\rangle^2 \tag{1.27}$$

and the average of any frequency function, $g(\omega)$, is

$$\langle g(\omega)\rangle = \int g(\omega)\,|S(\omega)|^2\,d\omega \tag{1.28}$$

What Does the Energy Density Spectrum Tell Us? The energy density spectrum tells us which frequencies existed during the total duration of the signal. It gives us no indication as to when these frequencies existed. The mathematical and physical ideas needed to understand and describe how the frequencies are changing in time is the subject of this book.

1.4 SIMPLE CALCULATION TRICKS

Suppose we want to calculate the average frequency. From the definition, Eq. (1.25), it appears that we first have to obtain the spectrum. But that is not so. There is an important method or "trick" that avoids the calculation of the spectrum, simplifies the algebra immensely, and moreover will be central to our development in the later chapters for deriving time-frequency representations. In this chapter we merely

discuss the method in the context of its significant calculational merit: To calculate averages of frequency functions we do not have to calculate the Fourier transform, $S(\omega)$. It can be done directly from the signal and done simply. We first state the result for average frequency and average square frequency, give a few examples to display the power of the method, and then discuss the general case.

The average frequency and average square frequency are given by

$$\langle \omega \rangle = \int \omega \, |S(\omega)|^2 \, d\omega = \int s^*(t) \, \frac{1}{j} \frac{d}{dt} \, s(t) \, dt \tag{1.29}$$

$$\langle \omega^2 \rangle = \int \omega^2 \, |S(\omega)|^2 \, d\omega = \int s^*(t) \left(\frac{1}{j} \frac{d}{dt} \right)^2 s(t) \, dt \tag{1.30}$$

$$= - \int s^*(t) \, \frac{d^2}{dt^2} \, s(t) \, dt \tag{1.31}$$

$$= \int \left| \frac{d}{dt} \, s(t) \right|^2 dt \tag{1.32}$$

That is, to calculate the average frequency, we differentiate the waveform and carry out the integration as per Eq. (1.29). For the average square frequency we have a choice. We can calculate the second derivative and use Eq. (1.31) or calculate just the first derivative and use Eq. (1.32). Either way we get the same answer. The bandwidth is given by

$$\sigma_\omega^2 = \int (\omega - \langle \omega \rangle)^2 \, |S(\omega)|^2 \, d\omega \tag{1.33}$$

$$= \int s^*(t) \left(\frac{1}{j} \frac{d}{dt} - \langle \omega \rangle \right)^2 s(t) \, dt \tag{1.34}$$

$$= \int \left| \left(\frac{1}{j} \frac{d}{dt} - \langle \omega \rangle \right) s(t) \right|^2 dt \tag{1.35}$$

These results are far from obvious, although they are easy to prove and we will do so shortly. First, a few examples.

Example 1.3: Quadratic Phase with Gaussian Envelope.

Take

$$s(t) = (\alpha/\pi)^{1/4} \, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \tag{1.36}$$

Now

$$\frac{1}{j} \frac{d}{dt} \, s(t) = [\, j\alpha t + \beta t + \omega_0 \,] \, s(t) \tag{1.37}$$

and therefore

$$\langle \omega \rangle = \int s^*(t) \{ j\alpha t + \beta t + \omega_0 \} s(t)\, dt = \sqrt{\frac{\alpha}{\pi}} \int \{ j\alpha t + \beta t + \omega_0 \} e^{-\alpha t^2}\, dt = \omega_0 \tag{1.38}$$

Also,

$$\langle \omega^2 \rangle = \sqrt{\frac{\alpha}{\pi}} \int | j\alpha t + \beta t + \omega_0 |^2 e^{-\alpha t^2}\, dt = \frac{\alpha^2 + \beta^2}{2\alpha} + \omega_0^2 \tag{1.39}$$

which gives

$$\sigma_\omega^2 = \frac{\alpha^2 + \beta^2}{2\alpha} \tag{1.40}$$

For this case the spectrum and energy density spectrum may be obtained without difficulty,

$$S(\omega) = \sqrt{\frac{\sqrt{\alpha}}{\sqrt{\pi}(\alpha - j\beta)}}\, e^{-(\omega-\omega_0)^2/2(\alpha - j\beta)} \tag{1.41}$$

$$| S(\omega) |^2 = \sqrt{\frac{\alpha}{\pi(\alpha^2 + \beta^2)}}\, e^{-\alpha(\omega-\omega_0)^2/(\alpha^2+\beta^2)} \tag{1.42}$$

and the averages may be calculated directly as a check.

Example 1.4: Sinusoidal Modulation.

This example is chosen to impress the reader with the simplicity and power of the method. Suppose we have the signal

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + jm \sin \omega_m t + j\omega_0 t} \tag{1.43}$$

and we want to calculate the average frequency and bandwidth.

The hard way: The reader is welcome to try to calculate the Fourier transform of $s(t)$ and then use it to calculate the average frequency and bandwidth.

The easy way: Taking the derivative of the signal we have

$$\frac{1}{j}\frac{d}{dt} s(t) = [j\alpha t + \beta t + m\omega_m \cos \omega_m t + \omega_0]\, s(t) \tag{1.44}$$

and therefore, using Eq. (1.29)

$$\langle \omega \rangle = \int s^*(t) \frac{1}{j}\frac{d}{dt} s(t)\, dt \tag{1.45}$$

$$= \sqrt{\frac{\alpha}{\pi}} \int \{ j\alpha t + \beta t + m\omega_m \cos \omega_m t + \omega_0 \} e^{-\alpha t^2}\, dt \tag{1.46}$$

$$= m\omega_m e^{-\omega_m^2/(4\alpha)} + \omega_0 \tag{1.47}$$

For the average square frequency we immediately get, using Eq. (1.32),

$$\langle \omega^2 \rangle = \sqrt{\frac{\alpha}{\pi}} \int \alpha t + j\beta t + jm\omega_m \cos \omega_m t + j\omega_0 \,|^2 \, e^{-\alpha t^2} \, dt \tag{1.48}$$

This integral is easy to evaluate since all the terms in the integrand that are linear in time drop out due to symmetry. All complex terms drop out also since we know that the answer has to be real; therefore the complex terms must add up to zero and do not have to be calculated. The remaining terms are simple integrals. Evaluation leads to

$$\langle \omega^2 \rangle = \frac{\alpha^2 + \beta^2}{2\alpha} + \omega_0^2 + 2\, m\, \omega_m\, \omega_0\, e^{-\omega_m^2/4\alpha} + \frac{m^2\omega_m^2}{2} \left(1 + e^{-\omega_m^2/\alpha}\right) \tag{1.49}$$

which gives

$$\sigma_\omega^2 = \frac{\alpha^2 + \beta^2}{2\alpha} + \frac{m^2\omega_m^2}{2} \left(1 - e^{-\omega_m^2/2\alpha}\right)^2 \tag{1.50}$$

The Frequency Operator and the General Case. For convenience one defines the frequency operator by

$$\mathcal{W} = \frac{1}{j} \frac{d}{dt} \tag{1.51}$$

and it is understood that repeated use, denoted by $\mathcal{W}^n$, is to mean repeated differentiation,

$$\mathcal{W}^n s(t) = \left(\frac{1}{j}\right)^n \frac{d^n}{dt^n} s(t) \tag{1.52}$$

We are now in a position to state and prove the general result that the average of a frequency function can be calculated directly from the signal by way of

$$\langle g(\omega) \rangle = \int g(\omega) \,|\, S(\omega) \,|^2 \, d\omega \tag{1.53}$$

$$= \int s^*(t)\, g(\mathcal{W})\; s(t)\, dt \tag{1.54}$$

$$= \int s^*(t)\, g\left(\frac{1}{j}\frac{d}{dt}\right) s(t)\, dt \tag{1.55}$$

In words: Take the function $g(\omega)$ and replace the ordinary variable ω by the operator $\frac{1}{j}\frac{d}{dt}$; operate on the signal, multiply by the complex conjugate signal, and integrate. Before we prove this we must face a side issue and discuss the meaning of $g(\mathcal{W})$ for an arbitrary function. If g is ω^n, then the procedure is clear, as indicated by Eq. (1.52). If g is the sum of powers, then it is also clear. For a general function we first expand the function in a Taylor series and then substitute the operator $\mathcal{W}$ for ω. That is,

$$\text{if} \quad g(\omega) = \sum g_n\, \omega^n \quad \text{then} \quad g(\mathcal{W}) = \sum g_n \mathcal{W}^n \tag{1.56}$$

To prove the general result, Eq. (1.54), we first prove it for the case of the average frequency:

$$\langle \omega \rangle = \int \omega \, |S(\omega)|^2 \, d\omega = \frac{1}{2\pi} \iiint \omega s^*(t)\, s(t')\, e^{j(t-t')\omega}\, d\omega\, dt'\, dt \tag{1.57}$$

$$= \frac{1}{2\pi j} \iiint s^*(t)\, s(t') \frac{\partial}{\partial t} e^{j(t-t')\omega}\, d\omega\, dt'\, dt \tag{1.58}$$

$$= \frac{1}{j} \iint s^*(t) \frac{\partial}{\partial t} \delta(t-t')\, s(t')\, dt'\, dt \tag{1.59}$$

$$= \int s^*(t) \frac{1}{j} \frac{d}{dt} s(t)\, dt \tag{1.60}$$

These steps can be repeated as often as necessary to prove it for $g = \omega^n$. Hence

$$\langle \omega^n \rangle = \int s^*(t) \left(\frac{1}{j} \frac{d}{dt} \right)^n s(t)\, dt = \int s^*(t)\, \mathcal{W}^n s(t)\, dt \tag{1.61}$$

Having proved the general result for functions of the form $g = \omega^n$, we now prove it for an arbitrary function, $g(\omega)$, by expanding the function in a Taylor series

$$\langle g(\omega) \rangle = \int g(\omega)\, |S(\omega)|^2\, d\omega = \int \sum g_n \omega^n\, |S(\omega)|^2\, d\omega \tag{1.62}$$

$$= \sum g_n \int s^*(t)\, \mathcal{W}^n\, s(t)\, dt \tag{1.63}$$

$$= \int s^*(t)\, g(\mathcal{W}^n)\, s(t)\, dt \tag{1.64}$$

Manipulation Rules. The frequency operator is a Hermitian operator, which means that for any two signals, $s_1(t)$ and $s_2(t)$,

$$\int s_1^*(t)\, \mathcal{W}\, s_2(t)\, dt = \int s_2(t)\, \{\mathcal{W}\, s_1(t)\}^*\, dt \tag{1.65}$$

This is readily proved by integrating by parts. Also, a real function, $g(\omega)$, of a Hermitian operator, $g(\mathcal{W})$, is also Hermitian. That is,

$$\int s_1^*(t) g(\mathcal{W})\, s_2(t) = \int s_2(t)\, \{g(\mathcal{W})\, s_1(t)\}^* \qquad [\text{if } g(\omega) \text{ is real}] \tag{1.66}$$

An important property of Hermitian operators is that their average value as defined by Eq. (1.64) must be real, so in the manipulation of averages we can simply discard the imaginary terms since we are assured that they add up to zero.

We now derive the second simplification, Eq. (1.32), for the average square of frequency. We have

$$\langle \omega^2 \rangle = \int s^*(t)\, \mathcal{W}^2\, s(t)\, dt = \int s^*(t)\, \mathcal{W}\mathcal{W}\, s(t)\, dt \tag{1.67}$$

$$= \int \mathcal{W}\, s(t)\, \{\mathcal{W} s(t)\}^*\, dt \tag{1.68}$$

$$= \int |\, \mathcal{W}\, s(t)\,|^2\, dt \tag{1.69}$$

This is an immense simplification since not only do we not have to find the spectrum, we also avoid a double differentiation.

The Time Operator. In the above discussion we emphasized that we can avoid the necessity of calculating the spectrum for the calculation of averages of frequency functions. Similarly, if we have a spectrum and want to calculate time averages, we can avoid the calculation of the signal. The time operator is defined by

$$\mathcal{T} = -\frac{1}{j}\frac{d}{d\omega} \tag{1.70}$$

and the same arguments and proofs as above lead to

$$\langle g(t) \rangle = \int g(t)\, |\, s(t)\,|^2\, dt = \int S^*(\omega)\, g(\mathcal{T})\, S(\omega)\, d\omega \tag{1.71}$$

In particular,

$$\langle t \rangle = \int t\, |\, s(t)\,|^2\, dt = \int S^*(\omega) \left(-\frac{1}{j}\frac{d}{d\omega}\right) S(\omega)\, d\omega \tag{1.72}$$

$$\langle t^2 \rangle = \int t^2\, |\, s(t)\,|^2\, dt = \int S^*(\omega) \left(-\frac{1}{j}\right)^2 \frac{d^2}{d\omega^2}\, S(\omega)\, d\omega \tag{1.73}$$

$$= -\int S^*(\omega)\, \frac{d^2}{d\omega^2} S(\omega)\, d\omega \tag{1.74}$$

$$= \int \left| \frac{d}{d\omega}\, S(\omega) \right|^2 d\omega \tag{1.75}$$

Example 1.5: Mean Time and Duration from the Spectrum.

Consider the normalized spectrum

$$S(\omega) = \sqrt{\frac{\alpha^{2n+1}}{(2n)!}}\,\omega^n\, e^{-\alpha\omega/2 - jt_0\omega} \qquad \omega \geq 0 \tag{1.76}$$

We want to calculate the mean time and duration directly from the spectrum. We have

$$\mathcal{T}S(\omega) = -\frac{1}{j}\frac{d}{d\omega}S(\omega) \;=\; [\,t_0 - j\alpha/2 + jn/\omega\,]\,S(\omega) \tag{1.77}$$

and therefore

$$\langle t\rangle \;=\; \int_0^\infty [\,t_0 - j\alpha/2 + jn/\omega\,]\,|\,S(\omega)\,|^2\,d\omega = t_0 \tag{1.78}$$

In evaluating this we don't have to do any algebra. The complex terms must be zero and hence need no evaluation. Also, since the spectrum is normalized, the real part integrates to t_0. Similarly

$$\langle t^2\rangle \;=\; \int_0^\infty |\,t_0 - j\alpha/2 + jn/\omega\,|^2\,|\,S(\omega)\,|^2\,d\omega = t_0^2 + \frac{1}{4}\frac{\alpha^2}{2n-1} \tag{1.79}$$

which gives

$$T^2 = \langle t^2\rangle - \langle t\rangle^2 = \frac{1}{4}\frac{\alpha^2}{2n-1} \tag{1.80}$$

The signal corresponding to this spectrum is

$$s(t) = n!\sqrt{\frac{\alpha^{2n+1}}{2\pi(2n)!}}\;\frac{1}{[\,\alpha/2 - j(t-t_0)\,]^{n+1}} \tag{1.81}$$

and the above average may be checked directly. Also, for this signal

$$\langle\omega\rangle = \frac{2n+1}{\alpha} \qquad ; \qquad \langle\omega^2\rangle = \frac{(2n+2)(2n+1)}{\alpha^2} \qquad ; \qquad B^2 = \frac{2n+1}{\alpha^2} \tag{1.82}$$

The Translation Operator. Many results in signal analysis are easily derived by the use of the translation operator, $e^{j\tau\mathcal{W}}$, where τ is a constant. Its effect on a function of time is

$$e^{j\tau\mathcal{W}}f(t) \;=\; f(t+\tau) \tag{1.83}$$

That is, the translation operator translates functions by τ. Note that it is not Hermitian. To prove Eq. (1.83) consider

$$e^{j\tau\mathcal{W}}f(t) \;=\; \sum_{n=0}^{\infty}\frac{(j\tau)^n\mathcal{W}^n}{n!}\,f(t) \;=\; \sum_{n=0}^{\infty}\frac{\tau^n}{n!}\frac{d^n}{dt^n}f(t) \tag{1.84}$$

But this is precisely the Taylor expansion of $f(t+\tau)$ and hence Eq. (1.83) follows. Similarly, the operator $e^{-j\theta\mathcal{T}}$ translates frequency functions,

$$e^{-j\theta\mathcal{T}}S(\omega) \;=\; S(\omega+\theta) \tag{1.85}$$

The Operator Method. We have shown how the properties of the time and frequency operators can be used to simplify calculations. Indeed, it is a powerful method for that purpose of which, hopefully by now, the reader is convinced. However, the operator method is not only a calculational tool but one of fundamental significance which will be developed and discussed in the later chapters of this book.

1.5 BANDWIDTH EQUATION

We now aim at expressing the mean frequency and bandwidth in terms of the time waveform.[184, 135, 353] The results obtained will give a strong motivation for time-frequency analysis and for the introduction of the complex signal.

Mean Frequency. Consider first

$$\mathcal{W}s(t) = \mathcal{W}A(t)\,e^{j\varphi(t)} = \frac{1}{j}\frac{d}{dt}A(t)\,e^{j\varphi(t)} \tag{1.86}$$

$$= \left(\varphi'(t) - j\frac{A'(t)}{A(t)}\right) s(t) \tag{1.87}$$

Therefore, the mean frequency is

$$\langle\omega\rangle = \int \omega\,|\,S(\omega)\,|^2\,d\omega = \int s^*(t)\,\frac{1}{j}\frac{d}{dt}\,s(t)\,dt \tag{1.88}$$

$$= \int \left(\varphi'(t) - j\frac{A'(t)}{A(t)}\right) A^2(t)\,dt \tag{1.89}$$

The second term is zero. This can be seen in two ways. First, since that term is purely imaginary it must be zero for $\langle\omega\rangle$ to be real. Alternatively, we note that the integrand of the second term is a perfect differential that integrates to zero. Hence

$$\langle\omega\rangle = \int \varphi'(t)\,|\,s(t)\,|^2\,dt = \int \varphi'(t)\,A^2(t)\,dt \tag{1.90}$$

This is an interesting and important result because it says that the average frequency may be obtained by integrating "something" with the density over all time. This something must be the instantaneous value of the quantity for which we are calculating the average. In this case the something is the derivative of the phase, which may be appropriately called the frequency at each time or the instantaneous frequency, $\omega_i(t)$,

$$\omega_i(t) = \varphi'(t) \tag{1.91}$$

Instantaneous frequency, as an empirical phenomenon, is experienced daily as changing colors, changing pitch, etc. Whether or not the derivative of the phase meets our intuitive concept of instantaneous frequency is a central issue and is addressed in subsequent chapters. In addition, this brings up the question that if

instantaneous frequency is the derivative of the phase, what phase are we to use? According to this definition the instantaneous frequency of a real signal is zero, which is clearly an absurd result. The means to get around these difficulties are developed in the next chapter with the introduction of the complex signal which corresponds to the real signal.

Bandwidth Equation.[184, 145] Now consider the bandwidth

$$B^2 = \sigma_\omega^2 = \int (\omega - \langle\omega\rangle)^2 \, |S(\omega)|^2 \, d\omega \tag{1.92}$$

$$= \int s^*(t) \left(\frac{1}{j}\frac{d}{dt} - \langle\omega\rangle\right)^2 s(t)\, dt \tag{1.93}$$

$$= \int \left| \left(\frac{1}{j}\frac{d}{dt} - \langle\omega\rangle\right) s(t) \right|^2 dt \tag{1.94}$$

$$= \int \left| \frac{1}{j}\frac{A'(t)}{A(t)} + \varphi'(t) - \langle\omega\rangle \right|^2 A^2(t)\, dt \tag{1.95}$$

or

$$B^2 = \int \left(\frac{A'(t)}{A(t)}\right)^2 A^2(t)\, dt + \int (\varphi'(t) - \langle\omega\rangle)^2 \, A^2(t)\, dt \tag{1.96}$$

While $A^2(t)$ can be canceled in the first term, the expression as written is preferable because it explicitly shows that the bandwidth is the average of two terms, one depending on the amplitude and the other depending only on the phase. What is the meaning of this formula? What is the significance of the two terms? The explanation will come with the ideas of describing a signal jointly in time and frequency, which is the topic of this book. The same steps lead to

$$\langle\omega^2\rangle = \int \omega^2 \, |S(\omega)|^2 \, d\omega \tag{1.97}$$

$$= \int \left(\frac{A'(t)}{A(t)}\right)^2 A^2(t)\, dt + \int \varphi'^2(t)\, A^2(t)\, dt \tag{1.98}$$

Calculation Techniques. These equations, besides being inherently interesting and offering a challenge for their interpretation, are very useful for practical calculations, as the following example shows.

Example 1.6: Cubic and Quadratic Phase Modulation.

Consider

$$s(t) = (\alpha/\pi)^{1/4} \, e^{-\alpha t^2/2 + j\gamma t^3/3 + j\beta t^2/2 + j\omega_0 t} \tag{1.99}$$

The derivative of the phase is given by $\varphi' = \omega_0 + \beta t + \gamma t^2$ and hence

$$\langle \omega \rangle = \int \varphi'(t) \, | \, s(t) \, |^2 \, dt = \int (\omega_0 + \beta t + \gamma t^2) \, | \, s(t) \, |^2 \, dt = \frac{\gamma}{2\alpha} + \omega_0 \qquad (1.100)$$

For the average square frequency we note that $A'/A = -\alpha t$ and therefore

$$\langle \omega^2 \rangle = \int (-\alpha t)^2 \, | \, s(t) \, |^2 + \int (\omega_0 + \beta t + \gamma t^2)^2 \, | \, s(t) \, |^2 \qquad (1.101)$$

$$= \frac{\alpha^2 + \beta^2}{2\alpha} + \omega_0^2 + \frac{3\gamma^2}{4\alpha^2} + \frac{\gamma\omega_0}{\alpha} \qquad (1.102)$$

from which we obtain

$$\sigma_\omega^2 = \frac{\alpha^2 + \beta^2}{2\alpha} + \frac{\gamma^2}{2\alpha^2} \qquad (1.103)$$

1.6 AM AND FM CONTRIBUTIONS TO THE BANDWIDTH[135, 145]

What contributes to the bandwidth? The bandwidth is an indication of the spread in frequencies for the duration of the signal. If a sound is produced at 1000 Hz and increased in frequencies to 1200 Hz at more or less constant amplitude, we expect that the spread in frequencies will be about 200 Hz, and that is indeed the case. However, if we have a signal of constant frequency at 1000 Hz, then we can also achieve the same bandwidth by making it a short duration signal or by varying the amplitude rapidly. Therefore the bandwidth does not give us a good indication of whether the spread of frequencies is due to deviations from the average frequency or to fast amplitude change or a to combination of both. We now develop a measure of these two qualitatively different contributions. But first we illustrate with an example. Examine the set of signals shown in Fig. 1.1. They all have the same bandwidth but they are qualitatively different. In (a) the variation in frequency from the mean is zero while the variation in amplitude is large. For the signal (d) the variation in frequency is high but has low amplitude variation. The signals (a) – (d) are progressions from one extreme to the other. These two contributions to the bandwidth are apparent in the bandwidth equation, Eq. (1.96). Since the first term averages an amplitude term over all time and the second term averages a phase dependent term, it is natural to define the AM and FM contributions by

$$B_{AM}^2 = \int A'^2(t) \, dt \quad ; \quad B_{FM}^2 = \int (\varphi'(t) - \langle \omega \rangle)^2 \, A^2(t) \, dt \qquad (1.104)$$

with

$$B^2 = B_{AM}^2 + B_{FM}^2 \qquad (1.105)$$

We also define the fractional contributions and the ratio of the two contributions by

$$r_{FM} = \frac{B_{FM}}{B} \qquad r_{AM} = \frac{B_{AM}}{B} \qquad (1.106)$$

To gain some insight we consider a few examples.

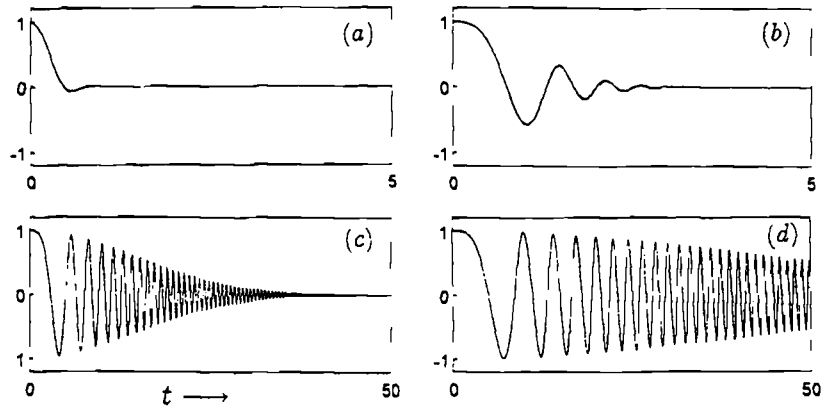

Fig. 1.1 The real part of the signal $s(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2 + j\beta t^2/2}$ for various values of α and β. All signals have the same bandwidth. There is a large amplitude modulation in (a) and a small amplitude modulation in (d) but large frequency deviations from the mean frequency. The AM and FM contributions as defined by Eq. (1.104) are a measure of the two qualitatively different effects that are contributing to the bandwidth. The AM contributions are 71, 18, 1.3 and 0.4 % respectively, as we progress from (a) to (d). The values for α, β are (16,16), (1, 5.56), (0.005, 0.4), and (0.0005, 0.127) and each pair gives a bandwidth of 4.

Example 1.7: Arbitrary Amplitude with Constant Frequency.

Consider the signal

$$s(t) = A(t)\, e^{j\omega_0 t} \tag{1.107}$$

where the amplitude, $A(t)$, is arbitrary. Since the instantaneous frequency is a constant we immediately see that B_{FM} is zero and all the contributions to the bandwidth come from the amplitude modulation,

$$r_{AM} = 1 \quad ; \quad r_{FM} = 0 \tag{1.108}$$

This is reasonable because a constant frequency does not contribute to the bandwidth, since the bandwidth indicates the spread in frequency.

Example 1.8: Linear Frequency Modulation.

Take

$$s(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \tag{1.109}$$

Direct calculation yields that

$$B_{AM} = \sqrt{\frac{\alpha}{2}} \qquad B_{FM} = \frac{\beta}{\sqrt{2\alpha}} \tag{1.110}$$

$$r_{AM} = \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} \qquad r_{FM} = \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} \tag{1.111}$$

A "large" bandwidth can be achieved in two qualitatively different ways. Since the instantaneous frequency is $\omega_i = \omega_0 + \beta t$, we can get a large spread by waiting long enough. We achieve this by making the duration long, which is accomplished by taking a small α. However, we can also get a large bandwidth by making the duration very small, that is by taking a large α.

Example 1.9: Sinusoidal FM.

As a further example consider a sinusoidal modulated frequency with a Gaussian amplitude modulation,

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + jm \sin \omega_m t + j\omega_0 t} \tag{1.112}$$

The amplitude and frequency modulation contributions are

$$B_{AM} = \sqrt{\frac{\alpha}{2}} \quad ; \quad B_{FM} = \frac{m\omega_m}{\sqrt{2}} \left(1 - e^{-\omega_m^2/2\alpha}\right) \tag{1.113}$$

The instantaneous frequency is $m\omega_m \cos \omega_m t + \omega_0$ and we see that we can indeed achieve a large bandwidth if we have a large modulation index, m, or large ω_m, since by making either one large we get a large instantaneous frequency change if we wait long enough. As with the previous example we can achieve the same bandwidth by making the duration small. Qualitatively we have two very different effects and the AM and FM expressions measure these two effects.

1.7 DURATION AND MEAN TIME IN TERMS OF THE SPECTRUM

We have written the bandwidth and mean frequency in terms of the phase and amplitude of the signal. The identical derivations can be used to write the mean time and duration in terms of the amplitude and phase of the spectrum. In particular,

$$\langle t \rangle = -\int \psi'(\omega) \, |S(\omega)|^2 \, d\omega \tag{1.114}$$

and

$$T^2 = \sigma_t^2 = \int \left(\frac{B'(\omega)}{B(\omega)}\right)^2 B^2(\omega)\, d\omega + \int \left(\psi'(\omega) + \langle t \rangle\right)^2 B^2(\omega)\, d\omega \tag{1.115}$$

Examine Eq. (1.114). It says that if we average $-\psi'(\omega)$ over all frequencies we will get the average time. Therefore we may consider $-\psi'(\omega)$ to be the average time for a particular frequency. This is called the group delay and we shall use the following notation for it

$$t_g(\omega) = -\psi'(\omega) \tag{1.116}$$

In Section 1.6 we showed how amplitude and frequency modulation contribute to the bandwidth. Similarly, the amplitude and phase variations of the spectrum

contribute to the duration. We define the spectral amplitude modulation (SAM) and spectral phase modulation (SPM) contributions to the duration by

$$T_{SAM}^2 = \int B'^2(\omega)\, d\omega \qquad T_{SPM}^2 = \int (\psi'(\omega) + \langle t \rangle)^2\, B^2(\omega)\, d\omega \tag{1.117}$$

with

$$T^2 = T_{SAM}^2 + T_{SPM}^2 \tag{1.118}$$

Example 1.10: Linear Frequency Modulation.

Consider the spectrum of the signal

$$s(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \tag{1.119}$$

which we have given in Eq. (1.41) and which we rewrite here in terms of the spectral phase and amplitude,

$$S(\omega) = \frac{(\alpha/\pi)^{1/4}}{\sqrt{\alpha - j\beta}}\, \exp\left[-\frac{\alpha(\omega - \omega_0)^2}{2(\alpha^2 + \beta^2)} - j\frac{\beta(\omega - \omega_0)^2}{2(\alpha^2 + \beta^2)} \right] \tag{1.120}$$

Therefore,

$$t_g = \frac{\beta}{\alpha^2 + \beta^2}(\omega - \omega_0) \tag{1.121}$$

Also,

$$T_{SAM}^2 = \frac{\alpha}{2(\alpha^2 + \beta^2)} \qquad ; \qquad T_{SPM}^2 = \frac{\beta^2}{2\alpha(\alpha^2 + \beta^2)} \tag{1.122}$$

1.8 THE COVARIANCE OF A SIGNAL

If we want to determine, in a crude way, whether there is a relationship between height and weight of a population, we do it by calculating the covariance or the correlation coefficient. In the same way, we want a measure of how time and instantaneous frequency are related. To see this we introduce the concept of covariance or correlation for signals. Consider the quantity

$$\langle t\, \varphi'(t) \rangle = \int t\, \varphi'(t)\, |\, s(t)\, |^2\, dt \tag{1.123}$$

which may be thought of as the average of time multiplied by the instantaneous frequency. Now if time and frequency have nothing to do with each other then we would expect $\langle t\, \varphi'(t) \rangle$ to equal $\langle t \rangle \langle \phi'(t) \rangle = \langle t \rangle \langle \omega \rangle$. Therefore the excess of $\langle t\, \varphi'(t) \rangle$ over $\langle t \rangle \langle \omega \rangle$ is a good measure of how time is correlated with instantaneous frequency. This is precisely what is called the covariance for variables such as height and weight, and we similarly define the covariance of a signal by

$$\text{Cov}_{t\omega} = \langle t\,\varphi'(t)\rangle - \langle t\rangle\langle\omega\rangle \tag{1.124}$$

The correlation coefficient is the normalized covariance

$$r = \frac{\text{Cov}_{t\omega}}{\sigma_t\sigma_\omega} \tag{1.125}$$

The reason for defining the correlation coefficient in the standard considerations such as for height and weight is that it ranges from minus one to one and hence gives an absolute measure. That is not the case here, but nonetheless it does give a good indication of the relationship between time and frequency.

Covariance in Terms of the Spectrum. Suppose we place ourselves in the frequency domain so that time is t_g and frequency is ω. It is reasonable to define the covariance by

$$\text{Cov}_{t\omega} = \langle t_g\,\omega\rangle - \langle t\rangle\langle\omega\rangle \tag{1.126}$$

with

$$\langle t_g\,\omega\rangle = -\langle\omega\,\psi'(\omega)\rangle = -\int \omega\,\psi'(\omega)\,|\,S(\omega)\,|^2\,d\omega \tag{1.127}$$

Are these two definitions, Eq. (1.124) and Eq. (1.127), identical? For them to be identical we must have

$$\int t\,\varphi'(t)\,|\,s(t)\,|^2\,dt = -\int \omega\,\psi'(\omega)\,|\,S(\omega)\,|^2\,d\omega \tag{1.128}$$

In fact, this equation is true but not obvious. It can be proven by brute force, but a very simple proof is given in Section 15.4. It is an interesting identity because it connects the phases and amplitudes of the signal and spectrum.

When Is the Covariance Equal to Zero? If the covariance is to be an indication of how instantaneous frequency and time are related, then when the instantaneous frequency does not change the covariance should be zero. That is indeed the case. Consider

$$s(t) = A(t)\,e^{j\omega_0 t} \tag{1.129}$$

where the amplitude modulation is arbitrary. Now

$$\langle t\,\varphi'(t)\rangle = \int t\,\omega_0\,|\,A(t)\,|^2\,dt = \omega_0\langle t\rangle \tag{1.130}$$

But since $\langle\omega\rangle = \omega_0$, we have

$$\langle\omega\rangle\langle t\rangle = \omega_0\langle t\rangle \tag{1.131}$$

and therefore the covariance and correlation coefficient are equal to zero. Similarly, if we have a spectrum of the form $S(\omega) = B(\omega)e^{j\omega t_0}$, then there is no correlation

between time and frequency. In general,

$$\text{Cov}_{t\omega} = 0 \quad ; \quad r = 0 \qquad \text{for } s(t) = A(t)\, e^{j\omega_0 t} \tag{1.132}$$

or

$$S(\omega) = B(\omega)\, e^{j\omega t_0} \tag{1.133}$$

Covariance of a Real Signal or Spectrum. Since the phase of a real signal is zero, the derivative is zero and hence the covariance of a real signal is always zero. This result misrepresents the physical situation and is another reason for defining a complex signal as we discuss in the next chapter. Similarly, signals that are symmetric in time have real spectra and their covariance is zero.

Example 1.11: Chirp.

For the signal

$$s(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \tag{1.134}$$

the average time is zero and therefore the covariance is

$$\text{Cov}_{t\omega} = \langle t\varphi'(t) \rangle = \int t\,(\beta t + \omega_0)\,|\, s(t)\,|^2\, dt = \beta \int t^2\,|\, s(t)\,|^2\, dt \tag{1.135}$$

This is readily evaluated to give

$$\text{Cov}_{t\omega} = \frac{\beta}{2\alpha} \quad ; \quad r = \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} \tag{1.136}$$

When $\beta \to 0$ the correlation coefficient goes to zero, in conformity with our discussion that for a constant frequency signal, time and frequency are not correlated. As $\alpha \to 0$ the correlation coefficient goes to ± 1, depending on the sign of β. That is, for constant amplitude we have total correlation. If β is positive, the instantaneous frequency increases with increasing time and that is what the value of $+1$ for r tells us. There is high positive correlation in the sense that when time is large, the value of the instantaneous frequency is also large. For β negative we have negative correlation, which is also reasonable, since as time is increasing the instantaneous frequency is decreasing. For $\alpha \to \infty$ we have the correlation coefficient going to zero, which is also reasonable because for that case we have a short duration signal and the frequencies bear no relation to the chirping but are simply due to the short duration.

1.9 THE FOURIER TRANSFORM OF THE TIME AND FREQUENCY DENSITIES

Both $|\, s(t)\,|^2$ and $S(\omega)\,|^2$ are densities. The Fourier transform of a density is called the characteristic function. It is a powerful method for studying densities, as we will see in the later chapters. Here we present a simple way to calculate the characteristic function of $|\, s(t)\,|^2$ and $|\, S(\omega)\,|^2$ and show the relationship to the translation

operator. The characteristic function for the energy density spectrum is

$$R(\tau) = \int |S(\omega)|^2 e^{j\tau\omega}\, d\omega = \int s^*(t)\, e^{j\tau\mathcal{W}} s(t)\, dt \tag{1.137}$$

But we know from Section 1.4 that $e^{j\tau\mathcal{W}}$ is the translation operator and therefore

$$R(\tau) = \int s^*(t)\, s(t+\tau)\, dt \tag{1.138}$$

Since this function compares or correlates the signal at two different times it is commonly called the (deterministic) autocorrelation function. Inversely, we have

$$|S(\omega)|^2 = \frac{1}{2\pi} \int R(\tau)\, e^{-j\omega\tau}\, d\tau \tag{1.139}$$

The generalization of this result to random signals is the Wiener-Khinchin theorem.

Similarly, the characteristic function in the frequency domain is

$$R(\theta) = \int |s(t)|^2 e^{j\theta t}\, dt = \int S^*(\omega)\, e^{j\theta\mathcal{T}} S(\omega)\, d\omega = \int S^*(\omega)\, S(\omega-\theta)\, d\omega \tag{1.140}$$

and hence

$$|s(t)|^2 = \frac{1}{2\pi} \int R(\theta)\, e^{-jt\theta}\, d\theta \tag{1.141}$$

1.10 NONADDITIVITY OF SPECTRAL PROPERTIES

Many of the conceptual difficulties associated with time-frequency analysis are a reflection of the basic properties of signals and spectra. If these properties are understood in their basic form, then the curiosities encountered later will not be so paradoxical. The fundamental idea to appreciate and always keep in mind is that the frequency content is not additive. Suppose we have a signal composed of two parts, the spectrum will be the sum of the corresponding spectrum of each part of the signal,

$$s = s_1 + s_2 \qquad ; \qquad S = S_1 + S_2 \tag{1.142}$$

However, the energy density is not the sum of the energy densities of each part

$$|S|^2 = |S_1 + S_2|^2 = |S_1|^2 + |S_2|^2 + 2\,\mathrm{Re}\,\{S_1^* S_2\} \tag{1.143}$$

$$\neq |S_1|^2 + |S_2|^2 \tag{1.144}$$

Thus the frequency content is not the sum of the frequency content of each signal. One cannot think of frequency content of signals as something we have in a bucket that we add to the frequency content or bucket of another signal. The physical reason is that when we add two signals, the waveforms may add and interfere in all sort of ways to give different weights to the original frequencies. Mathematically

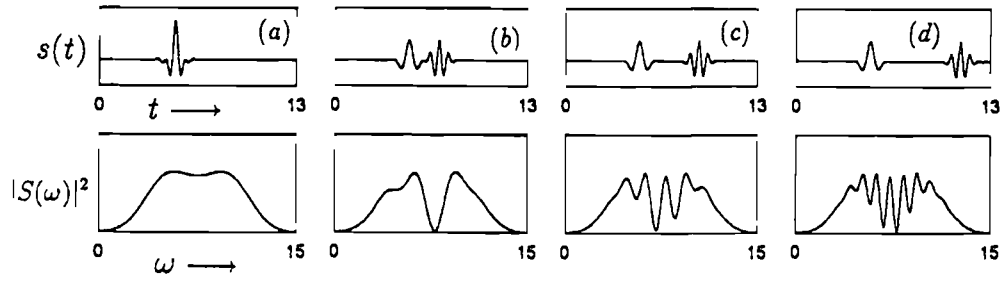

Fig. 1.2 The energy density spectrum for the sum of two signals, each localized in time. The signal is given by Eq. (1.145), where $\alpha_1 = \alpha_2$, $\omega_{1,2} = 5, 10$, and $t_1 = 5$. The values for t_2 are 5, 7, 9, and 11 in (a)-(d). Even for large time separations as in (c) and (d) the energy density spectrum is affected. The density of frequency is not just the sum of the frequencies of each part. The implications for time-frequency analysis are explained in the next figure.

this is reflected by the fact that the energy density spectrum is the absolute square of the sum of the spectra, which results in nonlinear effects. How the intensities change is taken into account by Eq. (1.143).

Even if the signals are well localized and greatly separated in time we still cannot add frequency content. For example, consider two signals each lasting a second but one happening today and the other a million years from now. Suppose we add the two signals. Because of the large separation in time we have the intuitive sense that physical properties should somehow also be separable and simply add, but that is not the case. Even though the two signals have in some sense nothing to do with each other since they are greatly separated in time, nonetheless the sum of the two signals produces an energy density spectrum that is not simply the sum of the two energy density spectra. Let us take a particular example:

$$s(t) = A_1 e^{-\alpha_1(t-t_1)^2/2+j\omega_1(t-t_1)} + A_2 e^{-\alpha_2(t-t_2)^2/2+j\omega_2(t-t_2)} \tag{1.145}$$

If we take α_1 and α_2 to be large in relation to $|t_2 - t_1|$ the two parts will be well separated in time. A few typical situations are illustrated in Fig. 1.2, where the energy density spectrum is also shown. The energy density spectrum changes with signal separation even when the separation is large. This is an important consideration since it leads to seemingly paradoxical results in the distribution of intensity in the time-frequency plane. Even though we have not begun our discussion of time-frequency distributions it is important to understand this nonadditivity property. Suppose we have two signals, each concentrated around the time-frequency points ω_1, t_1 and ω_2, t_2 and where these two points are well separated in both time and frequency. If we add these signals we may think that since they are well separated in time and frequency the resulting time-frequency density would be the sum of the same two clumps. However, that cannot be the case since it does not take into account the term $S_1^* S_2 + S_1 S_2^*$ in the energy density spectrum. This is illustrated in Fig. 1.3.

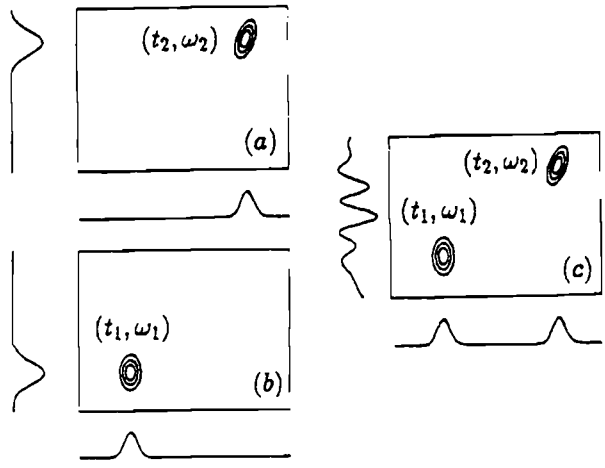

Fig. 1.3 In (a) we have a signal that is localized at the time-frequency point (t_2, ω_2) and in (b) we have another signal localized at (t_1, ω_1). Since (t_1, ω_1) and (t_2, ω_2) are separated and the densities localized, a seemingly plausible representation of the energy density in the time-frequency plane for the sum of the two signals of (a) and (b) is shown in (c). However, this cannot be correct because that would imply that the energy density spectrum is the sum of the density spectra of each signal, which is not the case, as discussed in the previous figure.

1.11 CLASSIFICATION OF SIGNALS

The types of signals found in nature vary greatly and a rich terminology has arisen to characterize them in broad terms. If a signal does not change in some sense then one says it is stationary; otherwise it is nonstationary. If a signal lasts a short time it is generally called a transient, burst, or wave packet. Short is a relative term that may mean a million years for astronomical signals or a billionth of a second in the case of atomic physics.

If the signal is explicitly known, we say we have a deterministic signal. Very often, because of our ignorance or because the physical process producing the signal is governed by random events, we have many possible signals, in which case we say we have a collection or ensemble of signals, a random signal, or a stochastic signal. A particular signal of the collection is said to be a realization. For example, if we produce sinusoids where the frequency is determined by some random event, then we would have a random signal.

The spectral content is sometimes used to classify signals. Signals whose spectra are concentrated in a small band relative to the mean frequency are called narrow band; otherwise they are called broadband. However, these classifications are crude. For signals whose spectrum is changing, they do not give a true sense of what is going on and can be misleading. We have already seen, for example, that the bandwidth can be caused by two physically different mechanisms. For exam-

ple, if a signal varies from 100 to 5000 Hz in 10 seconds in a pure and steady way, then to classify it as broadband does not present a complete picture, since at each time there may have been only one frequency. On the other hand we can produce a signal that has the same bandwidth where indeed at each time there is a broad range of frequencies. Time-frequency analysis enables us to classify signals with a considerably greater reflection of the physical situation than can be achieved by the spectrum alone.

Chapter 2

Instantaneous Frequency and the Complex Signal

2.1 INTRODUCTION

Signals in nature are real. Nevertheless, it is often advantageous to define a complex signal that in some sense or other corresponds to the real signal. In this chapter we describe the motivations for seeking a complex signal representation and its relation to the concept of instantaneous frequency. We have seen in Chapter 1 that it is natural to define instantaneous frequency as the derivative of the phase because its average over time is the average frequency. Thus far we have left open the question of how to get the phase. One of the motives for defining the complex signal is that it will allow us to define the phase, from which we can obtain the instantaneous frequency.

We seek a complex signal, $z(t)$, whose real part is the "real signal", $s_r(t)$, and whose imaginary part, $s_i(t)$, is our choice, chosen to achieve a sensible physical and mathematical description,

$$z(t) = s_r + j s_i = A(t)\, e^{j\varphi(t)} \tag{2.1}$$

If we can fix the imaginary part we can then unambiguously define the amplitude and phase by

$$A(t) = \sqrt{s_r^2 + s_i^2} \quad ; \quad \varphi(t) = \arctan s_r / s_i \tag{2.2}$$

which gives

$$\omega_i(t) = \varphi'(t) = \left(s_i' s_r - s_r' s_i\right) / A^2 \tag{2.3}$$

for the instantaneous frequency.

The issue is then how to define the imaginary part. Interest in the proper definition of instantaneous frequency first arose with the advent of frequency modulation for radio transmission in the 1920s. Historically there have been two methods, the quadrature method and the analytic signal method. Before the introduction of the analytic signal by Gabor, the main idea for forming a complex signal was the sense that for a signal of the form $s(t) = A(t)\cos\varphi(t)$, the complex counterpart should simply be $A(t)\,e^{j\varphi(t)}$. That begs the question because it requires first writing the signal in the form $A(t)\cos\varphi(t)$ and there are an infinite number of ways that can be done, although in some situations it is intuitively obvious. This idea is called the quadrature procedure and is discussed in Section 2.6. In 1946 the fundamental issues were crystallized by Gabor with the introduction of the analytic signal.[210]

As we will see, the "analytic signal procedure" devised by Gabor results in a complex signal that has a spectrum identical to that of the real signal for positive frequencies and zero for the negative frequencies. Because of this fact there has been a tendency in the recent past to introduce the analytic signal by merely saying that the negative frequencies do not exist anyway so let's get rid of them. However, "lets drop the negative frequencies", is neither the historical nor the physical reason for seeking a complex signal. The reasons for doing so are that a complex signal offers a way to overcome difficulties that arise when considering only real signals. Similarly, there has been a tendency lately to define instantaneous frequency as the derivative of the phase of the analytic signal. However, instantaneous frequency is a primitive concept and not a question of mere mathematical definition. The issue is whether any particular idea or definition does indeed match our intuitive sense and adequately represents that concept, and whether it leads to further fruitful ideas. As we will see, the derivative of the phase of the analytic signal does meet our intuitive sense of instantaneous frequency for many cases but also produces many counterintuitive consequences. That is all right because it is the counterintuitive situations that test the ideas. Alternatively, if an idea works well in some cases but apparently not in others, then perhaps it is our interpretation of these apparently curious cases that is wanting. In fact it will turn out that time-frequency analysis offers a framework which explains many of the curiosities and difficulties. We point out some of the difficulties with instantaneous frequency in this chapter. One should keep an open mind regarding the proper definition of the complex signal, that is, the appropriate way to define phase, amplitude, and instantaneous frequency. Probably the last word on the subject has not yet been said.

2.2 REASONS FOR THE COMPLEX SIGNAL

First and most importantly, for a real signal the spectrum satisfies $S(-\omega) = S^*(\omega)$ and therefore the energy density spectrum $|S(\omega)|^2$ is always symmetric about the origin. Fig. 2.1 symbolically draws such a density, which is a perfectly good density. Because of the symmetry, the average frequency will always come out to be zero! That is not what we want because it does not give us a sense of what is really going

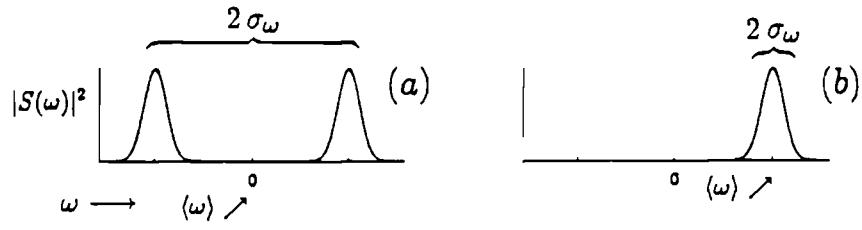

Fig. 2.1 A real signal has a symmetrical density in frequency, as illustrated in (a). Therefore, the average frequency is zero and the spread in frequency is roughly the distance between the two bumps, as illustrated in (a). Neither of these results is indicative of the physical situation. The analytic signal is defined to give the identical spectrum for the positive frequencies and zero for the negative frequencies, as in (b). This results in an average frequency and bandwidth, which better reflect the physical situation in that the average frequency falls somewhere in the middle of the bump and the bandwidth is the spread of the bump.

on. We want the answer to come out to be somewhere in the middle of the right-hand bump. Also, the spread in frequencies will be roughly the distance between the two bumps, while what we want is the spread of one bump. What can we do to obtain a value for average frequency that is roughly centered in the middle of the right hand bump? We can achieve this by simply neglecting the left bump in the averaging,

$$\langle \omega \rangle = \int_0^\infty \omega \, | \, S(\omega) \, |^2 \, d\omega \tag{2.4}$$

There are now two approaches we can take. First, we can continue to consider real signals and when taking spectral averages integrate from zero to infinity rather than $-\infty$ to ∞. Or, we can define a new signal that has the same spectrum for the positive frequencies and a zero spectrum for the negative frequencies. The advantage of the second approach is that we can calculate frequency averages directly from the signal and therefore it is advantageous to have the signal once and for all. In particular, the new signal, $z(t)$, as yet unknown, will assure that

$$\langle \omega \rangle = \int_0^\infty \omega \, | \, S(\omega) \, |^2 \, d\omega = \int z^*(t) \, \frac{1}{j} \frac{d}{dt} \, z(t) \, dt \qquad [\; z(t) = ? \;] \tag{2.5}$$

The second reason for wanting to form a complex signal is that it will allow us to obtain the phase and amplitude of a signal unambiguously, and that allows us to obtain an expression for instantaneous frequency.

2.3 THE ANALYTIC SIGNAL[1]

If the real signal, $s(t)$, has the spectrum, $S(\omega)$, then the complex signal, $z(t)$, whose spectrum is composed of the positive frequencies of $S(\omega)$ only, is given by the inverse transform of $S(\omega)$, where the integration goes only over the positive frequencies,

$$z(t) = 2\frac{1}{\sqrt{2\pi}}\int_0^\infty S(\omega)\, e^{j\omega t}\, dt \tag{2.6}$$

The factor of 2 is inserted so that the real part of the analytic signal will be $s(t)$; otherwise it would be one half of that. We now obtain the explicit form for $z(t)$ in terms of the real signal $s(t)$. Since

$$S(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty s(t)\, e^{-j\omega t}\, dt \tag{2.7}$$

we have

$$z(t) = 2\frac{1}{2\pi}\int_0^\infty \int s(t')\, e^{-j\omega t'}\, e^{j\omega t}\, dt'\, d\omega \tag{2.8}$$

$$= \frac{1}{\pi}\int_0^\infty \int s(t')\, e^{j\omega(t-t')}\, dt'\, d\omega \tag{2.9}$$

and using

$$\int_0^\infty e^{j\omega x}\, d\omega = \pi\,\delta(x) + \frac{j}{x} \tag{2.10}$$

we obtain

$$z(t) = \frac{1}{\pi}\int s(t')\left[\pi\,\delta(t-t') + \frac{j}{t-t'}\right] dt' \tag{2.11}$$

yielding

$$\mathcal{A}[s] = z(t) = s(t) + \frac{j}{\pi}\int \frac{s(t')}{t-t'}\, dt' \tag{2.12}$$

We use the notation $\mathcal{A}[s]$ to denote the analytic signal corresponding to the signal s. The reason for the name *analytic* is that these types of complex functions satisfy the Cauchy-Riemann conditions for differentiability and have been traditionally called analytic functions. The second part of Eq. (2.12) is the Hilbert transform of the signal and there are two conventions to denote the Hilbert transform of a function, $\hat{s}(t)$ and $H[s(t)]$

[1] In a series of seminal papers Vakman[548, 549, 551] has addressed the concepts of instantaneous frequency and the analytic signal and has brought forth the fundamental issues regarding these subjects. The classical review article on the subject is by Vakman and Vainshtein.[550]

$$H[s(t)] = \hat{s}(t) = \frac{1}{\pi} \int \frac{s(t')}{t - t'} dt' \tag{2.13}$$

The integration in the integral implies taking the principle part.

The Analytic Signal of a Complex Signal. We have emphasized the physical motivation for defining a complex signal that corresponds to a real signal. However, there is no reason why $s(t)$ in Eq. (2.12) cannot be any function, real or complex. For manipulative purposes it is important to allow for that possibility.

Energy of the Analytic Signal. Because we have insisted that the real part of the complex signal be the original signal, normalization is not preserved. Recall that the spectrum of the original real signal satisfies $|S(\omega)| = |S(-\omega)|$ and therefore the energy of the original signal is

$$E_s = \int |S(\omega)|^2 d\omega = 2\int_0^\infty |S(\omega)|^2 d\omega = \frac{1}{2}\int_0^\infty |2S(\omega)|^2 d\omega = \tfrac{1}{2}E_z \tag{2.14}$$

That is, the energy of the analytic signal is twice the energy of the original signal. In addition, the energy of the real part is equal to the energy of the imaginary part

$$E_s = E_{H[s]} \tag{2.15}$$

which can be seen by considering $|z(t)|^2 = |s(t) + jH[s]|^2$. When this is expanded the middle term is

$$\iint \frac{s^*(t)\, s(t') + s(t)\, s^*(t')}{t - t'} dt'\, dt = 0 \tag{2.16}$$

since the integrand is a two dimensional odd function.

2.4 CALCULATING THE ANALYTIC SIGNAL

It might seem reasonable to now study the properties of the Hilbert transform so that we may better understand the properties of the analytic signal. However, that is not the case. It is easier to study the analytic signal directly than to first develop theorems and results for the Hilbert transform. The main point to keep in mind is that the analytic signal is formed from the positive part of the spectrum of the real signal and multiplied by two. This will carry us far.

Consider $e^{j\omega t}$ whose spectrum is a delta function at ω. If ω is negative then there is no positive frequency to retain and the answer is zero. If it is positive then we just multiply by two. Therefore

$$\mathcal{A}[e^{j\omega t}] = \begin{cases} 0 & \text{if } \omega < 0 \\ 2\, e^{j\omega t} & \text{if } \omega > 0 \end{cases} \tag{2.17}$$

This simple result is very important because if we can express the signal in terms of exponentials then all we have to do to form the analytic signal is drop the terms with negative frequency and multiply the positive frequency terms in the expansion by two. A few examples will make this clear.

Example 2.1: The Analytic Signal of $\cos|\omega|t$.

Write $\cos|\omega|t$ in terms of exponentials and then use Eq. (2.17),

$$\mathcal{A}[\cos|\omega|t] = \tfrac{1}{2}\mathcal{A}[e^{j|\omega|t} + e^{-j|\omega|t}] \tag{2.18}$$

$$= \tfrac{1}{2}\mathcal{A}[e^{j|\omega|t}] + \tfrac{1}{2}\mathcal{A}[e^{-j|\omega|t}] \tag{2.19}$$

$$= \tfrac{1}{2}\mathcal{A}[e^{j|\omega|t}] \tag{2.20}$$

$$= e^{j|\omega|t} \tag{2.21}$$

Similarly,

$$\mathcal{A}[\sin|\omega|t] = \frac{1}{2j}\mathcal{A}[e^{j|\omega|t} - e^{-j|\omega|t}] = -j\,e^{j|\omega|t} \tag{2.22}$$

Example 2.2: The Analytic Signal of $s(t) = \cos\omega_1 t\,\cos\omega_2 t$.

For definiteness we take $0 \le \omega_1 \le \omega_2$. Rewrite $s(t)$ in terms of exponentials,

$$s(t) = \cos\omega_1 t\,\cos\omega_2 t \tag{2.23}$$

$$= \tfrac{1}{4}\left(e^{j\omega_2 t} + e^{-j\omega_2 t}\right)\left(e^{j\omega_1 t} + e^{-j\omega_1 t}\right) \tag{2.24}$$

$$= \tfrac{1}{4}\left(e^{j(\omega_2+\omega_1)t} + e^{j(\omega_2-\omega_1)t} + e^{-j(\omega_2+\omega_1)t} + e^{-j(\omega_2-\omega_1)t}\right) \tag{2.25}$$

The last two terms have negative frequencies and hence only the first two terms remain, giving

$$z(t) = 2\tfrac{1}{4}\left(e^{j(\omega_2+\omega_1)t} + e^{j(\omega_2-\omega_1)t}\right) = \tfrac{1}{2}\left(e^{j\omega_1 t} + e^{-j\omega_1 t}\right)e^{j\omega_2 t} \tag{2.26}$$

$$= \cos\omega_1 t\,e^{j\omega_2 t} \tag{2.27}$$

Notice that the analytic signal procedure chooses the higher frequency for the instantaneous frequency.

The Analytic Signal of an Analytic Signal. If we start with an analytic signal then its spectrum is nonzero only on the positive frequency axis and hence there is nothing to drop. Therefore we get back the same signal except for the factor of two,

$$\mathcal{A}[z(t)] = 2z(t) \qquad [\text{ if } z(t) \text{ is analytic }] \tag{2.28}$$

Analytic Signal of the Derivative. Suppose we want the analytic signal of the derivative of a function. If we have the analytic signal of the function then all we

have to do is differentiate it, because the analytic signal of the derivative of a function is the derivative of the analytic signal. More generally, the analytic signal of the nth derivative of a function is the nth derivative of its analytic signal. To see this consider

$$\mathcal{A}\left[\frac{d^n s}{dt^n}\right] = \mathcal{A}\left[\frac{d^n}{dt^n}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} S(\omega)\, e^{j\omega t}\, d\omega\right] \tag{2.29}$$

$$= \mathcal{A}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} (j\omega)^n\, S(\omega)\, e^{j\omega t}\, d\omega\right] \tag{2.30}$$

$$= 2\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} (j\omega)^n\, S(\omega)\, e^{j\omega t}\, d\omega \tag{2.31}$$

$$= \frac{d^n}{dt^n}\, 2\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} S(\omega)\, e^{j\omega t}\, d\omega \tag{2.32}$$

But this is just the nth derivative of the analytic signal of $s(t)$. Therefore

$$\mathcal{A}\left[\frac{d^n s}{dt^n}\right] = \frac{d^n}{dt^n}\,\mathcal{A}[s] \tag{2.33}$$

We note that in going from Eq. (2.30) to Eq. (2.31) we have used the fact the analytic signal operator is linear and hence can be taken inside the integration because integration is a sum.

Convolution. Suppose we have a signal, $s(t)$, whose spectrum is $S(\omega)$ and we form a new signal by way of

$$z(t) = 2\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} F(\omega)\, S(\omega)\, e^{j\omega t}\, d\omega \tag{2.34}$$

where $F(\omega)$ is any function. It is clear that $z(t)$ is analytic since we have made it so by construction. By specializing the function $F(\omega)$ many important results can be obtained. In particular, if we consider $F(\omega)$ to be the spectrum of a function, $f(t)$, and recall that the product of two spectra is the convolution of the two signals, then we have in general that

$$\int \mathcal{A}[s(t')]\, f(t-t')\, dt' \quad \text{is analytic for arbitrary } s \text{ and } f \tag{2.35}$$

That is, the convolution of an analytic signal with an arbitrary function results in an analytic signal.

Imposed Modulation. If we have a signal that is bandlimited between two frequencies, it is often necessary to shift the frequencies to higher values for the purpose of transmission. For example, a speech signal, which is limited to a few thousand

hertz, is raised to the megahertz range because we can build transmitters and receivers at those frequencies and because electromagnetic waves propagate well in our atmosphere at those frequencies. What is done is to multiply the original waveform by $e^{j\omega_0 t}$ with ω_0 positive to form a new signal,

$$s_{\text{new}}(t) = s(t)\, e^{j\omega_0 t} \qquad [\ \omega_0 \geq 0\ ] \tag{2.36}$$

In such a case one says that $e^{j\omega_0 t}$ is the imposed modulation and ω_0 is the carrier frequency. The spectrum of the new signal is the spectrum of the old signal shifted by an amount ω_0

$$S_{\text{new}}(\omega) = S(\omega - \omega_0) \tag{2.37}$$

When will the resulting signal be analytic? If the spectrum of the original signal extends from minus infinity to infinity then the spectrum S_{new} will likewise extend from minus infinity to infinity and not be analytic. But suppose the spectrum $S(\omega)$ is single sided, which means that it is zero below a certain value. Then if ω_0 is greater than that value, the spectrum of the new signal will be shifted to the positive part of the frequency axis and the signal will be analytic. Therefore

$$s(t)\ e^{j\omega_0 t} \qquad \text{is analytic if the spectrum of } s(t) \text{ is zero for } \omega \leq -\omega_0 \tag{2.38}$$

Note that we have not specified whether $s(t)$, the original signal, is real or complex. It can be complex. If it is real then this result implies that it must be bandlimited to the interval $(-\omega_0, \omega_0)$.

Analytic Signal of the Sum of Two Signals. Suppose we have a signal and want to add to it another signal so that the resulting sum will be analytic. All we have to do is choose a signal whose spectrum for negative frequencies is identical to the first signal but differs in sign. That way, the spectrum of the resulting signal will be zero for negative frequencies. Therefore

$$s = s_1 + s_2 \qquad \text{is analytic if } S_1(\omega) = -S_2(\omega) \text{ for } \omega \leq 0 \tag{2.39}$$

Factorization Theorem. An important situation is when we want to find the analytic signal of the product of two signals. Let us first address when one of the functions can be factored out. That is, when does $\mathcal{A}[\,s_1 s_2\,] = s_1 \mathcal{A}[\,s_2\,]$? If the spectrum of s_1 is nonzero for all frequencies, then there is nothing we can do to raise its spectrum to be only positive. The only possibility for a factorization theorem is if the spectrum of s_1 is zero below some frequency. Let us say that frequency is $-\omega_1$. Therefore, $\mathcal{A}[\,s_2\,]$ must raise the spectrum of s_1 so that it is on the positive side of the frequency axis. Think of the spectrum of s_1 as a sum of exponentials. The worst case is the lowest frequency, $e^{-j\omega_1 t}$. Now $\mathcal{A}[\,s_2\,]$ is analytic and again think of it as a sum of exponentials. If there is a frequency below ω_1, then the product of this

exponential with the worst case of s_1 will result in a negative frequency. Therefore s_2 cannot have any frequencies below ω_1 and hence

$$\mathcal{A}[s_1 s_2] = s_1 \mathcal{A}[s_2] \quad \text{if the spectrum of } s_1 \text{ is zero below the value } -\omega_1 \text{ and } s_2 \text{ is any signal whose analytic function has a spectrum that is zero below } \omega_1. \tag{2.40}$$

Notice that the spectrum of s_2 does not have to be zero below ω_1; it is the spectrum of $\mathcal{A}[s_2]$ that has to be zero below ω_1. If s_2 is real, then the condition is that the spectrum of s_2 must be zero outside the range $(-\omega_1, \omega_1)$. These results are due to Bedrosian[64] and Nuttall.[402]

Product of Two Analytic Signals. If s_1 and s_2 are both analytic, then they are both single sided and we have

$$\mathcal{A}[s_1 s_2] = s_1 \mathcal{A}[s_2] = 2 s_1 s_2 \qquad s_1 \text{ and } s_2 \text{ analytic} \tag{2.41}$$

Real Signals. Suppose s_1 and s_2 are real. Their spectra are then symmetric. Since the spectrum of s_1 vanishes below $-\omega_1$ it must vanish above ω_1. Therefore the spectrum of s_1 must be zero for $|\omega| \geq \omega_1$. Similarly, the spectrum of s_2 must vanish for $|\omega| \leq \omega_1$. Hence

$$\mathcal{A}[s_1 s_2] = s_1 \mathcal{A}[s_2] \quad \text{for real signal if spectrum of } s_1 \text{ is zero for } |\omega| \geq \omega_1 \text{ and the spectrum of } s_2 \text{ is zero for } |\omega| \leq \omega_1 \tag{2.42}$$

2.5 PHYSICAL INTERPRETATION OF THE ANALYTIC SIGNAL

Since the analytic signal is complex it can always be put into polar form

$$\mathcal{A}[s(t)] = A(t)\, e^{j\varphi(t)} \tag{2.43}$$

We now ask what the analytic signal procedure has done in terms of choosing the particular amplitude and phase, that is, what is special about the amplitude and phase to make it an analytic signal? Generally speaking, the answer is that the spectral content of the amplitude is lower than the spectral content of $e^{j\varphi(t)}$. We illustrate this first with a simple case where

$$s(t) = A(t)\, e^{j\omega_0 t} \tag{2.44}$$

Call the spectrum of the amplitude $S_A(\omega)$,

$$S_A(\omega) = \frac{1}{\sqrt{2\pi}} \int A(t)\, e^{-j\omega t}\, dt \tag{2.45}$$

The spectrum of $s(t)$ is then $S_A(\omega - \omega_0)$. For $S_A(\omega - \omega_0)$ to be analytic, that is, zero for negative frequencies, $S_A(\omega)$, it must be zero $\omega \leq \omega_0$. Therefore

$A(t)\, e^{j\omega_0 t}$ is analytic

if the spectrum of $A(t)$ is contained within $(-\omega_0, \omega_0)$ (2.46)

Hence all the low frequencies are in the amplitude, and the high frequency is in the cosine. One can generalize this result in the following way. Call $S_\varphi(\omega)$ the spectrum of $e^{j\varphi(t)}$,

$$S_\varphi(\omega) = \frac{1}{\sqrt{2\pi}} \int e^{j\varphi(t)}\, e^{-j\omega t} \tag{2.47}$$

The spectrum of $A(t)\, e^{j\varphi(t)}$ is then

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int S_A(\omega - \omega')\, S_\varphi(\omega')\, d\omega' \tag{2.48}$$

We can consider this the sum of shifted spectra of $A(t)$ with coefficients $S_\varphi(\omega')$. Now suppose $S_A(\omega)$ is bandlimited in the interval $(-\omega_1, \omega_1)$. A sufficient condition to shift $S(\omega)$ to the positive axis is if the lowest value of the range of ω' is greater than ω_1. That is, $S_\varphi(\omega')$ is zero for values less than ω_1,

$A(t)\, e^{j\varphi(t)}$ is analytic

if the spectrum of $A(t)$ is contained in $(-\omega_1, \omega_1)$

and the spectrum of $e^{j\varphi(t)}$ is zero for $\omega \leq \omega_1$ (2.49)

Therefore what the analytic procedure does, at least for signals that result in the above forms, is to put the low frequency content in the amplitude and the high frequency content in the term $e^{j\varphi(t)}$.

2.6 THE QUADRATURE APPROXIMATION

It seems natural that if we write a signal in the form $s(t) = A(t)\, \cos\varphi(t)$, the complex signal ought to be

$$s_q(t) = A(t)\, e^{j\varphi(t)} \qquad \text{for } s(t) = A(t)\, \cos\varphi(t) \tag{2.50}$$

at least for some situations. The $s_q(t)$ thus formed is called the quadrature model signal. This idea was used before the introduction of the analytic signal. As we mentioned in the introduction, it begs the question since the procedure does not tell us how to write the signal in the form $A(t)\, \cos\varphi(t)$ to start with. Nonetheless, in many situations we may think we know the phase and amplitude. Or we may have an $A(t)$ and $\varphi(t)$ and want to *construct* a signal and ask whether or not it is analytic. More importantly, since calculating the analytic signal is difficult, if we

can approximate it with the quadrature model a considerable simplification would be achieved.

We therefore want to know when the quadrature model will agree with the analytic signal. If the spectrum of the quadrature model is on the positive frequency axis only, then we have total agreement. In general, that will not be the case and we want a measure of the error.

Energy Criteria. The less of the spectrum of the quadrature signal there is on the negative frequency axis, the better is the agreement. To examine this question more precisely we define the following three signals:

$$s(t) = A(t)\cos\varphi(t) \qquad \text{real signal} \tag{2.51}$$

$$s_q(t) = A(t)\,e^{j\varphi(t)} \qquad \text{quadrature model} \tag{2.52}$$

$$s_a(t) = \frac{2}{\sqrt{2\pi}}\int_0^\infty S(\omega)\,e^{j\omega t}\,d\omega \qquad \text{analytic signal} \tag{2.53}$$

where $S(\omega)$ is the spectrum of the real signal

$$S(\omega) = \frac{1}{\sqrt{2\pi}}\int A(t)\,\cos\varphi(t)\,e^{-j\omega t}\,dt \tag{2.54}$$

Using $e^{j\varphi(t)} = \cos\varphi + j\sin\varphi$, the spectrum of the quadrature signal is seen to be

$$S_q(\omega) = S(\omega) + \frac{j}{\sqrt{2\pi}}\int A(t)\,\sin\varphi(t)\,e^{-j\omega t}\,dt \tag{2.55}$$

Take the complex conjugate of this equation, let $\omega \to -\omega$, and add, to obtain

$$2\,S(\omega) = S_q(\omega) + S_q^*(-\omega) \tag{2.56}$$

where we have used the fact that $S(\omega) = S^*(-\omega)$, since $s(t)$ is real. Also, the spectrum of the analytic signal is

$$S_a(\omega) = \begin{cases} 0 & \text{if } \omega < 0 \\ 2\,S(\omega) = S_q(\omega) + S_q^*(-\omega) & \text{if } \omega > 0 \end{cases} \tag{2.57}$$

As a criterion on closeness between the quadrature signal and analytic signal, Nuttall[402] considered the energy of the difference between the two signals,

$$\Delta E = \int |\,s_a - s_q\,|^2\,dt = \int |\,S_a(\omega) - S_q(\omega)\,|^2\,d\omega \tag{2.58}$$

$$= \int_{-\infty}^{0} |\,0 - S_q(\omega)\,|^2\,d\omega + \int_0^\infty |\,S_a(\omega) - S_q(\omega)\,|^2\,d\omega \tag{2.59}$$

$$= \int_{-\infty}^{0} |\,S_q(\omega)\,|^2\,d\omega + \int_0^\infty |\,S_q^*(-\omega)\,|^2\,d\omega \tag{2.60}$$

The two terms are equal and therefore

$$\Delta E = 2\int_{-\infty}^{0} |\, S_q(\omega)\,|^2\, d\omega \tag{2.61}$$

which is twice the energy of the quadrature model in the negative part of the spectrum.

Point by Point Comparison. A stronger condition, which compares s_q and s_a at each time, was given by Vakman and Vainshtein.[550] Using Eq. (2.57), consider

$$s_a(t) - s_q(t) = \frac{1}{\sqrt{2\pi}}\int S_a(\omega)e^{j\omega t}\,d\omega - \frac{1}{\sqrt{2\pi}}\int S_q(\omega)\,e^{j\omega t}\,d\omega \tag{2.62}$$

$$= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} S_q(\omega)\,e^{j\omega t}\,d\omega + \frac{1}{\sqrt{2\pi}}\int_{0}^{-\infty} S_q^*(-\omega)\,e^{j\omega t}\,d\omega \tag{2.63}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\left[\, S_q^*(\omega)\,e^{-j\omega t} - S_q(\omega)\,e^{j\omega t}\,\right] d\omega \tag{2.64}$$

This compares the signal at every time. If we take absolute values, then

$$|\, s_a(t) - s_q(t)\,| = \left|\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\left[\, S_q^*(\omega)\,e^{-j\omega t} - S_q(\omega)\,e^{j\omega t}\,\right] d\omega\right| \tag{2.65}$$

$$\leq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\left|\, S_q^*(\omega)\,e^{-j\omega t} - S_q(\omega)\,e^{j\omega t}\,\right| d\omega \tag{2.66}$$

and therefore

$$|\, s_a(t) - s_q(t)\,| \leq \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{0} |\, S_q(\omega)\,|\, d\omega \tag{2.67}$$

This gives the absolute possible deviation for any time.

Imposed Amplitude and Phase Modulation. Often we have a carrier and impose an amplitude and phase, $A(t)$ and $\varphi(t)$, to produce the complex signal

$$s(t) = A(t)\,e^{j\omega_0 t + j\varphi(t)} \tag{2.68}$$

The spectrum of $A(t)\,e^{j\varphi(t)}$ is shifted upward by an amount ω_0 and therefore, using Eq. (2.61),

$$\Delta E = 2\int_{-\infty}^{0} |\, S_q(\omega-\omega_0)\,|^2\, d\omega \tag{2.69}$$

$$= 2\int_{-\infty}^{-\omega_0} |\, S_q(\omega)\,|^2\, d\omega \tag{2.70}$$

Note that $S_q(\omega)$ is still the spectrum of $A(t)\,e^{j\varphi(t)}$.

Example 2.3: Chirp.

Take the real signal

$$s(t) = e^{-\alpha t^2/2}\cos(\omega_0 t + \beta t^2/2) \tag{2.71}$$

The quadrature model is

$$s_q(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2}\, e^{j\omega_0 t + j\beta t^2/2} \tag{2.72}$$

How does this compare with the analytic signal of $s(t)$? Using Eq. (1.42) for the spectrum of s_q we have

$$\Delta E = 2\int_{-\infty}^{0} |S(\omega)|^2\, d\omega = 1 - \frac{2}{\sqrt{\pi}}\int_0^{\sqrt{\frac{\alpha\omega_0^2}{\alpha^2+\beta^2}}} e^{-x^2}\, dx \tag{2.73}$$

This will be close to zero when the upper limit is close to infinity and therefore

$$s_a \sim s_q \qquad \text{when } \sqrt{\frac{\alpha}{\alpha^2+\beta^2}}\,\omega_0 = \frac{\omega_0}{\sqrt{2}\sigma_\omega} \text{ is large} \tag{2.74}$$

where σ_ω is the bandwidth of $s_q(t)$ as per Eq. (1.40). This is reasonable because for a given ω_0, the energy density spectrum will not spill over much into the negative frequency region if it is relatively narrow.

2.7 INSTANTANEOUS FREQUENCY

Instantaneous frequency is one of the most intuitive concepts, since we are surrounded by light of changing color, by sounds of varying pitch, and by many other phenomena whose periodicity changes. The exact mathematical description and understanding of the concept of changing frequency is far from obvious and it is fair to say that it is not a settled question. In Chapter 1 we discussed a fundamental reason why a good definition of instantaneous frequency is the derivative of the phase: If we do define instantaneous frequency that way, then its time average with the energy density gives the average frequency. Note, however, that this result does not depend on a specific method of getting the phase; it is true for any complex signal. We have in the previous sections given some plausibility arguments for why the phase should be determined by way of the analytic signal. In fact, instantaneous frequency is often defined as the derivative of the phase of the analytic signal. This is a bad idea because we should keep an open mind as to whether that is indeed the most suitable definition of the intuitive concept. The issue is whether the derivative of the phase of the analytic signal does indeed satisfy our intuition regarding instantaneous frequency. Although in many cases it does, in other cases it produces results that at first sight seem paradoxical. Of course, it is the paradoxes and unusual results that lead to abandonment of ideas, adjustment of our intuition, or the discovery of new ideas.

Paradoxes Regarding the Analytic Signal. There are five paradoxes or difficulties regarding the notion of instantaneous frequency if it is defined as the derivative of the phase of the analytic signal. To some extent we will be able to understand and resolve some of these paradoxes when we study the concept of instantaneous bandwidth in Section 13.2 and with the introduction of the idea that instantaneous frequency is a conditional average frequency. It is important to understand these difficulties in the most basic terms because in more complicated situations their reflection may cause difficulties in interpretation.

First, instantaneous frequency may not be one of the frequencies in the spectrum. That is strange because if instantaneous frequency is an indication of the frequencies that exist at each time, how can it not exist when we do the final bookkeeping by way of the spectrum? Second, if we have a line spectrum consisting of only a few sharp frequencies, then the instantaneous frequency may be continuous and range over an infinite number of values. Third, although the spectrum of the analytic signal is zero for negative frequencies, the instantaneous frequency may be negative. Fourth, for a bandlimited signal the instantaneous frequency may go outside the band. All these points are illustrated by the following simple example.

Example 2.4: Instantaneous Frequency for the Sum of Two Sinusoids.

Consider

$$s(t) = s_1(t) + s_2(t) \tag{2.75}$$

$$= A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \tag{2.76}$$

$$= A(t) e^{j\varphi(t)} \tag{2.77}$$

where the amplitudes A_1 and A_2 are taken to be constants and ω_1 and ω_2 are positive. The spectrum of this signal consists of two delta functions at ω_1 and ω_2,

$$S(\omega) = A_1\delta(\omega - \omega_1) + A_2\delta(\omega - \omega_2) \tag{2.78}$$

Since we take ω_1 and ω_2 to be positive, the signal is analytic. Solving for the phase and amplitude,

$$\varphi(t) = \arctan \frac{A_1 \sin\omega_1 t + A_2 \sin\omega_2 t}{A_1 \cos\omega_1 t + A_2 \cos\omega_2 t} \tag{2.79}$$

$$A^2(t) = A_1^2 + A_2^2 + 2A_1A_2\cos(\omega_2 - \omega_1)t \tag{2.80}$$

and taking the derivative of the phase we obtain

$$\omega_i = \varphi'(t) = \tfrac{1}{2}(\omega_2 + \omega_1) + \tfrac{1}{2}(\omega_2 - \omega_1)\,\frac{A_2^2 - A_1^2}{A^2(t)} \tag{2.81}$$

By taking different values of the amplitudes and frequency we can illustrate the points above. This is done in Fig. 2.2.

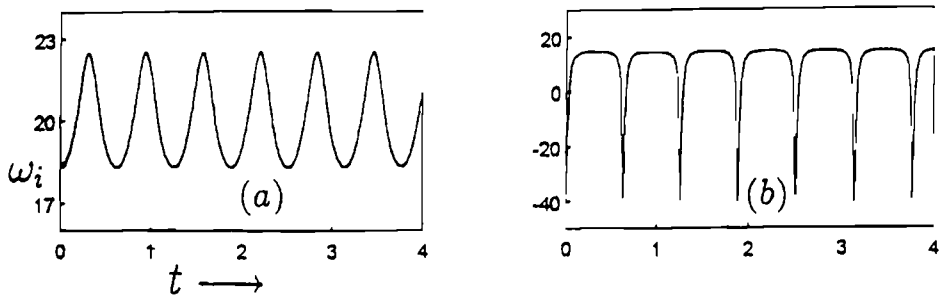

Fig. 2.2 The instantaneous frequency for the signal $s(t) = A_1 e^{j10t} + A_2 e^{j20t}$. The spectrum consists of two frequencies, at $\omega = 10$ and $\omega = 20$. In (a) $A_1 = .2$ and $A_2 = 1$. The instantaneous frequency is continuous and ranges outside the bandwidth. In (b) $A_1 = -1.2$ and $A_2 = 1$. Although the signal is analytic the instantaneous frequency may become negative.

One last paradox regarding the analytic signal. If instantaneous frequency is an indication of the frequencies that exist at time t, one would presume that what the signal did a long time ago and is going to do in the future should be of no concern; only the present should count. However, to calculate the analytic signal at time t we have to know the signal for all time. This paradox has been analyzed by Vakman[549] who makes the following analogy. Before Maxwell's equations, light was considered localized rays. We now know that light is electromagnetic waves which are highly nonlocal. This discovery forced a fruitful enlargement of the understanding of what light is. Moreover, from Maxwell's equations we can actually understand in which circumstances light behaves as rays. In the same sense, then, while we started out thinking of instantaneous frequency as a local concept, it may be the case that for a full explanation of the phenomenon we must enlarge the idea and context and accept its nonlocal nature.

2.8 DENSITY OF INSTANTANEOUS FREQUENCY[135]

We now ask for the density of instantaneous frequency in contrast to the density of frequency, that is, the energy density spectrum, $|S(\omega)|^2$. The method for finding densities is described in detail in Chapter 4, but we use the method here and make it plausible for the case of instantaneous frequency. The density of instantaneous frequency, $P(\omega_i)$, is given by

$$P(\omega_i) = \int \delta(\omega_i - \varphi'(t)) \, |s(t)|^2 \, dt \tag{2.82}$$

This says that for a given ω_i, choose only the values for which $\omega_i = \varphi'(t)$. If there is only one such value then this simplifies to

$$P(\omega_i) = \left. \frac{|s(t)|^2}{\varphi''(t)} \right|_{t=t_\omega} \tag{2.83}$$

where t_ω is a function of ω_i obtained by solving $\omega_i = \varphi'(t_\omega)$.

Instantaneous Frequency Spread. We already know that the average of instantaneous frequency is the mean frequency, that is, $\langle \omega_i \rangle = \langle \omega \rangle$. Therefore the spread of instantaneous frequency is

$$\sigma_{IF}^2 = \int (\varphi'(t) - \langle \omega_i \rangle)^2 \, |s(t)|^2 \, dt \tag{2.84}$$

$$= \int (\varphi'(t) - \langle \omega \rangle)^2 \, |s(t)|^2 \, dt \tag{2.85}$$

This is precisely the first term appearing in the bandwidth equation, Eq. (1.96), and therefore we can write that equation as

$$B^2 = \sigma_{IF}^2 + \int \left(\frac{A'(t)}{A(t)} \right)^2 A^2(t) \, dt \tag{2.86}$$

But the second term is positive and hence we conclude that the spread in instantaneous frequency is always smaller than the bandwidth,[184]

$$\sigma_{IF} \leq B \tag{2.87}$$

This may seem paradoxical since in the example above we showed that the spread of instantaneous frequency can be outside the range of the frequencies that exist in the spectrum. What this must mean is that while the instantaneous frequency can range widely, the occurrences when it ranges outside the bandwidth of the signal have small duration and hence do not contribute significantly to σ_{IF}. Note that the spreads are equal when the amplitude of the signal is constant.

Example 2.5: Chirp.

We calculate the distribution of instantaneous frequency for the signal

$$s(t) = (\alpha/\pi)^{1/4} \, e^{\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \tag{2.88}$$

The instantaneous frequency is $\omega_i = \omega_0 + \beta t$ and its derivative is simply β. Hence

$$P(\omega_i) = \left. \frac{|s(t)|^2}{\varphi''(t)} \right|_{t=(\omega_i - \omega_0)/\beta} \tag{2.89}$$

$$= \sqrt{\frac{\alpha}{\pi}} \, \frac{1}{\beta} \, e^{-\alpha(\omega_i - \omega_0)^2/\beta^2} \tag{2.90}$$

On the other hand the energy density spectrum is

$$|S(\omega)|^2 = \sqrt{\frac{\alpha}{\pi(\alpha^2 + \beta^2)}} \, e^{-\alpha(\omega - \omega_0)^2/(\alpha^2 + \beta^2)} \tag{2.91}$$

The spread of the spectrum, that is, the bandwidth, and the spread of the instantaneous frequency density are respectively

$$\sigma_\omega = \sqrt{\frac{\alpha^2 + \beta^2}{2\alpha}} \qquad ; \qquad \sigma_{IF} = \sqrt{\frac{\beta^2}{2\alpha}} \tag{2.92}$$

The distribution of instantaneous frequency is narrower than the distribution of frequency in conformity with our discussion.

Chapter 3

The Uncertainty Principle

3.1 INTRODUCTION[1,2]

The time-bandwidth product theorem, or uncertainty principle, is a fundamental statement regarding Fourier transform pairs. We are going to be particularly careful in our discussion since the uncertainty principle has played a prominent role in discussions, metaphysical and otherwise, of joint time-frequency analysis. The discovery of the uncertainty principle in physics and chemistry is one of the great achievements of the century. Unfortunately, it has generated many pseudo ideas, which are rivaled only by the number of pseudo ideas generated by relativity, the other great discovery of this century. The pseudo ideas in relativity are perhaps characterized by "everything is relative ", and the uncertainty principle by "everything is uncertain", neither view being remotely related to the physical or mathematical truth.

[1]The uncertainty principle was first derived by W Heisenberg in the paper "Uber den anschhaulichen Inhalt der quantentheoretichen Kinematic und Mechanik" ("On the Conceptual Content of Quantum Theoretical Kinematics and Mechanics") in 1927. In that paper he derived the uncertainty relations on the basis of a thought experiment involving resolution of an idealized microscope. In fact, Heisenberg presented it as an equality rather than as an inequality. It was Weyl who subsequently saw that uncertainty can be defined by the standard deviation and gave the proof commonly used today which is based on the Schwarz inequality. The physicist C. G. Darwin (grandson of Charles Darwin) made the connection between the uncertainty principle and Fourier transform pairs. E. U. Condon and H. P. Robertson extended the uncertainty principle for arbitrary variables. In 1930 Schrödinger grasped the full generality of the uncertainty principle, derived it for arbitrary variables and saw the fundamental connection with the commutator and anticommutator (see Eq. (15.87)). While the mathematics of the uncertainty principle was settled within five years of the discovery of quantum mechanics (1925-1926) by Schrödinger, Heisenberg, Born, and Dirac and within three years of Heisenberg's first paper on the subject, the interpretation and consequences of the uncertainty principle in quantum mechanics remains a subject of great interest and activity and the number of papers written on it in the intervening sixty years or so is truly phenomenal. After all, it is one of the most important discoveries of mankind.

[2]The uncertainty principle is discussed further in Sections 6.9 and 15.5.

For signal analysis, the meaning of the uncertainty principle and its importance have been clearly stated often enough, although a mystery still persists for some. Right on the mark is the statement by Skolnik[514]: "The use of the word 'uncertainty' is a misnomer, for there is nothing uncertain about the 'uncertainty relation'... . It states the well-known mathematical fact that a narrow waveform yields a wide spectrum and a wide waveform yields a narrow spectrum and both the time waveform and frequency spectrum cannot be made arbitrarily small simultaneously." Equally clear is Lerner[331]: The uncertainty principle " . . . has tempted some individuals to draw unwarranted parallels to the uncertainty principle in quantum mechanics. . . . The analogy is formal only." Ackroyd[4] has emphasized that "There is a misconception that it is not possible to measure the $t - f$ energy density of a given waveform and that this is a consequence of Gabor's uncertainty relation. However, the uncertainty principle of waveform analysis is not concerned with the measurement of $t - f$ energy density distributions: instead it states that if the effective bandwidth of a signal is W then the effective duration cannot be less than about 1/W (and conversely) . . . "

Although we can hardly improve on the above we try. Here is the point: The density in time is $|s(t)|^2$ and the density in frequency is $|S(\omega)|^2$; but $s(t)$ and $S(\omega)$ are related and hence we should not be surprised to find that there is a relation between the densities. The relation is such that if one density is narrow then the other is broad. That's it and no more. It is not that both time and frequency cannot arbitrarily be made narrow, but that the densities of time and frequency cannot both be made narrow.

If the interpretation is so straightforward in signal analysis, why is it one of the most profound discoveries in physics? Here is why: In classical physics and in every day life it seemed clear that we can choose the position and velocity of objects at will. No one imagined that we cannot place a ball at a given spot with a given velocity. However, present day physics, quantum mechanics, says precisely that and it is one of the great discoveries regarding the behavior of matter.

Quantum mechanics is inherently probabilistic. When we speak of densities in quantum mechanics we mean probability densities and the word "uncertainty" is appropriate because we are dealing with probability. In signal analysis the word "uncertainty" is highly misleading.

Let us be very clear that in both physics and signal analysis the uncertainty principle never applies to a single variable. It is always a statement about two variables. Furthermore, it does not apply to *any* two variables, but only to variables whose associated operators do not commute. In this chapter we deal with the uncertainty principle for time and frequency and in Section 15.5 we consider its generalization to other variables. There is a very elegant and simple way to derive the uncertainty principle for arbitrary quantities and this is done in Section 15.5. Here we use a simple brute force approach.

One of the reasons that there has been considerable confusion about the uncertainty principle in signal analysis is that one very often modifies a signal, e.g. filters it, windows it, etc. Once that is done we have two different signals, the original and the modified one. Therefore we have two uncertainty principles, one relating to the

original signal and the other to the modified signal. Very often these are confused. In Section 3.4 we will be particularly careful in making this distinction by deriving the uncertainty principle for modified signals.

3.2 THE UNCERTAINTY PRINCIPLE

The proof of the uncertainty principle is easy, but it is important to understand what goes into the derivation. We have defined duration, σ_t, and bandwidth, σ_ω, and have shown that they are good measures of the broadness of a signal in time and frequency. For convenience we repeat the definitions here:

$$T^2 = \sigma_t^2 = \int (t - \langle t \rangle)^2 \mid s(t) \mid^2 dt \tag{3.1}$$

$$B^2 = \sigma_\omega^2 = \int (\omega - \langle \omega \rangle)^2 \mid S(\omega) \mid^2 d\omega \tag{3.2}$$

We emphasize that T and B are standard deviations defined in the usual manner and no more. The uncertainty principle is

$$TB \geq \tfrac{1}{2} \tag{3.3}$$

Therefore one cannot have or construct a signal for which both T and B are arbitrarily small.

A More General Uncertainty Principle. A stronger version of the uncertainty principle is

$$\sigma_t \, \sigma_\omega \geq \tfrac{1}{2} \sqrt{1 + 4 \, \mathrm{Cov}_{t\omega}^2} \tag{3.4}$$

where $\mathrm{Cov}_{t\omega}$ is the covariance as defined by Eq. (1.124).

What Does the Proof of the Uncertainty Principle Depend On? It is important to have a clear picture of what the proof depends on so that no misleading interpretations creep in. The proof depends on only four things: first, on $\mid s(t) \mid^2$ being the density in time; second, on taking $\mid S(\omega) \mid^2$ as the density in frequency; third that $s(t)$ and $S(\omega)$ are Fourier pairs; and fourth, on defining T and B as standard deviations of time and frequency.

Notation. Very often the notation used to write the uncertainty principle is $\Delta t \Delta \omega \geq \frac{1}{2}$. There is nothing wrong with this notation as long we understand that Δ means standard deviation and nothing more. However, because Δ is typically used for the differential element of the calculus or to signify "error" there is a tendency to think of the uncertainty principle as having something to do with differential elements, smallness, measurement, or resolution. The Δ of the uncertainty principle means only one thing: the standard deviation. If this is kept in mind then no difficulties arise in interpretation or philosophy.

3.3 PROOF OF THE UNCERTAINTY PRINCIPLE

First, let us note that no loss of generality occurs if we take signals that have zero mean time and zero mean frequency. The reason is that the standard deviation does not depend on the mean because it is defined as the broadness about the mean. If we have a signal s_{old}, then a new signal defined by

$$s_{\text{new}}(t) = e^{-j\langle\omega\rangle(t+\langle t\rangle)}\, s_{\text{old}}(t+\langle t\rangle) \tag{3.5}$$

has the same shape both in time and frequency as s_{old} except that it has been translated in time and frequency so that the means are zero.[3] Conversely, if we have a signal $s_{\text{new}}(t)$ that has zero mean time and zero mean frequency and we want a signal of the same shape but with particular mean time and frequency, then

$$s_{\text{old}}(t) = e^{j\langle\omega\rangle t}\, s_{\text{new}}(t-\langle t\rangle) \tag{3.6}$$

The bandwidth expressed in terms of the signal is as per Eq. (1.35):

$$\sigma_\omega^2 = \int \omega^2 \,|\, S(\omega)\,|^2\, d\omega = \int |\, s'(t)\,|^2\, dt \tag{3.7}$$

The duration is

$$\sigma_t^2 = \int t^2 \,|\, s(t)\,|^2\, dt \tag{3.8}$$

and therefore

$$\sigma_t^2\, \sigma_\omega^2 = \int |\, t\, s(t)\,|^2\, dt \;\times\; \int |\, s'(t)\,|^2\, dt \tag{3.9}$$

Equation (3.9) is it; no other assumptions or ideas are used. The fact that s and S are Fourier transform pairs is reflected in Eq. (3.7).

Now, for any two functions (not only Fourier transform pairs)

$$\int |\, f(x)\,|^2\, dx \int |\, g(x)\,|^2\, dx \;\geq\; \left| \int f^*(x)\, g(x)\, dx \right|^2 \tag{3.10}$$

which is commonly known as the Schwarz inequality.[4] Taking $f = ts$ and $g = s'$ gives

$$\sigma_t^2\, \sigma_\omega^2 \;\geq\; \left| \int t\, s^*(t) s'(t)\, dt \right|^2 \tag{3.11}$$

[3] The constant phase factor $e^{-j\langle\omega\rangle\langle t\rangle}$ is irrelevant and can be left out or in for the sake of symmetry.

[4] There are many proofs of the inequality. A simple one is to note that for any two functions

$$\int |\, f(x)\,|^2\, dx \int |\, g(x)\,|^2\, dx - \left| \int f^*(x)\, g(x)\, dx \right|^2 = \tfrac{1}{2} \iint |\, f(x)\, g(y) - f(y)\, g(x)\,|^2\, dx\, dy$$

which is readily verified by direct expansion of the right hand side. Since the right hand side is manifestly positive, we have Eq. (3.10).

The integrand, written in terms of amplitude and phase, is

$$t\,s^*(t)s'(t) = t\,A'A + jt\varphi' A^2 \tag{3.12}$$

$$= \frac{1}{2}\frac{d}{dt}tA^2 - \tfrac{1}{2}A^2 + j\,t\varphi'(t) \tag{3.13}$$

The first term is a perfect differential and integrates to zero. The second term gives one half since we assume the signal is normalized and the third term gives j times the covariance of the signal. Hence

$$\sigma_t^2\,\sigma_\omega^2 \;\geq\; \left|\int t\,s^*(t)\,s'(t)\,dt\right|^2 = \left|-\tfrac{1}{2} + j\,\mathrm{Cov}_{t\omega}\right|^2 = \tfrac{1}{4} + \mathrm{Cov}_{t\omega}^2 \tag{3.14}$$

Therefore we have the uncertainty principle as given by Eq. (3.4). Since $\mathrm{Cov}_{t\omega}^2$ is always positive, it can, if we so choose, be dropped to obtain the more usual form, Eq. (3.3).

Minimum Uncertainty Product Signals. Since the minimum value for the uncertainty product is one half, we can ask what signals have that minimum value. The Schwarz inequality becomes an equality when the two functions are proportional to each other. Hence, we take $g = -cf$, where c is a constant and the -1 has been inserted for convenience. We therefore have

$$-\,c\,t\,s(t) = s'(t) \tag{3.15}$$

This is a necessary condition for the uncertainty product to be the minimum. But it is not sufficient since we must also have the covariance equal to zero, because by Eq. (3.4) we see that is the only way we can actually get the value of 1/2. Since c is arbitrary we can write it in terms of its real and complex parts, $c = c_r + jc_i$. The solution of Eq. (3.15) is hence

$$s(t) \sim e^{-ct^2/2} = e^{-(c_r + jc_i)t^2/2} \tag{3.16}$$

The covariance is the average value of t multiplied by the derivative of the phase. The derivative of the phase is $-c_i t$ and, remembering that we are considering a signal whose time and frequency means are zero, we have

$$\mathrm{Cov}_{t\omega} = -\int t\,c_i\,t|s(t)|^2\,dt - \langle\omega\rangle\langle t\rangle = -c_i\int t^2\,e^{-c_r t^2} \tag{3.17}$$

The only way this can be zero is if c_i is equal to zero and hence c must be a real number. If we take $c = \alpha/2$ we then have

$$s(t) = (\alpha/\pi)^{1/4}\,e^{-\alpha t^2/2} \tag{3.18}$$

where we have included the appropriate normalization. Reinstating the arbitrary mean time and frequency by way of Eq. (3.6) we have

$$s(t) = (\alpha/\pi)^{1/4}\,e^{-\alpha(t-\langle t\rangle)^2/2 + j\langle\omega\rangle t} \tag{3.19}$$

This is the most general signal that has a time-bandwidth product equal to one half.

Example 3.1: Chirp.

The standard deviations of time and frequency for the signal

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \tag{3.20}$$

are given by Eq. (1.12) and (1.40). Using those values gives

$$\sigma_t \, \sigma_\omega = \sqrt{\frac{1}{2\alpha}} \sqrt{\frac{\alpha^2 + \beta^2}{2\alpha}} = \frac{1}{2}\sqrt{1 + \frac{\beta^2}{\alpha^2}} = \tfrac{1}{2}\sqrt{1 + 4\mathrm{Cov}_{t\omega}} \tag{3.21}$$

For this case the stronger version of the uncertainty principle yields an equality.

Example 3.2: Sinusoid Modulation.

Using the results of Eq. (1.12) and (1.50) we have for

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + jm \sin \omega_m t + j\omega_0 t} \tag{3.22}$$

that

$$\sigma_t \, \sigma_\omega = \sqrt{\frac{1}{2\alpha}} \sqrt{\frac{\alpha}{2} + \frac{m^2\omega_m^2}{2} \left(1 - e^{-\omega_m^2/(2\alpha)}\right)^2} \tag{3.23}$$

$$= \frac{1}{2}\sqrt{1 + \frac{m^2\omega_m^2}{\alpha} \left(1 - e^{-\omega_m^2/(2\alpha)}\right)^2} \tag{3.24}$$

Example 3.3: Damped Exponential.

For the signal

$$s(t) = \sqrt{\frac{\alpha^{2n+1}}{(2n)!}} \, t^n \, e^{-\alpha t/2 + j\omega_0 t} \qquad t \geq 0, \quad n \geq 1 \tag{3.25}$$

the spectrum is

$$S(\omega) = n! \sqrt{\frac{\alpha^{2n+1}}{2\pi(2n)!}} \, \frac{1}{[\,-\alpha/2 + j(\omega - \omega_0)\,]^{n+1}} \tag{3.26}$$

and

$$\langle t \rangle = \frac{2n+1}{\alpha} \quad ; \quad \langle t^2 \rangle = \frac{(2n+2)(2n+1)}{\alpha^2} \quad ; \quad \sigma_t^2 = \frac{2n+1}{\alpha^2} \tag{3.27}$$

$$\langle \omega \rangle = \omega_0 \quad ; \quad \langle \omega^2 \rangle = \omega_0^2 + \frac{\alpha^2}{2n-1} \quad ; \quad \sigma_\omega^2 = \frac{\alpha^2}{2n-1} \tag{3.28}$$

Therefore

$$\sigma_t \, \sigma_\omega = \frac{1}{2}\sqrt{\frac{2n+1}{2n-1}} \tag{3.29}$$

3.4 THE UNCERTAINTY PRINCIPLE FOR THE SHORT-TIME FOURIER TRANSFORM

There are many things one can do to signals to study them. However, if we do something to a signal that modifies it in some way, one should not confuse the uncertainty principle applied to the modified signal with the uncertainty principle as applied to the original signal. One of the methods used to estimate properties of a signal is to take only a small piece of the signal around the time of interest and study that piece while neglecting the rest of the signal. In particular, we can take the Fourier transform of the small piece of the signal to estimate the frequencies at that time. If we make the time interval around the time t small, we will have a very high bandwidth. This statement applies to the modified signal, that is, to the short interval that we have artificially constructed for the purpose of analysis. What does the uncertainty principle as applied to a small time interval have to do with the uncertainty principle of the original signal? Very often nothing and statements about the chopped up signal should not be applied to the original signal. The process of chopping up a signal for the purpose of analysis is called the short-time Fourier transform procedure. Although we will be studying the short-time Fourier transform in Chapter 7, this is an appropriate place to consider the uncertainty principle for it.

From the original signal $s(t)$ one defines a short duration signal around the time of interest, t, by multiplying it by a window function that is peaked around the time, t, and falls off rapidly. This has the effect of emphasizing the signal at time, t, and suppressing it for times far away from that time. In particular, we define the normalized short duration signal at time, t, by

$$\eta_t(\tau) = \frac{s(\tau)\,h(\tau - t)}{\sqrt{\int |\, s(\tau)\,h(\tau - t)\,|^2\, d\tau}} \tag{3.30}$$

where $h(t)$ is the window function, t is the fixed time for which we are interested, and τ is now the running time. This normalization ensures that

$$\int |\, \eta_t(\tau)\,|^2\, d\tau = 1 \tag{3.31}$$

for any t. Now $\eta_t(\tau)$ as a function of the time τ is of short duration since presumably we have chosen a window function to make it so. The time, t, acts as a parameter. The Fourier transform of the small piece of the signal, the modified signal, is[5]

$$F_t(\omega) = \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} \eta_t(\tau)\, d\tau \tag{3.32}$$

[5]If the denominator in Eq. (3.30) is omitted then $F_t(\omega)$ would become what is traditionally called the short-time Fourier transform. In calculating quantities such as conditional averages, the normalization must come in at some point and it is a matter of convenience as to when to take it into account. In Chapter 7 we use the more conventional definition and omit it from the definition. There we use $S_t(\omega)$ to denote the short-time Fourier transform and the relation between the two is $S_t(\omega) = F_t(\omega)\{\int |\, s(\tau)\,h(\tau - t)\,|^2\, d\tau\}^{1/2}$.

$F_t(\omega)$ gives us an indication of the spectral content at the time t. For the modified signal we can define all the relevant quantities such as mean time, duration, and bandwidth in the standard way, but they will be time dependent. The mean time and duration for the modified signal are

$$\langle \tau \rangle_t = \int \tau \, | \, \eta_t(\tau) \, |^2 \, d\tau = \frac{\int \tau \, |s(\tau)\, h(\tau - t)|^2 \, d\tau}{\int \, | \, s(\tau)\, h(\tau - t) \, |^2 \, d\tau} \tag{3.33}$$

$$T_t^2 = \int (\tau - \langle \tau \rangle)^2 \, | \, \eta_t(\tau) \, |^2 \, d\tau = \frac{\int (\tau - \langle \tau \rangle_t)^2 \, |s(\tau)\, h(\tau - t)|^2 \, d\tau}{\int \, | \, s(\tau)\, h(\tau - t) \, |^2 \, d\tau} \tag{3.34}$$

Similarly, the mean frequency and bandwidth for the modified signal are

$$\langle \omega \rangle_t = \int \omega \, | \, F_t(\omega) \, |^2 \, d\omega = \int \eta_t^*(\tau) \, \frac{1}{j} \frac{d}{d\tau} \, \eta_t(\tau) \, d\tau \tag{3.35}$$

$$B_t^2 = \int (\omega - \langle \omega \rangle_t)^2 \, | \, F_t(\omega) \, |^2 \, d\omega \tag{3.36}$$

Time-Dependent and Window-Dependent Uncertainty Principle. Since we have used a normalized signal to calculate the duration and bandwidth, we can immediately write that

$$B_t \; T_t \geq \tfrac{1}{2} \tag{3.37}$$

This is the uncertainty principle for the short-time Fourier transform. It is a function of time, the signal, and the window. It should not be confused with the uncertainty principle applied to the signal. It is important to understand this uncertainty principle because it places limits on the technique of the short-time Fourier transform procedure. However, it places no constraints on the original signal.

We know that for infinitely short duration signals the bandwidth becomes infinite. Hence we expect that $B_t \to \infty$ as we narrow the window, which is indeed the case. This is shown in Chapter 7 where we obtain explicit expressions for the above quantities in terms of the amplitude and phases of the signal and window. The point that we are making here is that we must be very clear with regard to the uncertainty principle for the original signal and modified signal. It is true that if we modify the signal by the technique of the short-time Fourier transform we limit our abilities in terms of resolution and so forth. This is a limitation of the technique. The uncertainty principle of the original signal does not change because we have decided to modify it by windowing.

Example 3.4: Chirp with a Gaussian Window.

Consider a chirp and use a Gaussian window,

$$s(t) = (\alpha/\pi)^{1/4} \, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \quad ; \quad h(t) = (a/\pi)^{1/4} \, e^{-at^2/2} \tag{3.38}$$

We first find the normalization factor, the denominator in Eq. (3.30),

$$\int |\, s(\tau)\, h(\tau - t)\,|^2\, d\tau = \left(\frac{a\alpha}{\pi(a+\alpha)}\right)^{1/2} \exp\left[-\frac{a\alpha}{a+\alpha}t^2\right] \tag{3.39}$$

and therefore the modified signal is

$$\eta_t(\tau) = \left(\frac{\alpha+a}{\pi}\right)^{1/4} \exp\left[-\tfrac{1}{2}(\alpha+a)\tau^2 + a\tau t + j\omega_0\tau + j\beta\tau^2/2 - \frac{a^2}{a+\alpha}t^2\right] \tag{3.40}$$

from which we get

$$\langle \tau \rangle_t = \frac{a}{\alpha+a}t \qquad ; \qquad T_t^2 = \frac{1}{2(\alpha+a)} \tag{3.41}$$

$$\langle \omega \rangle_t = \frac{a}{\alpha+a}\,\beta t + \omega_0 \qquad ; \qquad B_t^2 = \tfrac{1}{2}\,(\alpha+a) + \tfrac{1}{2}\,\frac{\beta^2}{\alpha+a} \tag{3.42}$$

and therefore

$$B_t\ T_t = \frac{1}{2}\sqrt{1 + \frac{\beta^2}{(\alpha+a)^2}} \tag{3.43}$$

For this case the local duration and bandwidth are time independent and so is the uncertainty product. This is a peculiarity of the Gaussian window and will not be the case with another window and/or signal. Now notice that as we narrow the window, that is, as $a \to \infty$, the bandwidth of the modified signal, B_t, goes to infinity. Of course, this has nothing to do with the bandwidth of the original signal.

Example 3.5: Example : Quadratic FM Signal.

Consider a signal whose phase is cubic in time with a Gaussian envelope,

$$s(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2 + j\gamma t^3/3} \qquad ; \qquad h(t) = (a/\pi)^{1/4}\, e^{-at^2/2} \tag{3.44}$$

The mean time and durations are the same as in the preceding Example. The mean frequency and bandwidth of the modified signal are

$$\langle \omega \rangle_t = \frac{1}{2}\,\frac{\gamma}{(\alpha+a)} + \frac{\gamma a^2}{(\alpha+a)^2}t^2 \tag{3.45}$$

$$B_t^2 = \tfrac{1}{2}(\alpha+a) + \frac{\gamma^2}{2(\alpha+a)^2} + \frac{2\gamma^2 a^2}{(\alpha+a)^3}\,t^2 \tag{3.46}$$

The uncertainty principle is then

$$B_t\ T_t = \frac{1}{2}\sqrt{1 + \frac{\beta^2}{(\alpha+a)^2} + \frac{\gamma^2 a^2}{(\alpha+a)^4}\,t^2} \tag{3.47}$$

For large times and short windows, the time dependent bandwidth goes as $B_t \sim t^2/a^2$; hence for large times or small windows we get a very large bandwidth. Again, this has nothing to do with the original signal. The bandwidth of the original signal is fixed.

Chapter 4

Densities and Characteristic Functions

4.1 INTRODUCTION

Densities are the basic quantities we deal with, since we seek the energy density of a signal at a particular time and frequency. We have already encountered densities in the previous chapters, namely, the absolute square of the signal is the energy density in time and the square magnitude of the Fourier transform is the energy density per unit frequency. In this chapter we develop the basic methods and concepts used to study and construct densities. We present the ideas in a manner that is useful in studying energy densities in time, frequency, and joint time-frequency densities.

By density we mean nothing more than the common usage, such as the density of mass or the density of trees. Many of the ideas such as "standard deviation" will have a probabilistic ring to them since they are usually associated with probability densities. However, we emphasize that the methods presented in this chapter apply to any density. For the one dimensional case we should simply think of a wire that may stretch from minus infinity to infinity and where the mass density varies with location. Similarly, for the two dimensional case one should think of a sheet where again the density varies with location on the sheet.

4.2 ONE DIMENSIONAL DENSITIES

A one dimensional density is the amount of something per unit something else. For example, the number of people per unit height, the amount of mass per unit length, the intensity per unit frequency, or the intensity per unit time. We use $P(x)$

to denote a one dimensional density of the quantity x:

$$P(x)\,\Delta x \;=\; \text{the amount in the interval } \Delta x \text{ at } x \tag{4.1}$$

Since $P(x)\Delta x$ is the amount in the interval Δx, the total amount is the sum,

$$\text{Total amount} \;=\; \int P(x)\,dx \;=\; 1 \tag{4.2}$$

Densities are often normalized so that the total amount is equal to one. The effect of doing that, is, for example, if $P(x)$ represents the actual number of people of a certain height, then by normalizing it to one it represents the ***fraction*** of people at that height.

Densities have to be single valued and positive. The positive requirement comes from the fact that a density is the amount of something per unit something else and both quantities are positive. It should be emphasized that the numerator is the amount of something and while that something can be negative, the amount is still positive. For example, while charge can be positive or negative the amount of positive or negative charge is positive.

Terminology. Density function is universally understood; in some fields "distribution" is used interchangeably with density because $P(x)$ indicates how something is distributed. In mathematics, however, distribution usually denotes the amount up to a certain value, that is, the integral of the density from $-\infty$ to x,

$$F(x) \;=\; \int_{-\infty}^{x} P(x)\,dx \tag{4.3}$$

which is also called the cumulative distribution function because it accumulates the density up to x. We will use density and distribution interchangeably and will have no occasion to use the concept of cumulative distribution function. Incidentally, the advantage of defining the cumulative distribution function is that often densities do not exist in the strict mathematical sense but cumulative distributions always do. Note that the density is related to the cumulative distribution function by $P(x) \;=\; F'(x)$ when indeed the derivative exists. Another reason for introducing a cumulative distribution function is that it allows one to handle in a mathematically smooth manner both continuous and discrete distributions. However if we are not particularly interested in mathematical sophistication the delta function provides a very effective way to do this.

Averages. From the density simpler quantities are obtained which sometimes give a gross indication of the basic characteristics of the density. The average, $\langle\, x \,\rangle$, is

$$\langle\, x \,\rangle \;=\; \int x\,P(x)\,dx \tag{4.4}$$

The average of any function, $f(x)$ can be obtained in two different ways: Either from the density of x or from the density of u, $P(u)$, where $u = f(x)$,

$$\langle f(x) \rangle = \int f(x) P(x) dx = \int u P(u) du \tag{4.5}$$

How to obtain the density $P(u)$ from $P(x)$ is described in Section 4.7. Two general properties of averages should be noted. First, the average of the sum of two functions is the sum of the average of each function, $\langle f(x) + g(x) \rangle = \langle f(x) \rangle + \langle g(x) \rangle$. Second, the average of a constant times a function is the constant times the average of the function, $\langle cf(x) \rangle = c\langle f(x) \rangle$.

Standard Deviation. While the average gives the place where things are balanced, it gives no indication as to whether the density is concentrated there or not. A measure of that is the variance, which is the average of $(x - \langle x \rangle)^2$. Variance is a good measure of concentration because if the density is concentrated near $\langle x \rangle$ then $(x - \langle x \rangle)^2$ will be relatively small. The weighting of that particular value is, of course, taken with the density. The standard deviation, σ_x, is defined as the square root of the variance and given by

$$\sigma_x^2 = \int (x - \langle x \rangle)^2 P(x) dx \tag{4.6}$$

$$= \langle x^2 \rangle - \langle x \rangle^2 \tag{4.7}$$

Similar considerations apply to functions of x,

$$\sigma_f^2 = \int (f(x) - \langle f(x) \rangle)^2 P(x) dx \tag{4.8}$$

$$= \langle f^2(x) \rangle - \langle f(x) \rangle^2 \tag{4.9}$$

Some general properties of the standard deviation are that the standard deviation of a constant is zero, $\sigma_c = 0$, and the standard deviation of cx is given by $\sigma_{cx} = c\sigma_x$

Moments. The moment of order n is the average of x^n,

$$\langle x^n \rangle = \int x^n P(x) dx \tag{4.10}$$

The significance of the moments are many fold. First, the first few moments give an indication of the general properties of the distribution. Second, in some sense, the more moments we know, the more we know about the distribution. Third, for well behaved distributions the moments uniquely determine the distribution. The procedure for constructing the distribution from the moments is presented in the next section. In Chapter 10 we see how these methods can be used for the construction of time-frequency distributions.

4.3 ONE DIMENSIONAL CHARACTERISTIC FUNCTIONS

The characteristic function is a powerful tool for the study and construction of densities. It is the Fourier transform of the density

$$M(\theta) = \int e^{j\theta x} P(x)\, dx = \langle e^{j\theta x} \rangle \tag{4.11}$$

The characteristic function is the average of $e^{j\theta x}$, where θ is a parameter. By expanding the exponential we have

$$M(\theta) = \int e^{j\theta x} P(x)\, dx = \int \sum_{i=1}^{\infty} \frac{(j\theta x)^n}{n!} P(x)\, dx = \sum_{i=1}^{\infty} \frac{j^n \theta^n}{n!} \langle x^n \rangle \tag{4.12}$$

which is a Taylor series in θ with coefficients $j^n \langle x^n \rangle$. Since the coefficients of a Taylor series are given by the nth derivative of the function evaluated at zero, we have

$$\langle x^n \rangle = \frac{1}{j^n} \frac{\partial^n M(\theta)}{\partial \theta^n} \bigg|_{\theta=0} \tag{4.13}$$

The fact that moments can be calculated by differentiation rather than by integration is one of the advantages of the characteristic function, since differentiation is always easier than integration. Of course, one has to first obtain the characteristic function and that may be hard.

Fourier transform pairs are uniquely related and hence the characteristic function determines the distribution,

$$P(x) = \frac{1}{2\pi} \int M(\theta)\, e^{-j\theta x}\, d\theta \tag{4.14}$$

Some general properties that a function must posses if it is a proper characteristic function are easily obtained. By proper we mean a characteristic function that comes from a normalized positive density. First, taking $\theta = 0$ we see that

$$M(0) = \int P(x)\, dx = 1 \tag{4.15}$$

Taking the complex conjugate of Eq. (4.11) and using the fact that densities are real, we have

$$M^*(\theta) = \int e^{-j\theta x} P^*(x)\, dx = M(-\theta) \tag{4.16}$$

or

$$M^*(-\theta) = M(\theta) \tag{4.17}$$

The absolute value of the characteristic function is always less than or equal to one,

$$|\, M(\theta)\,| \leq 1 \tag{4.18}$$

This follows from

$$|M(\theta)| = \left| \int e^{j\theta x} P(x)\, dx \right| \leq \int |e^{j\theta x}|\, |P(x)|\, dx = \int P(x)\, dx = 1 \quad (4.19)$$

We know that the characteristic function at the origin is equal to one and therefore

$$|M(\theta)| \leq M(0) \quad (4.20)$$

Example 4.1: Gaussian Distribution.

For the Gaussian

$$P(x) = \sqrt{\frac{1}{2\pi\sigma^2}}\, \exp\left[-(x-x_0)^2/2\sigma^2\right] \quad (4.21)$$

the characteristic function is

$$M(\theta) = \sqrt{\frac{1}{2\pi\sigma^2}} \int \exp\left[-(x-x_0)^2/2\sigma^2\right] e^{j\theta x}\, dx = e^{-\sigma^2\theta^2/2 + jx_0\theta} \quad (4.22)$$

The first and second moments can be obtained by differentiation,

$$\langle x \rangle = \frac{1}{j}\frac{\partial M(\theta)}{\partial \theta}\bigg|_{\theta=0} = \frac{1}{j}\left(-\sigma^2\theta + jx_0\right) M(\theta)\bigg|_{\theta=0} = x_0 \quad (4.23)$$

$$\langle x^2 \rangle = \frac{1}{j^2}\frac{\partial^2 M(\theta)}{\partial \theta^2}\bigg|_{\theta=0} = \frac{1}{j^2}\left[\left(-\sigma^2\theta + jx_0\right)^2 - \sigma^2\right] M(\theta)\bigg|_{\theta=0} = x_0^2 + \sigma^2 \quad (4.24)$$

Note that the advantage of pulling out $M(\theta)$ is that it is always equal to one for $\theta = 0$.

Example 4.2: Exponential Density.

For the exponential density

$$P(x) = \lambda e^{-\lambda x} \qquad 0 \leq x \leq \infty \quad (4.25)$$

the characteristic function is

$$M(\theta) = \int_0^\infty \lambda e^{-\lambda x} e^{j\theta x} = \frac{\lambda}{\lambda - j\theta} \quad (4.26)$$

which by differentiation gives, for the first and second moments,

$$\langle x \rangle = \frac{1}{j}\frac{\partial M(\theta)}{\partial \theta}\bigg|_{\theta=0} = \frac{1}{j}\frac{j\lambda}{(\lambda - j\theta)^2}\bigg|_{\theta=0} = 1/\lambda \quad (4.27)$$

$$\langle x^2 \rangle = \frac{1}{j^2}\frac{\partial^2 M(\theta)}{\partial \theta^2}\bigg|_{\theta=0} = \frac{1}{j^2}\frac{j^2 2\lambda}{(\lambda - j\theta)^3}\bigg|_{\theta=0} = 2/\lambda^2 \quad (4.28)$$

Relation of Moments to Density. Generally speaking, the knowledge of all the moments determines the distribution. This can be seen in the following way. From the moments we can calculate the characteristic function as per Eq. (4.12). Once the characteristic function is obtained the density is obtained by Fourier inversion. There are cases where the moments do not determine a unique characteristic function or density, but these exceptions will not concern us.

Example 4.3: Distribution from the Moments.

Suppose we have the set of moments

$$\langle x^n \rangle = \frac{(n+1)!}{\alpha^n} \tag{4.29}$$

We construct the characteristic function

$$M(\theta) = \sum \frac{j^n \theta^n}{n!} \langle x^n \rangle = \sum \frac{j^n \theta^n}{n!} \frac{(n+1)!}{\alpha^n} = \sum \frac{j^n \theta^n (n+1)}{\alpha^n} = \left(\frac{\alpha}{\alpha - j\theta} \right)^2 \tag{4.30}$$

Therefore the density is

$$P(x) = \frac{1}{2\pi} \int M(\theta) e^{-j\theta x} \, d\theta = \frac{1}{2\pi} \int \left(\frac{\alpha}{\alpha - j\theta} \right)^2 e^{-j\theta x} \, d\theta \tag{4.31}$$

$$= \alpha^2 x e^{-\alpha x^2} \qquad [\text{if } 0 \leq \alpha \leq \infty] \tag{4.32}$$

and zero otherwise.

When Is a Function a Characteristic Function? A characteristic function is a complex function, but not every complex function is a characteristic function, since the function must be the Fourier transform of a density. The conditions just derived, Eqs. (4.15), (4.17), and (4.20), are necessary but not sufficient. The reason is that nowhere have we used the fact that the density is positive. A necessary and sufficient condition for a function to be a characteristic function was given by Khinchin.[304] A function $M(\theta)$, is a characteristic function if and only if there exists another function $g(\theta)$ such that

$$M(\theta) = \int g^*(\theta') \, g(\theta' + \theta) \, d\theta' \tag{4.33}$$

The function g is to be normalized to one,

$$\int |g(\theta)|^2 \, d\theta = 1 \tag{4.34}$$

This is a very basic result, and will be important to our considerations since the g's, will turn out to be "signals".

We show the sufficiency here because it produces a result that is revealing. Assuming that $M(\theta)$ is a proper characteristic function, the density is given by

$$P(x) = \frac{1}{2\pi}\int M(\theta)\, e^{-j\theta x}\, d\theta = \frac{1}{2\pi}\int\int g^*(\theta')g(\theta'+\theta)\, e^{-j\theta x}\, d\theta'\, d\theta \qquad (4.35)$$

Making a change of variables $\theta'' = \theta' + \theta \;\; ; \;\; d\theta'' = d\theta$, we have

$$P(x) = \frac{1}{2\pi}\left|\int g(\theta)\, e^{-j\theta x}\, d\theta\right|^2 \qquad (4.36)$$

which shows that $P(x)$ is a proper density since it is positive. By Parceval's theorem we know that $P(x)$ is normalized to one if g is normalized to one as per Eq. (4.34). Note that while there is a one-to-one relationship of the density to the characteristic function, that is not the case with the function g. Many g's produce the same characteristic function and hence the same density.

4.4 TWO DIMENSIONAL DENSITIES

Consideration of two dimensional densities, that is, densities that depend on two variables, forces the introduction of a number of new concepts not encountered with one dimensional densities. These new ideas, such as conditional average and correlation remain intact for densities of more than two variables.

A two dimensional density $P(x,y)$ is the amount of something per unit x and per unit y at the point x, y. The total amount is

$$\text{Total amount} = \iint P(x,y)\, dy\, dx = 1 \qquad (4.37)$$

Again we normalize to one so that, for example, if $P(x,y)$ is the number of people of a certain weight and height, then when normalized to one it becomes the fraction of people of that weight and height.

Marginals. If we have a joint density we may want the density of just one of the variables irrespective of the value of the other variable. This is achieved by integrating out the other variable,

$$P(x) = \int P(x,y)\, dy \qquad ; \qquad P(y) = \int P(x,y)\, dx \qquad (4.38)$$

These densities are called marginal densities or marginals. The origins of the word "marginal" came about as follows. When the basic concepts of probability and statistics were being developed, joint densities were written on a piece of paper in columns and rows, in the fashion of the modern spreadsheet. For example, if we wanted to make a table indicating how many individuals there are of a certain height and weight (the joint density), the weight axis would be on top of the page and the height axis would run down the page as in a typical spreadsheet. At the

intersection of each row and column one would write the number of people of that weight and height. Now suppose one wanted the number of people of a certain weight irrespective of height (the weight marginal) and the number of people of a certain height, irrespective of weight (the height marginal). Then one adds up the columns and rows. A natural place to put the sums is in the margins of the paper, that is on the right margin and the bottom of the page – hence the terminology marginals.

Global Averages. For a function $g(x, y)$, the global average is

$$\langle g(x,y) \rangle = \iint g(x,y)\, P(x,y)\, dy\, dx \tag{4.39}$$

Two Dimensional Characteristic Functions and Moments. The two dimensional characteristic function $M(\theta, \tau)$, is the average of $e^{j\theta x + j\tau y}$,

$$M(\theta,\tau) = \langle e^{j\theta x + j\tau y} \rangle = \iint e^{j\theta x + j\tau y} P(x,y)\, dx\, dy \tag{4.40}$$

and the distribution function may be obtained from $M(\theta, \tau)$ by Fourier inversion,

$$P(x,y) = \frac{1}{4\pi^2} \iint M(\theta,\tau)\, e^{-j\theta x - j\tau y}\, d\theta\, d\tau \tag{4.41}$$

Similar to the one dimensional case, expanding the exponential in Eq. (4.40) results in a two dimensional Taylor series,

$$M(\theta,\tau) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{(j\theta)^n (j\tau)^m}{n!m!} \langle x^n y^m \rangle \tag{4.42}$$

and therefore

$$\langle x^n y^m \rangle = \frac{1}{j^n j^m} \frac{\partial^{n+m}}{\partial\theta^n \partial\tau^m} M(\theta,\tau) \Big|_{\theta,\tau=0} \tag{4.43}$$

Relation Between the Joint Characteristic Function and the Characteristic Function of the Marginals. The characteristic function of the marginal for x is

$$M(\theta) = \int e^{j\theta x} P(x)\, dx = \int e^{j\theta x} P(x,y)\, dx\, dy = M(\theta, 0) \tag{4.44}$$

Similarly, the characteristic function for y is

$$M(\tau) = \int e^{j\tau y} P(y)\, dy = \int e^{j\tau y} P(x,y)\, dx\, dy = M(0, \tau) \tag{4.45}$$

Therefore if we know the characteristic function of the joint distribution we can trivially obtain the characteristic functions of the marginals by taking zero for one of the variables.

Example 4.4: Two Dimensional Gaussian.

The standard form for a two dimensional Gaussian is

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_x^2}+\frac{(y-b)^2}{\sigma_y^2}-2r\frac{(x-a)(y-b)}{\sigma_x\sigma_y}\right]\right\} \quad (4.46)$$

with $|r| \leq 1$. The characteristic function, calculated by way of Eq. (4.40), is

$$M(\theta,\tau) = \exp\left[ja\theta + jb\tau - \tfrac{1}{2}(\sigma_x^2\theta^2 + 2r\sigma_x\sigma_y\theta\tau + \sigma_y^2\tau^2)\right] \quad (4.47)$$

The characteristic function of the marginal in x is therefore

$$M(\theta) = M(\theta,0) = \exp\left[ja\theta - \sigma_x^2\theta^2/2\right] \quad (4.48)$$

This is the same form as the characteristic function of a one dimensional Gaussian, Eq. (4.22), and hence the marginal is Gaussian with mean a and standard deviation σ_x.

Example 4.5: Non-Gaussian Joint Density with Gaussian Marginals.

The distribution given by Eq. (4.46) is the standard joint Gaussian distribution. As we just showed, it has Gaussian marginals. However, an infinite number of two dimensional non Gaussian distributions have Gaussian marginals. A method for readily constructing such distributions is discussed in Section 14.2.

Example 4.6: Characteristic Function of the Sum.

One of the main uses of characteristic functions is that they allow us to obtain new densities in an easy way. This is discussed in subsequent sections, but a simple example is appropriate here. Suppose we want to obtain the density of the sum of two variables,

$$z = x + y \quad (4.49)$$

The characteristic function of z is

$$M_z(\theta) = \langle e^{j\theta(x+y)} \rangle = M(\theta,\theta) \quad (4.50)$$

from which the distribution of z may be obtained by inversion. For example, for the two dimensional Gaussian discussed above

$$M_z(\theta) = M(\theta,\theta) = \exp\left[j(a+b)\theta - \theta^2\left(\sigma_x^2 + 2r\sigma_x\sigma_y + \sigma_y^2\right)/2\right] \quad (4.51)$$

This is recognized to be the characteristic function of a Gaussian, Eq. (4.22), with mean $a+b$ and standard deviation $\sigma_x^2 + 2r\sigma_x\sigma_y + \sigma_y^2$.

Independence. If fixing a particular variable has no effect on the density of the other variable, then one says that the joint distribution is independent. That is the case if the joint distribution is factorable *and* the factors are the marginals,

$$P(x,y) = P(x)\,P(y) \tag{4.52}$$

For this situation the characteristic function and the joint moments are also factorable,

$$M(\theta,\tau) = M_x(\theta)M_y(\tau) \quad ; \quad \langle x^n y^m \rangle = \langle x^n \rangle \langle y^m \rangle \tag{4.53}$$

However, one should be careful because sometimes a density may appear factorable without being independent. For independence the factors must be the marginals. The examples below illustrate this point.

Example 4.7: Rectangular and Circular Two Dimensional Uniform Distributions.

Consider first

$$P(x,y) = \begin{cases} \frac{4}{a^4}xy & \text{if } 0 \le x,y \le a \\ 0 & \text{otherwise} \end{cases} \tag{4.54}$$

The marginals are

$$P(x) = \frac{4}{a^4}\int_0^a xy\,dy = \frac{2}{a^2}x \quad ; \quad P(y) = \frac{4}{a^4}\int_0^a xy\,dx = \frac{2}{a^2}y \tag{4.55}$$

Thus $P(x,y) = P(x)P(y)$ and hence the distribution is an independent one. Now consider

$$P(x,y) = \begin{cases} \frac{8}{a^4}xy & \text{if } x^2+y^2 \le a^2 \quad ; \quad 0 \le x,y \\ 0 & \text{otherwise} \end{cases} \tag{4.56}$$

We have

$$P(x) = \frac{8}{a^4}x\int_0^{\sqrt{a^2-x^2}} y\,dy = \frac{4}{a^4}x\,(a^2-x^2) \tag{4.57}$$

$$P(y) = \frac{8}{a^4}y\int_0^{\sqrt{a^2-y^2}} x\,dx = \frac{4}{a^4}y\,(a^2-y^2) \tag{4.58}$$

We see that even though the joint density is a product in some sense it is not independent because the product is not the product of the marginals. In fact, this joint distribution is not really factorable for all values of x and y. Examples like these have historically been presented as a warning to use caution. In reality it is just a failure of notation. If we rewrite the joint density as

$$P(x,y) = \frac{4}{a^4}x\,y\,\epsilon(a^2-x^2-y^2)\,\epsilon(x)\,\epsilon(y) \tag{4.59}$$

where $\epsilon(x)$ is the step function, then it is clear that $P(x,y)$ is not factorable and the issue never arises.

Covariance and Correlation. Suppose we want to know how strongly one variable depends on the other. This is what a joint distribution tells us in precise terms. However, if we want a gross measure that reflects in a simple way the strength of the dependence, we can consider the expected value of the first mixed moment, $\langle x\,y\rangle$. If the joint distribution is independent, the first mixed moment will equal the product of the moments,

$$\langle x\,y\rangle = \langle x\rangle\langle y\rangle \qquad [\,x, y \text{ independent}\,] \tag{4.60}$$

The dependency can be measured by the excess of the first mixed moment over $\langle x\rangle\langle y\rangle$. That is the covariance

$$\mathrm{Cov}_{xy} = \langle x\,y\rangle - \langle x\rangle\langle y\rangle \tag{4.61}$$

The covariance is a number that can be positive or negative and of any magnitude. To standardize the correlation coefficient is defined by

$$r = \frac{\mathrm{Cov}_{xy}}{\sigma_x \sigma_y} \tag{4.62}$$

where σ_x and σ_y are the standard deviations of x and y. The correlation coefficient ranges from -1 to $+1$ for all densities and hence standardizes the strength of the correlation between x and y. The variables x and y are said to be uncorrelated if the correlation coefficient is zero, strongly correlated at $r = +1$, and strongly oppositely correlated when $r = -1$. However, one should be cautious in applying these ideas because they do not necessarily reflect the intuitive notion as to whether the two variables have something to do with each other. One must not confuse dependence with correlation. In fact, we can have a distribution where it is clear that the two variables are strongly dependent and yet the correlation coefficient is zero. The correlation coefficient is a single number and we cannot expect too much from just one number. Nonetheless, the correlation coefficient is often a good gross indicator of the dependence of two variables. The simplest way to calculate the first mixed moment is from the characteristic function

$$\langle x\,y\rangle = -\left.\frac{\partial^2 M(\theta,\tau)}{\partial\theta\partial\tau}\right|_{\theta,\tau=0} \tag{4.63}$$

if it is available.

4.5 LOCAL QUANTITIES

Suppose we have the density of height and weight of a population and want to study the weight of people who are 6 ft. tall. The density of weight for that subpopulation is the joint density, but we fix the height at 6 ft. Specifically, if we have two variables then we use the notation $P(y \mid x)$ to mean the density of y for a fixed x. Such a density is called a conditional density since it depends on the value of x

chosen. It can be thought of as a one dimensional density with x as a parameter. If we insist, as we should, that the conditional densities be normalized, then they are

$$P(y \mid x) = \frac{P(x,y)}{P(x)} \quad ; \quad P(x \mid y) = \frac{P(x,y)}{P(y)} \tag{4.64}$$

where $P(x)$ and $P(x)$ are the marginals. If $P(y \mid x)$ does not depend on x, then this is exactly what we would want to mean by independence. And if that is the case, then $P(y \mid x) = P(y)$, which implies that $P(y,x) = P(y)P(x)$, in conformity with our previous discussion of an independent distribution.

Conditional Averages and Their Standard Deviation. Suppose we want the average of y for a given x, for example, the average weight of those people who are 6 ft. tall. That is called the conditional average and is denoted by $\langle y \rangle_x$. Since the density of y for a given x is $P(x \mid y)$, we have

$$\langle y \rangle_x = \int y P(y \mid x)\, dy = \frac{1}{P(x)} \int y P(x,y)\, dy \tag{4.65}$$

More generally, the conditional average of any function (the local average) is

$$\langle g(y) \rangle_x = \frac{1}{P(x)} \int g(y)\, P(x,y)\, dy \tag{4.66}$$

Conditional Standard Deviation. Suppose a person weighs 250 lb. Is he overweight? If he compares himself to the mean of the total population he may be exceptional. But he should be comparing himself with individuals of the same height. Therefore, one defines the conditional standard deviation by the deviations from the conditional mean,

$$\sigma^2_{y|x} = \frac{1}{P(x)} \int (\, y - \langle y \rangle_x \,)^2\, P(x,y)\, dy \tag{4.67}$$

$$= \langle y^2 \rangle_x - \langle y \rangle_x^2 \tag{4.68}$$

4.6 RELATION BETWEEN LOCAL AND GLOBAL AVERAGES

If we integrate the conditional average, $\langle y \rangle_x$, over all values of x, we expect to obtain the global average of y and indeed that is the case since

$$\langle y \rangle = \iint y\, P(x,y)\, dy\, dx = \iint y P(y \mid x) P(x)\, dy\, dx \tag{4.69}$$

The inner integration is precisely the conditional average and therefore

$$\langle y \rangle = \int \langle y \rangle_x P(x)\, dx \tag{4.70}$$

This equation shows that the global average is the average of the conditional average.

We now seek the relationship between the global standard deviation and the conditional standard deviation. The global standard deviation is the not average of the local standard deviation, but rather

$$\sigma_y^2 = \int \sigma_{y|x}^2 \, P(x)\,dx + \int (\langle y \rangle_x - \langle y \rangle)^2 \, P(x)\,dx \tag{4.71}$$

To prove this we average $\sigma_{y|x}^2$,

$$\int \sigma_{y|x}^2 \, P(x)\,dx = \iint (y - \langle y \rangle_x)^2 \, P(x,y)\,dy\,dx \tag{4.72}$$

$$= \langle y^2 \rangle - \int \langle y \rangle_x^2 \, P(x)\,dx \tag{4.73}$$

Now subtract and add $\langle y \rangle^2$ to the right hand side of Eq. (4.73),

$$\int \sigma_{y|x}^2 \, P(x)\,dx = \langle y^2 \rangle - \langle y \rangle^2 - \int (\langle y \rangle_x^2 - \langle y \rangle^2) \, P(x)\,dx \tag{4.74}$$

$$= \sigma_y^2 - \int (\langle y \rangle_x - \langle y \rangle)^2 \, P(x)\,dx \tag{4.75}$$

which is Eq. (4.71).

We see that there are *always* two contributions to the global standard deviation. One is the average of the local standard deviation and the other the deviations of the conditional mean about the global mean. This equation is particularly important in our considerations of time-frequency analysis. The reader is invited now to compare this equation with the bandwidth equation derived in Chapter 1, Eq. (1.96), and to draw his own conclusions.

4.7 DISTRIBUTION OF A NEW VARIABLE

One Dimensional Case. Suppose we have a density, $P(x)$, and want the density of a new variable, say u, which is a function of x,

$$u = f(x) \tag{4.76}$$

If we knew the characteristic function of u, $M_u(\theta)$, we could obtain the distribution by

$$P(u) = \frac{1}{2\pi} \int e^{-j\theta u} M_u(\theta)\,d\theta \tag{4.77}$$

But the characteristic function of u is the average value of $e^{j\theta f(x)}$ and therefore can be obtained from $P(x)$ by

$$M_u(\theta) = \langle e^{j\theta f(x)} \rangle = \int e^{j\theta f(x)} P(x)\,dx \tag{4.78}$$

This solves the problem because from $P(x)$ we obtain $M_u(\theta)$ and then $P(u)$ by Eq. (4.77). We can actually carry out the procedure explicitly and obtain an interesting form for $P(u)$. Substitute Eq. (4.78) into Eq. (4.77), to obtain

$$P(u) = \frac{1}{2\pi} \iint e^{j\theta f(x)} e^{-j\theta u} P(x)\, d\theta\, dx \tag{4.79}$$

or

$$P(u) = \int \delta(u - f(x))\, P(x)\, dx \tag{4.80}$$

The physical interpretation is clear. The delta function picks up from $P(x)$ only the values for which $u = f(x)$. This form avoids the calculation of the characteristic function. Further simplification is possible. One of the properties of the delta function is

$$\delta(g(x)) = \sum_i \frac{1}{|g'(x_i)|} \delta(x - x_i) \tag{4.81}$$

where x_i's are the solutions to $g(x_i) = 0$, that is, the zeros of g. In our case we have $g(x) = f(x) - u$ and since the derivative of g is the derivative of f we have

$$\delta(f(x) - u) = \sum_i \delta(x - x_i) \frac{1}{|f'(x_i)|} \tag{4.82}$$

where now the x_i values run over the solutions of $f(x_i) = u$. Therefore,

$$P(u) = \int \delta(u - f(x))\, P(x)\, dx \tag{4.83}$$

$$= \int \sum_i \delta(x - x_i) \frac{1}{|f'(x_i)|} P(x)\, dx \tag{4.84}$$

or

$$P(u) = \sum_i \frac{P(x_i)}{|f'(x_i)|} \tag{4.85}$$

[where x_i's are the solutions of $f(x_i) = u$]

Suppose there is only one solution to $f(x_i) = u$, which we might as well call x. Then

$$P(u) = \left. \frac{P(x)}{|f'(x)|} \right|_{x = f^-(u)} \tag{4.86}$$

where $x = f^-(u)$ signifies that we must solve for x in terms of u by solving $f(x) = u$. Equivalently, we note that for this case $P(u)\, du = P(x)\, dx$, which is the easiest way to do it.

Example 4.8: One Zero.

As a specific example consider

$$P(x) = \frac{1}{\pi} \qquad 0 \le x \le \pi \tag{4.87}$$

We wish to find the density of u where

$$u = f(x) = A\cos x \tag{4.88}$$

Since for the range considered the function is single valued, we have $f'(x) = -A\sin x$ and therefore

$$P(u) = \frac{1}{\pi}\frac{1}{|A\sin x|}\bigg|_{x=\arccos(u/A)} = \frac{1}{\pi}\frac{1}{\sqrt{A^2-u^2}} \qquad \text{if } -A \le u \le A \tag{4.89}$$

Example 4.9: Quadratic Transformation.

In the particular case where we transform $u = x^2$ we have two solutions

$$x_1 = u, \quad x_2 = -u \tag{4.90}$$

and the derivative of u is $2x$. Therefore

$$P(u) = \frac{1}{2u}\left[P_x(-u) + P_x(u)\right] \qquad 0 \le u \le \infty \tag{4.91}$$

Example 4.10: Scaled and Translated Densities.

Suppose we wish to scale a variable and also translate it. We take $u = ax + b$. The characteristic function of u is

$$M_u(\theta) = \langle e^{j\theta(ax+b)} \rangle = \iint e^{j\theta(ax+b)} P(x)\,dx = e^{j\theta b} M(a\theta) \tag{4.92}$$

where $M(a\theta)$ is the characteristic function of x. The distribution for u is therefore

$$P(u) = \frac{1}{2\pi}\int e^{-j\theta u} M_u(\theta)\,d\theta = \frac{1}{a}P_x\left(\frac{u-b}{a}\right) \tag{4.93}$$

Alternatively, we can use Eq. (4.86). Solving for the zero we have $x = (u-b)/a$ and hence

$$P(u) = \frac{1}{a}P(x)\Big|_{x=(u-b)/a} = \frac{1}{a}P_x\left(\frac{u-b}{a}\right) \tag{4.94}$$

Two Dimensional to One. Suppose we have a two dimensional density $P(x,y)$ and wish to find the density of a single variable which is a function of x and y,

$$u = f(x,y) \tag{4.95}$$

The characteristic function for u is

$$M_u(\theta) = \langle e^{j\theta f(x,y)} \rangle = \iint e^{j\theta f(x,y)} P(x,y)\, dx\, dy \tag{4.96}$$

Taking the inverse transform leads directly to

$$P(u) = \iint \delta(u - f(x,y))\, P(x,y)\, dx\, dy \tag{4.97}$$

Example 4.11: Sum of Two Variables.

Consider the case where

$$u = x + y \tag{4.98}$$

The density u is then

$$P(u) = \iint \delta(u - (x+y))\, P(x,y)\, dx\, dy \tag{4.99}$$

$$= \int P(x, u-x)\, dx \tag{4.100}$$

Two Dimensional to Two Dimensional. In the most general case we start with a density of two variables and wish to find the density of two new variables functionally related by

$$u = f(x,y) \quad ; \quad v = g(x,y) \tag{4.101}$$

Using the characteristic function approach we have, for the characteristic function for the variable u and v,

$$M_{uv}(\theta,\tau) = \iint e^{j\theta f(x,y) + j\tau g(x,y)}\, P(x,y)\, dx\, dy \tag{4.102}$$

The density for u, v is then

$$P(u,v) = \frac{1}{4\pi^2} \iint M_{uv}(\theta,\tau)\, e^{-j\theta u - j\tau v}\, d\theta\, d\tau \tag{4.103}$$

$$= \frac{1}{4\pi^2} \iiiint e^{j\theta[u - f(x,y)] + j\tau[v - g(x,y)]}\, P(x,y)\, dx\, dy\, d\theta\, d\tau \tag{4.104}$$

The θ and τ integrations give delta functions and hence

$$P(u,v) = \iint \delta(u - y(x,y))\, \delta(v - g(x,y))\, P(x,y)\, dx\, dy \tag{4.105}$$

4.8 NEGATIVE DENSITIES

Densities by definition are positive. However, it is of interest to examine which ideas and results regarding positive densities still hold true if we relax the positivity requirement. The reason we concern ourselves with this issue is that some of the densities we will be dealing with will not be manifestly positive. Perhaps they should not be called densities but nonetheless they will be treated and manipulated as densities. As specific cases arise we will examine the consequences of not having a manifestly positive density, but we make some general remarks now.

Most of the concepts of the characteristic function and moments present no difficulty. Specifically, the necessary requirements for a function to be a characteristic function obtained in Section 4.3 remain valid. However, they will not have the physical interpretation commonly associated with densities. For example the standard deviation may turn out to be negative. In addition, the correlation coefficient may not be bound by $(-1, 1)$.

The results that do not go through are those pertaining to conditions on the characteristic function to assure that it came from a manifestly positive density. The only time we have addressed that issue thus far is in Khinchin's theorem in Section 4.3. Hence, questions that address the positivity issue must be reexamined, and we will do so as the occasions arise.

Chapter 5

The Need for Time-Frequency Analysis

5.1 INTRODUCTION

The aim of this chapter is to show, by way of examples, the need for a combined time-frequency representation and to clarify why time analysis and frequency analysis by themselves do not fully describe the nature of signals. We also describe some of the physical reasons why spectra change in time.

Suppose we have the individual densities of height and weight of a particular type of animal. These individual densities tell us everything we want to know about the distribution of height and the distribution of weight. From these distributions can one determine how height and weight are related? Can one determine whether the tall individuals are the heavy ones? The answer is no. Can one determine whether an individual with weight of 250 lb. is exceptional? The answer is yes. But can one determine whether he is exceptional for his height? The answer is no. The individual densities of height and weight are not a full description of the situation because from these densities we cannot ascertain how height and weight are related. What is needed is the joint density of height and weight. In the same way, the time energy density and the frequency energy density are not sufficient to describe the physical situation because they do not *fully* describe what is happening. In particular, from the spectrum we know which frequencies were present in the signal but we do not know when those frequencies existed, hence the need to describe how the spectral content is changing in time and to develop the physical and mathematical ideas to understand what a time-varying spectrum is. We wish to devise a distribution that represents the energy or intensity of a signal simultaneously in time and frequency.

Time varying spectra are common in ordinary life. During a sunset, the fre-

quency composition of the light changes quickly and dramatically. In saying that the sky is getting redder we are conveying a time-frequency description because we are describing how the frequencies are changing in time. The pitch, which is the common word for frequency of human speech, changes as we speak and produces the richness of language. Similarly the pitch of animal sounds changes during vocalization. Standard musical notation is a time-frequency representation since it shows the player what notes, or frequency, should be played as time progresses.

In this chapter we give a number of simple analytic and real examples of time varying spectra so that we may become familiar with the need for a time-frequency description, the variety of applications, and the language involved. We leave aside the methods used to construct the joint time-frequency representations. That will be the subject of the subsequent chapters. The reader should think of these representations as no different from a joint density of height and weight of a population that gives the relative concentration of people at particular heights and weights. Similarly we should think of a time-frequency distribution as telling us the intensity or energy concentration at particular times and frequencies.

We first take a real example to illustrate the basic idea before embarking on a more systematic discussion. Using a whale sound, Fig. 5.1 shows three plots. Running up the page is the sound (the air pressure) as a function of time. By examining it visually we cannot tell much, although we can clearly tell how the intensity or loudness varies with time. Below the main figure is the energy density spectrum, that is, the absolute square of the Fourier transform. It indicates which frequencies existed and what their relative strengths were. For this sound the spectrum tells us that the frequencies ranged from about 175 to about 325 cycles per second. This information is interesting and important, but does not fully describe what happened, because from the spectrum we cannot know when these frequencies existed. For example, we cannot know just by looking at the spectrum when the 300 Hz sound was made, or whether it was made for the total duration of the sound or just at certain times. The main figure is a time versus frequency plot, that is, a joint time-frequency distribution. From it we can determine the frequencies and their relative intensities as time progresses. It allows us to understand what is going on: At the start the frequency was about 175 Hz and increased more or less linearly to about 325 Hz in about half a second, stayed there for about a tenth of a second, and so forth. In answer to the question of when the 300 Hz sound occurred, we can now see the answer. It occurred twice, at 0.6 and 1.3 seconds.

The difference between the spectrum and a joint time-frequency representation is that the spectrum allows us to determine which frequencies existed, but a combined time-frequency analysis allows us to determine which frequencies existed at a particular time.

5.2 SIMPLE ANALYTIC EXAMPLES

We now examine some simple analytic examples which will progressively develop the main ideas.

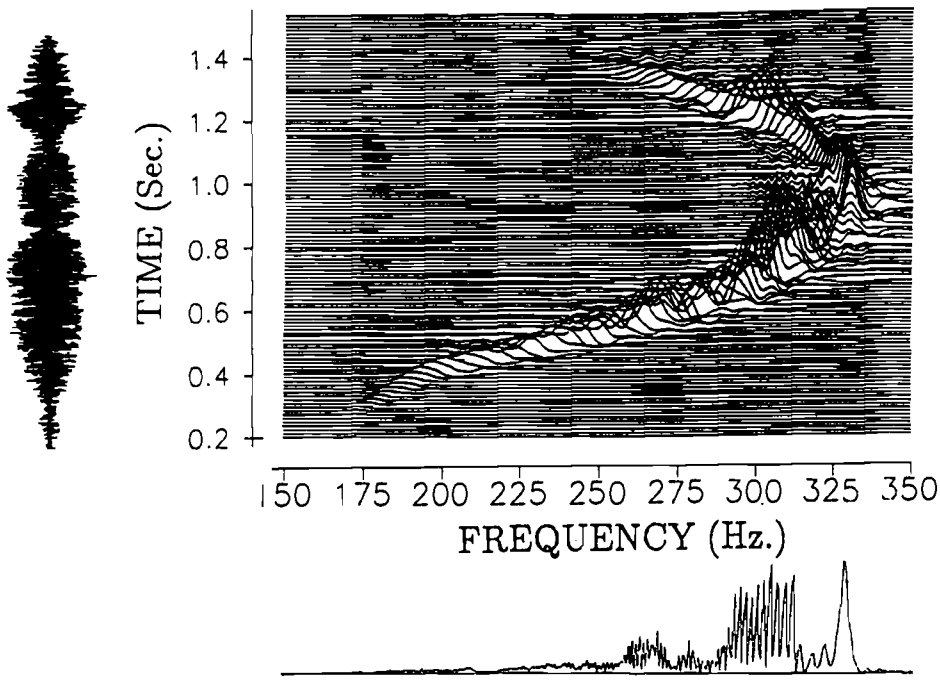

Fig. 5.1 A time-frequency plot of a sound made by a Bowhead whale. The wave form is on the left plot with time increasing upwards. The energy density spectrum is below the main figure and indicates which frequencies existed for the duration. The main figure is a joint time-frequency plot and shows how the frequencies change with time. (From the work of C. Rosenthal and L. Cohen.)

Finite Duration Sine Waves. In Fig. 5.2 we show examples of signals composed of finite duration sine waves at three frequencies. All three examples have basically the same spectrum, which is reasonable since for all cases we had three frequencies. The cases are different in regard to when the frequencies existed and this can be easily ascertained from the time-frequency plots.

Chirp. Consider the signal

$$s(t) = \frac{1}{\sqrt{T}} e^{j\beta t^2/2 + j\omega_0 t} \qquad 0 \le t \le T \tag{5.1}$$

From our discussion in Chapter 2 the instantaneous frequency is $\omega_0 + \beta t$. Here it ranges from ω_0 to $\omega_0 + \beta T$. In Fig. 5.3 (a) we have plotted the energy spectrum and the time-frequency plot. The spectrum is basically flat, telling us that these frequencies existed with equal intensity. The time-frequency plot tells us precisely when they existed.

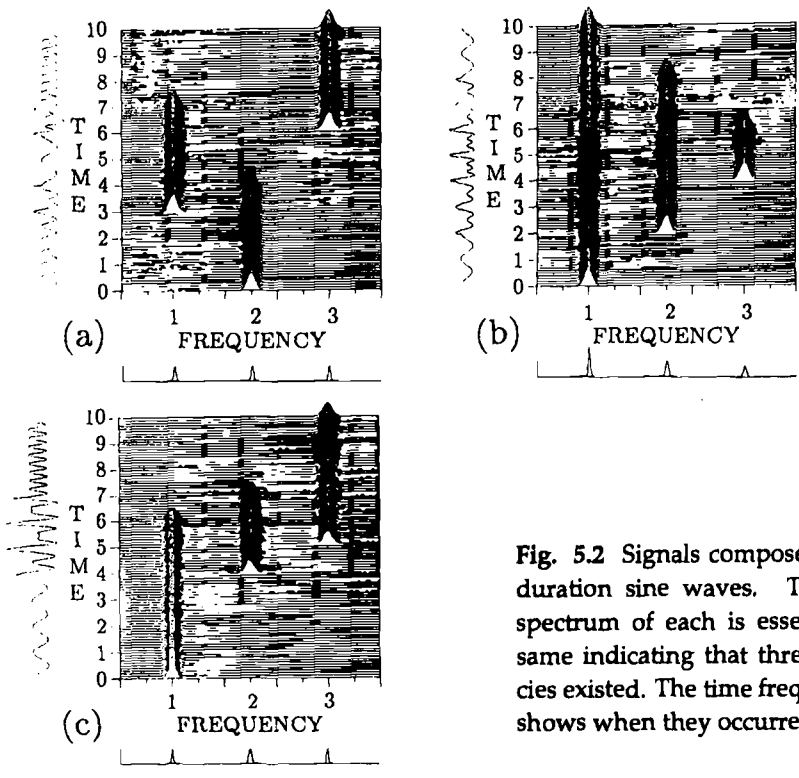

Fig. 5.2 Signals composed of finite duration sine waves. The energy spectrum of each is essentially the same indicating that three frequencies existed. The time frequency plot shows when they occurred.

Sinusoidal Modulation. Take

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + jm \sin \omega_m t + j\omega_0 t} \tag{5.2}$$

where m, ω_m, and ω_0 are constants. The spectrum is given by

$$S(\omega) = \sqrt{4\pi}\,(\pi/\alpha)^{1/4} \sum_{n=-\infty}^{\infty} J_n(\beta)\, e^{-(\omega - n\omega_m - \omega_0)^2/2\alpha} \tag{5.3}$$

where J_n is the nth-order Bessel function. Fig. 5.3(b) plots the signal energy density spectrum and the time-frequency plot. There is a considerable difference between what we learn from the two.

Other Examples. The cases of a cubic phase and a log phase are shown in Figs. 5.3 (c) and (d).

Multicomponent Signals. One of the significant advantages of time-frequency analysis is that it allows us to determine whether a signal is multicomponent or not. A multicomponent signal is one that has well delineated regions in the time-frequency plane. A few examples are illustrated in Fig. 5.4. Multicomponent signals are common in nature as will be seen from some of the examples shown subsequently.

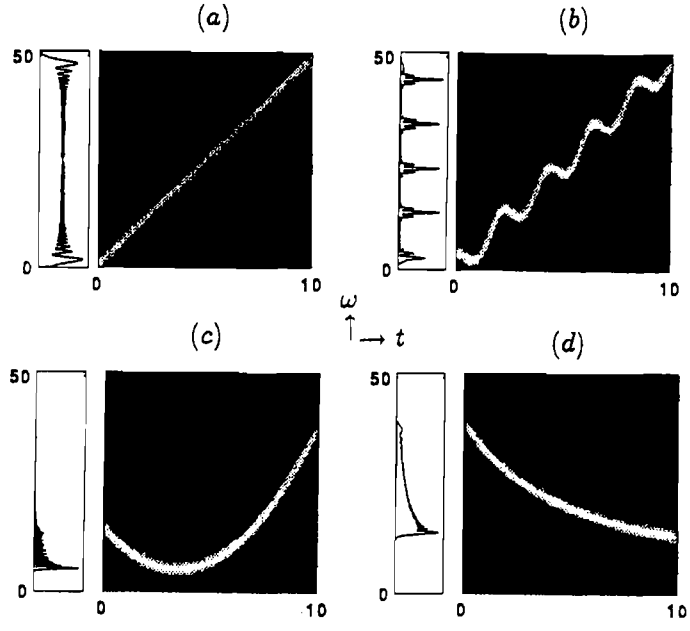

Fig. 5.3 The time-frequency plots for various signals. In (a) we have a chirp, $s(t) = e^{j\beta t^2/2+j\omega_0 t}$. The instantaneous frequency, $\omega_i = \omega_0 + \beta t$, is increasing linearly. The energy density spectrum is more or less flat, indicating that each frequency existed, but gives no indication of when it existed. The time-frequency plot does. In (b) we have a sinusoidal modulation, $e^{j\beta t^2/2+jm\sin\omega_m t+j\omega_0 t}$, and the energy spectrum shows a concentration at certain frequencies but gives no indication of the times. The reason that there is concentration at those frequencies is that those frequencies last longer, as can be clearly seen from the time-frequency plot. In (c) we have a signal with a cubic phase, $s(t) = e^{j\gamma t^3/3+\beta t^2/2+j\omega_0 t}$. In (d) we have a hyperbolic signal, $s(t) = e^{-j\ln(t-t_0)}$, which gives an instantaneous frequency of $\omega_i = 1/(t-t_0)$, which is a hyperbola in the time-frequency plane.

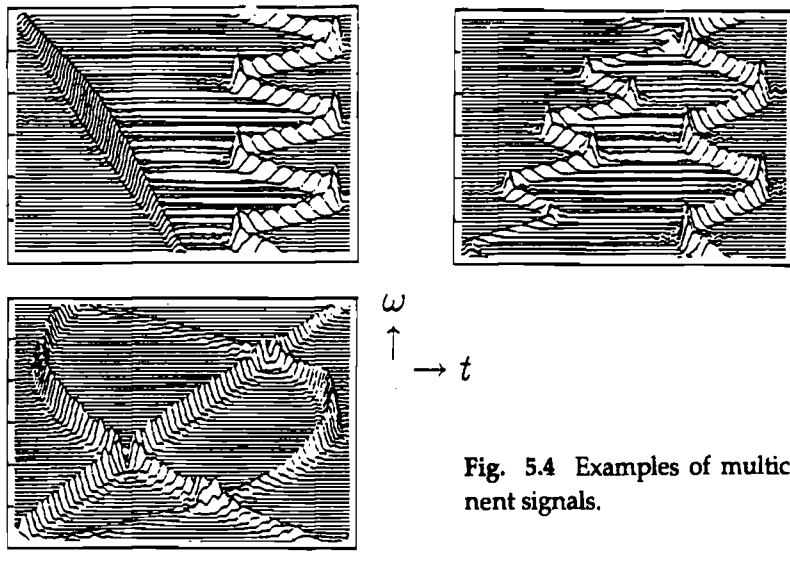

Fig. 5.4 Examples of multicomponent signals.

5.3 REAL SIGNALS

We now give some examples of real signals.

Human speech. The analysis of human speech was the main reason for the practical development in the 1940s of time-frequency analysis. The main method was and still is the short-time Fourier transform, which we will study in Chapter 7. This approach gave dramatic new understanding of the production and classification of speech. An example is shown in Fig. 5.5.

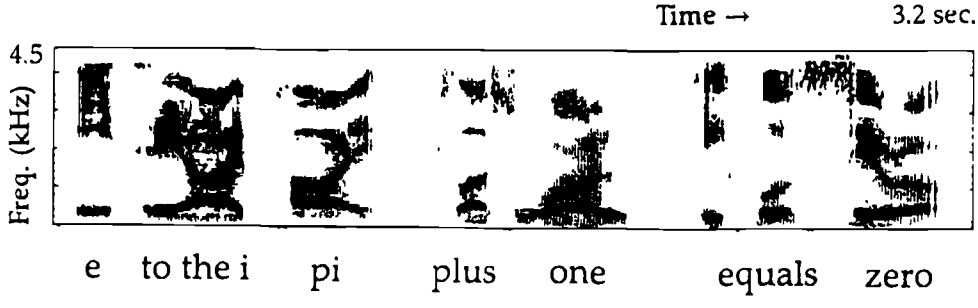

Fig. 5.5 Time-freqency plot of a speech signal. The utterance is: $e^{i\pi} + 1 = 0$

Whale Sounds. In Fig. 5.6 we show a number of whale sounds.

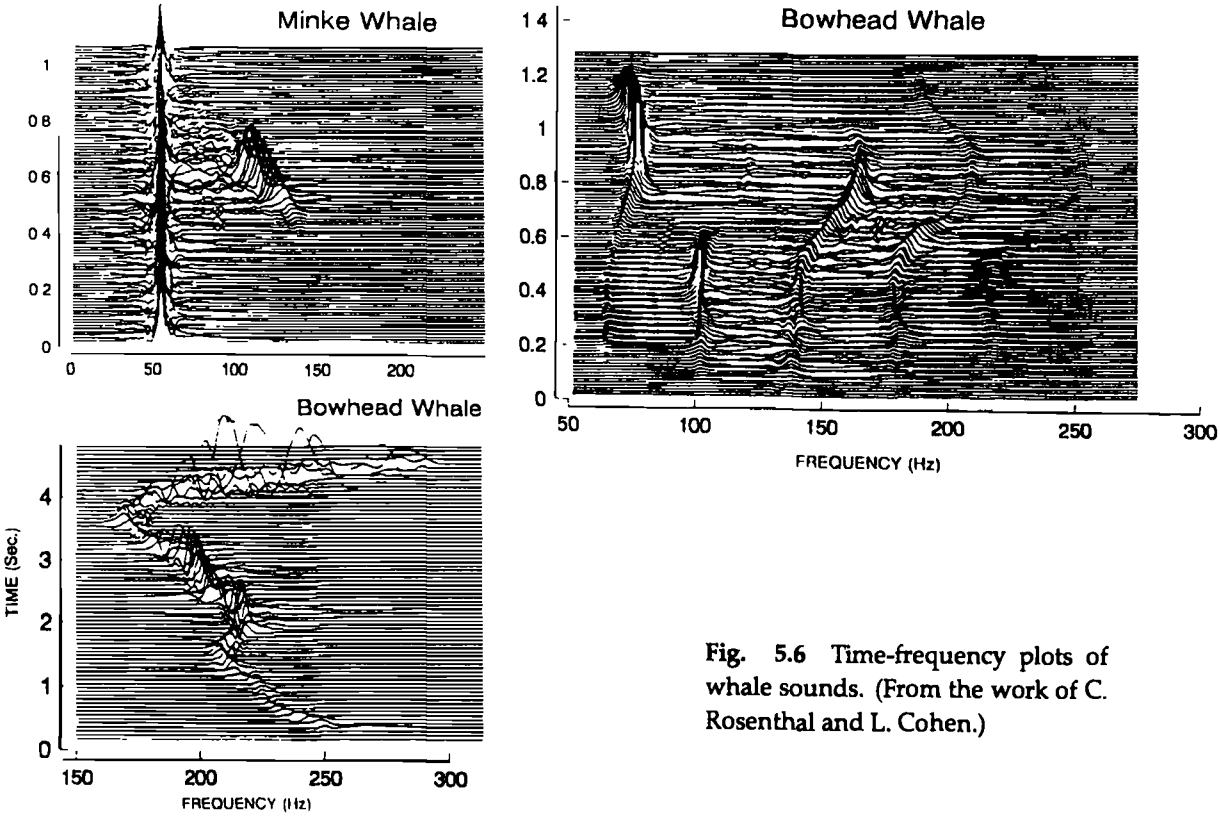

Fig. 5.6 Time-frequency plots of whale sounds. (From the work of C. Rosenthal and L. Cohen.)

Propeller Sounds. When a propeller or fan turns it pushes the air at a repetitive rate depending on the rotation. If the propeller has, for example, three symmetrical blades, then the frequency of repetition will be three times the number of revolutions per unit time. For a constant revolution rate this will show up as an acoustic pressure wave of constant frequency. However, if the ship decreases or increases its speed by changing the revolution rate this will result in changing frequencies.

Inelastic Acoustic Scattering. When a rigid object is hit with a sound wave, the wave gets reflected, producing an echo. The characteristics of the reflected wave depend on the shape of the object, its surface properties, the medium of propagation, and location. This reflection is called specular reflection. However, if the object is deformable, then the initial wave that hits the object produces waves in the object no differently than if the object had been hit with a hammer. The waves travel within the object and get reradiated. Hence for an elastic object hit by a wave not only do we get the usual echo, but we get additional waves due to the elastic response. Fig. 5.7 gives an example of a toy submarine which is a thin metallic shell with internal supporting structures. The submerged submarine was irradiated with an acoustic wave (sonar). The first part of the returned signal is the specular reflections. The rest of the signal is due to the reradiated acoustic elastic response. The advantage of a time-frequency plot is that we can immediately see at what times the reradiation occurred and what frequencies were radiated out.

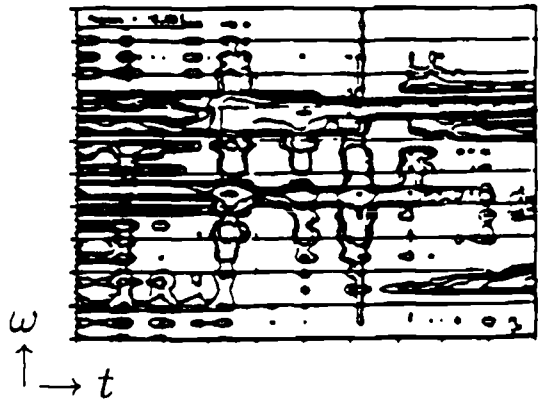

Fig. 5.7 A time-frequency plot of the pressure wave at a particular point in space coming from a model submarine after it was irradiated by a sonar pulse. The first return is the specular reflection (not shown) and the other returns are due to the reradiation of waves induced in the structure by the impeding wave. (From the work of C. Rosenthal and L. Cohen.)

Windshield Wipers. The sounds that windshield wipers make are often annoying. Fig. 5.8 shows a time-frequency plot of a good blade and an annoying one. In the annoying case a web-like structure is seen at higher frequencies. The frequency band appearing in both is due to the motor sound.

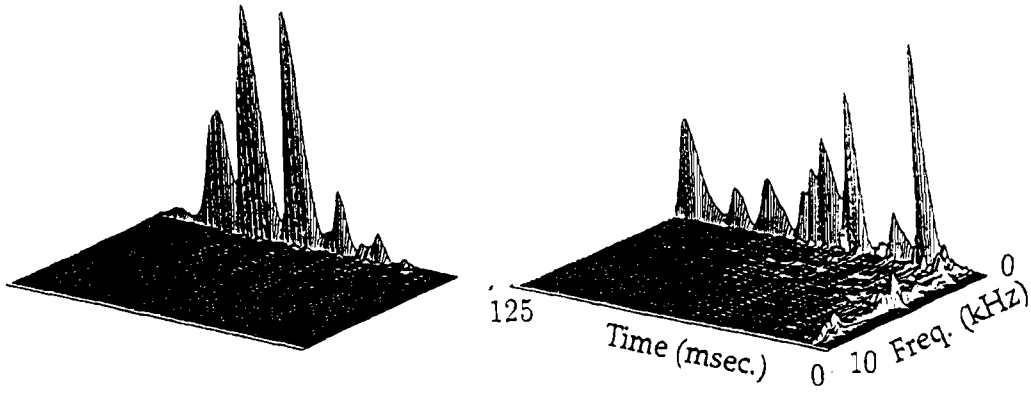

Fig. 5.8 Time frequency plots of windshield blades of two different models of cars. The first (a) is judged to be annoying, the second (b) is not. (Courtesy of J. Feng, Ford Motor Co.)

Car Door Slams. The slamming of a car door has historically been a major concern for many reasons, Madison Avenue notwithstanding. We have all experienced the rich sounding door with its solid thump and the tinny door that does not engender room for confidence in the rest of the construction. The sound produced by a slamming car door is quite involved because there are a number of diverse events occurring including the coupling with the latch, the hinging, and the seals around the frame. Fig. 5.9 shows a time-frequency plot for a good sounding car door (a) and a tinny one (b). Note that in the poor sounding door there is considerable acoustic energy at a wide spread in frequencies immediately after the contact of door with latch. It has been found that this acoustic energy, which lasts for about a twentieth of a second, is the main reason people say a door sounds tinny and hence poor. The main frequency concentration is the vibration of the automobile as a whole caused by the slamming of the door.

Fig. 5.9 Time-frequency plots of two car door slams. The one in (a) is characterized by a tinny sound and the one in (b) by a rich sound. The tinny sound is caused by the generation of a wide spectrum of frequencies in the initial contact of the door with the latch. (Courtesy of J. Feng, Ford Motor Co.)

Fault Analysis. It is important to have methods for the earliest possible detection of when a machine is starting to go bad. This is particularly important for machines such as in airplanes and ships whose failure may lead to critical situations. It is also critical in manufacturing since a bad manufacturing machine may produce defective products. We illustrate such situations in Figs. 5.10 and 5.11.

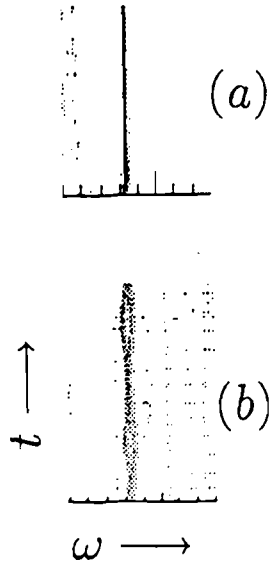

Fig. 5.10 A time-frequency plot of a region near a spectral line of a machine that is about to go bad (a) and one that has gone bad (b). (Courtesy of R. A. Rohrbaugh.)

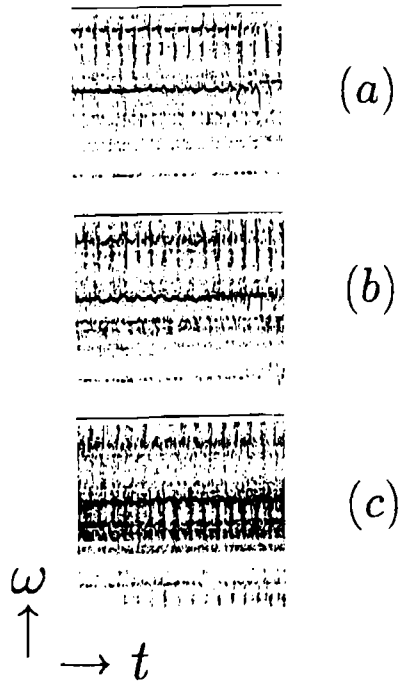

Fig. 5.11 Drilling of certain metal parts often requires the resulting hole to be of high precision. Therefore it is of importance to have an effective means for the determination as to when a drill is beginning to get dull. In (a) we a have sharp drill, in (b) the drill is somewhat dull and in (c) the drill is very dull. The data is obtained from an accelerometer on the spindle. (Courtesy of P. Loughlin, J. Pitton, L. Atlas and G. Bernard. Data courtesy of Boeing Commerical Airplane Group.)

Trumpet. In Fig. 5.12 we show a time-frequency plot of the first half second of a trumpeter initiating a note. The ridges in the frequency direction are the harmonics of the fundamental. To achieve the steady state desired the trumpeter has to adjust in a split second. The frequency of each harmonic gets adjusted in the same way, that is, the percentage change is the same. While the absolute change is insignificant for the lower frequencies, it is significant at the higher frequencies.

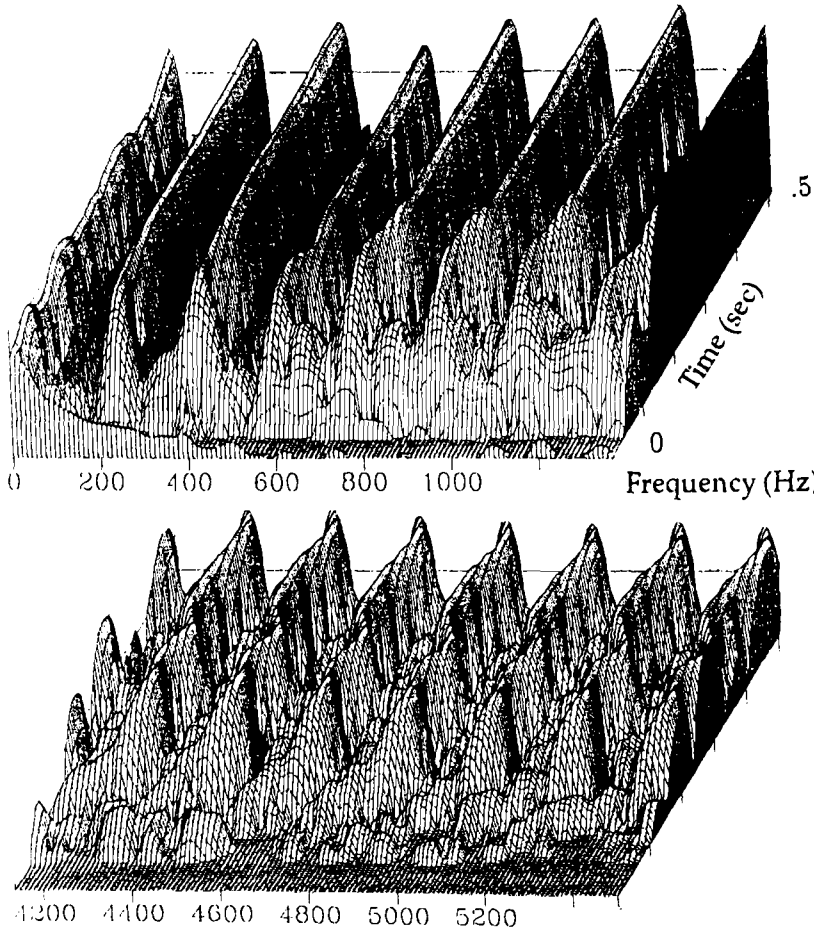

Fig. 5.12 Time-frequency plots of a trumpet sound at its initiation. (Courtesy of W. J. Pielemeier and G. H. Wakefield.)

Bottlenose Dolphin. Fig. 5.13 shows the time-frequency plot of a dolphin sound.

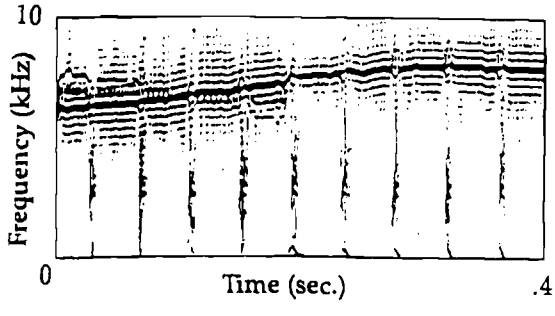

Fig. 5.13 Bottlenose Dolphin. The time-frequency plot shows that the sound consists of linearly increasing frequency and periodic clicks. (Courtesy of Williams and Tyack.)

Heart Sound. While the metaphysical importance of the heart has been recognized since ancient times, the notion that the heart is responsible for the heart sound was not accepted until the 1700s. There was considerable debate as to whether the heart produces a sound at all. Indeed, the invention of signal analysis can be traced to Hook[1] :"I have been able to hear very plainly the beating of a Man's heart... Who knows , I say, but that it may be possible to discover the Motions of the Internal Parts of Bodies... by the sound they make, that one may discover the Works performed in the several Offices and Shops of a Man's Body, and thereby discover what Instruments or Engine is out of order".[608] Fig. 12.4 shows a time-frequency plot of a heart sound.

5.4 WHY SPECTRA CHANGE

There are many causes for time-varying spectra, but two broad physical mechanisms encompass most situations. The primary reason is that the production of particular frequencies depends on physical parameters that may change in time. For example, a string of a fixed length and tension produces a particular frequency if disturbed. The mechanical oscillation of the string beats the air at the mechanical vibration rate. If the length or tension changes with time, different frequencies will be produced in time because the string will vibrate with different frequencies and beat the air accordingly. When violinists play, for example, they continually change the length of strings to produce different vibrations of the strings, which in turn produce different vibrations in the air. If a ship is going at constant speed by turning the propeller at a constant speed, then the sound of the propeller is at a constant frequency, namely the number of times it turns per minute times the number of blades. If the ship accelerates, then the propeller will change the revolution rate, beating the water at a changing rate. That shows up as increasing frequency. Related to this question of changing the properties of the parameters that are causing the vibration is the possible changing of the immediate physical surroundings. Suppose we produce sound waves at the end of a pipe. If the length of the pipe is constant then the frequency spectrum of the output will be constant in time. But if the shape and length of the pipe change with time, then the output will also vary with time. That is the case with human speech. As we speak we are continually changing the physical shape of our tongue, mouth, nose, etc.

The other broad reason for changing spectra is that the propagation of waves in a medium is generally frequency dependent. That is why we can see through glass but not through wood, while if our eyes where sensitive to X-rays we would be able to see through wood but not so well through glass. The propagation of waves is governed by a wave equation. If the velocity of propagation, v, does not

[1] In textbooks Robert Hook is remembered for Hook's law. However, he was one of the greatest scientists of his century Hook discovered that plants were made up of cells, the red spot of Jupiter, the wave nature of light, and Boyle's law of gases (Boyle published this law in a book on the subject and gave full credit to Hook, but Boyle's name stuck), among many other discoveries. He was about ten years older than Newton and their lives were intertwined both positively and negatively.

depend on frequency, for example, electromagnetic waves in a vacuum, then the disturbance obeys $v^2\,\partial^2 u(x,t)/\partial x^2 = \partial^2 u(x,t)/\partial t^2$, where u is the physical quantity that is changing (pressure, electric field, etc.), and x, t are the position and time. The unique feature of this wave equation is that if we start with a signal at $t = 0$ given by $u(x,0)$, then at a later time the wave will be the same function but displaced according to $u(x - vt, 0)$. The shape remains the same and so does the frequency content. That is why a person 5 feet away sees and hears pretty much the same as a person 50 feet away from the source. It is because light and sound propagate in air without any substantial change in the frequencies that our ears and eyes can detect.

However, it is generally the case that the propagation of waves in a medium is a frequency dependent phenomenon. This is reflected in the wave equation by additional terms that do not admit solutions of the form $u(x - vt)$. The functional form changes in time and we have distortion or filtering. This phenomenon, that waves of different frequencies propagate with different velocities, is called dispersion. The reason for the name is that a prism "disperses" light and that was the earliest discovered manifestation of the effect. The velocity may decrease or increase with frequency, depending on the materials, and the two situations are described by the phrase "normal" and "anomalous" dispersion, the former being the most common. In addition, absorption, dying out, and attenuation, all meaning the same thing, are also generally frequency dependent. In normal conditions here on earth there is almost no attenuation of sound waves at the frequencies we hear and that is why we can hear from far away. On the other hand, light gets attenuated when passing though water and the attenuation is dependent on frequency. Similarly, high frequency electromagnetic waves are damped within a short distance when entering the surface of a conductor. Also, as a wave propagates from one medium to another, part of it gets reflected and part of it is transmitted, which is generally frequency dependent.

The above situations are generally described by "filtering", which usually means filtering of frequencies. Thus glass allows the visible portion of the spectrum to go through but is a pretty good filter of the X-rays. Paper, on the other hand, filters visible light out but lets through X-rays. The sun's maximum output is at frequencies we call the visible spectrum, which may seem to be the reason why evolution gave us the ability to see at those frequencies. But what is important for evolution is not what the sun produces, but what gets through our atmosphere. It is a remarkable coincidence that our atmosphere lets through the visible part. If we lived on a planet whose atmosphere filtered everything but the X-rays, we might have developed X-ray vision, assuming that a good fraction of the light coming from the star was in the X-ray region.

Chapter 6

Time-Frequency Distributions: Fundamental Ideas

6.1 INTRODUCTION

The basic objective of time-frequency analysis is to devise a function that will describe the energy density of a signal simultaneously in time and frequency, and that can be used and manipulated in the same manner as any density. If we had such a distribution we could ask for the fraction of the energy in a certain frequency and time range, we could calculate the density of frequency at a particular time, we could calculate the global and local moments of the distribution such as the mean conditional frequency and its local spread, and so on. We now begin our study of how to construct such distributions and in this chapter we describe the main ideas. To crystalize our aim we recall that the instantaneous power or intensity in time is

$$|s(t)|^2 = \text{intensity per unit time at time } t, \quad \text{or}$$

$$|s(t)|^2 \Delta t = \text{the fractional energy in the time interval } \Delta t \text{ at time } t$$

and the density in frequency, the energy density spectrum, is

$$|S(\omega)|^2 = \text{intensity per unit frequency at } \omega, \quad \text{or}$$

$$|S(\omega)|^2 \Delta\omega = \text{the fractional energy in the frequency interval } \Delta\omega \text{ at frequency } \omega$$

What we seek is a joint density, $P(t,\omega)$, so that

$$P(t,\omega) = \text{the intensity at time } t \text{ and frequency } \omega, \quad \text{or}$$

$$P(t,\omega)\,\Delta t\,\Delta\omega = \text{the fractional energy in the time-frequency cell } \Delta t\,\Delta\omega \text{ at } t,\ \omega$$

Do there exist joint time-frequency distributions that satisfy our intuitive ideas of a time-varying spectrum? How can they be constructed? Can they be interpreted as true densities? Do they represent the correlations between time and frequency? What reasonable conditions can be imposed to obtain such densities? The hope is that they do exist, but if they don't in the full sense of true densities, what is the best we can do? Are there inherent limitations to such a development? This is the scope of time-frequency analysis.

Marginals. Summing up the energy distribution for all frequencies at a particular time should give the instantaneous energy, and summing up over all times at a particular frequency should give the energy density spectrum. Therefore, ideally, a joint density in time and frequency should satisfy

$$\int P(t,\omega)\,d\omega \;=\; |\,s(t)\,|^2 \tag{6.1}$$

$$\int P(t,\omega)\,dt \;=\; |\,S(\omega)\,|^2 \tag{6.2}$$

which are called the time and frequency marginal conditions.

Total Energy. The total energy of the distribution should be the total energy of the signal

$$E \;=\; \iint P(t,\omega)\,d\omega\,dt \;=\; \int |\,s(t)\,|^2\,dt \;=\; \int |\,S(\omega)\,|^2\,d\omega \tag{6.3}$$

Note that if the joint density satisfies the marginals, it automatically satisfies the total energy requirement, but the converse is not true. It is possible that a joint density can satisfy the total energy requirement without satisfying the marginals. The spectrogram that we study in the next section is one such example. The total energy requirement is a weak one and that is why many distributions that do not satisfy it may nonetheless give a good representation of the time-frequency structure.

Characteristic Functions. We have seen in Chapter 4 that characteristic functions are a powerful way to study distributions. The joint characteristic function of a time-frequency density is

$$M(\theta,\tau) \;=\; \langle\, e^{j\theta t + j\tau\omega}\,\rangle \;=\; \iint P(t,\omega)\,e^{j\theta t+j\tau\omega}\,dt\,d\omega \tag{6.4}$$

However, many of the distributions that we will be studying are not proper. By proper we mean a well-defined manifestly positive density function. Hence, for a particular distribution, the characteristic function may not satisfy all the standard attributes of characteristic functions that we discussed in Chapter 4.

6.2 GLOBAL AVERAGES

Global Averages. The average value of any function of time and frequency is to be calculated in the standard way

$$\langle g(t,\omega) \rangle = \iint g(t,\omega)\, P(t,\omega)\, d\omega\, dt \tag{6.5}$$

and of course the answers should be meaningful and reasonable. If the marginals are satisfied then we are guaranteed that averages of the form

$$\langle g_1(t) + g_2(\omega) \rangle = \iint \{g_1(t) + g_2(\omega)\}\, P(t,\omega)\, d\omega\, dt \tag{6.6}$$

$$= \int g_1(t)\, |\, s(t)\, |^2\, dt + \int g_2(\omega)\, |\, S(\omega)\, |^2\, d\omega \tag{6.7}$$

will be correctly calculated since the calculation requires only the satisfaction of the marginals.

But what about averages that are mixed, that is, average values of arbitrary time-frequency functions? An example is the covariance which involves the calculation of $\langle t\,\omega \rangle$. For mixed averages we do not know what we want and can only be guided by intuition and by the plausibility of possible guesses. That is not surprising. Mathematics cannot answer what is essentially an issue of science. Consider, for example, the question of height and weight of an unknown species and suppose we knew the distribution of height and weight, that is, the marginals. Can we know how height and weight are related, how the correlation coefficient behaves? No. Similarly from knowing the time-frequency marginals we cannot know how time and frequency should be related and mathematics cannot help since mathematics allows all possibilities. In Chapter 1 we discussed the covariance of a signal and showed how that is related to the first mixed moment $\langle t\,\omega \rangle$. We argued that a plausible guess to $\langle t\,\omega \rangle$ is the average of $t\,\varphi'(t)$ and we showed that this quantity behaves reasonably well in the sense that it meets our intuition for the cases considered. Therefore, a plausible requirement of a joint distribution is

$$\langle t\,\omega \rangle = \iint t\,\omega\, P(t,\omega)\, d\omega\, dt = \int t\,\varphi'(t)\, |\, s(t)\, |^2\, dt \tag{6.8}$$

However, for other mixed moments we have nothing to guide us.

6.3 LOCAL AVERAGES

Treating a joint time-frequency density as any other density, we can immediately argue, as was done in Chapter 4, that the density of frequency for a given time and the density of time for a given frequency are respectively given by

$$P(\omega\,|\,t) = \frac{P(t,\omega)}{P(t)} \quad ; \quad P(t\,|\,\omega) = \frac{P(t,\omega)}{P(\omega)} \tag{6.9}$$

where $P(t)$ and $P(\omega)$ are the marginal distributions

$$P(t) = \int P(t,\omega)\, d\omega \quad ; \quad P(\omega) = \int P(t,\omega)\, dt \tag{6.10}$$

Notice that we have used $P(t)$ and $P(\omega)$ rather than $|\, s(t)\,|^2$ and $|\, S(\omega)\,|^2$ to allow for the possibility that the marginals are not satisfied, as will be the case with the spectrogram discussed in the next chapter.

The conditional average value of a function at a given time or frequency is

$$\langle\, g(\omega)\,\rangle_t = \frac{1}{P(t)} \int g(\omega)\; P(t,\omega)\, d\omega \tag{6.11}$$

$$\langle\, g(t)\,\rangle_\omega = \frac{1}{P(\omega)} \int g(t)\; P(t,\omega)\, dt \tag{6.12}$$

In Chapter 2 we saw that a good candidate for the average frequency at a given time is the derivative of the phase, the instantaneous frequency. Similarly, we argued that the average time for a given frequency should be the derivative of the spectral phase. If we accept these results, then a joint time-frequency density should satisfy

$$\langle\, \omega\,\rangle_t = \frac{1}{P(t)} \int \omega\, P(t,\omega)\, d\omega = \varphi'(t) \tag{6.13}$$

$$\langle\, t\,\rangle_\omega = \frac{1}{P(\omega)} \int t\, P(t,\omega)\, dt = -\psi'(\omega) \tag{6.14}$$

From this point of view instantaneous frequency is an average, the average frequency at a particular time. We should also be able to define the spread of frequencies at a given time and the spread of time for a given frequency, that is, the conditional standard deviations. These are

$$\sigma^2_{\omega|t} = \frac{1}{P(t)} \int (\,\omega - \langle\,\omega\,\rangle\,)^2\; P(t,\omega)\, d\omega \tag{6.15}$$

$$\sigma^2_{t|\omega} = \frac{1}{P(\omega)} \int (\,t - \langle\, t\,\rangle\,)^2\; P(t,\omega)\, dt \tag{6.16}$$

What should these quantities be? These quantities will play a crucial role in the development of the basic theory, so we hope they turn out to be sensible. For some signals we know what the answer should be. For example, for a pure sinusoid we should get $\sigma_{\omega|t} = 0$, since a pure sinusoid has one definite frequency for all time.

6.4 TIME AND FREQUENCY SHIFT INVARIANCE

Suppose we have a signal $s(t)$ and another signal that is identical to it but translated in time by t_0. We want the distribution corresponding to each signal to be identical in form, but that the one corresponding to the time shifted signal be translated by t_0. That is

$$\text{if} \quad s(t) \rightarrow s(t-t_0) \quad \text{then} \quad P(t,\omega) \rightarrow P(t-t_0,\omega) \tag{6.17}$$

Similarly, if we shift the spectrum by a constant frequency we expect the distribution to be shifted by that frequency,

$$\text{if} \quad S(\omega) \rightarrow S(\omega-\omega_0) \qquad \text{then} \quad P(t,\omega) \rightarrow P(t,\omega-\omega_0) \tag{6.18}$$

Both of these cases can be handled together. If $s(t)$ is the signal, then a signal that is translated in time by t_0 and translated in frequency by ω_0 is given by $e^{j\omega_0 t}s(t-t_0)$. Accordingly, we expect the distribution to be shifted in time and frequency in the same way,

$$\text{if} \quad s(t) \rightarrow e^{j\omega_0 t}s(t-t_0) \quad \text{then} \quad P(t,\omega) \rightarrow P(t-t_0,\omega-\omega_0) \tag{6.19}$$

6.5 LINEAR SCALING

For a signal $s(t)$, the signal given by $s_{sc}(t) = \sqrt{a}s(at)$ is a scaled version of $s(t)$. The new signal is blown up or reduced, depending on whether a is less or greater than one. The square root factor keeps the normalization the same as the original signal. The spectrum of the scaled signal is

$$S_{sc}(\omega) = \frac{1}{\sqrt{a}}\,S(\omega/a) \quad \text{if} \quad s_{sc}(t) = \sqrt{a}\,s(at) \tag{6.20}$$

We see that if the signal is compressed then the spectrum is expanded, and conversely. If we want these relations to hold for the joint distribution, then we must have

$$P_{sc}(t,\omega) = P(at,\omega/a) \tag{6.21}$$

The scaled distribution satisfies the marginals of the scaled signal

$$\int P_{sc}(t,\omega)\,d\omega = |\,s_{sc}(t)\,|^2 = a\,|\,s(at)\,|^2 \tag{6.22}$$

$$\int P_{sc}(t,\omega)\,dt = |\,S_{sc}(\omega)\,|^2 = \frac{1}{a}\,|\,S(\omega/a)\,|^2 \tag{6.23}$$

6.6 WEAK AND STRONG FINITE SUPPORT

Suppose a signal doesn't start until t_1. We want the joint distribution also to not start until t_1. Similarly, if the signal stops after time t_2 we expect the distribution to be zero after that time. If that is the case we say the distribution has weak finite time support. The reason for the word weak will be apparent shortly. Similarly, if

the spectrum is zero outside a frequency band, then the distribution should also be zero outside the band. In such a case we say that the distribution has weak finite spectral support. We can express these requirements mathematically as

$$P(t,\omega) = 0 \quad \text{for } t \text{ outside } (t_1,t_2) \qquad \text{if } s(t) \text{ is zero outside } (t_1,t_2) \tag{6.24}$$

$$P(t,\omega) = 0 \quad \text{for } \omega \text{ outside } (\omega_1,\omega_2) \qquad \text{if } S(\omega) \text{ is zero outside } (\omega_1,\omega_2) \tag{6.25}$$

Now suppose we have a signal that stops for a half hour and then starts again. We would expect the distribution to be zero for that half hour. Similarly, if we have a gap in the spectrum, then we expect the distribution to be zero in that gap. If a distribution satisfies these requirements, namely that it is zero whenever the signal is zero or is zero whenever the spectrum is zero, then we say the distribution has strong finite support:

$$P(t,\omega)=0 \qquad \text{if } s(t) = 0 \text{ for a particular time} \tag{6.26}$$

$$P(t,\omega)=0 \qquad \text{if } S(\omega) = 0 \text{ for a particular frequency} \tag{6.27}$$

Strong finite support implies weak finite support, but not conversely.

Distributions Concentrated in a Finite Region. A signal cannot be both of finite duration and bandlimited in frequency. Therefore if a distribution satisfies the weak finite support property it cannot be limited to a finite region of the time-frequency plane. If it were, it would be both time and frequency limited, which is impossible. If it turns out that a distribution is limited in a finite region, then it does not satisfy the finite support properties and/or the marginals.

6.7 UNCERTAINTY PRINCIPLE

In Chapter 3 we emphasized that the uncertainty principle depends on only three statements. First and second are that the time and frequency standard deviations are calculated using $|\, s(t)\,|^2$ and $|\, S(\omega)\,|^2$ as the respective densities,

$$T^2 = \int (\, t - \langle t \rangle\,)^2 \,|\, s(t)\,|^2 \, dt \tag{6.28}$$

$$B^2 = \int (\, \omega - \langle \omega \rangle\,)^2 \,|\, S(\omega)\,|^2 \, d\omega \tag{6.29}$$

and the third is that $s(t)$ and $S(\omega)$ are Fourier transform pairs. From a joint distribution the standard deviations are obtained by

$$\sigma_t^2 = \iint (\, t - \langle t \rangle\,)^2 \, P(t,\omega) \, dt \, d\omega = \int (\, t - \langle t \rangle\,)^2 \, P(t) \, dt \tag{6.30}$$

$$\sigma_\omega^2 = \iint (\, \omega - \langle \omega \rangle\,)^2 \, P(t,\omega) \, dt \, d\omega = \int (\, \omega - \langle \omega \rangle\,)^2 \, P(\omega) \, d\omega \tag{6.31}$$

When do we get the correct uncertainty principle? When the standard deviations calculated using the joint distribution give the same answer as when calculated by Eqs. (6.28) – (6.29). This will be the case when the marginals are correctly given,

$$P(t) = |s(t)|^2 \quad \text{and} \quad P(\omega) = |S(\omega)|^2 \qquad \text{for uncertainty principle} \tag{6.32}$$

Therefore, any joint distribution that yields the correct *marginals* will yield, and is totally consistent with, the uncertainty principle.

6.8 THE UNCERTAINTY PRINCIPLE AND JOINT DISTRIBUTIONS

We have emphasized that the uncertainty principle depends only on the time and frequency marginals, that is, on $|s(t)|^2$ and $|S(\omega)|^2$ being the density in time and frequency. Any joint distribution that has these marginals will satisfy the uncertainty principle. For any two marginals there are an infinite number of joint distributions, but the fact that the marginals are related imposes certain constraints on the possible joint distributions. For example, since the marginals are the absolute square of Fourier transform pairs and since we know that Fourier transform pairs cannot both be of finite extent, we cannot possibly have a joint distribution that is confined to a finite region in the time-frequency plane. We now examine how the uncertainty principle constrains the possible time-frequency distributions.

Let us crystalize the essence of the issue in the following way. Suppose that someone has arranged the world so that no matter what marginals the two variables x and y have, they are rigged so that the standard deviations of the marginals satisfy

$$\sigma_y \sigma_x \geq \eta \tag{6.33}$$

where η is a universal constant.[1]

This type of situation is easy to set up. So as not to get too abstract let us concentrate on the following example. Suppose x and y have the following densities

$$P(x) = \frac{1}{\sqrt{2\pi\eta/l}} \exp\left[-\frac{(x-a)^2}{2\eta/l}\right] \tag{6.34}$$

$$P(y) = \frac{1}{\sqrt{2\pi\eta(l^2+k^2)/l}} \exp\left[-\frac{(y-b)^2}{2\eta(l^2+k^2)/l}\right] \tag{6.35}$$

where l and k are any two positive numbers at our disposal. The universal constant η is fixed and not at our disposal.

These two marginals are perfectly good densities. But no matter what we do or try to do we have the uncertainty principle, $\sigma_y \sigma_x \geq \eta$. That is the case since

[1] The more general case is where η is a functional of the signal. In the case of time-frequency it is a constant equal to half. For time and scale, for example, it is functional of the signal. See Section 15.5. To keep things simple we take η as a constant here but our discussions apply generally

$$\sigma_x = \sqrt{\eta/l} \quad ; \quad \sigma_y = \sqrt{\eta(l^2+k^2)/l} \tag{6.36}$$

and therefore

$$\sigma_x\sigma_y = \eta\sqrt{1+l^2/k^2} \geq \eta \tag{6.37}$$

No matter how we choose l and k we will always have $\sigma_x\sigma_y \geq \eta$.

Related Marginals, Independence, and Joint Distributions. The reason we have an uncertainty principle is that the marginals are functionally related. In this case that appears by way of the parameters l and k, which occur in both distributions. Because the marginals are functionally related, if we change one marginal we change the other. However, one must not conclude from this that the variables are dependent, or that there is any correlation between them. Functional relationship between the marginals does not mean that the variables are dependent. That can be decided only by the joint distribution and not the marginals.

In Section 14.2 we show how to construct, for any two marginals, an infinite number of joint distributions. Here we want to examine what constraints marginals that satisfy the uncertainty principle put on the possible joint representations. Or, phrasing the question another way, what are the properties of joint distributions that produce marginals that satisfy the uncertainty principle? Let us first consider some examples of joint densities that satisfy the above marginals.

Example 1. Consider the following perfectly well behaved joint distribution which satisfies the marginal and hence the uncertainty principle,

$$P(x,y) = \frac{1}{2\pi\eta\sqrt{1+l^2/k^2}} \exp\left[-\frac{(x-a)^2}{2\eta/l} - \frac{(y-b)^2}{2\eta(l^2+k^2)/l}\right] = P(x)\,P(y) \tag{6.38}$$

This distribution is the independent joint distribution. Even though the marginals are related, the variables are not correlated. For any value of x the density of y is the same.

Example 2. Another joint density, which satisfies the same marginals and has correlations, is

$$P(x,y) = \frac{1}{2\pi\eta\sqrt{1+l^2/k^2}\sqrt{1-r^2}} \tag{6.39}$$

$$\exp\left[-\frac{1}{2(1-r^2)}\left\{\frac{(x-a)^2}{\eta/l} - 2\frac{r(x-a)(y-b)}{\eta\sqrt{1+l^2/k^2}} + \frac{(y-b)^2}{\eta(l^2+k^2)/l}\right\}\right] \tag{6.40}$$

where r is any number between -1 and 1. For this case there is correlation between the variables. The correlation is positive or negative, depending on the sign of r. There is a rich variety of possibilities, but the marginals are always satisfied

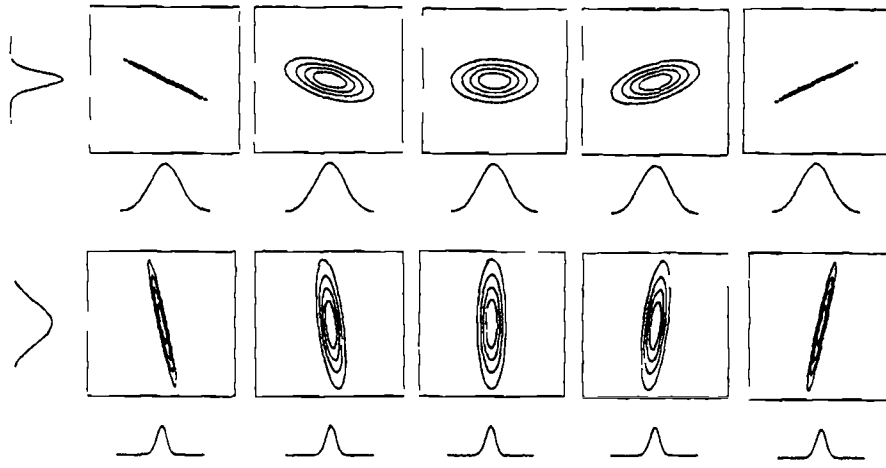

Fig. 6.1 The uncertainty principle depends only on the marginals. It constrains the type of joint distribution that is possible. A joint distribution cannot be narrow in both variables, but this still leaves an infinite number of possibilities. The marginals give no indication as to whether, or how, two variables may be correlated.

and so is the uncertainty principle. In particular, when one marginal is broad, the other is narrow. A proper joint distribution has no difficulty accommodating such marginals and although we have just demonstrated two of them, there are an infinite number of them.[127] Lest the reader think this is an artificial example, we point out that it is precisely the Wigner distribution for a chirp with a Gaussian envelope. Also, constructing other joint distributions is very easy and we do so in Section 14.2.

In Fig. 6.1 we plot the distribution for two different sets of marginals and different values of l and k. For a given set of marginals, one can always find an infinite number of joint distributions with various correlations between the variables. From the marginals, nothing can be concluded about the correlation between the variables.

6.9 UNCERTAINTY PRINCIPLE AND CONDITIONAL STANDARD DEVIATION

What general and specific statements can be made regarding joint distributions having marginals that satisfy the uncertainty principle? Basically, the only general property is that the joint distribution cannot be concentrated in both directions. If it was, then it would produce narrow marginals, which would be a violation, since both marginals cannot be narrow. The fact that both marginals cannot be narrow eliminates a number of possible joint distributions but still leaves an infinite number to choose from. We can fix this idea somewhat better by considering the conditional standard deviations of a joint distribution. First, let us recall that the uncertainty principle relates the global standard deviations. But we know from Section 4.6 that

there is a relationship between the global standard deviation and the conditional standard deviation and the conditional average. In particular, writing Eq. (4.3) for both x and y we have

$$\sigma_y^2 = \int \sigma_{y|x}^2 \, P(x)\, dx + \int (\langle y \rangle_x - \langle y \rangle)^2 \, P(x)\, dx \tag{6.41}$$

$$\sigma_x^2 = \int \sigma_{x|y}^2 \, P(y)\, dy + \int (\langle x \rangle_y - \langle x \rangle)^2 \, P(y)\, dy \tag{6.42}$$

Note that for each variable there are two ways to control the standard deviation and hence when we multiply σ_x^2 by σ_y^2 we will have four terms, the sum of which must be greater than η^2. We can satisfy the uncertainty principle by making any one of those terms greater than η^2 and have freedom with the rest of them. Or, in general, we can choose any value for each term as long as the sum adds up to greater than η^2. These four terms are a gross characterization of a joint distribution. Because they are not individually constrained we can have a wide variety of qualitatively different joint distributions and still satisfy the uncertainty principle.

Signal Analysis. The case of signal analysis is no different, although historically the uncertainty principle was viewed as presenting unsurmountable difficulty for the construction of joint distributions. Quite the contrary, it is trivially easy to construct joint distributions consistent with the uncertainty principle, as the above considerations have shown. All we have to do is satisfy the marginals. In fact, the real problem is that there are an infinite number of such joint distributions and we do not fully understand what other considerations besides the uncertainty principle are required to construct a sensible and comprehensive theory. Satisfying the uncertainty principle is easy; the rest is challenging.

6.10 THE BASIC PROBLEMS AND BRIEF HISTORICAL PERSPECTIVE

The fundamental issue is how to construct joint distributions that satisfy the rather mild conditions we have set forth. From a mathematical point of view there are an infinite number of joint distributions satisfying these requirements. However, writing them down explicitly is not easy and which ones out of the infinite number are "right" is not obvious. Mathematics cannot solve the problem, although it can guide us in constraining what is possible and what is not. The basic issue is not one of mathematics, since the conditions do not define the problem uniquely. The issue is one of science, engineering, and desire to achieve particular goals. The methods that have been developed to obtain distributions are not strict mathematical derivations but are based on physical considerations, mathematical suggestiveness, and the time honored scientific method of guessing, intelligent and otherwise. As of this writing the method of characteristic function operators, which is discussed in the later chapters, is the only consistent method for deriving distributions.

The current state of affairs is that we do not have a complete theory. Nonetheless, the ideas and techniques that have been developed thus far are powerful, give us considerable insight into the nature of signals, meet our intuition to a very large extent, and have been applied with immense success to practical problems. This is demonstrated in the subsequent chapters. Often some of the predicted results are clearly not plausible and that is what makes the subject challenging and fascinating. Because we do not have a complete and comprehensive theory, it is important to understand what is known with certainty, what is speculation, and what hidden assumptions go into any particular proof.

Since we have not yet studied any distributions it would not be appropriate at this juncture to discuss the mathematical or physical issues in any great detail. To appreciate these issues fully, we should first see how they arise. We do mention one idea that has a checkered history. Because many distributions have been proposed over the last fifty years, something that should be considered in a positive light, there have been attempts to prove that a particular one is best. This has been done by listing a set of desirable conditions and trying to prove that only one distribution, usually the writer's favorite, fits them. Typically, however, the list presented is not complete with the obvious requirements, because the author knows the added desirable properties would not be satisfied by the distribution he or she is advocating. Also these lists very often contain conditions that are taken out of thin air and are obviously put in to force a particular result. Moreover, there are usually hidden assumptions in the proofs that are glossed over or hidden in the methodologies of the proofs. However, there is one benefit to these impossibility and uniqueness proofs. The searching of the errors and hidden assumptions sometimes leads to a reexamination of the basic ideas. There are time-varying spectra in nature and therefore their description and properties can be described and understood. Someone will eventually come up with the right theory and when that happens we will not have to prove it mathematically correct; it will be obviously correct.

Terminology. For the type of distributions we will be studying, a rich variety of terminology has arisen. We first address the general usage of density and distribution. When we say that a function represents the number of things per unit something, that is called a density and hence density function. In certain fields density functions are called distributions because they indicate how things are distributed. For example, one says the "Maxwell distribution of velocities." As discussed in Chapter 4, we shall use the words density and distribution interchangeably. These types of distributions first arose in quantum mechanics where the term "probability density" or "distribution" is properly applied, since quantum mechanics is inherently probabilistic. For deterministic signals where no probabilistic considerations enter, the reader should think of distributions as "intensities" or "densities" in the common usage of the words, or simply as how the energy is "distributed" over the time-frequency plane. As we will see, many of the known distributions may become negative or even complex. Hence they are sometimes called quasi or pseudo distributions. Also, a joint time-frequency distribution is, of course, dependent on the signal and can be said to represent the signal in time and frequency; hence the phrase "time-frequency representation."

Chapter 7

The Short-Time Fourier Transform

7.1 INTRODUCTION

The short-time Fourier transform is the most widely used method for studying non-stationary signals. The concept behind it is simple and powerful. Suppose we listen to a piece of music that lasts an hour where in the beginning there are violins and at the end drums. If we Fourier analyze the whole hour, the energy spectrum will show peaks at the frequencies corresponding to the violins and drums. That will tell us that there were violins and drums but will not give us any indication of when the violins and drums were played. The most straightforward thing to do is to break up the hour into five minute segments and Fourier analyze each interval. Upon examining the spectrum of each segment we will see in which five minute intervals the violins and drums occurred. If we want to localize even better, we break up the hour into one minute segments or even smaller time intervals and Fourier analyze each segment. That is the basic idea of the short-time Fourier transform: break up the signal into small time segments and Fourier analyze each time segment to ascertain the frequencies that existed in that segment. The totality of such spectra indicates how the spectrum is varying in time.

Can this process be continued to achieve finer and finer time localization? Can we make the time intervals as short as we want? The answer is no, because after a certain narrowing the answers we get for the spectrum become meaningless and show no relation to the spectrum of the original signal. The reason is that we have taken a perfectly good signal and broken it up into short duration signals. But short duration signals have inherently large bandwidths, and the spectra of such short duration signals have very little to do with the properties of the original signal. This should be attributed not to any fundamental limitation, but rather to a limitation of the technique which makes short duration signals for the purpose of estimating

the spectrum. Sometimes this technique works well and sometimes it does not. It is not the uncertainty principle as applied to the signal that is the limiting factor; it is the uncertainty principle as applied to the small time intervals that we have created for the purpose of analysis. The distinction between the uncertainty principle for the small time intervals created for analysis and the uncertainty principle for the original signal should be clearly kept in mind and the two should not be confused.

We should always keep in mind that in the short-time Fourier transform the properties of the signal are scrambled with the properties of the window function, the window function being the means of chopping up the signal. Unscrambling is required for proper interpretation and estimation of the original signal.

The above difficulties notwithstanding, the short-time Fourier transform method is ideal in many respects. It is well defined, based on reasonable physical principles, and for many signals and situations it gives an excellent time-frequency structure consistent with our intuition. However, for certain situations it may not be the best method available in the sense that it does not always give us the clearest possible picture of what is going on. Thus other methods have been developed, which are discussed in subsequent chapters.

7.2 THE SHORT-TIME FOURIER TRANSFORM AND SPECTROGRAM

To study the properties of the signal at time t, one emphasizes the signal at that time and suppresses the signal at other times. This is achieved by multiplying the signal by a window function, $h(t)$, centered at t, to produce a modified signal,

$$s_t(\tau) = s(\tau)\, h(\tau - t) \tag{7.1}$$

The modified signal is a function of two times, the fixed time we are interested in, t, and the running time, τ. The window function is chosen to leave the signal more or less unaltered around the time t but to suppress the signal for times distant from the time of interest. That is,

$$s_t(\tau) \sim \begin{cases} s(\tau) & \text{for } \tau \text{ near } t \\ 0 & \text{for } \tau \text{ far away from } t \end{cases} \tag{7.2}$$

The term "window" comes from the idea that we are seeking to look at only a small piece of the signal as when we look out of a real window and see only a relatively small portion of the scenery. In this case we want to see only a small portion.

Since the modified signal emphasizes the signal around the time t, the Fourier transform will reflect the distribution of frequency around that time,

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s_t(\tau)\, d\tau \tag{7.3}$$

$$= \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s(\tau)\, h(\tau - t)\, d\tau \tag{7.4}$$

The energy density spectrum at time t is therefore

$$P_{SP}(t,\omega) = |S_t(\omega)|^2 = \left| \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s(\tau)\, h(\tau - t)\, d\tau \right|^2 \tag{7.5}$$

For each different time we get a different spectrum and the totality of these spectra is the time-frequency distribution, P_{SP}. It goes under many names, depending on the field; we shall use the most common phraseology, "spectrogram."

Since we are interested in analyzing the signal around the time t, we presumably have chosen a window function that is peaked around t. Hence the modified signal is short and its Fourier transform, Eq. (7.4), is called the short-time Fourier transform. However, it should be emphasized that often we will not be taking narrow windows – which is done when we want to estimate time properties for a particular frequency. When we want to estimate time properties for a given frequency we do not take short times but long ones, in which case the short-time Fourier transform may be appropriately called the long-time Fourier transform or the short-frequency time transform.

The Short-Frequency Time Transform. In motivating the short-time Fourier transform we emphasized the desire to study frequency properties at time t. Conversely, we may wish to study time properties at a particular frequency. We then window the spectrum, $S(\omega)$, with a frequency window function, $H(\omega)$, and take the time transform, which, of course, is the inverse Fourier transform. In particular, we define the short-frequency time transform by

$$s_\omega(t) = \frac{1}{\sqrt{2\pi}} \int e^{j\omega' t}\, S(\omega')\, H(\omega - \omega')\, d\omega' \tag{7.6}$$

If we relate the window function in time $h(t)$ with the window function in frequency $H(\omega)$ by

$$H(\omega) = \frac{1}{\sqrt{2\pi}} \int h(t)\, e^{-j\omega t}\, dt \tag{7.7}$$

then

$$S_t(\omega) = e^{-j\omega t}\, s_\omega(t) \tag{7.8}$$

The short-time Fourier transform is the same as the short-frequency time transform except for the phase factor $e^{-j\omega t}$. Since the distribution is the absolute square, the phase factor $e^{-j\omega t}$ does not enter into it and either the short-time Fourier transform or short-frequency time transform can be used to define the joint distribution,

$$P(t,\omega) = |S_t(\omega)|^2 = |s_\omega(t)|^2 \tag{7.9}$$

This shows that the spectrogram can be used to study the behavior of time properties at a particular frequency. This is done by choosing an $H(\omega)$ that is narrow, or equivalently by taking an $h(t)$ that is broad.

Narrowband and Wideband Spectrogram. As just discussed, if the time window is of short duration the frequency window, $H(\omega)$, is broad; in that case the spec-

trogram is called a broadband spectrogram. If the window is of long duration then $H(\omega)$ is narrow, and we say we have a narrowband spectrogram.

Characteristic Function. The characteristic function of the spectrogram is straightforwardly obtained,

$$M_{SP}(\theta,\tau) = \iint |S_t(\omega)|^2 e^{j\theta t + j\tau\omega} dt d\omega \tag{7.10}$$

$$= A_s(\theta,\tau) A_h(-\theta,\tau) \tag{7.11}$$

where

$$A_s(\theta,\tau) = \int s^*(t-\tfrac{1}{2}\tau)\, s(t+\tfrac{1}{2}\tau)\, e^{j\theta t}\, dt \tag{7.12}$$

is the ambiguity function of the signal, and A_h is the ambiguity function of the window defined in the identical manner, except that we use $h(t)$ instead of $s(t)$. Note that $A(-\theta,\tau) = A^*(\theta,-\tau)$, a relation we will use later.

Notation. The results we will obtain are revealing when expressed in terms of the phases and amplitudes of the signal and window and their transforms. The notation we use is

$$s(t) = A(t)\, e^{j\varphi(t)} \quad ; \quad h(t) = A_h(t)\, e^{j\varphi_h(t)} \tag{7.13}$$

$$S(\omega) = B(\omega)\, e^{j\psi(\omega)} \quad ; \quad H(\omega) = B_H(\omega)\, e^{j\psi_H(\omega)} \tag{7.14}$$

In the calculation of global averages (e.g., mean frequency, bandwidth) we will have to indicate which density function is being used. We will use the superscript $^{(s)}$, for example, to indicate that the signal is being used. In particular, the mean of frequency with respect to the spectrogram, signal, and window will be indicated respectively by

$$\langle \omega \rangle^{(SP)} = \iint \omega\, |S_t(\omega)|^2\, d\omega\, dt \tag{7.15}$$

$$\langle \omega \rangle^{(s)} = \int \omega\, |S(\omega)|^2\, d\omega \tag{7.16}$$

$$\langle \omega \rangle^{(h)} = \int \omega\, |H(\omega)|^2\, d\omega \tag{7.17}$$

and similar notation will be used for other quantities. When it is clear from the context which density is being used the superscript notation will be omitted.

7.3 GENERAL PROPERTIES

Total Energy. The total energy is obtained by integrating over all time and frequency. However, we know that it is given by the characteristic function evaluated at zero. Using Eqs. (7.11) and (7.12) we have

$$E_{SP} = \iint P_{SP}(t,\omega)\,dt\,d\omega = M_{SP}(0,0) \tag{7.18}$$

$$= A_s(0,0)\,A_h(0,0) \tag{7.19}$$

$$= \int |s(t)|^2\,dt \times \int |h(t)|^2\,dt \tag{7.20}$$

Therefore, we see that if the energy of the window is taken to be one, then the energy of the spectrogram is equal to the energy of the signal.

Marginals. The time marginal is obtained by integrating over frequency,

$$P(t) = \int |S_t(\omega)|^2\,d\omega \tag{7.21}$$

$$= \frac{1}{2\pi}\int s(\tau)\,h(\tau-t)s^*(\tau')\,h^*(\tau'-t)\,e^{-j\omega(\tau-\tau')}\,d\tau\,d\tau'\,d\omega \tag{7.22}$$

$$= \int s(\tau)\,h(\tau-t)s^*(\tau')\,h^*(\tau'-t)\,\delta(\tau-\tau')\,d\tau\,d\tau' \tag{7.23}$$

$$= \int |s(\tau)|^2\,|h(\tau-t)|^2\,d\tau \tag{7.24}$$

$$= \int A^2(\tau)\,A_h^2(\tau-t)\,d\tau \tag{7.25}$$

Similarly, the frequency marginal is

$$P(\omega) = \int B^2(\omega')\,B_H^2(\omega-\omega')\,d\omega' \tag{7.26}$$

As can be seen from these equations, the marginals of the spectrogram generally do not satisfy the correct marginals, namely $|s(t)|^2$ and $|S(\omega)|^2$,

$$P(t) \neq A^2(t) = |s(t)|^2 \tag{7.27}$$

$$P(\omega) \neq B^2(\omega) = |S(\omega)|^2 \tag{7.28}$$

The reason is that the spectrogram scrambles the energy distributions of the window with those of the signal. This introduces effects unrelated to the properties of the original signal.

Notice that the time marginal of the spectrogram depends only on the magnitude of the signal and window and not on their phases. Similarly, the frequency marginal depends only on the amplitudes of the Fourier transforms.

Averages of Time and Frequency Functions. Since the marginals are not satisfied, averages of time and frequency functions will never be correctly given,

$$\langle g_1(t) + g_2(\omega) \rangle = \iint \{g_1(t) + g_2(\omega)\} P_{SP}(t,\omega)\, d\omega\, dt \tag{7.29}$$

$$\neq \int g_1(t) \,|\, s(t) \,|^2 dt + \int g_2(\omega) \,|\, S(\omega) \,|^2 d\omega \tag{7.30}$$

never correctly given

This is in contrast to other distributions we will be studying where these types of averages are always correctly given.

Finite Support. Recall from our discussion in Chapter 6 that for a finite duration signal we expect the distribution to be zero before the signal starts and after it ends. This property was called the finite support property. Let us see whether the spectrogram satisfies this property. Suppose one chooses a time t before the signal starts. Will the spectrogram be zero for that time? Generally, no, because the modified signal as a function of t will not necessarily be zero since the window may pick up some of the signal. That is, even though $s(t)$ may be zero for a time t, $s(\tau)\, h(\tau - t)$ may not be zero for that time. This will always be the case for windows that are not time limited. But even if a window is time limited we will still have this effect for time values that are close to the beginning or end of the signal. Similar considerations apply to the frequency domain. Therefore the spectrogram does not possess the finite support property in either time or frequency.

Localization Trade-off. If we want good time localization we have to pick a narrow window in the time domain, $h(t)$, and if we want good frequency localization we have to pick a narrow window, $H(\omega)$, in the frequency domain. But both $h(t)$ and $H(\omega)$ cannot be made arbitrarily narrow; hence there is an inherent trade-off between time and frequency localization in the spectrogram for a particular window. The degree of trade-off depends on the window, signal, time, and frequency. The uncertainty principle for the spectrogram quantifies these trade off dependencies, as we discussed in Section 3.4.

One Window or Many? We have just seen that one window, in general, cannot give good time and frequency localization. That should not cause any problem of principle as long as we look at the spectrogram as a tool at our disposal that has many options including the choice of window. There is no reason why we cannot change the window depending on what we want to study. That can sometimes be done effectively, but not always. Sometimes a compromise window does very well. One of the advantages of other distributions that we will be studying is that both time and frequency localization can be done concurrently.

Entanglement and Symmetry Between Window and Signal. The results obtained using the spectrogram generally do not give results regarding the signal solely, because the short-time Fourier transform entangles the signal and window. Therefore we must be cautious in interpreting the results and we must attempt to disentan-

gle the window. That is not always easy. In fact, because of the basic symmetry in the definition of the short-time Fourier transform between the window and signal, we have to be careful that we are not using the signal to study the window. The mathematics makes no distinction. The distinction must come only from a judicious choice of window, which is totally under our control. The basic symmetry between the window and signal does have a mathematical advantage in that results obtained should show a symmetry and therefore act as a mathematical check.

7.4 GLOBAL QUANTITIES

The mean time and frequency in the spectrogram are given by

$$\langle t \rangle^{(SP)} = \iint t \, |\, S_t(\omega) \,|^2 \, dt \, d\omega \quad ; \quad \langle \omega \rangle^{(SP)} = \iint \omega \, |\, S_t(\omega) \,|^2 \, dt \, d\omega \tag{7.31}$$

Direct evaluation leads to

$$\langle t \rangle^{(SP)} = \langle t \rangle^{(s)} - \langle\, t \,\rangle^{(h)} \quad ; \quad \langle \omega \rangle^{(SP)} = \langle \omega \rangle^{(s)} + \langle\, \omega \,\rangle^{(h)} \tag{7.32}$$

If the window is chosen so that its mean time and frequency are zero, which can be done by choosing a window symmetrical in time and whose spectrum is symmetrical in frequency, then the mean time and frequency of the spectrogram will be that of the signal.

The second conditional moments are calculated to be

$$\langle \omega^2 \rangle^{(SP)} = \langle \omega^2 \,\rangle^{(s)} + \langle \omega^2 \rangle^{(h)} + 2 \, \langle \omega \rangle^{(s)} \, \langle \omega \rangle^{(h)} \tag{7.33}$$

$$\langle t^2 \rangle^{(SP)} = \langle t^2 \,\rangle^{(s)} + \langle t^2 \rangle^{(h)} - 2 \, \langle t \rangle^{(s)} \, \langle t \rangle^{(h)} \tag{7.34}$$

By combining these with Eqs. (7.32) we find that the durations and bandwidths are related by

$$T_{SP}^2 = T_s^2 + T_h^2 \quad ; \quad B_{SP}^2 = B_s^2 + B_h^2 \tag{7.35}$$

which indicates how the duration of the windowed signal is related to the durations of the signal and window.

Covariance and Correlation Coefficient. For the first mixed moment

$$\langle t \, \omega \rangle^{(SP)} = \iint t \, \omega \, |\, S_t(\omega) \,|^2 \, dt \, d\omega \tag{7.36}$$

$$= \langle t \, \varphi' \rangle^{(s)} - \langle t \, \varphi_h' \rangle^{(h)} - \langle t \rangle^{(h)} \, \langle \varphi' \,\rangle^{(s)} + \langle t \rangle^{(s)} \, \langle \varphi_h' \,\rangle^{(h)} \tag{7.37}$$

Subtracting $\langle\, t \,\rangle^{(SP)}\langle\, \omega \,\rangle^{(SP)}$, as given by Eqs. (7.32), from both sides we have that the covariance of the spectrogram is

$$\mathrm{Cov}_{t\omega}^{(SP)} = \langle t\omega \rangle^{(SP)} - \langle\, t \,\rangle^{(SP)}\langle\, \omega \,\rangle^{(SP)} = \mathrm{Cov}_{t\omega}^{(s)} - \mathrm{Cov}_{t\omega}^{(h)} \tag{7.38}$$

We know from Chapter 1 that the covariance of a real signal is zero and hence if we take real windows, the covariance of the spectrogram will be the covariance of the signal

$$\mathrm{Cov}_{t\omega}^{(SP)} = \mathrm{Cov}_{t\omega}^{(s)} \qquad \text{for real windows} \tag{7.39}$$

7.5 LOCAL AVERAGES

Method of Calculation. Much of the analysis and calculations using the spectrogram is simplified significantly if we keep in mind that the modified signal, s_h, and the short-time Fourier transform, $S_t(\omega)$, form a Fourier pair with respect to the running time, τ,

$$s(\tau)\, h(\tau - t) \Longleftrightarrow S_t(\omega) \qquad [\text{ Fourier pair between } \tau, \omega\ ] \tag{7.40}$$

The real time, t, is considered a parameter. The modified signal expressed in terms of the phases and amplitudes is

$$s_t(\tau) = s(\tau)\, h(\tau - t) = A(\tau)\ A_h(\tau - t)\, e^{j[\varphi(\tau) + \varphi_h(\tau - t)]} \tag{7.41}$$

A fruitful way to look at the situation is that we are dealing with a signal in the variable τ whose amplitude is $A(\tau)\ A_h(\tau - t)$ and whose phase is $\varphi(\tau) + \varphi_h(\tau - t)$. However, $s_t(\tau)$ is not normalized in the variable τ and therefore we define, as in Section 3.4, the normalized modified signal by

$$\eta_t(\tau) = \frac{s(\tau)\, h(\tau - t)}{\sqrt{\int\, |\, s(\tau)\, h(\tau - t)\, |^2\, d\tau}} = \frac{s(\tau)\, h(\tau - t)}{\sqrt{P(t)}} \tag{7.42}$$

Everything we have done in Chapter 1 can be translated to the calculation of conditional values for the spectrogram by replacing the signal s with η and considering τ as the time. In particular, the conditional average of any function is

$$\langle\, g(\omega) \,\rangle_t = \frac{1}{P(t)} \int g(\omega)\ |\, S_t(\omega)\, |^2\ d\omega = \int \eta_t^*(\tau)\, g\Big(\frac{1}{j}\frac{d}{d\tau}\Big)\, \eta_t(\tau)\, d\tau \tag{7.43}$$

Local Frequency: Estimate of Instantaneous Frequency. Look at Eq. (1.90) and simply let $A^2 \to A^2(\tau) A_h^2(\tau - t)$ and $\varphi \to \varphi'(\tau) + \varphi_h'(\tau - t)$ to obtain

$$\langle\, \omega \,\rangle_t = \frac{1}{P(t)} \int \omega\, |\, S_t(\omega)\, |^2\, d\omega = \int \eta_t^*(\tau)\, \frac{1}{j}\frac{d}{d\tau}\, \eta_t(\tau)\, d\tau \tag{7.44}$$

$$= \frac{1}{P(t)} \int A^2(\tau)\ A_h^2(\tau - t)\, \{\varphi'(\tau) + \varphi_h'(\tau - t)\}\, d\tau \tag{7.45}$$

Local Square Frequency. Similarly, transcribing Eq. (1.98) yields

$$\langle \omega^2 \rangle_t = \int \eta_t^*(\tau) \left(\frac{1}{j} \frac{d}{d\tau} \right)^2 \eta_t(\tau)\, d\tau = \int \left| \frac{d}{d\tau} \eta_t(\tau) \right|^2 d\tau \tag{7.46}$$

$$= \frac{1}{P(t)} \int \left(\frac{d}{d\tau} A(\tau)\, A_h(\tau - t) \right)^2 d\tau + \frac{1}{P(t)} \int A^2(\tau) A_h^2(\tau - t) \{ \varphi'(\tau) + \varphi_h'(\tau - t) \}^2 d\tau \tag{7.47}$$

The Conditional or Instantaneous Bandwidth. Using $\sigma^2_{\omega|t} = \langle \omega^2 \rangle_t - \langle \omega \rangle_t^2$ with $\langle \omega^2 \rangle_t$ and $\langle \omega \rangle_t$ as given by Eqs. (7.45) and (7.47) we can obtain the conditional or instantaneous bandwidth. However, the resulting expression does not explicitly show the manifestly positive nature of the standard deviation. A way to obtain an alternative but manifestly positive expression is to use Eq. (1.96), the bandwidth equation, and do the appropriate replacement as above,

$$B_t^2 = \sigma^2_{\omega|t} = \frac{1}{P(t)} \int (\omega - \langle \omega \rangle_t)^2 \, | S_t(\omega) |^2 \, d\omega \tag{7.48}$$

$$= \frac{1}{P(t)} \int \left(\frac{d}{d\tau} A(\tau)\, A_h(\tau - t) \right)^2 d\tau + \frac{1}{2P_1^2(t)} \iint A^2(\tau_1)\, A^2(\tau_2)\, A_h^2(\tau_1 - t)\, A_h^2(\tau_2 - t) \times [\varphi'(\tau_1) - \varphi'(\tau_2) + \varphi_h'(\tau_1 - t) - \varphi_h'(\tau_2 - t)]^2 \, d\tau_1 \, d\tau_2 \tag{7.49}$$

For convenience we define

$$\langle \omega^2 \rangle_t^0 = \frac{1}{P(t)} \int \left(\frac{d}{d\tau} A(\tau)\, A_h(\tau - t) \right)^2 d\tau \tag{7.50}$$

7.6 NARROWING AND BROADENING THE WINDOW

All the above results show that the physical quantities obtained using the spectrogram result in an entanglement of signal and window. We can view these quantities as estimates which unfortunately depend on the window function chosen. However, one would hope that in some sense or other the basic results are window independent and that when the window is narrowed we will get better and better estimates. Let us first consider the conditional frequency and suppose that we

narrow the window so that $A_h^2(t)$ approaches a delta function,

$$A_h^2(t) \rightarrow \delta(t) \tag{7.51}$$

Also, we consider real windows so as not to introduce phase due to the window. In this limit, Eq. (7.45) goes as

$$\langle \omega \rangle_t = \frac{\int A^2(\tau)\, A_h^2(\tau - t)\, \varphi'(\tau)\, d\tau}{\int A^2(\tau)\, A_h^2(\tau - t)\, d\tau} \tag{7.52}$$

$$\rightarrow \frac{\int A^2(\tau)\, \delta(\tau - t)\, \varphi'(\tau)\, d\tau}{\int A^2(\tau)\, \delta(\tau - t)\, d\tau} = \frac{A^2(t)\, \varphi'(t)}{A^2(t)} \tag{7.53}$$

or

$$\langle \omega \rangle_t \rightarrow \varphi'(t) \tag{7.54}$$

That is, if the window is narrowed to get increasing time resolution, the limiting value of the estimated instantaneous frequency is the derivative of the phase, which is the instantaneous frequency. This a very pleasing and important result, but there is a penalty to pay. As we narrow the window we broaden the standard deviation and, therefore, as the estimate goes to the right answer it becomes harder to see and determine. The reason for this can be seen by direct mathematical substitution of the delta function in Eq. (7.49), which results in

$$\sigma_{\omega|t} \rightarrow \infty \qquad [\text{for } A_h^2(t) \rightarrow \delta(t)\,] \tag{7.55}$$

A simple explanation is available: we have taken a signal and made it of narrow duration at time t; in fact, we have made it infinitely narrow by letting the magnitude square of the window go to a delta function. The standard deviation of the modified signal must therefore go to infinity because of the time-bandwidth product theorem. Even though the average is correctly predicted, it is hard to discern. A compromise is discussed in Section 7.11.

7.7 GROUP DELAY

We now give the average time for a given frequency. Because of the symmetry we can immediately write

$$\langle t \rangle_\omega = \frac{1}{P(\omega)} \int B^2(w)\, B_h^2(\omega - w)\, [\, \psi'(w) - \psi_h'(\omega - w)\,]\, dw \tag{7.56}$$

where $P(\omega)$ is the marginal in frequency given by Eq. (7.26). If the window is narrowed in the frequency domain then a similar argument as before shows that the estimated group delay goes as

$$\langle t \rangle_\omega \rightarrow -\psi'(\omega) \qquad [\text{for } H^2(\omega) \rightarrow \delta(\omega)\,] \tag{7.57}$$

The second moment and conditional standard deviation are given by

$$\langle t^2 \rangle_\omega = \frac{1}{P(\omega)} \int \left(\frac{d}{dw} B(w)\, B_h(\omega - w) \right)^2 dw + \frac{1}{P(\omega)} \int B^2(w)\, B_h^2(\omega - w)\, [\, \psi'(w) - \psi_h'(\omega - w)\,]^2\, dw \tag{7.58}$$

$$\sigma_{t|\omega}^2 = \frac{1}{P(\omega)} \int \left(\frac{d}{dw} B(w)\, B_h(\omega - w) \right)^2 dw + \frac{1}{2P^2(\omega)} \iint B^2(w_1)\, B^2(w_2)\, B_h^2(\omega - w_1)\, B_h^2(\omega - w_2) \times [\, \psi'(w_1) - \psi'(w_2) - \psi_h'(\omega - w_1) + \psi_h'(\omega - w_2)\,]^2\, dw_1\, dw_2 \tag{7.59}$$

The standard deviation of time at a given frequency may be interpreted as the duration of the signal for that frequency. The physical significance can be thought of in the following way. We envision a signal as composed of many frequencies, each frequency having an envelope. The duration at each frequency is the time width of the envelope for that frequency. Equation (7.59) is an estimate of that squared duration. All the remarks we made about the estimated instantaneous frequency apply to the group delay, except we now have to narrow the window in the frequency domain.

7.8 EXAMPLES

Example 7.1: Sinusoid.

As a first example we consider a sinusoid and use a Gaussian window

$$s(t) = e^{j\omega_0 t} \qquad h(t) = (a/\pi)^{1/4}\, e^{-at^2/2} \tag{7.60}$$

The short-time Fourier transform is

$$S_t(\omega) = \frac{1}{(a\pi)^{1/4}}\, e^{-j(\omega - \omega_0)t} \exp\left[-\frac{(\omega - \omega_0)^2}{2a} \right] \tag{7.61}$$

which yields the following time-frequency distribution

$$P_{SP}(t, \omega) = |\, S_t(\omega)\, |^2 = \frac{1}{(a\pi)^{1/2}} \exp\left[-\frac{(\omega - \omega_0)^2}{a} \right] \tag{7.62}$$

Using it we have

$$\langle \omega \rangle_t = \omega_0 \quad ; \quad \sigma_{\omega|t}^2 = \tfrac{1}{2} a \tag{7.63}$$

The average value of frequency for a given time is always ω_0, but the width about that frequency is dependent on the window width.

Example 7.2: Impulse.

For an impulse at $t = t_0$, with the same window as above

$$s(t) = \sqrt{2\pi}\,\delta(t - t_0) \qquad h(t) = (a/\pi)^{1/4}\, e^{-at^2/2} \tag{7.64}$$

we have

$$S_t(\omega) = (a/\pi)^{1/4}\, e^{-j\omega t_0}\, e^{-a(t-t_0)^2/2} \tag{7.65}$$

$$P_{SP}(t,\omega) = |\, S_t(\omega)\,|^2 = (a/\pi)^{1/2}\, e^{-a(t-t_0)^2} \tag{7.66}$$

Example 7.3: Sinusoid Plus Impulse.

Now consider the sum of a sinusoid and impulse,

$$s(t) = e^{j\omega_0 t} + \sqrt{2\pi}\delta(t - t_0) \qquad h(t) = (a/\pi)^{1/4}\, e^{-at^2/2} \tag{7.67}$$

We have

$$S_t(\omega) = \frac{1}{(a\pi)^{1/4}}\, e^{-j(\omega-\omega_0)t}\, \exp\left[-\frac{(\omega-\omega_0)^2}{2a}\right] + (a/\pi)^{1/4}\, e^{-j\omega t_0}\, e^{-a(t-t_0)^2/2} \tag{7.68}$$

and

$$P_{SP}(t,\omega) = |\, S_t(\omega)\,|^2 = \frac{1}{(a\pi)^{1/2}}\, e^{-(\omega-\omega_0)^2/a} + (a/\pi)^{1/2}\, e^{-a(t-t_0)^2} + \frac{2}{\sqrt{\pi}}\, e^{-(\omega-\omega_0)^2/a - a(t-t_0)^2} \cos[\omega(t-t_0) - \omega_0 t] \tag{7.69}$$

This example illustrates one of the fundamental difficulties with the spectrogram. For one window we cannot have high resolution in time and frequency. The broadness of the self terms, the first two terms of Eq. (7.69), depends on the window size in an inverse relation. If we try to make one narrow, then the other must be broad. That will not be the case with other distributions. In Fig. 7.1 we plot the spectrogram for different window sizes. The cosine term is an example of the so called cross terms. We discuss these in greater detail later. Note that they essentially fall on the self terms in the spectrogram.

Example 7.4: Linear FM.

Consider a chirp with a Gaussian envelope and a Gaussian window,

$$s(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \quad ; \quad h(t) = (a/\pi)^{1/4}\, e^{-at^2} \tag{7.70}$$

A straightforward but lengthy calculation gives

$$P_{SP}(t,\omega) = |\, S_t(\omega)\,|^2 = \frac{P(t)}{\sqrt{2\pi\sigma^2_{\omega|t}}}\, \exp\left[-\frac{(\omega - \langle\omega\rangle_t)^2}{2\sigma^2_{\omega|t}}\right] \tag{7.71}$$

$$= \frac{P(\omega)}{\sqrt{2\pi\sigma^2_{t|\omega}}}\, \exp\left[-\frac{(t - \langle t\rangle_\omega)^2}{2\sigma^2_{t|\omega}}\right] \tag{7.72}$$

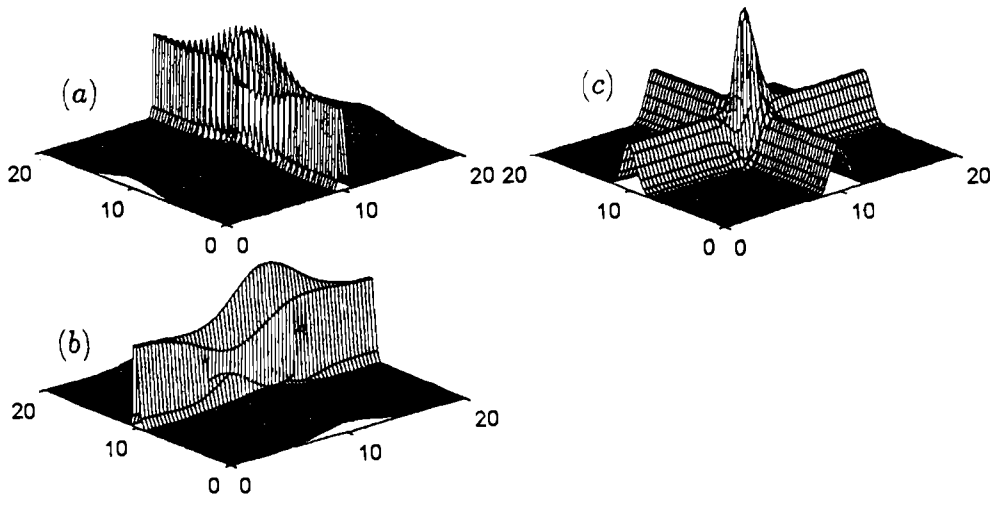

Fig. 7.1 Spectrograms of a signal composed of a constant frequency plus an impulse, $s(t) = e^{j10t} + \delta(t-10)$. In (a) we use a short duration window, which gives a good indication of when the impulse occurred but gives a broad localization for the frequency. In (b) we use a long duration window, which gives the opposite effect. In (c) a compromise window is used. For a comparison with the Wigner distribution see Fig. 8.5.

where $P(t)$ and $P(\omega)$ are the marginal distributions

$$P(t) = \sqrt{\frac{a\alpha}{\pi(\alpha+a)}}\ \exp\left[-\frac{a\alpha}{\alpha+a}\,t^2\right] \tag{7.73}$$

$$P(\omega) = \sqrt{\frac{a\alpha/\pi}{\alpha a^2 + a(\alpha^2+\beta^2)}}\ \exp\left[-\frac{a\alpha}{\alpha a^2 + a(\alpha^2+\beta^2)}\,\omega^2\right] \tag{7.74}$$

and

$$\langle\omega\rangle_t = \frac{a}{\alpha+a}\,\beta t + \omega_0 \quad ; \quad \sigma^2_{\omega|t} = \tfrac{1}{2}\,(\alpha+a) + \tfrac{1}{2}\,\frac{\beta^2}{\alpha+a} \tag{7.75}$$

$$\langle t\rangle_\omega = \frac{a\beta}{\alpha a^2 + a(\alpha^2+\beta^2)}\,\omega \tag{7.76}$$

$$\sigma^2_{t|\omega} = \frac{1}{2}\,\frac{(\alpha+a)^2+\beta^2}{\alpha a^2 + a(\alpha^2+\beta^2)} \tag{7.77}$$

The concentration of energy for a given time is along the estimated instantaneous frequency, and for a given frequency it is along the estimated time delay. As the window becomes narrow, that is, as $a \to \infty$, the estimate for the instantaneous frequency approaches $\beta t + \omega_0$. However, in this limit the estimated group delay approaches zero, which is a very bad estimate. In fact, this answer is just the average time for the signal. It is an understandable answer because in that limit we have a flat window in the frequency domain. Conversely, if we want to focus in on time properties for a given frequency we must take a broad window in the time domain, which is achieved by taking $a \to \infty$. In that case $\langle t\rangle_\omega \to \beta/(\alpha^2+\beta^2)$, which is the correct answer (see Eq.

(1.121)). On the other hand, at this limit the estimate of the instantaneous frequency approaches ω_0, which is not the instantaneous frequency at all but just the average frequency. Again, the reason is in that limit the window in the time domain is flat for all time. This example shows that for the spectrogram one window cannot give good localization in both time and frequency. In contrast, other distributions will allow us to do that. We also note that for this case, the local bandwidth is a constant in time. That would not be the case if a different window were chosen.

Example 7.5: Single Sided Sinusoid with Rectangular Window.

Take

$$s(t) = e^{j\omega_0 t}, \qquad t \geq 0 \qquad ; \qquad h(t) = \frac{1}{\sqrt{T}} \quad -T/2 \leq t \leq T/2 \tag{7.78}$$

A straightforward calculation yields

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{(\omega - \omega_0)} \begin{cases} j\left(e^{-j(\omega-\omega_0)(t+T/2)} - 1\right) & -T/2 \leq t \leq T/2 \\ 2e^{-j(\omega-\omega_0)t} \sin[(\omega-\omega_0)T/2] & T/2 \leq t \end{cases} \tag{7.79}$$

and

$$|\, S_t(\omega)\,|^2 = \frac{2}{\pi} \frac{1}{(\omega - \omega_0)^2} \begin{cases} \sin^2[\frac{1}{2}(\omega-\omega_0)(t+T/2)] & -T/2 \leq t \leq T/2 \\ \sin^2[(\omega-\omega_0)T/2] & T/2 \leq t \end{cases} \tag{7.80}$$

Even though the signal is zero before $t = 0$, the spectrogram indicates that there is a value in the range $-T/2 \leq t \leq T/2$. The reason is that for that range the window is picking up values from times greater than zero. The longer we make the window, the sharper the spectrogram will be around $\omega = \omega_0$, but then we will have poor localization in time. If we take a small window then the spectrum will be broad. This is illustrated Fig. 7.2. This example shows that the spectrogram does not satisfy the finite support property.

Example 7.6: Chirp With Rectangular Window.

For the signal of Eq. (7.70), if a rectangular window is used instead of a Gaussian, we obtain

$$P(t) = \frac{1}{2T}\left\{\mathrm{erf}\left[\sqrt{\alpha}(t+\tfrac{1}{2}T)\right] - \mathrm{erf}\left[\sqrt{\alpha}(t-\tfrac{1}{2}T)\right]\right\} \tag{7.81}$$

$$\langle \omega_t \rangle = \frac{1}{P(t)} \sqrt{\frac{1}{\alpha\pi}} \frac{\beta}{T} e^{-\alpha(t^2+T^2/4)} \sinh \alpha t T \tag{7.82}$$

where $\mathrm{erf}(x)$ is the error function. It can be readily verified that $\langle \omega_t \rangle \rightarrow \omega_i$ as the window size becomes small. If we now attempt to calculate the second conditional moment and the standard deviation, infinity is obtained. This is due to the sharp cutoff of the rectangular window. Hence a new measure of spread is needed for some windows or the window has to be modified to avoid divergences.

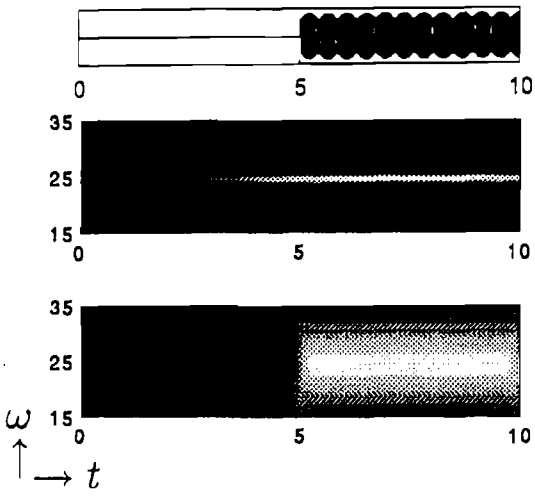

Fig. 7.2 Spectrograms of a sine wave that starts at $t = 5$. In (b) a very broad window is used to get good frequency localization. This spectrogram does not give a good indication of where the signal is zero. In (c) a very narrow window is used. The spectrum is very broad because it is the spectrum of a short duration signal, the modified signal, Eq. (7.1).

Example 7.7: Sinusoidal FM Signal/Chirp with a Gaussian Window.

Consider

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t + jm\sin\omega_m t} \quad ; \quad h(t) = (a/\pi)^{1/4} e^{-at^2/2} \tag{7.83}$$

The instantaneous frequency is

$$\omega_i = m\omega_m \cos\omega_m t + \beta t + \omega_0 \tag{7.84}$$

which indicates that the frequency oscillates sinusoidally around the straight line given by $\beta t + \omega_0$. To calculate the estimated instantaneous frequency we use Eq. (7.45) to obtain

$$\langle\omega\rangle_t = \frac{a}{\alpha+a}\beta t + \omega_0 + m\omega_m e^{-\frac{\omega_m^2}{4(\alpha+a)}} \cos\left(\frac{a\omega_m}{\alpha+a}\right) t \tag{7.85}$$

$$\langle\omega^2{}_t\rangle = \tfrac{1}{2}(\alpha+a) + \tfrac{1}{2}m^2\omega_m^2 \left(e^{-\frac{\omega_m^2}{\alpha+a}} \cos\left(\frac{a\omega_m}{\alpha+a}\right) t + 1 \right) \tag{7.86}$$

The standard deviation is calculated from $\sigma^2_{t|\omega} = \langle\omega^2\rangle_t - \langle\omega\rangle_t^2$. In the limit as $a \to 0$, $\langle\omega\rangle_t \to \omega_i$, as expected. The short-time Fourier transform can be done exactly. Substituting the signal and window in the definition we have

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} \left(\frac{a\alpha}{\pi^2}\right)^{1/4} e^{-at^2/2} \int e^{-(a+\alpha-j\beta)\tau^2/2 - (\omega-\omega_0)\tau + jm\sin(\omega_m\tau)} \tag{7.87}$$

But

$$e^{jm\sin(\omega_m\tau)} = \sum_{n=-\infty}^{\infty} J_n(m)\, e^{jn\omega_m\tau} \tag{7.88}$$

where J_n are the Bessel functions. Substituting this into Eq. (7.87) and evaluating the Gaussian integral we have

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} \left(\frac{a\alpha}{\pi^2(a+\alpha-j\beta)} \right)^{1/4} e^{-at^2/2} \times \sum_{n=-\infty}^{\infty} J_n(m) \exp\left[\frac{\{at - j(\omega - \omega_0 - n\omega_m)\}^2}{2(a+\alpha-j\beta)} \right] \tag{7.89}$$

7.9 INVERSION

Can the signal be recovered from the spectrogram? Since the spectrogram is the absolute square frequency of the short-time Fourier transform it is often stated that we have lost the phase and hence the signal cannot be recovered. This argument is not correct, because the phase we have lost is the phase of the short-time Fourier transform, not the signal. In fact, the phase and amplitude of the signal appear in both the phase and amplitude of the short-time Fourier transform. Therefore, having the amplitude of the short-time Fourier transform may be sufficient to recover the signal. We now see under what conditions this is the case. From Eq. (7.11) the ambiguity function of the signal is

$$A_s(\theta,\tau) = \frac{M_{SP}(\theta,\tau)}{A_h(-\theta,\tau)} \tag{7.90}$$

$$= \frac{1}{A_h(-\theta,\tau)} \iint |S_t(\omega)|^2 e^{j\theta t + i\tau\omega}\, dt\, d\omega \tag{7.91}$$

But the ambiguity function determines the signal. To see this, we Fourier invert the ambiguity function of the signal, Eq. (7.12), to obtain

$$s^*(t - \tfrac{1}{2}\tau)\, s(t + \tfrac{1}{2}\tau) = \frac{1}{2\pi} \int A_s(\theta,\tau)\, e^{-j\theta t}\, d\theta \tag{7.92}$$

and taking a particular value, $t = \tau/2$, we have

$$s^*(0)\, s(t) = \frac{1}{2\pi} \int A_s(\theta,t)\, e^{-j\theta t/2}\, d\theta \tag{7.93}$$

where $s^*(0)$ is a number. Therefore

$$s(t) = \frac{1}{2\pi s^*(0)} \int \frac{M_{SP}(\theta,t)}{A_h(-\theta,t)}\, e^{-j\theta t/2}\, d\theta \tag{7.94}$$

which shows that the signal can be recovered from the spectrogram. However, there is an important caveat. The division by the ambiguity function of the window must be possible and that will be the case if it has no finite regions where it is zero in the θ, τ plane. That depends on the window function.

7.10 EXPANSION IN INSTANTANEOUS FREQUENCY

We now show that the local quantities have interesting expansions in terms of the instantaneous frequency and its derivatives. For simplicity we consider real windows.

Local Mean Frequency. We rewrite Eq. (7.45) as

$$\langle \omega \rangle_t = \frac{\int A^2(\tau + t)\, A_h^2(\tau)\, \varphi'(\tau + t)\, d\tau}{\int A^2(\tau + t)\, A_h^2(\tau)\, d\tau} \tag{7.95}$$

Expanding $\varphi'(\tau + t)$ as a power series in τ

$$\varphi'(\tau + t) = \sum_{n=0}^{\infty} \varphi^{(n+1)}(t)\, \frac{\tau^n}{n!} \tag{7.96}$$

and substituting into Eq. (7.95), we have

$$\langle \omega \rangle_t = \sum_{n=0}^{\infty} \frac{M_n(t)}{n!}\, \varphi^{(n+1)}(t) \tag{7.97}$$

where

$$M_n(t) = \frac{\int A^2(\tau + t)\, A_h^2(\tau)\, \tau^n\, d\tau}{\int A^2(\tau + t)\, A_h^2(\tau)\, d\tau} \tag{7.98}$$

For $n = 0$ the denominator and numerator are equal, and hence $M_0(t)$ is always equal to one. Writing out the first few terms for $\langle \omega \rangle_t$ we have

$$\langle \omega \rangle_t \sim \varphi' + M_1(t)\, \varphi'' + \tfrac{1}{2}\, M_2(t)\, \varphi''' \tag{7.99}$$

Similar considerations lead to the following expansions for the mean square and conditional standard deviation,

$$\langle \omega^2 \rangle_t = \langle \omega^2 \rangle_t^0 + \sum_{n=0}^{\infty} \left[\sum_{k=0}^{n} \frac{\varphi^{(n-k+1)}\, \varphi^{(k+1)}}{(n-k)!\, k!} \right] M_n(t) \tag{7.100}$$

$$\sim \langle \omega^2 \rangle_t^0 + \varphi'^2 + 2\, \varphi'\, \varphi''\, M_1 + [\, \varphi'\, \varphi''' + \varphi''^2\,]\, M_2 \tag{7.101}$$

$$\sigma_{\omega|t}^2 = \langle \omega^2 \rangle_t^0 + \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} \varphi^{(n-k+1)}\, \varphi^{(k+1)}\, \frac{M_n - M_{n-k} M_k}{(n-k)!\, k!} \tag{7.102}$$

$$\sim \langle \omega^2 \rangle_t^0 + (M_2 - M_1^2)\, \varphi''^2 + (M_3 - M_2 M_1)\, \varphi''\, \varphi''' \tag{7.103}$$

Note that the derivative of the phase of the signal is absent from the expansion of $\sigma^2_{\omega|t}$. Also, $M_2 - M_1^2$ is the local duration as defined by Eq. (3.34),

$$T_t^2 = M_2 - M_1^2 \tag{7.104}$$

Example 7.8: Arbitrary Frequency with Gaussian Window.

We take an arbitrary phase but impose a Gaussian amplitude modulation

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\varphi(t)} \qquad A_h^2(t) = (a/\pi)^{1/2} e^{-at^2} \tag{7.105}$$

in which case the integrals can be readily done to obtain

$$\langle \omega \rangle_t = \varphi' - \frac{\alpha t}{\alpha + a}\varphi'' + \frac{1}{2(\alpha+a)}\left(\frac{1}{2} + \frac{\alpha^2 t^2}{\alpha + a}\right)\varphi''' \tag{7.106}$$

$$\sigma^2_{\omega|t} \sim \tfrac{1}{2}(\alpha+a) + \frac{1}{2(\alpha+a)}\varphi''^2 - \frac{\alpha t}{(\alpha+a)^2}\varphi''\varphi''' + \frac{1}{2(\alpha+a)}\left(\frac{1}{2} + \frac{\alpha^2 t^2}{\alpha+a}\right)\varphi''\varphi'''' \tag{7.107}$$

7.11 OPTIMAL WINDOW

To obtain the finest resolution of the instantaneous frequency, we minimize the spread about the estimated instantaneous frequency. That is, a window should be chosen so that the local standard deviation, $\sigma_{\omega|t}$, is a minimum. Since the local standard deviation depends on the signal, the optimal window will also depend on the signal. How the dependence comes in may be seen in rough terms as follows. As an approximation we drop all terms higher than the second in the expansion of $\sigma_{\omega|t}$ in Eq. (7.102),

$$\sigma^2_{\omega|t} \sim \langle \omega^2 \rangle^0_t + T_t^2 \, \varphi''^2(t) \tag{7.108}$$

Now, roughly, by the time dependent uncertainty principle we have that

$$\langle \omega^2 \rangle^0_t \; T_t^2 \sim \tfrac{1}{4} \tag{7.109}$$

and therefore

$$\sigma^2_{\omega\,|\,t} \sim \frac{1}{4T_t^2} + T_t^2 \varphi''^2(t) \tag{7.110}$$

The minimum is achieved at

$$T_t^2\big|_{\min} \sim \frac{1}{2\,|\,\varphi''(t)\,|} \tag{7.111}$$

At that value the local standard deviation is

$$\sigma^2_{\omega\,|\,t} \sim |\,\varphi''(t)\,| \tag{7.112}$$

We emphasize that T_t^2 is not the width of the window. It is the width of the modified signal at time t, which involves both the window and the signal. However, if the signal is purely frequency modulated then T_t^2 is the width of the window, in which case we have

$$\text{Optimal width of window} \sim \frac{1}{2\,|\,\varphi''(t)\,|} \qquad (7.113)$$

[for purely frequency modulated signals]

Example 7.9: Exactly Solvable.

Equation (7.75) gives the exact answer for $\sigma^2_{\omega|t}$ and it is revealing to do the minimization exactly. To minimize $\sigma^2_{\omega|t}$ we set the derivative with respect to a to zero

$$\frac{d\sigma^2_{\omega|t}}{da} = \tfrac{1}{2} - \tfrac{1}{2}\,\frac{\beta^2}{(\alpha+a)^2} = 0 \qquad (7.114)$$

which gives

$$a_{\min} = |\beta| - \alpha \qquad (7.115)$$

Note that $a_{\min}$ does not have to be positive. If it is negative the window will not be normalizable but nonetheless the STFT will be normalizable as long as $a_{\min} + \alpha$ is positive. Hence divergent windows may be used and indeed are desirable and necessary to obtain a high resolution. The standard deviation of the window is $\sigma_W^2 = 1/(2a)$ and we have that the optimal window is

$$\sigma_W^2 = \frac{1}{2a_{\min}} = \frac{1}{2(\,|\beta| - \alpha)} \qquad (7.116)$$

If we have a constant amplitude signal ($\alpha = 0$) then, $a_{\min} = |\beta|$ in accordance with Eq. (7.113).

Example 7.10: Gaussian Envelope with Arbitrary Phase.

Let

$$s(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2 + j\varphi(t)} \qquad (7.117)$$

and suppose we use the window

$$A_h^2(t) = \sqrt{\frac{1}{2\pi\sigma_W^2}}\, e^{-t^2/(2\sigma_W^2)} \qquad (7.118)$$

We obtain

$$a_{\min} = \frac{1}{2\sigma_W^2} \sim |\varphi''| - \alpha + \frac{2c_1\varphi''^2 + 3c_2\varphi''}{\varphi''^4 + 6c_1\varphi'' + 12c_2} \qquad (7.119)$$

where

$$c_1 = \tfrac{1}{4}\,\varphi''\,\varphi'''' + \tfrac{1}{8}\,\varphi'''^2 - \alpha t\,\varphi''\,\varphi''' \qquad (7.120)$$

$$c_2 = \tfrac{1}{2}\,\alpha^2 t^2\,\left[\,\varphi''\,\varphi'''' + \varphi'''^2\,\right] \qquad (7.121)$$

For a chirp where $\varphi' = \beta t$, c_1 and c_2 are both zero and we have that $a_{\min} = |\beta| - \alpha$, which agrees with Eq. (7.115).

Example 7.11: Cubic Phase.

For the cubic case where the phase is $\phi(t) = \gamma t^3/3$, we get the following approximation,

$$a_{\text{mun}} \sim \begin{cases} 2\gamma t - 3\alpha - \frac{9\alpha^2}{4\gamma t} + \left(\frac{1}{4} - \frac{5\alpha^3}{\gamma^2}\right)\frac{1}{t^2} & t \sim \infty \\ (2\gamma^2)^{1/3} - \alpha + \left[\ 2\alpha^2 - \frac{8}{3}\alpha(2\gamma^2)^{1/3} + \frac{4}{3}\left(\frac{\gamma^2}{2}\right)^{1/3}\right] t^2 & t \sim 0 \end{cases} \tag{7.122}$$

Chapter 8

The Wigner Distribution

8.1 INTRODUCTION

The Wigner distribution[1] is the prototype of distributions that are qualitatively different from the spectrogram. The discovery of its strengths and shortcomings has been a major thrust in the development of the field. Although it has often been studied in contrast to the spectrogram, we will see in Chapter 9 that both are members of a general class of representations.

Wigner was aware that there were other joint densities but chose what has now become the Wigner distribution "because it seems to be the simplest". Nonetheless, Kirkwood,[305] a year later, and Terletsky,[541] a few years after, came up with a simpler candidate that is commonly called the Rihaczek or Margenau-Hill distribution.

The Wigner distribution was introduced into signal analysis by Ville,[561] some 15 years after Wigner's paper. Ville gave a plausibility argument for it and "derived" it by a method based on characteristic functions. Remarkably, the same type of derivation was used by Moyal[390] at about the same time. These derivations appeared to be from first principles; however, they contained an ambiguity and

[1] Wigner's original motivation for introducing it was to be able to calculate the quantum correction to the second virial coefficient of a gas, which indicates how it deviates from the ideal gas law. Classically, to calculate the second virial coefficient one needs a joint distribution of position and momentum. So Wigner devised a joint distribution that gave, as marginals, the quantum mechanical distributions of position and momentum. The quantum mechanics came in the distribution, but the distribution was used in the classical manner. It was a hybrid method. We discuss the analogy of quantum mechanics to signal analysis in Section 13.13. Also, Wigner was motivated in part by the work of Kirkwood and Margenau who had previously calculated this quantity but Wigner improved on it. Kirkwood subsequently developed what is now the standard theory for nonequilibrium statistical mechanics, the BBGKY Hierarchy (the theory was developed independently by Bogoliubov, Born and Green, Kirkwood, and Yvon). Kirkwood attempted to extend the classical theory to the quantum case and devised the distribution commonly called the Rihaczek or Margenau-Hill distribution to do that. Many years later Margenau and Hill derived the Margenau-Hill distribution. The importance of the Margenau-Hill work is not the distribution but the derivation. They were also the first to consider joint distributions involving spin.

the generalization of the method, which we shall call the characteristic operator method, can be used to derive all distributions. This is discussed in Chapters 10 and 17. In a comprehensive study, Mark[368] pointed out the cross term issue of the Wigner distribution and also showed the relation of the Wigner distribution to the spectrogram. In 1980, in an important set of papers, Claasen and Mecklenbräuker[118, 119, 120] developed a comprehensive approach and originated many new ideas and procedures uniquely suited to the time-frequency situation.

Calculating properties of the Wigner distribution is fairly straightforward. However, they can be easily determined by the methods we will develop to study the general class of distributions in Chapter 9. Hence in this chapter we do not delve into detailed proofs for every property, but emphasize the interpretation of the results.

8.2 THE WIGNER DISTRIBUTION

The Wigner distribution in terms of the signal, $s(t)$ or its spectrum, $S(\omega)$, is

$$W(t,\omega) = \frac{1}{2\pi}\int s^*(t-\tfrac{1}{2}\tau)\, s(t+\tfrac{1}{2}\tau)\, e^{-j\tau\omega}\, d\tau \tag{8.1}$$

$$= \frac{1}{2\pi}\int S^*(\omega+\tfrac{1}{2}\theta)\, S(\omega-\tfrac{1}{2}\theta)\, e^{-jt\theta}\, d\theta \tag{8.2}$$

The equivalence of the two expressions is easily checked by writing the signal in terms of the spectrum and substituting into Eq. (8.1). The Wigner distribution is said to be bilinear in the signal because the signal enters twice in its calculation.

Notice that to obtain the Wigner distribution at a particular time we add up pieces made up of the product of the signal at a past time multiplied by the signal at a future time, the time into the past being equal to the time into the future. Therefore, to determine the properties of the Wigner distribution at a time t we mentally fold the left part of the signal over to the right to see if there is any overlap. If there is, then those properties will be present now, at time t. If this simple point is kept in mind, many issues and results regarding the Wigner distribution become clear. For example, suppose we have a finite duration signal with noise appearing only for a small part of the time. (We use noise only for the purpose of illustration and our remarks hold for any other property of the signal.) Now let us pick a time and ask whether noise will appear at that time. Fold the signal at that time and if in the folding the noise is overlapped then noise will appear at the present time even though there is no noise in the signal at this time. This is illustrated in Fig. 8.1.

Everything we have said for the time domain holds for the frequency domain because the Wigner distribution is basically identical in form in both domains. Another important point is that the Wigner distribution weighs the far away times equally to the near times. Hence the Wigner distribution is highly nonlocal.

Range of the Wigner Distribution. The above argument shows that for an infinite duration signal, the Wigner distribution will be nonzero for all time, since no matter

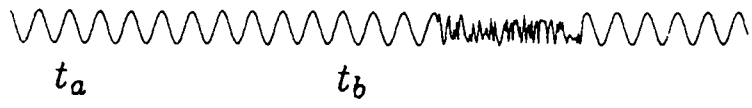

Fig. 8.1 An easy way to ascertain the behavior of the Wigner distribution is to mentally fold over the signal about the time being considered and determine whether the overlap includes properties of interest. We have illustrated a finite duration signal where noise appears for a small time interval. At time t_a no noise will appear in the Wigner distribution because if we fold over the signal about t_a there is no noise in the overlap. Now consider time t_b. Folding over does result in overlap with the noise and thus noise will apear at time t_b even though there is no noise at that time. Although we have used noise for illustration, the same arguments apply to any other property.

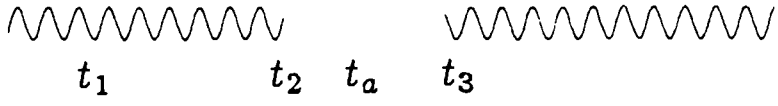

Fig. 8.2 The Wigner distribution is not generally zero when the signal is zero and similarly it is not zero for values of frequency at places where the spectrum is zero. To see this, consider the signal illustrated, which is zero for the interval (t_2, t_3). Now focus at $t = t_a$. Mentally fold over the signal and note that we have an overlap. Hence the Wigner distribution will not be zero at t_a even though the signal is zero then.

what time we choose, the folding of the right with the left parts will result in a nonzero answer. Now let us consider a signal that has a beginning in time and calculate the Wigner distribution for a time before the signal starts. Mentally folding over the right with the left we will get zero, since there is nothing to the left to fold. Hence the Wigner distribution will be zero for times before the signal starts. Also, for a signal that stops, the Wigner distribution will be zero after that time. For a finite duration signal the Wigner distribution will be zero before the start of the signal and after the end. The same arguments in the spectral domain show that for a bandlimited signal the Wigner distribution will be zero for frequencies outside the band. Therefore, the Wigner distribution satisfies the finite support properties in time and frequency,

$$W(t,\omega) = 0 \quad \text{for } t \text{ outside } (t_1, t_2) \qquad \text{if } s(t) \text{ is zero outside } (t_1, t_2) \tag{8.3}$$

$$W(t,\omega) = 0 \quad \text{for } \omega \text{ outside } (\omega_1, \omega_2) \qquad \text{if } S(\omega) \text{ is zero outside } (\omega_1, \omega_2) \tag{8.4}$$

Now consider a signal that is turned off for a finite time and then turned on and let us focus on a time for which the signal is zero, as illustrated in Fig. 8.2. Will the Wigner distribution be zero? No, because if we fold over the right side with the left side we do not obtain zero. Similar consideration applies to the spectrum. Therefore, generally, the Wigner distribution is not necessarily zero at times when the signal is zero and it is not necessarily zero for frequencies that do not exist in the spectrum. Manifestations of this phenomenon have sometimes been called interference or cross terms and the cause for this behavior has very often been attributed to the fact that the Wigner distribution is bilinear in the signal. It is not bilinearity as such that is doing it since there are other joint distributions that are bilinear, satisfy the marginals, but are always zero when the signal is zero. Particular illustrations of this effect are presented in the examples and also when we discuss multicomponent signals.

The Characteristic Function of the Wigner Distribution. We have

$$M(\theta,\tau) = \iint e^{j\theta t + j\tau\omega}\, W(t,\omega)\, dt\, d\omega \tag{8.5}$$

$$= \frac{1}{2\pi}\iiint e^{j\theta t + j\tau\omega} s^*(t-\tfrac{1}{2}\tau')\, s(t+\tfrac{1}{2}\tau')\, e^{-j\tau'\omega}\, d\tau'\, dt\, d\omega \tag{8.6}$$

$$= \iint e^{j\theta t}\delta(\tau-\tau') s^*(t-\tfrac{1}{2}\tau')\, s(t+\tfrac{1}{2}\tau')\, d\tau'\, dt \tag{8.7}$$

$$= \int s^*(t-\tfrac{1}{2}\tau)\, s(t+\tfrac{1}{2}\tau)\, e^{j\theta t}\, dt \tag{8.8}$$

$$= A(\theta,\tau) \tag{8.9}$$

This function and variants of it have played a major role in signal analysis. This particular form is called the symmetric ambiguity function. It was first derived by Ville and Moyal and its relation to matched filters was developed by Woodward. We have previously discussed it in the calculation of the characteristic function of the spectrogram, Eq. (7.12). In terms of the spectrum the characteristic function is

$$M(\theta,\tau) = \int S^*(\omega+\tfrac{1}{2}\theta)\, S(\omega-\tfrac{1}{2}\theta)\, e^{j\tau\omega}\, d\omega \tag{8.10}$$

Nonpositivity. We mentioned in Chapter 6 that a bilinear distribution that satisfies the marginals cannot be positive throughout the time-frequency plane; it must go negative somewhere. As we will see, the Wigner distribution does satisfy the marginals and hence we expect it to always have regions of negative values for any signal. That is indeed the case, with one glaring exception, the signal given by Eq. (8.43). How is it possible that there can be an exception? The reason is that the Wigner distribution for that signal is not really bilinear and it belongs to the class of positive distributions that are not bilinear. These distributions are considered in Section 14.2.

8.3 GENERAL PROPERTIES

We now discuss the basic properties of the Wigner distribution.

Reality. The Wigner distribution is always real, even if the signal is complex. This can be verified by considering the complex conjugate of $W(t,\omega)$,

$$W^*(t,\omega) = \frac{1}{2\pi}\int s(t-\tfrac{1}{2}\tau)\, s^*(t+\tfrac{1}{2}\tau)\, e^{j\tau\omega}\, d\tau \tag{8.11}$$

$$= -\frac{1}{2\pi}\int_{\infty}^{-\infty} s(t+\tfrac{1}{2}\tau)\, s^*(t-\tfrac{1}{2}\tau)\, e^{-j\tau\omega}\, d\tau \tag{8.12}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} s(t+\tfrac{1}{2}\tau)\, s^*(t-\tfrac{1}{2}\tau)\, e^{-j\tau\omega}\, d\tau \tag{8.13}$$

$$= W(t,\omega) \tag{8.14}$$

The fact that the Wigner distribution is real for any signal can also be seen from the characteristic function. Recall that $M^*(-\theta,-\tau) = M(\theta,\tau)$ is the condition for a distribution to be real. But the characteristic function of the Wigner distribution is the ambiguity function, $A(\theta,\tau)$, Eq. (8.8), which does satisfy this property.

Symmetry. Substituting $-\omega$ for ω into the Wigner distribution we see that we obtain the identical form back if the signal is real. But real signals have symmetrical spectra. Therefore, for symmetric spectra the Wigner distribution is symmetrical in the frequency domain. Similarly for real spectra the time waveform is symmetrical and the Wigner distribution is symmetric in time. Therefore,

$$W(t,\omega) = W(t,-\omega) \text{ for real signals} \equiv \text{symmetrical spectra, } S(\omega) = S(-\omega) \tag{8.15}$$

$$W(t,\omega) = W(-t,\omega) \text{ for real spectra} \equiv \text{symmetrical signals, } s(t) = s(-t) \tag{8.16}$$

Marginals. The Wigner distribution satisfies the time-frequency marginals

$$\int W(t,\omega)\, d\omega = |\, s(t)\,|^2 \tag{8.17}$$

$$\int W(t,\omega)\, dt = |\, S(\omega)\,|^2 \tag{8.18}$$

Both of these equations can be readily verified by examining $M(\theta,0)$ and $M(0,\tau)$. By inspection of Eq. (8.8) and Eq. (8.10) we have

$$M(\theta,0) = \int |\, s(t)\,|^2 e^{i\theta t}\, dt \quad ; \quad M(0,\tau) = \int |\, S(\omega)\,|^2 e^{i\tau\omega}\, d\omega \tag{8.19}$$

But these are the characteristic functions of the marginals and hence the marginals are satisfied. To do it directly,

$$P(t) = \int W(t,\omega)\,d\omega = \frac{1}{2\pi}\iint s^*(t-\tfrac{1}{2}\tau)\,s(t+\tfrac{1}{2}\tau)\,e^{-j\tau\omega}\,d\tau d\omega \quad (8.20)$$

$$= \int s^*(t-\tfrac{1}{2}\tau)\,s(t+\tfrac{1}{2}\tau)\,\delta(\tau)\,d\tau \quad (8.21)$$

$$= |\,s(t)\,|^2 \quad (8.22)$$

and similarly for the marginal in frequency. Since the marginals are satisfied, the total energy condition is also automatically satisfied,

$$E = \iint W(t,\omega)\,d\omega\,dt = \int |\,s(t)\,|^2\,d\tau = 1 \quad (8.23)$$

Time and Frequency Shifts. If we time shift the signal by t_0 and/or shift the spectrum by ω_0, then the Wigner distribution is shifted accordingly,

$$\text{if} \quad s(t) \rightarrow e^{j\omega_0 t}s(t-t_0) \quad \text{then} \quad W(t,\omega) \rightarrow W(t-t_0,\omega-\omega_0) \quad (8.24)$$

To see this we replace the signal by $e^{j\omega_0 t}\,s(t-t_0)$ in the Wigner distribution and call W_{sh} the shifted distribution,

$$W_{sh}(t,\omega) = \frac{1}{2\pi}\int e^{-j\omega_0(t-\tau/2)}\,s^*(t-t_0-\tfrac{1}{2}\tau)$$

$$\times \quad e^{j\omega_0(t+\tau/2)}\,s(t-t_0+\tfrac{1}{2}\tau)\,e^{-j\tau\omega}\,d\tau \quad (8.25)$$

$$= \frac{1}{2\pi}\int s^*(t-t_0-\tfrac{1}{2}\tau)\,s(t-t_0+\tfrac{1}{2}\tau)\,e^{-j\tau(\omega-\omega_0)}\,d\tau \quad (8.26)$$

$$= W(t-t_0,\omega-\omega_0) \quad (8.27)$$

8.4 GLOBAL AVERAGES

The global average of a function g is

$$\langle\, g(t,\omega)\,\rangle = \iint g(t,\omega)\,W(t,\omega)\,d\omega\,dt \quad (8.28)$$

Since the Wigner distribution satisfies the marginals it will give the correct answer for averages that are only functions of time or frequency or the sum of two such functions

$$\langle\, g_1(t) + g_2(\omega)\,\rangle = \iint \{g_1(t) + g_2(\omega)\}\,W(t,\omega)\,d\omega\,dt \quad (8.29)$$

$$= \int g_1(t)\,|\,s(t)\,|^2\,dt + \int g_2(\omega)\,|\,S(\omega)\,|^2\,d\omega \qquad (8.30)$$

always correctly given since averages

are calculated with correct marginals

Mean Time, Mean Frequency, Duration, Bandwidth, and the Uncertainty Principle. All these quantities are automatically satisfied since all these quantities depend on the marginals only, which are correctly given by the Wigner distribution.

Correlation and Covariance. The first mixed moment of the Wigner distribution is

$$\langle t\omega \rangle = \iint t\,\omega\,W(t,\omega)\,dt\,d\omega = \int t\,\varphi'(t)\,|\,s(t)\,|^2\,dt \qquad (8.31)$$

Therefore the Wigner distribution gives the covariance of a signal as defined by Eq. (1.124).

8.5 LOCAL AVERAGES

The first conditional moments of time and frequency are

$$\langle \omega \rangle_t = \frac{1}{|\,s(t)\,|^2}\int \omega\,W(t,\omega)\,d\omega \quad ; \quad \langle t \rangle_\omega = \frac{1}{|\,S(\omega)\,|^2}\int t\,W(t,\omega)\,dt \qquad (8.32)$$

When these are evaluated we obtain,

$$\langle \omega \rangle_t = \varphi'(t) \quad ; \quad \langle t \rangle_\omega = -\psi'(\omega) \qquad (8.33)$$

where ϕ and ψ are the phase and spectral phase of the signal. But $\varphi'(t)$ and $-\psi'(\omega)$ are the instantaneous frequency and group delay. These are important results because they are always true for any signal. Recall that for the spectrogram they were never correctly given, although we could approximate one or the other. Furthermore, if we tried to make one of the two relations approximately true by narrowing the window, the other relation would become very poor.

Local Spread. The result just obtained shows that instantaneous frequency is an average, the conditional average for a particular time. We now consider the spread about that average, the conditional standard deviation. First consider the second conditional moment in frequency,

$$\langle \omega^2 \rangle_t = \frac{1}{|\,s(t)\,|^2}\int \omega^2\,W(t,\omega)\,d\omega \qquad (8.34)$$

$$= \frac{1}{2}\left[\left(\frac{A'(t)}{A(t)}\right)^2 - \frac{A''(t)}{A(t)}\right] + \varphi'^2(t) \qquad (8.35)$$

where A is the amplitude of the signal. The conditional spread in frequency is

$$\sigma^2_{\omega|t} = \langle \omega^2 \rangle_t - \langle \omega \rangle_t^2 \tag{8.36}$$

$$= \frac{1}{2}\left[\left(\frac{A'(t)}{A(t)}\right)^2 - \frac{A''(t)}{A(t)}\right] \tag{8.37}$$

This expression for $\sigma^2_{\omega|t}$ may go negative and hence cannot be properly interpreted. In fact, for such a case $\sigma_{\omega|t}$ is imaginary. Therefore, while the Wigner distribution gives an excellent result for the average conditional frequency, it gives a very poor one for the spread of those frequencies. Are there distributions that give a plausible answer? The answer is yes and we consider this question in Section 13.2.

8.6 EXAMPLES

Before proceeding with the general properties of the Wigner distribution we consider some examples.

Example 8.1: Sinusoid and Impulse.

For the sinusoid and impulse the Wigner distribution gives an answer totally consistent with intuition,

$$s(t) = e^{j\omega_0 t} \qquad W(t,\omega) = \delta(\omega - \omega_0) \tag{8.38}$$

$$s(t) = \sqrt{2\pi}\delta(t - t_0) \qquad W(t,\omega) = \delta(t - t_0) \tag{8.39}$$

For a sinusoid the Wigner distribution is totally concentrated along the frequency of the sinusoid, and for the impulse it is totally concentrated at the time of occurrence.

Example 8.2: Sinusoid with Gaussian Amplitude.

If we modulate a pure sinusoid with a Gaussian envelope then

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\omega_0 t} \quad ; \quad W(t,\omega) = 1/\pi\, e^{-\alpha t^2 - (\omega-\omega_0)^2/\alpha} \tag{8.40}$$

We see that the Wigner distribution is still concentrated around the single frequency of the sinusoid but now there is spreading. In addition, for any particular value of frequency it falls off as the squared amplitude of the signal. To calculate the conditional spread, Eq. (8.37), we have

$$A'/A = -\alpha t \quad ; \quad A''/A = \alpha^2 t^2 - \alpha \quad ; \quad \sigma_{\omega|t} = \alpha/2 \tag{8.41}$$

Note that this expression for the conditional spread as given by the Wigner distribution is always the case for a Gaussian envelope since the conditional spread does not depend on the phase of the signal. However, other distributions will give different answers. See Section 13.2.

Example 8.3: Pure Chirp.

For the chirp with no amplitude modulation

$$s(t) = e^{j\beta t^2/2 + j\omega_0 t} \quad ; \quad W(t,\omega) = \delta(\omega - \beta t - \omega_0) \tag{8.42}$$

The instantaneous frequency is $\omega_i = \beta t + \omega_0$ and therefore, for this case, the Wigner distribution is totally concentrated along the instantaneous frequency.

Example 8.4: Chirp With Gaussian Amplitude.

Now place a Gaussian envelope on the chirp. The Wigner distribution is

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \quad ; \quad W(t,\omega) = \frac{1}{\pi} e^{-\alpha t^2 - (\omega - \beta t - \omega_0)^2/\alpha} \tag{8.43}$$

This is the most general case for which the Wigner distribution is positive and this case encompasses all previous examples. If α is small, then the distribution is concentrated along the $\omega = \omega_0 + \beta t$, which is the instantaneous frequency. In Fig. 8.3 we plot the Wigner distribution.

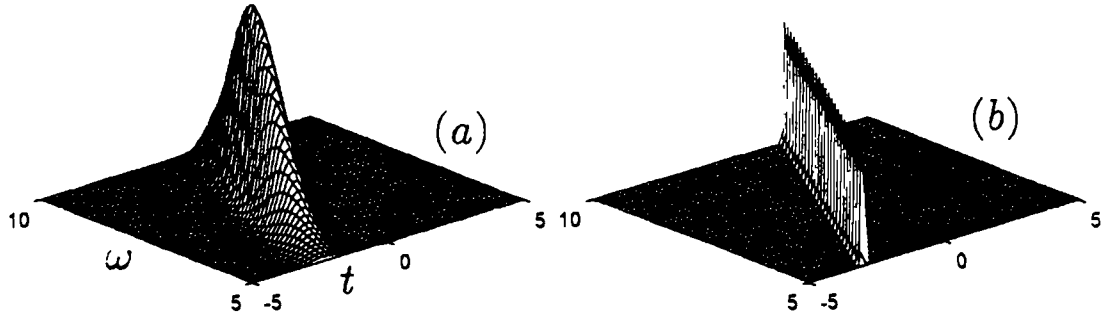

Fig. 8.3 The Wigner distribution for the signal $s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t}$. This is the most general signal for which the Wigner distribution is positive throughout the time-frequency plane. As the envelope becomes more and more flat the Wigner distribution gets more and more concentrated along the instantaneous frequency, as shown in (b). In (a) and (b) $\alpha = 1$ and $\alpha = 0.001$ respectively. In both figures $\omega_0 = 8$ and $\beta = 1$.

Example 8.5: The Signal: $t\, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t}$.

Now take the signal considered above and multiply it by time,

$$s(t) = (4\alpha^3/\pi)^{1/4}\, t\, e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \tag{8.44}$$

where we have also normalized it. The Wigner distribution is

$$W(t,\omega) = \frac{2}{\pi}\left[\alpha t^2 + (\omega - \beta t - \omega_0)^2/\alpha - \tfrac{1}{2}\right] e^{-\alpha t^2 - (\omega - \beta t - \omega_0)^2/\alpha} \tag{8.45}$$

and is plotted in Fig. 8.4. It is negative whenever

$$\alpha t^2 + (\omega - \beta t - \omega_0)^2/\alpha \leq \tfrac{1}{2} \tag{8.46}$$

This is the typical situation. The Wigner distribution goes negative somewhere for every signal, with the exception of the signal given by Eq. (8.43).

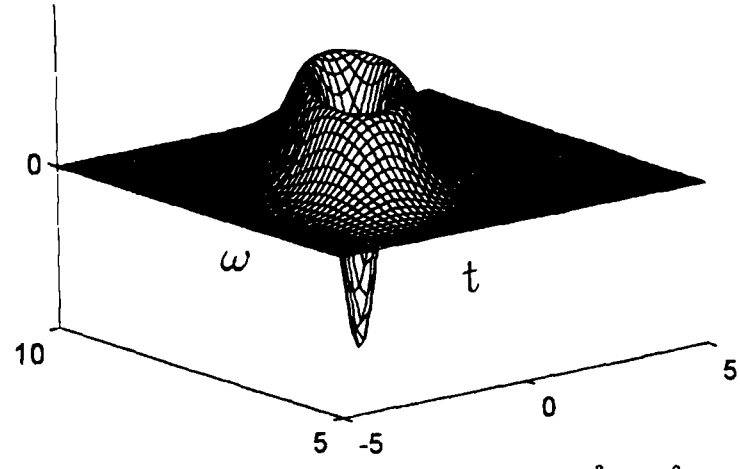

Fig. 8.4 The Wigner distribution for the signal $s(t) = te^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t}$. Everything below the plane is negative.

Example 8.6: Cubic Phase.

For the signal

$$s(t) = e^{j\varphi(t)} \qquad ; \qquad \varphi(t) = \gamma^3 t^3/3 + \beta t^2/2 + \omega_0 t \tag{8.47}$$

we have

$$-\varphi(t - \tfrac{1}{2}\tau) + \varphi(t + \tfrac{1}{2}\tau) = \gamma\tau^3/12 + (\gamma t^2 + \beta t + \omega_0)\tau \tag{8.48}$$

$$= \gamma\tau^3/12 + \varphi'(t)\tau \tag{8.49}$$

and therefore the Wigner distribution is

$$W(t,\omega) = \frac{1}{2\pi}\int e^{j\gamma\tau^3/12 + j(\varphi'(t)-\omega)\tau}\, d\tau \tag{8.50}$$

$$= \frac{2}{2\pi}\left(\frac{4}{\gamma}\right)^{1/3}\int \cos\left[\gamma\tau^3/12 + j(\varphi'(t)-\omega)\tau\right] dt \tag{8.51}$$

$$= \left(\frac{4}{\gamma}\right)^{1/3} \mathrm{Ai}\left((4/\gamma)^{1/3}[\varphi'(t)-\omega]\right) \tag{8.52}$$

where $\mathrm{Ai}(x)$ is the Airy function

$$\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^\infty \cos(u^3/3 + xu)\, du \tag{8.53}$$

Unlike the quadratic case, this distribution is not a delta function of $\omega - \varphi'(t)$. However, it is concentrated along $\omega = \varphi'(t)$.

Example 8.7: Finite Duration Chirp.

For the finite duration chirp

$$s(t) = e^{j\beta t^2/2 + j\omega_0 t} \qquad -T/2 \le t \le T/2 \tag{8.54}$$

the Wigner distribution is calculated to be

$$W(t,\omega) = \frac{1}{\pi(\omega - \beta t - \omega_0)} \begin{cases} \sin[(\omega - \beta t - \omega_0)(t + T/2)] & -T/2 \le t \le 0 \\ \sin[(\omega - \beta t - \omega_0)(T/2 - t)] & 0 \le t \le T/2 \\ 0 & \text{otherwise} \end{cases} \tag{8.55}$$

Note that the Wigner distribution goes to zero at the end points.

Example 8.8: Sinusoidal Frequency Modulation.

The signal

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + jm \sin \omega_m t + j\omega_0 t} \tag{8.56}$$

has an instantaneous frequency given by $\omega_i = \beta t + \omega_0 + m\omega_m \cos \omega_m t$. The Wigner distribution can be obtained analytically. Substituting the signal into the definition of the Wigner distribution we obtain

$$W(t,\omega) = \frac{1}{2\pi} \left(\frac{\alpha}{\pi}\right)^{1/2} e^{-\alpha t^2} \int e^{-\alpha\tau^2/4 + 2jm\cos(\omega_m t)\sin(\omega_m \tau/2) - j\tau(\omega - \omega_0 - \beta t)}\, d\tau \tag{8.57}$$

Now

$$e^{2jm\cos(\omega_m t)\sin(\omega_m \tau/2)} = \sum_{n=-\infty}^{\infty} J_n(2m\cos(\omega_m t))\, e^{jn\omega_m \tau/2} \tag{8.58}$$

where J_n is the Bessel function of order n. Substituting this into equation Eq. (8.57) and integrating we have

$$W(t,\omega) = \frac{1}{\pi} e^{-\alpha t^2} \sum_{n=-\infty}^{\infty} J_n(2m\cos(\omega_m t))\, e^{-(\omega - \omega_0 - \beta t - n\omega_m/2)^2/\alpha} \tag{8.59}$$

This can be interpreted as the Wigner distribution of the sum of Gaussians each with instantaneous frequency $\omega_i = \omega_0 + n\omega_m/2 + \beta t$.

Complex Signals and Real Signals. The reader may have noticed that all the examples presented thus far have been for complex signals. The reason is that there is an effective way to think of the Wigner distribution for the sum of two signals. A

real signal can always be thought of as the sum of a complex signal plus its complex conjugate. We therefore leave the consideration of real signals until we develop some simple results regarding the Wigner distribution for the sum of two signals.

8.7 THE WIGNER DISTRIBUTION OF THE SUM OF TWO SIGNALS

The Cross Wigner Distribution. Suppose we express a signal as the sum of two pieces,

$$s(t) = s_1(t) + s_2(t) \tag{8.60}$$

Substituting this into the definition, we have

$$W(t,\omega) = W_{11}(t,\omega) + W_{22}(t,\omega) + W_{12}(t,\omega) + W_{21}(t,\omega) \tag{8.61}$$

where

$$W_{12}(t,\omega) = \frac{1}{2\pi}\int s_1^*(t-\tfrac{1}{2}\tau)\, s_2(t+\tfrac{1}{2}\tau)\, e^{-j\tau\omega}\, d\tau \tag{8.62}$$

This is called the cross Wigner distribution. In terms of the spectrum it is

$$W_{12}(t,\omega) = \frac{1}{2\pi}\int S_1^*(\omega+\tfrac{1}{2}\theta)\, S_2(\omega-\tfrac{1}{2}\theta)\, e^{-jt\theta}\, d\theta \tag{8.63}$$

The cross Wigner distribution is complex. However, $W_{12} = W_{21}^*$, and therefore $W_{12}(t,\omega) + W_{21}(t,\omega)$ is real. Hence

$$W(t,\omega) = W_{11}(t,\omega) + W_{22}(t,\omega) + 2\,\mathrm{Re}\,\{W_{12}(t,\omega)\} \tag{8.64}$$

We see that the Wigner distribution of the sum of two signals is not the sum of the Wigner distribution of each signal but has the additional term $2\,\mathrm{Re}\,\{W_{12}(t,\omega)\}$. This term is often called the interference term or the cross term and it is often said to give rise to artifacts. However, one has to be cautious with the images these words evoke, because any signal can be broken up into an arbitrary number of parts and the so-called cross terms are therefore not generally unique and do not characterize anything but our own division of a signal into parts. Sometimes there is a natural decomposition where the self terms and cross terms take on special meaning. This question is addressed in the next section and in Section 13.3.

Example 8.9: Sum of Two Sinusoids.

We take

$$s(t) = A_1\, e^{j\omega_1 t} + A_2\, e^{j\omega_2 t} \tag{8.65}$$

The Wigner distribution is

$$W(t,\omega) = A_1^2\delta(\omega-\omega_1) + A_2^2\delta(\omega-\omega_2) + 2A_1A_2\,\delta(\omega-\tfrac{1}{2}(\omega_1+\omega_2))\,\cos(\omega_2-\omega_1)t \tag{8.66}$$

The three terms correspond to the ones just discussed for Eq. (8.64). Besides the concentration at ω_1 and ω_2, we also have nonzero values at the frequency $\frac{1}{2}(\omega_1+\omega_2)$.

This is an illustration of the cross term. It is sharp at $\omega = \frac{1}{2}(\omega_1 + \omega_2)$ because that is the only value for which there is an overlap, since the signal is sharp at both ω_1 and ω_2.

Example 8.10: Cosine.

For a cosine wave, $s(t) = \cos \omega_0 t$, we specialize Example 8.9 by taking $A_2 = A_1 = 1/2$ and $\omega_1 = -\omega_2 = \omega_0$, which gives

$$W(t,\omega) = \delta(\omega - \omega_0) + \delta(\omega + \omega_0) + 2\,\delta(\omega)\,\cos 2\omega_0 t \qquad [s(t) = \cos \omega_0 t] \tag{8.67}$$

Example 8.11: Sinusoid Plus Impulse.

Now consider the sum of a sinusoid and impulse,

$$s(t) = e^{j\omega_0 t} + \sqrt{2\pi}\delta(t - t_0) \tag{8.68}$$

Straightforwardly we obtain

$$W(t,\omega) = \delta(\omega - \omega_0) + \delta(t - t_0) + \frac{1}{\sqrt{2\pi}} \cos[2(\omega - \omega_0)(t - t_0) - \omega_0 t_0] \tag{8.69}$$

Comparing this equation to the spectrogram, Eq. (7.69), we see that the self terms are infinitely sharp – something the spectrogram cannot achieve. However, the cross terms in the Wigner distribution are not hidden by the self terms as is the case in the spectrogram, as per the discussion after Eq. (7.69). In Fig. 8.5 we plot this distribution. We also give the ambiguity function for this case,

$$A(\theta,\tau) = 2\pi e^{j\omega_0 \tau}\delta(\theta) + 2\pi e^{j\theta t_0}\delta(\tau) + \sqrt{2\pi}\, e^{-j\omega_0(t_0 - \tau) + j\theta(t_0 - \tau/2)} + \text{c.c.} \tag{8.70}$$

where c.c. stands for the complex conjugate of the preceding terms.

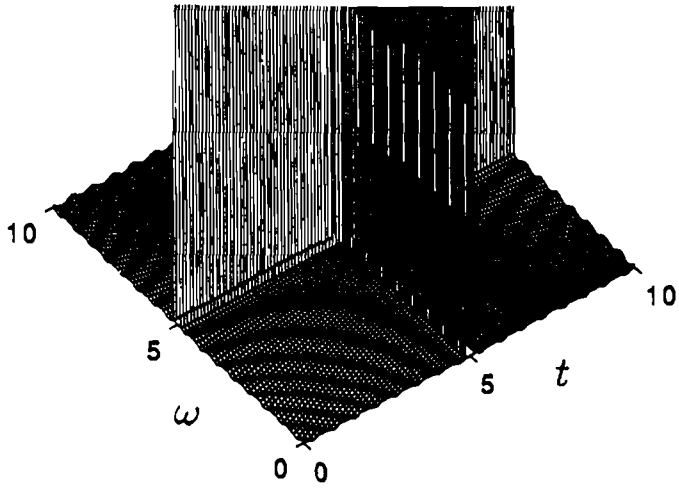

Fig. 8.5 The Wigner distribution of $s(t) = e^{j10t} + \delta(t - 10)$. It is infinitely peaked in time and frequency at the indicated values. The ripples are the cross terms. Compare this with the spectrogram, Fig. 7.1. Note the spectrogram also has cross terms but they are mostly hidden under the self terms.

Example 8.12: Sum of Two Tones with Amplitude Modulation.

Now if we put amplitude modulation on each part

$$s(t) = A_1 (\alpha_1/\pi)^{1/4} e^{-\alpha_1 t^2/2 + j\omega_1 t} + A_2 (\alpha_2/\pi)^{1/4} e^{-\alpha_2 t^2/2 + j\omega_2 t} \tag{8.71}$$

where the modulation extends over all time, the cross terms will extend over all times and frequencies. The Wigner distribution can be calculated exactly,

$$\begin{aligned} W(t,\omega) &= \frac{A_1^2}{\pi} e^{-\alpha_1 t^2 - (\omega-\omega_1)^2/\alpha_1} + \frac{A_2^2}{\pi} e^{-\alpha_2 t^2 - (\omega-\omega_2)^2/\alpha_2} \\ &\quad + 2\frac{A_1 A_2}{\pi} \sqrt{\frac{2(\alpha_1\alpha_2)^{1/2}}{\alpha_1+\alpha_2}} \cos\left[\frac{2t}{\alpha_1+\alpha_2}(\omega + \omega_2\alpha_1 - \omega_1\alpha_2)\right] \\ &\quad \times \exp\left[-\frac{2}{\alpha_1+\alpha_2}\left\{\alpha_1\alpha_2 t^2 + [\omega - \tfrac{1}{2}(\omega_1+\omega_2)]^2\right\}\right] \end{aligned} \tag{8.72}$$

Cross Terms, Interference Terms, Ghosts, Artifacts, Multicomponent Signals. From the examples we have just considered we see that the Wigner distribution sometimes places values in the middle of the two signals both in time and in frequency. Sometimes these values are in places in the time frequency plane at odds with what is expected. A typical case is illustrated in Fig. 8.6. Because of this a language has arisen regarding the cross terms and very often phrases like artifacts, ghosts, and interference terms are used, words chosen to describe something apparently undesirable. The implication is that since these effects are due to the cross terms, the cross terms themselves are undesirable.

It is important to clarify these issues. First let us make clear that it is not generally true that the cross terms produce undesirable effects. Any signal can be broken up into parts, in an infinite number of ways. If we have a signal for which we are pleased with the answer the Wigner distribution gives and someone comes along and expresses the signal in terms of two parts, will the cross terms be undesirable? Quite the contrary, they are highly desirable, for without them we would not get the pleasing answer. In fact, since any signal can be broken up into a sum of parts in an arbitrary way, the cross terms can be neither bad nor good since they are not uniquely defined; they are different for different decompositions. The Wigner distribution does not know about cross terms, since the breaking up of a signal into parts is not unique.

There are decompositions for which we think the parts are special in the sense that we think they should be concentrated in certain regions of the time-frequency plane, hence if we have two such signals we also sense that they should be well delineated in the time-frequency plane and that no other terms should appear. If a signal is indeed well delineated in a region we shall call it monocomponent, and if there is more than one well delineated region we shall call it a multicomponent signal. Very often, in the literature, a signal has been called multicomponent when it is expressed as the sum of two parts. This is not appropriate, since a signal can be broken up into parts in an infinite number of arbitrary ways. We shall reserve the

phrase "multicomponent signal" to mean a signal that in the time-frequency plane is well delineated into regions and not simply to mean that it has been expressed as the sum of parts. One of the strengths of the Wigner distribution is that it indeed does indicate when a signal is multicomponent. We emphasize that when we feed a signal into the Wigner distribution (and others) we do not tell it that it is multicomponent – it tells us, and that is one of the accomplishments of time-frequency analysis. However, these considerations do not answer the question as to what type of signals produce well delineated regions in the time-frequency plane. This is a fundamental question which is addressed in Section 13.3.

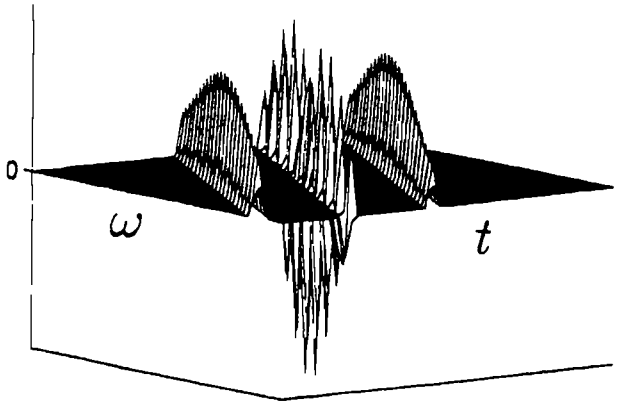

Fig. 8.6 The Wigner distribution of the sum of two chirps illustrating the cross terms.

8.8 ADDITIONAL PROPERTIES

Inversion and Uniqueness. The Wigner distribution is the Fourier transform of $s^*(t-\frac{1}{2}\tau)\ s(t+\frac{1}{2}\tau)$ with respect to τ. Inverting

$$s^*(t-\tfrac{1}{2}\tau)\ s(t+\tfrac{1}{2}\tau) = \int W(t,\omega)\, e^{j\tau\omega}\, d\omega \tag{8.73}$$

and taking the specific value, $t = \tau/2$, letting $k = 1/s^*(0)$, and then setting $\tau = t$, we have

$$s(t) = k \int W(t/2,\omega)\, e^{jt\omega}\, d\omega \tag{8.74}$$

which shows that the signal can be recovered from the Wigner distribution up to a constant. The constant can be obtained from the normalization condition up to an arbitrary constant phase factor. A constant phase factor in the signal clearly drops out when we calculate the Wigner distribution, since we are multiplying the signal times its complex conjugate. Therefore it can never be recovered.

Representability. Not every function of time and frequency is a proper Wigner distribution because there may not exist a signal that will generate it. If a two dimensional time-frequency function is generated from some signal, we shall say that it is

representable or realizable. It is not always easy to ascertain whether a function is representable or not, although sometimes it is. For example, any manifestly positive two dimensional function is not a representable Wigner function unless it happens to be of the form given by Eq. (8.43). This is because the Wigner distribution always goes negative somewhere, the only exception being Eq. (8.43). Generally, a way to ascertain whether a two dimensional function is a Wigner distribution is to assume that it is, use the inversion formula to find the signal, and then calculate the Wigner distribution from the derived signal. If we recover the same function then it is a representable Wigner function.

Example 8.13: Representability of the Wigner Distribution.

Take the two dimensional function

$$G(t,\omega) = e^{-t^2/\alpha-\omega^2/\beta+2\gamma\omega t} \tag{8.75}$$

For what values of α, β, γ can this be a Wigner distribution? Assume that it is and obtain the signal using the inversion formula, Eq. (8.74),

$$s(t) = k\int G(t/2,\omega)\, e^{jt\omega}\, d\omega \sim e^{-[1/\alpha-\beta^2(\gamma+j)]t^2/4} \tag{8.76}$$

Now calculate the Wigner distribution of this signal to obtain

$$W(t,\omega) \sim \exp\left[-(c+\beta^2\gamma^2/c)t^2 - \omega^2/c + 2\gamma\beta\omega t\right] \qquad [c = \tfrac{1}{2}(1/\alpha+\beta-\beta\gamma^2)] \tag{8.77}$$

When does $W(t,\omega) = G(t,\omega)$? Only when $c = \beta$ or

$$\frac{1}{\alpha\beta} = 1+\gamma^2 \tag{8.78}$$

Therefore the three constants must be chosen according to Eq. (8.78) if $G(t,\omega)$ is to be a Wigner distribution.

Example 8.14: Wigner Nonrepresentability.

Now consider the function

$$G(t,\omega) = t^2\omega^2\, e^{-\alpha t^2-\beta\omega^2+2\gamma\omega t} \tag{8.79}$$

If the above calculation is repeated we see that there is no way to choose the constants appropriately. There is a much easier way to ascertain that. Note that this G is always nonnegative. We know that cannot be unless the distribution is given by Eq. (8.42), and this G is not of that form. Hence it is not a representable Wigner function.

Overlap of Two Signals – Moyal Formula. In many problems the overlap of two signals enters. It can be expressed as the overlap of their respective Wigner distributions,

$$\left|\int s_1(t)\, s_2^*(t)\, dt\right|^2 = 2\pi\iint W_1(t,\omega)\, W_2(t,\omega)\, dt\, d\omega \tag{8.80}$$

This is easily verified and was first shown by Moyal.[390] Of course, if we have two signals it would make no sense to calculate their overlap by first calculating their respective Wigner distributions and then doing the integrals indicated. However, this expression is of theoretical interest and also has been applied to problems of detection by Flandrin.[191]

Range of the Cross Terms. We discussed the range of the Wigner distribution in Section 8.2. Similar considerations establish the range of the cross Wigner distribution. In particular, for two functions, $s_1(t)$ and $s_2(t)$, which are respectively zero outside the intervals (t_1, t_2) and (t_3, t_4), the cross Wigner distribution satisfies the following[253]

$$W_{12}(t,\omega)=0 \qquad \text{for } t \text{ outside } (\tfrac{1}{2}(t_1+t_3), \tfrac{1}{2}(t_2+t_4)) \qquad (8.81)$$
$$\text{if } s_1(t) \text{ is zero outside } (t_1,t_2) \text{ and } s_2(t) \text{ is zero outside } (t_3,t_4)$$

In the frequency domain, for two bandlimited functions, $S_1(\omega)$ and $S_2(\omega)$, which are zero outside the intervals (ω_1, ω_2) and (ω_3, ω_4), respectively, we have

$$W_{12}(t,\omega)=0 \quad \text{for } \omega \text{ outside } (\tfrac{1}{2}(\omega_1+\omega_3), \tfrac{1}{2}(\omega_2+\omega_4)) \qquad (8.82)$$
$$\text{if } S_1(\omega) \text{ is zero outside } (\omega_1,\omega_2) \text{ and } S_2(\omega) \text{ is zero outside } (\omega_3,\omega_4)$$

Product of Two Signals and Product of Two Spectra. The Wigner distribution of the product of two signals $s(t) = s_1(t)\, s_2(t)$ can be written in terms of the Wigner distribution of each signal. The way to do it is to substitute $s_1(t)\, s_2(t)$ into the definition of the Wigner distribution and then use the inversion formula for each signal. This results in

$$W(t,\omega) = \int W_1(t,\omega')W_2(t,\omega-\omega')\,d\omega' \qquad [\text{for } s(t) = s_1(t)\, s_2(t)] \qquad (8.83)$$

Similarly, if we have a spectrum that is the product of two spectra (in which case the signal is the convolution of the two signals), then

$$W(t,\omega) = \int W_1(t',\omega)W_2(t'-t,\omega)\,dt' \qquad [\text{for } S(\omega) = S_1(\omega)S_2(\omega)] \qquad (8.84)$$

Analytic Signal. Since the analytic signal eliminates the negative frequencies, the Wigner distribution of an analytic signal will be zero for the negative frequencies. Also, use of the analytic signal will eliminate the cross terms between the negative and positive parts of the spectrum.

Relation to the Spectrogram. If we convolve, in time and frequency, the Wigner

distribution of the signal with the Wigner distribution of a window, we get the spectrogram

$$P_{SP}(t,\omega) = \iint W_s(t',\omega')W_h(t'-t,\omega-\omega')dt'\,d\omega' \tag{8.85}$$

$$= \left| \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s(\tau)\, h(\tau - t)\, d\tau \right|^2 \tag{8.86}$$

This relation is interesting, but no particular significance should be attached to the fact that we are convolving the *Wigner* distributions, because all time-frequency distributions satisfy a similar relation. Indeed, the best way to prove this result will be from our general considerations in the next chapter.

Noise. The Wigner distribution is very noisy and even places noise at times where there was no noise in the signal. To understand this, consider an infinite duration signal where there was noise for only five minutes. Take any particular time and, using our usual argument, fold the future with the past. The five minutes of noise, no matter when it occurred, will contribute and therefore noise will appear even though there was no noise at the time being considered. This will hold true for any time. For a finite duration signal it will appear at those times when the folding over procedure includes the noise.

Fig. 8.7 illustrates these effects. An analysis of additive white noise and the Wigner distribution has been done by Nuttall.[403]

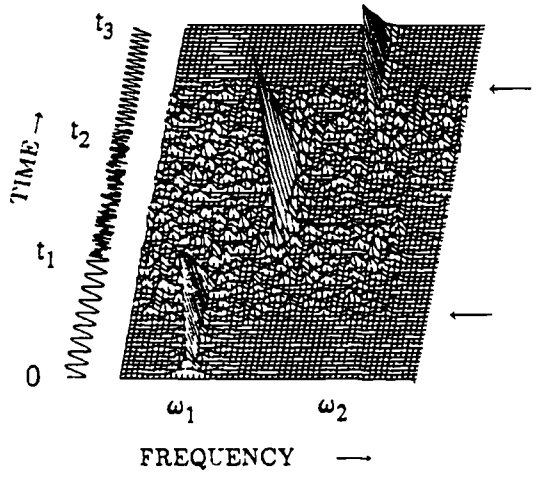

Fig. 8.7 The Wigner distribution is "noisy" because we have cross terms between the noise and between the noise and signal. Note that noise appears at times when there is no noise in the signal.

8.9 PSEUDO WIGNER DISTRIBUTION

Windowing the Lag. For a given time the Wigner distribution weighs equally all times of the future and past. Similarly, for a given frequency it weighs equally all

frequencies below and above that frequency. There are two reasons for wanting to modify this basic property of the Wigner distribution. First, in practice we may not be able to integrate from minus to plus infinity and so one should study the effects of limiting the range. Second, in calculating the distribution for a time t, we may want to emphasize the properties near the time of interest compared to the far away times. To achieve this, note that for a given time the Wigner distribution is the Fourier transform with respect to τ of the quantity $s^*(t-\frac{1}{2}\tau)\,s(t+\frac{1}{2}\tau)$. The variable τ is called the lag variable. Therefore if we want to emphasize the signal around time t, we multiply this product by a function that is peaked around $\tau = 0$, $h(\tau)$ say, to define the pseudo Wigner distribution

$$W_{PS}(t,\omega) = \int h(\tau)s^*(t-\tfrac{1}{2}\tau)\,s(t+\tfrac{1}{2}\tau)\,e^{-j\omega\tau}\,d\tau \tag{8.87}$$

The Wigner distribution is highly nonlocal and the effect of the windowing is to make it less so. One of the consequences of this is that the pseudo Wigner distribution suppresses, to some extent, the cross terms for multicomponent signals. This is because we have made the Wigner distribution local. While windowing the lag does suppress the cross terms, it also destroys many of the desirable properties of the Wigner distribution. For example the marginals and instantaneous frequency properties no longer hold.

Example 8.15: Pseudo Wigner Distribution for a Sine Wave.

Take as an example a pure sine wave and a Gaussian, $h(t)$,

$$s(t) = e^{j\omega_0 t} \qquad h(t) = e^{-at^2/2} \tag{8.88}$$

The pseudo Wigner distribution can be calculated analytically,

$$W_{PS}(t,\omega) = \frac{1}{\sqrt{2\pi a}}\,e^{-(\omega-\omega_0)/(2a)} \tag{8.89}$$

In the non windowed version, Eq. (8.36), the Wigner distribution was totally concentrated at $\omega = \omega_0$. It was $W(t,\omega) = \delta(\omega-\omega_0)$. That is no longer the case and the broadness depends on the window size.

Example 8.16: Windowed Wigner for the Sum of Two Sine Waves.

For the sum of two sine waves

$$s(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \tag{8.90}$$

with the same $h(t)$ as above we have

$$\begin{aligned} W(t,\omega) &= \frac{1}{\sqrt{2\pi\alpha}}\left[A_1^2\,e^{-(\omega-\omega_1)/(2a)} + A_2^2\,e^{-(\omega-\omega_2)/(2a)}\right] \\ &\quad + \frac{2A_1A_2}{\sqrt{2\pi\alpha}}\cos[(\omega_2-\omega_1)t]\,e^{-(\omega-(\omega_1+\omega_2)/2)^2/(2a)} \end{aligned} \tag{8.91}$$

If we choose a small α the cross terms can be made small. However, note that the self terms get spread out.

8.10 MODIFIED WIGNER DISTRIBUTIONS AND POSITIVITY

Historically the main motive for modifying the Wigner distribution was the attempt to achieve a positive distribution. One way to approach this problem is to attempt to smooth it by convolving it with a smoothing function $L(t,\omega)$ to obtain the smoothed Wigner distribution

$$W_{SM}(t,\omega) = \int L(t-t',\omega-\omega')\, W(t',\omega')\, dt'd\omega' \tag{8.92}$$

The first example of a smoothing function was devised by Cartwright,[109]

$$L(t,\omega) = e^{-t^2/\alpha-\omega^2/\beta} \tag{8.93}$$

Substituting this L and the definition of the Wigner distribution into Eq. (8.92) and integrating over ω' results in

$$W_{SM}(t,\omega) = \frac{\sqrt{\beta\pi}}{2\pi}\iint e^{-(t'-t)/\alpha-\beta\tau^2/4-j\omega\tau} s^*(t'-\tfrac{1}{2}\tau)\, s(t'+\tfrac{1}{2}\tau)\, dt'\, d\tau \tag{8.94}$$

Making the change of variables $y = t' - \frac{1}{2}\tau - t; x = t' + \frac{1}{2}\tau - t$, we obtain

$$W_{SM}(t,\omega) = \frac{\sqrt{\beta\pi}}{2\pi}\iint e^{(\beta-1/\alpha)xy/2} f^*(y)\, f(x)\, dx\, dy \quad [f(x) = e^{-(\beta+1/\alpha)x^2/4-j\omega x}] \tag{8.95}$$

Expanding $e^{(\beta-1/\alpha)xy/2}$ in a Taylor series we have

$$W_{SM}(t,\omega) = \frac{\sqrt{\beta\pi}}{2\pi}\sum_{n=0}^{\infty}\frac{1}{n!\,2^n}(\beta-1/\alpha)^n\left|\int f(x)x^n\, dx\right|^2 \tag{8.96}$$

If each term is positive then the sum will be positive. To assure that we must take $\beta - 1/\alpha \geq 0$ or equivalently

$$\alpha\,\beta \geq 1 \tag{8.97}$$

Notice that if $\alpha\beta = 1$, then only the first term survives. But with this condition L is a Wigner distribution, as we showed in Eq. (8.78). In that case, we know from Eq. (8.85) that the resulting distribution is the spectrogram. In the general case, Eq. (8.96) is a sum of spectrograms.

Nuttall[403, 404] has generalized this by considering the smoothing function

$$L(t,\omega) = e^{-t^2/\alpha-\omega^2/\beta-2\gamma\omega t} \tag{8.98}$$

The resulting distribution is positive if $\alpha\beta \geq (1+\gamma^2)^{-1}$.

Smoothing, Transforming, and Convolution. Smoothing or blurring is sometimes desirable, as, for example to eliminate the wrinkles of a portrait. More often, though,

it is undesirable, as exemplified by the blurring of celestial objects by the atmosphere. Since these phenomena are so common there is a vast methodology on the subject. One way to study blurring is to model it as a convolution of the original function with a blurring function. Generally, convolution is not reversible, which can be interpreted as our inability to unblur an image that has been blurred. However, convolution may be reversible. We mention this here because it will turn out that all bilinear distributions may be obtained from the Wigner distribution by a convolution. No significance should be attached to this, since all other distributions may be obtained by convolving with any distribution, not only the Wigner. A distribution that we will be studying is the Choi-Williams distribution which can be obtained by convolving the Wigner distribution with a certain function. It would be wrong to think of the Choi-Williams as a smoothed Wigner distribution because it is also true that the Wigner distribution can be obtained from the Choi-Williams by a convolution. Hence one could argue that the Wigner is a smoothed version of the Choi-Williams. It is better in such situations to think of the process as a reversible transformation. See section 9.7.

8.11 COMPARISON OF THE WIGNER DISTRIBUTION WITH THE SPECTROGRAM

It has often been said that one of the advantages of the Wigner distribution over the spectrogram is that we do not have to bother with choosing the window. This viewpoint misses the essence of the issue. The spectrogram is not one distribution, it is an infinite class of distributions and to say that an advantage is that one does not have to choose makes as much sense as saying one book is better than a library because we don't have to choose which book to read. Here is the point: The Wigner distribution in some respects is better than *any* spectrogram. It is not that we do not have to bother about choosing a window, it is that even if we bothered we wouldn't find one that produces a spectrogram that is better than the Wigner. In particular, the Wigner distribution gives a clear picture of the instantaneous frequency and group delay. In fact, the conditional averages are exactly the instantaneous frequency and group delay. This is always true for the Wigner distribution; it is never true for the spectrogram. We could search forever and never find a window that will produce a spectrogram that will give the instantaneous frequency and group delay, although sometimes a good approximation is achieved.

One of the advantages of the spectrogram is that it is a proper distribution in the sense that it is positive. Because it is manifestly positive, the results obtained from it can be interpreted, although they may be wrong or poor. The Wigner distribution is (with one exception) never manifestly positive, which sometimes leads to results that cannot be interpreted and indeed go against our sensibilities. For example, the conditional standard deviation may be negative.

Finally, the Wigner distribution and spectrograms allow us to ascertain, in most cases, whether a signal is multicomponent. But the Wigner distribution suffers from the fact that for multicomponent signals we get confusing artifacts. On the other

hand, the spectrogram often cannot resolve the components effectively. We mention here that other distributions we will consider keep the desirable properties of the Wigner distribution and have considerably better behavior with regard to the undesirable properties. In Figs. 8.8, 8.9, and 8.10 we give a few examples contrasting the Wigner distribution and spectrogram.

Contrast Regarding Global Averages and the Uncertainty Principle. Since the Wigner distribution satisfies the marginals, it always gives the correct answers for averages of functions of frequency or time and always satisfies the uncertainty principle of the signal. On the other hand, the spectrogram never gives the correct answers for these averages and never satisfies the uncertainty principle of the signal.

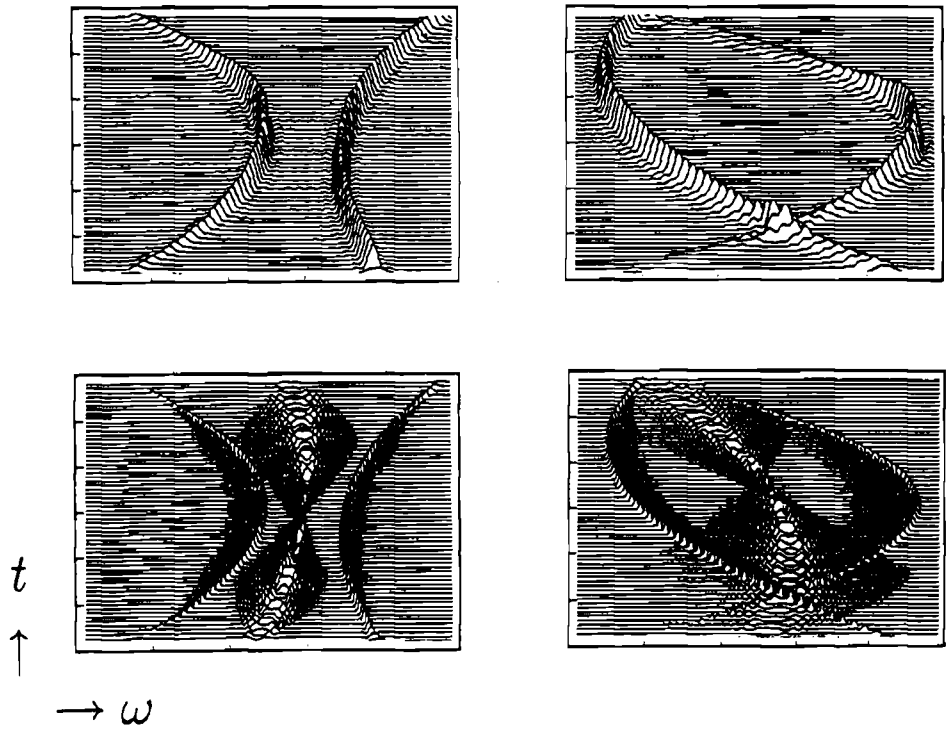

Fig. 8.8 Comparison of the Wigner distribution (bottom) with the spectrogram (top) for two multicomponent signals.

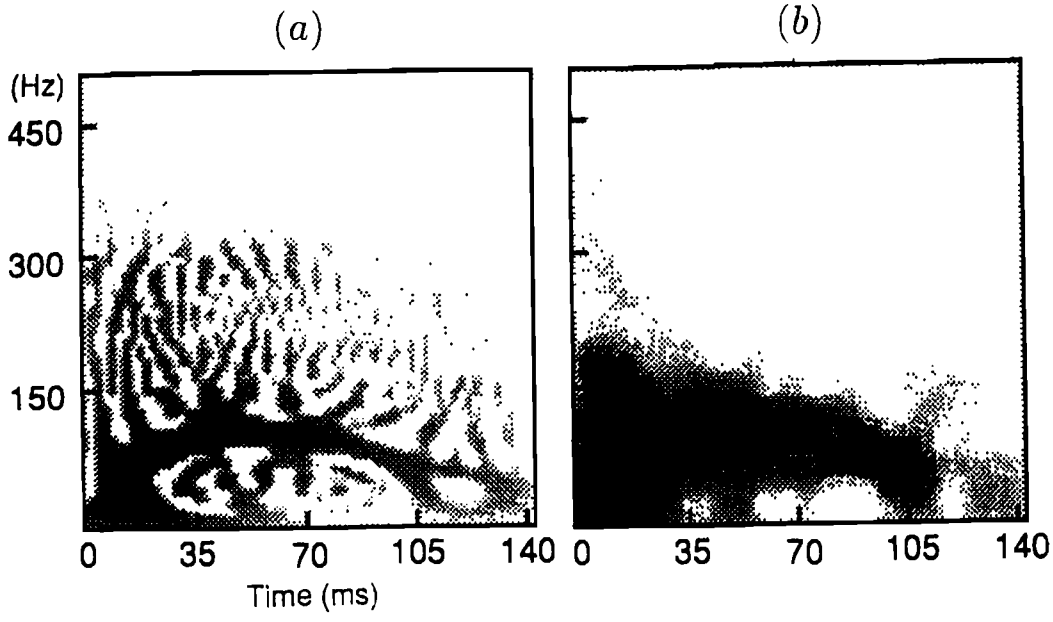

Fig. 8.9 When muscles contract they make a sound. In (a) and (b) we have the Wigner distribution and spectrogram of a typical case. (Courtesy of J. Wood, N. M. Cole and D. T. Barry)

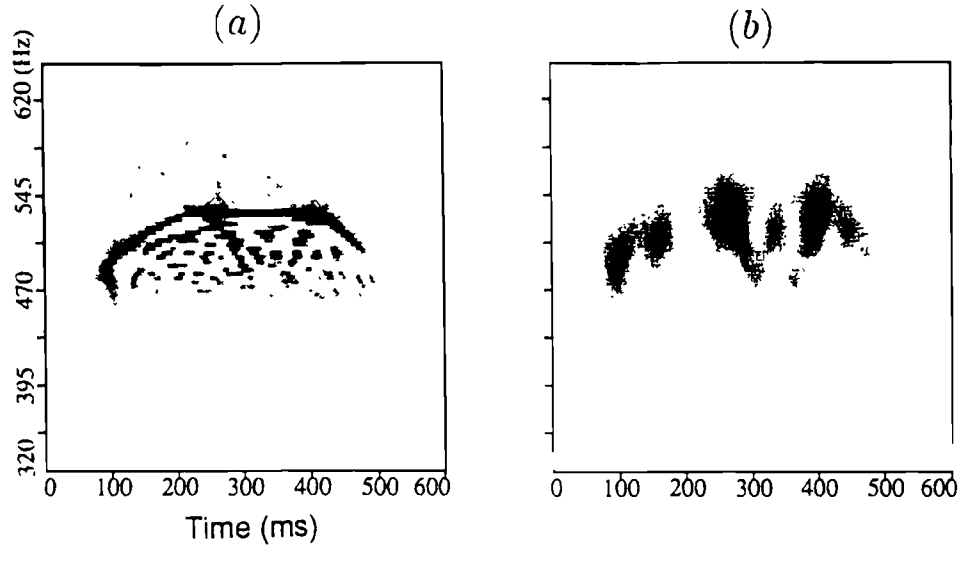

Fig. 8.10 Aneurysm signal. About 5% of the population has a small spherical enlargement of the wall of the cerebral artery (aneurysm). Some aneurysms emit a sound. In the above figure are the Wigner distribution (a) and spectrogram (b) of such a sound. (Courtesy of M. Sun and R. J. Sclabassi.)

Chapter 9

General Approach and the Kernel Method

9.1 INTRODUCTION

The Wigner distribution, as considered in signal analysis, was the first example of a joint time-frequency distribution that was qualitatively different form the spectrogram. The idea of the spectrogram crystallized in the 1940s. Independent of that development, there was considerable activity in the 1940s, 1950s, and 1960s, devising distributions which were similar in spirit to the Wigner distribution in the sense that they satisfied the marginals, the instantaneous frequency condition, and other desirable properties. Among the distributions proposed then (in signal analysis and quantum mechanics) were the Rihaczek, Page, and Margenau-Hill. In 1966 a method was devised that could generate in a simple manner an infinite number of new ones.[125] The approach characterizes time-frequency distributions by an auxiliary function, the kernel function. The properties of a distribution are reflected by simple constraints on the kernel, and by examining the kernel one readily can ascertain the properties of the distribution. This allows one to pick and choose those kernels that produce distributions with prescribed, desirable properties. This general class can be derived by the method of characteristic functions. In this chapter we explain the method and the general ideas associated with it and give the derivation in the next chapter.

9.2 GENERAL CLASS

All time-frequency representations can be obtained from

$$C(t,\omega) = \frac{1}{4\pi^2}\iiint s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, \phi(\theta,\tau)\, e^{-j\theta t - j\tau\omega + j\theta u}\, du\, d\tau\, d\theta \qquad (9.1)$$

where $\phi(\theta,\tau)$ is a two dimensional function called the kernel, a term coined by

Claasen and Mecklenbräuker[120] and whom, with Janssen[276], made many important contributions to the general understanding of the general class, particularly in the signal analysis context.

The kernel determines the distribution and its properties. In Table 9.1 we list some distributions and their corresponding kernels. For the Wigner distribution the kernel is one; however, no particular significance should be attached to that since it is possible to write the general form so that the kernel of any distribution is one, in which case the kernel of the Wigner distribution would be something else.

Spectrum. In terms of the spectrum the general class is

$$C(t,\omega) = \frac{1}{4\pi^2}\iiint S^*(u+\tfrac{1}{2}\theta)\,S(u-\tfrac{1}{2}\theta)\,\phi(\theta,\tau)\,e^{-j\theta t-j\tau\omega+j\tau u}\,d\theta\,d\tau\,du \quad (9.2)$$

as may be verified by expressing the signal in terms of the spectrum and substituting in Eq. (9.1).

Alternate Forms. There are a number of alternative ways of writing the general class of time-frequency distributions that are convenient and add considerable physical insight.

Characteristic Function Formulation. Recall that the characteristic function is the double Fourier transform of the distribution. By inspection of Eq. (9.1), we see that

$$C(t,\omega) = \frac{1}{4\pi^2}\iint M(\theta,\tau)\,e^{-j\theta t-j\tau\omega}\,d\theta\,d\tau \quad (9.3)$$

where

$$M(\theta,\tau) = \phi(\theta,\tau)\int s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,e^{j\theta u}\,du \quad (9.4)$$

$$= \phi(\theta,\tau)\,A(\theta,\tau) \quad (9.5)$$

and where $A(\theta,\tau)$ is the symmetrical ambiguity function. The characteristic function may be appropriately called the generalized ambiguity function.

Time Dependent Autocorrelation Function. In Chapter 1 we showed that the energy density can be written as the Fourier transform of the (deterministic) autocorrelation function. Now we generalize this idea by thinking of the time-frequency energy distribution as the Fourier transform of a local autocorrelation function,[117] $R_t(\tau)$,

$$C(t,\omega) = \frac{1}{2\pi}\int R_t(\tau)\,e^{-j\omega\tau}\,d\tau \quad (9.6)$$

By comparing with Eq. (9.1) we have

$$R_t(\tau) = \frac{1}{2\pi}\iint s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,\phi(\theta,\tau)\,e^{j\theta(u-t)}\,d\theta\,du \quad (9.7)$$

We shall call $R_t(\tau)$ the deterministic generalized local autocorrelation function.

Table 9.1. Some Distributions and Their Kernels

Name	Kernel: $\phi(\theta,\tau)$	Distribution: $C(t,\omega)$
General class (Cohen[125])	$\phi(\theta,\tau)$	$\frac{1}{4\pi^2}\iiint e^{-j\theta t-j\tau\omega+j\theta u}\,\phi(\theta,\tau)\, s^*(u-\frac{1}{2}\tau)\,s(u+\frac{1}{2}\tau)\,du\,d\tau\,d\theta$
Wigner[584]	1	$\frac{1}{2\pi}\int e^{-j\tau\omega}s^*(t-\frac{1}{2}\tau)\,s(t+\frac{1}{2}\tau)\,d\tau$
Margenau-Hill[358]	$\cos\frac{1}{2}\theta\tau$	$\mathrm{Re}\,\frac{1}{\sqrt{2\pi}}\,s(t)S^*(\omega)\,e^{-jt\omega}$
Kirkwood[305] Rihaczek[484]	$e^{j\theta\tau/2}$	$\frac{1}{\sqrt{2\pi}}\,s(t)S^*(\omega)\,e^{-jt\omega}$
Born-Jordan[1] (Cohen[125])	$\frac{\sin\frac{1}{2}\theta\tau}{\frac{1}{2}\theta\tau}$	$\frac{1}{2\pi}\int\frac{1}{\lvert\tau\rvert}\,e^{-j\tau\omega}\int_{t-\lvert\tau\rvert/2}^{t+\lvert\tau\rvert/2} s^*(u-\frac{1}{2}\tau)\,s(u+\frac{1}{2}\tau)\,du\,d\tau$
Page[419]	$e^{j\theta\lvert\tau\rvert}$	$\frac{\partial}{\partial t}\left\lvert\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}s(t')\,e^{-j\omega t'}\,dt'\right\rvert^2$
Choi-Williams[117]	$e^{-\theta^2\tau^2/\sigma}$	$\frac{1}{4\pi^{3/2}}\iint\frac{1}{\sqrt{\tau^2/\sigma}}\,e^{-\sigma(u-t)^2/\tau^2-j\tau\omega} s^*(u-\frac{1}{2}\tau)\,s(u+\frac{1}{2}\tau)\,du\,d\tau$
Spectrogram	$\int h^*(u-\frac{1}{2}\tau)\,e^{-j\theta u}\,h(u+\frac{1}{2}\tau)\,du$	$\left\lvert\frac{1}{\sqrt{2\pi}}\int e^{-j\omega\tau}\,s(\tau)h(\tau-t)\,d\tau\right\rvert^2$
Zhao-Atlas-Marks[626]	$g(\tau)\,\lvert\tau\rvert\,\frac{\sin a\theta\tau}{a\theta\tau}$	$\frac{1}{4\pi a}\int g(\tau)\,e^{-j\tau\omega}\int_{t-\lvert\tau\rvert a}^{t+\lvert\tau\rvert a} s^*(u-\frac{1}{2}\tau)\,s(u+\frac{1}{2}\tau)\,du\,d\tau$
Positive[2] (Cohen, Posch, Zaparovanny[127, 128])	see Chapter 14	$\lvert S(\omega)\rvert^2\,\lvert s(t)\rvert^2\,\Omega(u,v)$

[1] Derived in reference [125] using the ordering rule of Born and Jordan. The derivation is repeated in Section 10.6.

[2] These distributions and the spectrogram are manifestly positive, but only the 'Positive' ones satisfy the marginals.

Fourier Transform of the Kernel. Notice that in the general form θ does not appear in the signal. Hence if we define

$$r(t,\tau) = \frac{1}{2\pi}\int \phi(\theta,\tau)\, e^{-jt\theta}\, d\theta \tag{9.8}$$

the general class can be written as

$$C(t,\omega) = \frac{1}{2\pi}\iint r(t-u,\tau)\, s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, e^{-j\omega\tau}\, du\, d\tau \tag{9.9}$$

The local autocorrelation becomes

$$R_t(\tau) = \int r(t-u,\tau)\, s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, du \tag{9.10}$$

Bilinear Transformation.[585, 137] If in Eq. (9.1) we make the transformation

$$x' = u - \tfrac{1}{2}\tau \quad ; \quad x = u + \tfrac{1}{2}\tau \qquad \text{or} \qquad u = \tfrac{1}{2}(x+x') \quad ; \quad \tau = x - x' \tag{9.11}$$

then we have

$$C(t,\omega) = \iint K(t,\omega;x',x)\, s^*(x')\, s(x)\, dx\, dx' \tag{9.12}$$

with

$$K(t,\omega;x',x) = \frac{1}{2\pi}\, r(t-\tfrac{1}{2}(x+x'),x'-x)\, e^{-j\omega(x'-x)} \tag{9.13}$$

As we discuss in the subsequent sections, properties of the distribution are reflected in the kernel. These constraints may be imposed on ϕ or K but are generally simpler when expressed in terms of ϕ. This is a bilinear form only when K is signal independent.

Sum of Two Signals. For

$$s(t) = s_1(t) + s_2(t) \tag{9.14}$$

the distribution is

$$C(t,\omega) = C_{11}(t,\omega) + C_{22}(t,\omega) + C_{12}(t,\omega) + C_{21}(t,\omega) \tag{9.15}$$

where the cross distribution between two signals is

$$C_{kl}(t,\omega) = \frac{1}{4\pi^2}\iiint \phi(\theta,\tau)\, s_k^*(u-\tfrac{1}{2}\tau)\, s_l(u+\tfrac{1}{2}\tau)\, e^{-j\theta t - j\tau\omega + j\theta u}\, du\, d\tau\, d\theta \tag{9.16}$$

9.3 THE KERNEL METHOD

There are three basic advantages for characterizing time-frequency distributions by the kernel function. First, we can obtain and study the distributions with certain properties by constraining the kernel. For example, suppose we want all distributions that satisfy the marginals. We will see that to assure the satisfaction of the marginals the kernel must have the simple property $\phi(0,\tau) = \phi(\theta,0) = 1$. Therefore, if we want to study the distributions that satisfy the marginals then we consider only kernels that satisfy these conditions. This still leaves an infinite number of choices but we are assured that the marginals are satisfied.

Second, the properties of a distribution can be readily ascertained by a simple examination of the kernel. For example, if the kernel satisfies the condition just discussed, then we know the distribution satisfies the marginals and we do not have to do any further calculations.

Third, given a kernel, a distribution is easy to generate. Before we discuss the conditions on the kernel that guarantee that a distribution will have particular properties, it is worthwhile to classify and discuss the various possibilities and dependence of the kernel.

Functional Dependence of the Kernel. The kernel can be a functional of frequency and time and explicitly of the signal. If we were to be notationally explicit we would write the kernel as $\phi(\theta,\tau,t,\omega;s)$ to signify the possible dependencies. However, we write it as a function of θ and τ and discern from the context whether it depends on other variables.

Types of Kernels. An important subclass are those distributions for which the kernel is a function of the product $\theta\tau$,

$$\phi(\theta,\tau) = \phi_{PR}(\theta\tau) = \phi(\theta\tau) \qquad \text{product kernels} \tag{9.17}$$

For notational clarity, we will drop the subscript PR since we can tell whether we are talking about the general case or product case by the number of variables attributed to $\phi(\theta,\tau)$. Kernels that are a product of a function of each variable

$$\phi(\theta,\tau) = \phi_1(\theta)\,\phi_2(\tau) \qquad \text{separable kernels} \tag{9.18}$$

are called separable kernels.

Bilinearity. If the kernel does not explicitly depend on the signal then the signal enters into the general form twice, hence one says that the distribution is bilinear, a phrase used by Wigner. All the distributions listed in Table 9.1 are bilinear since the kernel does not depend on the signal explicitly. Distributions obtained by taking the kernel to be signal dependent are discussed in Sections 13.5 and Chapter 14. In this chapter we will assume that the kernel is bilinear and time and frequency independent. However, the methods developed for the bilinear case often apply directly to the nonbilinear case. We point that out when appropriate. Notice that

the marginals are bilinear in the signal. However, that in no way reflects on the functional dependence of the joint distribution. Having bilinear marginals does not imply that the joint distribution must be bilinear.

Determination of the Kernel. If a distribution is given we can usually pick out the kernel by putting the distribution in one of the standard forms. If that is not readily doable, we can calculate it by finding the characteristic function of the distribution and using Eq. (9.5). Explicitly,

$$\phi(\theta,\tau) = \frac{M(\theta,\tau)}{A(\theta,\tau)} = \frac{\iint C(t,\omega)\, e^{j\theta t + j\tau\omega}\, dt\, d\omega}{\int s^*(u-\frac{1}{2}\tau)\, s(u+\frac{1}{2}\tau)\, e^{j\theta u}\, du} \tag{9.19}$$

Example 9.1: The Spectrogram.

In Section 7.2 we derived the characteristic function for the spectrogram, Eq. (7.10). We showed that $M_{SP}(\theta,\tau) = A_h(-\theta,\tau)A(\theta,\tau)$. Therefore

$$\phi_{SP}(\theta,\tau) = \frac{M_{SP}(\theta,\tau)}{A(\theta,\tau)} = \frac{A_h(-\theta,\tau)\, A(\theta,\tau)}{A(\theta,\tau)} = A_h(-\theta,\tau) \tag{9.20}$$

We see that the kernel for the spectrogram is the ambiguity function of the window with θ replaced by $-\theta$.

9.4 BASIC PROPERTIES RELATED TO THE KERNEL

We now discuss some of the basic properties of distributions that have been deemed desirable and show how they get reflected as constraints on the kernel. Two important properties are left for consideration in other chapters, the conditions for positivity and the condition on the kernel to minimize the cross terms for multicomponent signals.

Marginals: Instantaneous Energy and Energy Density Spectrum. Integrating the basic form, Eq. (9.1), with respect to frequency gives $2\pi\delta(\tau)$ and therefore

$$\int C(t,\omega)\, d\omega = \frac{1}{2\pi}\iiint \delta(\tau)\, s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, \phi(\theta,\tau)\, e^{j\theta(u-t)}\, d\theta\, du\, d\tau \tag{9.21}$$

$$= \frac{1}{2\pi}\iint \phi(\theta,0)\, |\, s(u)\, |^2\, e^{j\theta(u-t)}\, d\theta\, du \tag{9.22}$$

If we want this to equal $|\, s(t)\, |^2$ then the θ integration must be made to give

$$\frac{1}{2\pi}\int \phi(\theta,0)\, e^{j\theta(u-t)}\, d\theta = \delta(t-u) \tag{9.23}$$

and the only way we can do that is to insist that

$$\phi(\theta, 0) = 1 \qquad [\text{ for time marginal }] \tag{9.24}$$

This is the condition for the time marginal to be satisfied. Similarly, for the frequency marginal to be satisfied

$$\int C(t, \omega)\, dt = |\, S(\omega)\,|^2 \tag{9.25}$$

we must take

$$\phi(0, \tau) = 1 \qquad [\text{ for frequency marginal }] \tag{9.26}$$

Total Energy. If the marginals are correctly given, then of course the total energy will be the energy of the signal. However, we can retain total energy conservation (normalization to one) without insisting on the marginals. Integrating the general form with respect to time and frequency shows that for the total energy to be preserved,

$$\iint C(t, \omega)\, d\omega\, dt = 1 \qquad [\text{ for total energy }] \tag{9.27}$$

the kernel must satisfy

$$\phi(0, 0) = 1 \qquad [\text{ for total energy }] \tag{9.28}$$

Uncertainty Principle. As we have discussed in Section 6.9, any joint distribution that yields the marginals will yield the uncertainty principle of the signal. Thus the condition for the satisfaction of the uncertainty principle is that both marginals must be correctly given. In terms of the kernel this is

$$\phi(\theta, 0) = 1 \quad \text{and} \quad \phi(0, \tau) = 1 \qquad [\text{ for uncertainty principle }] \tag{9.29}$$

Reality. For a distribution to be real the characteristic function must satisfy $M(\theta, \tau) = M^*(-\theta, -\tau)$. Look at Eq. (9.5) for the characteristic function and note that the ambiguity function satisfies $A(\theta, \tau) = A^*(-\theta, -\tau)$. Therefore the only way that $M(\theta, \tau) = M^*(-\theta, -\tau)$ is for the kernel to also satisfy the identical condition,

$$\phi(\theta, \tau) = \phi^*(-\theta, -\tau) \tag{9.30}$$

Time and Frequency Shifts. A signal that is shifted by a time t_0 and frequency ω_0 is $s_{sh} = e^{j\omega_0 t} s(t - t_0)$, where t_0 is the amount of time translation, and ω_0 is the translation in the frequency domain. Substituting this into the general form we have, using C_{sh} for the translated distribution,

$$C_{sh}(t,\omega) = \frac{1}{4\pi^2}\iiint e^{-j\omega_0(u-\tau/2-t_0)}\, e^{j\omega_0(u+\tau/2-t_0)} \times s^*(u-\tfrac{1}{2}\tau-t_0)\, s(u+\tfrac{1}{2}\tau-t_0)\, \phi(\theta,\tau)\, e^{-j\theta t-j\tau\omega+j\theta u}\, d\theta\, d\tau\, du \quad (9.31)$$

$$= \frac{1}{4\pi^2}\iiint \phi(\theta,\tau) s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, e^{-j\theta t-j\tau(\omega-\omega_0)+j\theta(u+t_0)}\, d\theta\, d\tau\, du \quad (9.32)$$

$$= \frac{1}{4\pi^2}\iiint \phi(\theta,\tau) s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, e^{-j\theta(t-t_0)-j\tau(\omega-\omega_0)+j\theta u}\, d\theta\, d\tau\, du \quad (9.33)$$

$$= C(t-t_0,\omega-\omega_0) \quad (9.34)$$

In going from step (9.31) to step (9.32) we assumed that the kernel is not a function of time and frequency. Therefore

$C(t,\omega)$ is time shift invariant if ϕ is independent of t (9.35)

$C(t,\omega)$ is frequency shift invariant if ϕ is independent of ω (9.36)

Note that the kernel can be signal dependent.

Scaling Invariance. In Section 6.5 we showed that if a signal is linearly scaled then the spectrum is inversely scaled and we argued that a plausible requirement on the joint distribution is that it scale the same way in time and frequency. In particular, if $C_{sc}(t,\omega)$ is the distribution of the scaled signal, s_{sc}, then the requirement is

$$C_{sc}(t,\omega) = C(at,\omega/a) \qquad \text{for } s_{sc}(t) = \sqrt{a}\, s(at) \quad (9.37)$$

For the scaled function

$$C_{sc}(t,\omega) = \frac{a}{4\pi^2}\iiint s^*(a(u-\tfrac{1}{2}\tau))\, s(a(u+\tfrac{1}{2}\tau))\, \phi(\theta,\tau)\, e^{-j\theta t-j\tau\omega+j\theta u}\, d\theta\, d\tau\, du \quad (9.38)$$

Transforming by $u \to u/a$, $\tau \to \tau/a$, and $\theta \to a\theta$, we have

$$C_{sc}(t,\omega) = \frac{1}{4\pi^2}\iiint s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, \phi(a\theta,\tau/a)\, e^{-j\theta at-j\tau\omega/a+j\theta u}\, d\theta\, d\tau\, du \quad (9.39)$$

We see that

$$C_{sc}(t,\omega) = C(at,\omega/a) \qquad \text{if} \qquad \phi(a\theta,\tau/a) = \phi(\theta,\tau) \quad (9.40)$$

The only way that a function of two variables can satisfy $\phi(a\theta,\tau/a) = \phi(\theta,\tau)$ is if it is a function of the product of the two variables. Therefore, for scale invariance the kernel must be a product kernel,

$$\phi(\theta,\tau) = \phi(\theta\tau) \qquad [\text{ for scale invariance }] \quad (9.41)$$

Note that we have implicitly assumed that the kernel is signal independent.

Weak Finite Support. Recall from Section 6.6 that weak finite time support means that for a finite duration signal we want the distribution to be zero before the signal starts and after the signal ends. The condition for that to be the case is

$$\int \phi(\theta,\tau)\, e^{-j\theta t}\, d\theta = 0 \qquad \text{for } |\tau| \leq 2|t| \tag{9.42}$$

[for weak finite time support]

Similarly for a bandlimited signal we want the distribution to be zero outside the band. The condition for that to be so is

$$\int \phi(\theta,\tau)\, e^{-j\tau\omega}\, d\tau = 0 \qquad \text{for } |\theta| \leq 2|\omega| \tag{9.43}$$

[for weak finite frequency support]

Strong Finite Support. Strong finite support means that the distribution is zero whenever the signal or spectrum is zero. The conditions for this to be the case were derived by Loughlin, Pitton, and Atlas[342] for the case of bilinear distributions. The conditions are

$$\int \phi(\theta,\tau)\, e^{-j\theta t}\, d\theta = 0 \qquad \text{for } |\tau| \neq 2|t| \tag{9.44}$$

[for strong finite time support]

$$\int \phi(\theta,\tau) e^{-j\tau\omega}\, d\tau = 0 \qquad \text{for } |\theta| \neq 2|\omega| \tag{9.45}$$

[for strong finite frequency support]

Example 9.2: Choi-Williams Kernel.

To see how the above results can be used to quickly ascertain the behavior of a particular distribution, consider the Choi-Williams kernel,

$$\phi_{CW}(\theta,\tau) = e^{-\theta^2\tau^2/\sigma} \tag{9.46}$$

where σ is a constant. It is clear that for either θ or τ being zero the kernel is one and hence the Choi-Williams distribution satisfies the marginals. Also, since it is time and frequency independent it satisfies the shift properties. For the finite support property we have

$$\int \phi(\theta,\tau)\, e^{-j\theta t}\, d\theta = \sqrt{\frac{\pi\sigma}{\tau^2}}\, \exp\left[-\frac{\sigma t^2}{4\tau^2}\right] \tag{9.47}$$

Therefore it does not satisfy the weak finite support properties but does so approximately for σ sufficiently large.

Example 9.3: Wigner Distribution.

Since the kernel for the Wigner distribution is one, it satisfies the marginals. Also, since it is signal independent it satisfies the shift properties. The finite support property is

$$\int \phi(\theta,\tau)\, e^{-j\theta t}\, d\theta \;=\; 2\pi\delta(t) \tag{9.48}$$

and this clearly satisfies the weak finite support property, Eq. (9.42).

Example 9.4: Spectrogram.

In the kernel for the spectrogram, Eq. (9.20) set $\theta = 0$ giving

$$\phi_{SP}(0,\tau) \;=\; \int |h(t)|^2\, e^{-j\theta t}\, dt \neq 1 \tag{9.49}$$

which cannot be equal to one unless we take a delta function for $|h|^2$. Therefore the spectrogram cannot satisfy the marginals.

Example 9.5: Sinc.

The sinc kernel is

$$\phi(\theta,\tau) \;=\; \frac{\sin(a\theta\tau)}{a\theta\tau} \tag{9.50}$$

This kernel satisfies the marginals, since $\sin 0/0 = 1$. Consider now the weak finite support property. Using Eq. (9.42) we have

$$\int \frac{\sin(a\theta\tau)}{a\theta\tau}\, e^{-j\theta t}\, dt \;=\; \begin{cases} \dfrac{\pi}{a\tau} & |t| \le a\tau \\ 0 & \text{otherwise} \end{cases} \tag{9.51}$$

Comparing to the condition for weak finite support, Eq. (9.42), we see that the condition is satisfied only if $a < \frac{1}{2}$.

Inversion. From Eq. (9.5) we have

$$A(\theta,\tau) = \frac{M(\theta,\tau)}{\phi(\theta,\tau)} = \int s^*(u - \tfrac{1}{2}\tau)\, s(u + \tfrac{1}{2}\tau)\, e^{j\theta u}\, du \tag{9.52}$$

and taking the Fourier inverse yields

$$s^*(u - \tfrac{1}{2}\tau)\, s(u + \tfrac{1}{2}\tau) \;=\; \frac{1}{2\pi}\int \frac{M(\theta,\tau)}{\phi(\theta,\tau)}\, e^{-j\theta u}\, d\theta \tag{9.53}$$

Letting $t = u + \frac{1}{2}\tau$ and $t' = u - \frac{1}{2}\tau$ gives

$$s^*(t')\, s(t) \;=\; \frac{1}{2\pi}\int \frac{M(\theta, t-t')}{\phi(\theta, t-t')}\, e^{-j\theta(t+t')/2}\, d\theta \tag{9.54}$$

and taking the specific value $t' = 0$ we have

$$s(t) = \frac{1}{2\pi s^*(0)} \int \frac{M(\theta,t)}{\phi(\theta,t)} e^{-j\theta t/2}\, d\theta \tag{9.55}$$

$$= \frac{1}{2\pi s^*(0)} \iiint \frac{C(t',\omega)}{\phi(\theta,t)} e^{jt\omega + j\theta(t'-t/2)}\, dt'\, d\omega\, d\theta \tag{9.56}$$

Hence the signal can be recovered from the distribution if the division by the kernel is uniquely defined for all values of θ, τ. If the kernel is zero in regions, we cannot recover the signal uniquely. For the spectrogram this will depend on the window.

Example 9.6: Wigner.

For the Wigner distribution

$$s(t) = \frac{1}{2\pi s^*(0)} \iiint W(t',\omega)\, e^{jt\omega + j\theta(t'-t/2)}\, dt'\, d\omega\, d\theta \tag{9.57}$$

$$= \frac{1}{s^*(0)} \iint W(t',\omega)\, e^{jt\omega} \delta(t'-t/2)\, dt'\, d\omega \tag{9.58}$$

$$= \frac{1}{s^*(0)} \int W(t/2,\omega)\, e^{jt\omega}\, d\omega \tag{9.59}$$

which agrees with the answer obtained in Eq. (8.74).

9.5 GLOBAL AVERAGES

For a function, $g(t,\omega)$, the global average is

$$\langle\, g(t,\omega)\,\rangle = \iint g(t,\omega)\, C(t,\omega)\, d\omega\, dt \tag{9.60}$$

If the marginals are satisfied, then averages of the form $g(t) = g_1(t) + g_2(\omega)$ will be automatically satisfied,

$$\langle\, g_1(t) + g_2(\omega)\,\rangle = \iint \{g_1(t) + g_2(\omega)\}\, C(t,\omega)\, d\omega\, dt \tag{9.61}$$

$$= \int g_1(t)\, |\, s(t)\,|^2\, dt + \int g_2(\omega)\, |\, S(\omega)\,|^2\, d\omega \tag{9.62}$$

correctly given if $\phi(\theta,0) = \phi(0,\tau) = 1$

Correlation Coefficient and Covariance. Let us now consider the first mixed moment,

$$\langle\, \omega t\,\rangle = -\left.\frac{\partial^2 M(\theta,\tau)}{\partial\theta\partial\tau}\right|_{\theta,\tau=0} \tag{9.63}$$

$$= -\frac{\partial^2}{\partial\tau\partial\theta}\,\phi(\theta,\tau)\Big|_{\theta,\tau=0}\int A^2(t)\,dt + \int t\,\varphi'(t)\,A^2(t)\,dt \qquad (9.64)$$

If we want the covariance to be equal to the form given in Section 1.8 then we must have that $\langle\, t\omega \,\rangle = \langle t\,\varphi'(t)\rangle$, which can be achieved by taking the mixed partial derivative of the kernel to be zero at the origin,

$$\langle\, t\omega \,\rangle = \int t\,\varphi'(t)|s(t)|^2\,dt \qquad \text{if} \quad \frac{\partial\phi(\theta,\tau)}{\partial\theta\partial\tau}\Big|_{\theta,\tau=0} = 0 \qquad (9.65)$$

Example 9.7: Choi-Williams.

For the Choi Williams kernel,

$$\frac{\partial\phi(\theta,\tau)}{\partial\theta\partial\tau}\Big|_{\theta,\tau=0} = (4\theta^3\tau^3/\sigma^2 - 4\theta\tau/\sigma)e^{-\theta^2\tau^2/\sigma}\Big|_{\theta,\tau=0} = 0 \qquad (9.66)$$

and we see that the Choi-Williams distribution gives the proper expression for the covariance.

9.6 LOCAL AVERAGES

The density of frequency for a given time, $C(\omega\,|\,t)$, and the density of time for a given frequency, $C(t\,|\,\omega)$, are

$$C(\omega\,|\,t) = \frac{C(t,\omega)}{P(t)} \quad ; \quad C(t\,|\,\omega) = \frac{C(t,\omega)}{P(\omega)} \qquad (9.67)$$

where $P(t)$ and $P(\omega)$ are the marginal distributions. Thus the expectation value of a function g at a given time or frequency is

$$\langle\, g(\omega)\,\rangle_t = \frac{1}{P(t)}\int g(\omega)\,C(t,\omega)\,d\omega \qquad (9.68)$$

$$\langle\, g(t)\,\rangle_\omega = \frac{1}{P(\omega)}\int g(t)\;C(t,\omega)\,dt \qquad (9.69)$$

Instantaneous Frequency and Its Spread. For the average frequency at a given time we calculate

$$\langle\,\omega\,\rangle_t = \frac{1}{P(t)}\int \omega\,C(t,\omega)\;d\omega \qquad (9.70)$$

$$= \frac{1}{2\pi P(t)}\iint A^2(u)\left[\,\phi(\theta,0)\,\varphi'(u) - j\frac{\partial\phi(\theta,\tau)}{\partial\tau}\Big|_{\tau=0}\right] e^{j\theta(u-t)}\,d\theta\,du \qquad (9.71)$$

$$= \frac{1}{P(t)}\left[\,\phi(0)\,A^2(t)\,\varphi'(t) + 2\,\phi'(0)\,A(t)\,A'(t)\,\right] \quad [\text{ product kernel }] \qquad (9.72)$$

Suppose we satisfy the time marginal in which case $P(t) = A^2(t)$. Then

$$\langle \omega \rangle_t = \varphi'(t) - j\frac{1}{P(t)}\frac{1}{2\pi}\iint A^2(u)\left.\frac{\partial\phi(\theta,\tau)}{\partial\tau}\right|_{\tau=0} e^{j\theta(u-t)}\,d\theta\,du \tag{9.73}$$

If we want this local average to be equal to the derivative of the phase we must take the integrand to be zero. Therefore

$$\langle \omega \rangle_t = \varphi'(t) \qquad \text{if } \left.\frac{\partial\phi(\theta,\tau)}{\partial\tau}\right|_{\tau=0} = 0 \tag{9.74}$$

$$\text{if } \phi'(0) = 0 \quad [\text{ product kernel }] \tag{9.75}$$

Second Conditional Moment of Frequency. For simplicity we give the result for product kernels:

$$\begin{aligned}\langle \omega^2 \rangle_t P(t) &= \tfrac{1}{2}[\phi(0) + 4\phi''(0)]\,A'^2(t) - \tfrac{1}{2}[\phi(0) - 4\phi''(0)]\,A(t)\,A''(t) \\ &+ \phi(0)\,A^2(t)\,\varphi'^2(t) + 2\phi'(0)\left[2\,A(t)\,A'(t)\,\varphi'(t) + A^2(t)\,\varphi''(t)\right]\end{aligned} \tag{9.76}$$

If we impose the conditions to obtain the derivative of the phase for the first conditional moment, that is, the marginal conditions and $\phi'(0) = 0$, this becomes

$$\langle \omega^2 \rangle_t = \tfrac{1}{2}[1 + 4\phi''(0)]\left(\frac{A'(t)}{A(t)}\right)^2 - \tfrac{1}{2}[1 - 4\phi''(0)]\frac{A''(t)}{A(t)} + \varphi'^2(t) \tag{9.77}$$

Standard Deviation. The local standard deviation is obtained from $\sigma^2_{\omega|t} = \langle \omega^2 \rangle_t - \langle \omega \rangle_t^2$. From Eqs. (9.77) and (9.74) we have

$$\sigma^2_{\omega|t} = \tfrac{1}{2}[1 + 4\phi''(0)]\left(\frac{A'(t)}{A(t)}\right)^2 - \tfrac{1}{2}[1 - 4\phi''(0)]\frac{A''(t)}{A(t)} \tag{9.78}$$

Positive Spread. In general, the conditional second moment and standard deviation become negative and cannot be interpreted. However, for the choice

$$\phi''(0) = \tfrac{1}{4} \tag{9.79}$$

we obtain

$$\langle \omega^2 \rangle_t = \left(\frac{A'(t)}{A(t)}\right)^2 + \varphi'^2(t) \tag{9.80}$$

$$\sigma^2_{\omega|t} = \left(\frac{A'(t)}{A(t)}\right)^2 \tag{9.81}$$

which are manifestly positive. Are these plausible results? We discuss this issue in detail in Section 13.3 where we develop the idea of instantaneous bandwidth and connect it to the nature of multicomponent signals.

Example 9.8: Wigner Distribution.

For the Wigner distribution $\phi'(0) = 0$ and hence

$$\sigma^2_{\omega|t} = \frac{1}{2}\left[\left(\frac{A'(t)}{A(t)}\right)^2 - \frac{A''(t)}{A(t)}\right] \tag{9.82}$$

which agrees with the expression given in Section 8.5.

Group Delay. If we want the conditional moment of time for a given frequency to be the derivative of the spectral phase then the identical approach as above gives

$$\langle t \rangle_\omega = -\psi'(\omega) \qquad \text{if} \qquad \left.\frac{\partial\phi(\theta,\tau)}{\partial\theta}\right|_{\theta=0} = 0 \tag{9.83}$$

where $\psi(\omega)$ is the spectral phase.

9.7 TRANSFORMATION BETWEEN DISTRIBUTIONS[125, 126]

We now address the relationship between distributions. First let us discuss whether indeed there should be a procedure that allows us to transform one distribution into another. In Section 9.4 we showed that a signal can be recovered from a particular distribution if the kernel is not zero in a finite region. Given a distribution for which the signal can be recovered, we may take the recovered signal and calculate any other distribution, so in these cases we do expect a relationship to exist between them. To obtain that relationship suppose we have two distributions, C_1 and C_2, with corresponding kernels, ϕ_1 and ϕ_2. Their characteristic functions are

$$M_1(\theta,\tau) = \phi_1(\theta,\tau)\int s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,e^{j\theta u}\,du \tag{9.84}$$

$$M_2(\theta,\tau) = \phi_2(\theta,\tau)\int s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,e^{j\theta u}\,du \tag{9.85}$$

Divide one equation by the other to obtain

$$M_1(\theta,\tau) = \frac{\phi_1(\theta,\tau)}{\phi_2(\theta,\tau)}\,M_2(\theta,\tau) \tag{9.86}$$

This is an important relationship because it connects the characteristic functions. Note that we have a division by the kernel. For the division to be proper the kernel cannot be zero in regions which is consistent with our discussion above.

To obtain the relationship between the distributions take the double Fourier transform of both sides and use Eq. (9.3) to obtain

$$C_1(t,\omega) = \frac{1}{4\pi^2}\iint \frac{\phi_1(\theta,\tau)}{\phi_2(\theta,\tau)} M_2(\theta,\tau)\, e^{-j\theta t - j\tau\omega}\, d\theta\, d\tau \tag{9.87}$$

Now express M_2 in terms of C_2 to obtain

$$C_1(t,\omega) = \frac{1}{4\pi^2}\iiiint \frac{\phi_1(\theta,\tau)}{\phi_2(\theta,\tau)} C_2(t',\omega')\, e^{j\theta(t'-t)+j\tau(\omega'-\omega)}\, d\theta\, d\tau\, dt'\, d\omega' \tag{9.88}$$

This relationship can be written as

$$C_1(t,\omega) = \iint g_{12}(t'-t,\omega'-\omega)\, C_2(t',\omega')\, dt'\, d\omega' \tag{9.89}$$

with

$$g_{12}(t,\omega) = \frac{1}{4\pi^2}\iint \frac{\phi_1(\theta,\tau)}{\phi_2(\theta,\tau)} e^{j\theta t+j\tau\omega}\, d\theta\, d\tau \tag{9.90}$$

Relation of the Spectrogram to Other Bilinear Representations. We now specialize to the case where we transform from an arbitrary representation to the spectrogram. In Eq. (9.89) we want C_1 to be the spectrogram and C_2 to be arbitrary. To simplify notation we set $\phi_{SP} = \phi_1$ and $\phi = \phi_2$ and $g_{SP} = g_{12}$ and write

$$C_{SP}(t,\omega) = \iint g_{SP}(t'-t,\omega'-\omega)\, C(t',\omega')\, dt'\, d\omega' \tag{9.91}$$

The kernel for the spectrogram with window, $h(t)$, is $A_h(-\theta,\tau)$ and therefore

$$g_{SP}(t,\omega) = \frac{1}{4\pi^2}\iint \frac{A_h(-\theta,\tau)}{\phi(\theta,\tau)} e^{j\theta t+j\tau\omega}\, d\theta\, d\tau \tag{9.92}$$

$$= \frac{1}{4\pi^2}\iiint \frac{1}{\phi(\theta,\tau)} h^*(u-\tfrac{1}{2}\tau)\, h(u+\tfrac{1}{2}\tau)\, e^{j\theta t+j\tau\omega-j\theta u}\, du\, d\tau\, d\theta \tag{9.93}$$

$$= \frac{1}{4\pi^2}\iiint h^*(u-\tfrac{1}{2}\tau)\, h(u+\tfrac{1}{2}\tau) \times \frac{\phi(\theta,\tau)}{\phi(\theta,\tau)\,\phi(-\theta,\tau)} e^{-j\theta t+j\tau\omega+j\theta u}\, du\, d\tau\, d\theta \tag{9.94}$$

If we take kernels for which $\phi(-\theta,\tau)\,\phi(\theta,\tau) = 1$, we see that $g_{SP}(t,\omega)$ is just the distribution of the window function, except that it is evaluated at $-\omega$. Therefore

$$g_{SP}(t,\omega) = C_h(t,-\omega) \tag{9.95}$$

for kernels that satisfy $\phi(-\theta,\tau)\,\phi(\theta,\tau) = 1$

and

$$C_{SP}(t,\omega) = \iint C_s(t',\omega')\, C_h(t'-t,\omega-\omega')\, dt'\, d\omega' \qquad (9.96)$$

for kernels that satisfy $\phi(-\theta,\tau)\,\phi(\theta,\tau) = 1$

This was first shown by Janssen.[276] For the case where $\phi(-\theta,\tau)\,\phi(\theta,\tau)$ does not equal one, then

$$C_{SP}(t,\omega) = \iiiint G(t'',\omega'')C_s(t',\omega')C_h(t''+t'-t,-\omega''+\omega-\omega')\, dt'\, dt''\, d\omega'\, d\omega'' \qquad (9.97)$$

where

$$G(t,\omega) = \frac{1}{4\pi^2}\iint \frac{e^{-j\theta t-j\tau\omega}}{\phi(-\theta,\tau)\,\phi(\theta,\tau)}\, d\theta\, d\tau \qquad (9.98)$$

Positive Bilinear Distributions. In Section 8.10 we saw that if the Wigner distribution is convolved with certain functions a positive distribution is obtained. We have also just seen that when the general class is convolved with certain $g's$ the spectrogram is obtained. The general question as to when one obtains a positive bilinear distribution by convolving the general class has been comprehensively addressed by Janssen[277] who derived a number of important results relating to this question.

Chapter 10

Characteristic Function Operator Method

10.1 INTRODUCTION

In the last chapter we showed that an infinite number of joint time-frequency distributions can be generated. We now describe a method for deriving them from first principles. In this chapter we develop the ideas for time and frequency; in Chapter 17 we generalize to arbitrary variables and subsequently specialize to the physical variable "scale".

10.2 CHARACTERISTIC FUNCTION METHOD

Recall that a distribution and its characteristic function are connected by

$$M(\theta,\tau) = \iint P(t,\omega)\, e^{j\theta t + j\tau\omega}\, dt\, d\omega = \langle e^{j\theta t + j\tau\omega} \rangle \tag{10.1}$$

and

$$C(t,\omega) = \frac{1}{4\pi^2} \iint M(\theta,\tau)\, e^{-j\theta t - j\tau\omega}\, d\theta\, d\tau \tag{10.2}$$

The aim is to find the distribution. If it were possible to find the characteristic function without first knowing the distribution, our problem would be solved because we would then obtain the distribution by way of Eq. (10.2).

Correspondence Rules. In Chapter 1 we showed that the average of a function of frequency can be directly obtained from the signal. In particular, the average of

$g(\omega)$ can be calculated by

$$\langle g(\omega) \rangle = \int g(\omega) \, |S(\omega)|^2 \, d\omega = \int s^*(t) \, g(\mathcal{W}) \, s(t) \, dt \tag{10.3}$$

where $\mathcal{W}$ is the frequency operator. We generalize this by assuming that we can find the average of any function of time and frequency, $G(t,\omega)$, by

$$\langle G(t,\omega) \rangle = \int s^*(t) \, \mathcal{G}(\mathcal{T},\mathcal{W}) \, s(t) \, dt \tag{10.4}$$

where $\mathcal{G}(\mathcal{T},\mathcal{W})$ is an operator, as yet unknown, that corresponds to the ordinary function G. A rule that associates an ordinary function with an operator is called a correspondence rule, a rule of association, or ordering rule.

Characteristic Function Operator. The characteristic function is an average, the average of $e^{j\theta t + j\tau\omega}$, and hence it seems plausible that we can find an operator $\mathcal{M}(\theta,\tau;\mathcal{T},\mathcal{W})$ so that

$$M(\theta,\tau) = \langle \mathcal{M}(\theta,\tau;\mathcal{T},\mathcal{W}) \rangle = \int s^*(t) \, \mathcal{M}(\theta,\tau;\mathcal{T},\mathcal{W}) \, s(t) \, dt \tag{10.5}$$

We shall call $\mathcal{M}$ the characteristic function operator. What can we take for $\mathcal{M}$? One possible choice is to substitute the time and frequency operators in the expression $e^{j\theta t + j\tau\omega}$ for the time and frequency variables,

$$e^{j\theta t + j\tau\omega} \rightarrow \mathcal{M}_W = e^{j\theta\mathcal{T} + j\tau\mathcal{W}} \qquad [\text{ Weyl}^{[568]}] \tag{10.6}$$

This is called the Weyl correspondence. Equally plausible is

$$e^{j\theta t + j\tau\omega} \rightarrow \mathcal{M}_N = e^{j\theta\mathcal{T}} \, e^{j\tau\mathcal{W}} \qquad [\text{ Normal }] \tag{10.7}$$

which is called the normal correspondence. Now, for ordinary variables $e^{j\theta t + j\tau\omega} = e^{j\theta t} \, e^{j\tau\omega}$, but for operators it is not true that $e^{j\theta\mathcal{T} + j\tau\mathcal{W}}$ equals $e^{j\theta\mathcal{T}} \, e^{j\tau\mathcal{W}}$ because the operators do not commute. Since these two choices are not equal they will give different answers for the characteristic function and therefore produce different distributions. In addition to the above two choices the following are also equally plausible,

$$e^{j\theta t + j\tau\omega} \rightarrow e^{j\tau\mathcal{W}} \, e^{j\theta\mathcal{T}} \qquad [\text{ Antinormal }] \tag{10.8}$$

$$\rightarrow \tfrac{1}{2} \left[e^{j\theta\mathcal{T}} \, e^{j\tau\mathcal{W}} + e^{j\tau\mathcal{W}} \, e^{j\theta\mathcal{T}} \right] \qquad [\text{ Symmetrization }] \tag{10.9}$$

$$\rightarrow \frac{\sin \theta\tau/2}{\theta\tau/2} \, e^{j\tau\mathcal{W} + j\theta\mathcal{T}} \qquad [\text{ Cohen}^{[125]}] \tag{10.10}$$

As we will show, we can generalize these rules by

$$e^{j\theta t + j\tau\omega} \rightarrow \phi(\theta,\tau) \, e^{j\theta\mathcal{T} + j\tau\mathcal{W}} \qquad [\text{ Cohen}^{[125]}] \tag{10.11}$$

There is one basic reason why all these correspondence rules are plausible and equally correct in some sense. Recall from Eqs. (4.44)-(4.45) that the joint characteristic function of two variables, x, y, is related to the individual characteristic functions, $M_x(\theta), M_y(\tau)$, by

$$M_x(\theta) = M(\theta, 0) \quad ; \quad M_y(\tau) = M(0, \tau) \tag{10.12}$$

and therefore the characteristic function operator is constrained by

$$\mathcal{M}(\theta, 0; \mathcal{T}, \mathcal{W}) = e^{j\theta\mathcal{T}} \tag{10.13}$$

$$\mathcal{M}(0, \tau; \mathcal{T}, \mathcal{W}) = e^{j\tau\mathcal{W}} \tag{10.14}$$

if we are to be assured of the right marginals. Moreover, this assures that averages of functions of time or frequency only are correctly given. All the correspondences given above satisfy these constraining equations.

To obtain a distribution we choose a characteristic function operator, calculate the characteristic function by way of Eq. (10.5), and then find the distribution via Eq. (10.2). These equations can be combined. Substituting Eq. (10.5) into Eq. (10.2) we have

$$C(t, \omega) = \frac{1}{4\pi^2} \iiint s^*(u)\, \mathcal{M}(\theta, \tau; \mathcal{T}_u, \mathcal{W}_u)\, s(u)\, e^{-j\theta t - j\tau\omega}\, d\theta\, d\tau\, du \tag{10.15}$$

where $\mathcal{T}_u$ and $\mathcal{W}_u$ signify the time and frequency operators operating on functions of u. We now address the evaluation of expressions like Eq. (10.5) for the various choices of the characteristic function operator.

10.3 EVALUATION OF THE CHARACTERISTIC FUNCTION

First let us note that the evaluation of the characteristic function can be done using the signal or spectrum,

$$M(\theta, \tau) = \langle \mathcal{M} \rangle = \int s^*(t)\, \mathcal{M}(\theta, \tau; \mathcal{T}, \mathcal{W})\, s(t)\, dt \tag{10.16}$$

$$= \int S^*(\omega) \mathcal{M}(\theta, \tau; \mathcal{T}, \mathcal{W})\, S(\omega)\, d\omega \tag{10.17}$$

If we are using the signal we will say that we are in the time domain, and if we are using the spectrum we will say we are in the frequency domain. We recall from Section 1.4 that the explicit forms of the time and frequency operators are

$$\mathcal{T} = t\,, \qquad \mathcal{W} = \frac{1}{j}\frac{d}{dt} \qquad \text{in the time domain} \tag{10.18}$$

$$\mathcal{T} = j\frac{d}{d\omega}\,, \qquad \mathcal{W} = \omega \qquad \text{in the frequency domain} \tag{10.19}$$

The fundamental relationship between these operators is

$$\mathcal{TW} - \mathcal{WT} = j \tag{10.20}$$

which is discussed and derived in Section 15.3.

Three important results are helpful for the simplification and evaluation of characteristic function operators. Two of them, already discussed in Chapter 1, are

$$e^{j\tau\mathcal{W}} s(t) = s(t+\tau) \quad ; \quad e^{j\theta\mathcal{T}} S(\omega) = S(\omega - \theta) \tag{10.21}$$

The third is an important result of operator theory, which is well known but far from trivial,

$$e^{j\theta\mathcal{T}+j\tau\mathcal{W}} = e^{-j\theta\tau/2}\, e^{j\tau\mathcal{W}}\, e^{j\theta\mathcal{T}} \tag{10.22}$$

$$= e^{j\theta\tau/2}\, e^{j\theta\mathcal{T}}\, e^{j\tau\mathcal{W}} \tag{10.23}$$

We will not prove this here because we return to it in our considerations of a more general case in Chapter 17.

In general,

$$e^{j\theta\mathcal{T}+j\tau\mathcal{W}} s(t) = e^{j\theta\tau/2}\, e^{j\theta\mathcal{T}}\, e^{j\tau\mathcal{W}} s(t) = e^{j\theta\tau/2}\, e^{j\theta t} s(t+\tau) \tag{10.24}$$

We are now ready to see, by way of examples, how to derive distributions using the characteristic function approach.

Example 10.1: The Wigner Distribution.

For the characteristic function operator take the Weyl ordering, in which case

$$M(\theta,\tau) = \int s^*(t)\, e^{j\tau\mathcal{W}+j\theta\mathcal{T}} s(t)\, dt \tag{10.25}$$

$$= \int s^*(t)\, e^{j\theta\tau/2}\, e^{j\theta\mathcal{T}}\, e^{j\tau\mathcal{W}} s(t)\, dt \tag{10.26}$$

$$= \int s^*(t)\, e^{j\theta\tau/2}\, e^{j\theta t} s(t+\tau)\, dt \tag{10.27}$$

$$= \int s^*(u - \tfrac{1}{2}\tau)\, e^{j\theta u} s(u + \tfrac{1}{2}\tau)\, du \tag{10.28}$$

The distribution for this ordering is

$$P(t,\omega) = \frac{1}{4\pi^2} \iint M(\theta,\tau)\, e^{-j\theta t - j\tau\omega}\, d\theta\, d\tau \tag{10.29}$$

$$= \frac{1}{4\pi^2} \iiint s^*(u - \tfrac{1}{2}\tau)\, e^{j\theta u} s(u + \tfrac{1}{2}\tau)\, e^{-j\theta t - j\tau\omega}\, d\theta\, d\tau\, du \tag{10.30}$$

$$= \frac{1}{2\pi}\iint s^*(u-\tfrac{1}{2}\tau)\,e^{-j\tau\omega}\,\delta(u-t)\,s(u+\tfrac{1}{2}\tau)\,d\tau\,du \tag{10.31}$$

$$= \frac{1}{2\pi}\int s^*(t-\tfrac{1}{2}\tau)\,e^{-j\tau\omega}\,s(t+\tfrac{1}{2}\tau)\,d\tau \tag{10.32}$$

which is the Wigner distribution.

Example 10.2: Margenau - Hill Distribution.

Using the ordering given by Eq. (10.7) we have for the characteristic function

$$M(\theta,\tau) = \int s^*(t)\,e^{j\theta\mathcal{T}}\,e^{j\tau\mathcal{W}}s(t) = \int s^*(t)\,e^{j\theta t}s(t+\tau)\,dt \tag{10.33}$$

The distribution is

$$P(t,\omega) = \frac{1}{4\pi^2}\iiint s^*(u)\;e^{j\theta u}s(u+\tau)\,e^{-j\theta t\,-j\tau\omega}\,d\theta\,d\tau\,du \tag{10.34}$$

$$= \frac{1}{\sqrt{2\pi}}\,s^*(t)\,e^{jt\omega}S(\omega) \tag{10.35}$$

If the correspondence of Eq. (10.8) is used, the complex conjugate of Eq. (10.35) is obtained and if the symmetrization rule is used, Eq. (10.9), we obtain the real part.

10.4 THE GENERAL CLASS

All possible distributions are obtained if we consider all possible correspondence rules. A way to approach this problem is to start with any one, say the Weyl rule, and any other correspondence is expressible as

$$\mathcal{M}_G(\mathcal{T},\mathcal{W}) \rightarrow \phi(\theta,\tau)\,e^{j\theta\mathcal{T}+j\tau\mathcal{W}} \tag{10.36}$$

For Eqs. (10.13) and (10.14) to be satisfied we must have

$$\phi(0,\tau) = \phi(\theta,0) = 1 \tag{10.37}$$

The general characteristic function is the average of the characteristic function operator

$$M_G(\theta,\tau) = \langle\,\mathcal{M}_G\,\rangle = \phi(\theta,\tau)\int s^*(t)\,e^{j\theta\mathcal{T}+j\tau\mathcal{W}}s(t)\,dt \tag{10.38}$$

$$= \phi(\theta,\tau)\int s^*(t-\tfrac{1}{2}\tau)\,e^{j\theta t}s(t+\tfrac{1}{2}\tau)\,dt \tag{10.39}$$

and the distribution is

$$C(t,\omega) = \frac{1}{4\pi^2}\iint M_G(\theta,\tau)\,e^{-j\theta t-j\tau\omega}\,d\theta\,d\tau \tag{10.40}$$

$$= \frac{1}{4\pi^2}\iiint s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,\phi(\theta,\tau)\,e^{-j\theta t-j\tau\omega+j\theta u}\,du\,d\tau\,d\theta \tag{10.41}$$

which is the general class of time-frequency distributions developed in Chapter 9.

10.5 AVERAGES

If we have a function of time and frequency, $g(t,\omega)$, its average can be calculated using the time-frequency distribution or directly from the signal,

$$\langle\, g(t,\omega)\,\rangle = \iint g(t,\omega)\,C(t,\omega)\,dt\,d\omega \tag{10.42}$$

$$= \int s^*(t)\,\mathcal{G}\,(\mathcal{T},\mathcal{W})\;s(t)\,dt \tag{10.43}$$

where $\mathcal{G}$ is the operator corresponding to the ordinary function g. The relation between g and $\mathcal{G}$ is

$$\mathcal{G}(\mathcal{T},\mathcal{W}) = \iint \gamma(\theta,\tau)\phi(\theta,\tau)\,e^{j\theta\mathcal{T}+j\tau\mathcal{W}}\,d\theta\,d\tau \tag{10.44}$$

where

$$\gamma(\theta,\tau) = \frac{1}{4\pi^2}\iint g\,(t,\omega)\,e^{-j\theta t-j\tau\omega}\,dt\,d\omega \tag{10.45}$$

Or, equivalently,

$$\mathcal{G}(\mathcal{T},\mathcal{W}) = \frac{1}{4\pi^2}\iiiint g(t,\omega)\,\phi(\theta,\tau)\,e^{j\theta(\mathcal{T}-t)+\,j\tau(\mathcal{W}-\omega)}\,d\theta\,d\tau\,dt\,d\omega \tag{10.46}$$

To prove the equivalence of the two methods of calculating averages, Eqs. (10.42) and (10.43), consider

$$\langle\, g(t,\omega)\,\rangle = \int s^*(t)\mathcal{G}\,(\mathcal{T},\mathcal{W})\;s(t)\,dt \tag{10.47}$$

$$= \iiint s^*(t)\,\gamma(\theta,\tau)\,\phi(\theta,\tau)\,e^{j\theta\mathcal{T}+j\tau\mathcal{W}}s(t)\,d\theta\,d\tau\,dt \tag{10.48}$$

$$= \iiint s^*(u-\tfrac{1}{2}\tau)\,\gamma(\theta,\tau)\,\phi(\theta,\tau)\,e^{j\theta u}s(u+\tfrac{1}{2}\tau)\,d\theta\,d\tau\,du \tag{10.49}$$

$$= \frac{1}{4\pi^2}\iiiiint g\,(t,\omega)\,e^{-j\theta t-j\tau\omega}\,\phi(\theta,\tau)\,e^{j\theta u}$$

$$\times\quad s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,d\theta\,d\tau\,dt\,du\,d\omega \tag{10.50}$$

$$= \iint g(t,\omega)\,C(t,\omega)\,dt\,d\omega \tag{10.51}$$

10.6 THE MOMENT METHOD

The moment method is another way to construct distributions. It is equivalent to the characteristic function method just described but presents a number of new features both conceptually and mathematically. In Section 4.3 we showed how the characteristic function can be constructed from the moments by way of

$$M(\theta,\tau) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{(j\theta)^n (j\tau)^m}{n!m!} \langle t^n \omega^m \rangle \tag{10.52}$$

If we had the moments we could calculate the characteristic function and the distribution. Moments are average values and hence should be calculable directly from the signal by an operator that corresponds to the moments. We symbolically write

$$t^n \omega^m \rightarrow C_{nm}(\mathcal{T},\mathcal{W}) \tag{10.53}$$

where $C_{nm}(\mathcal{T},\mathcal{W})$ indicates a correspondence. We expect to be able to calculate the moments from

$$\langle t^n \omega^m \rangle = \int s^*(t)\, C_{nm}(\mathcal{T},\mathcal{W})\, s(t)\, dt \tag{10.54}$$

For the nonmixed moments to be correct, we want to assure that

$$t^n \rightarrow C_{n0}(\mathcal{T},0) = \mathcal{T}^n \qquad \text{and} \qquad \omega^m \rightarrow C_{0m}(0,\mathcal{W}) = \mathcal{W}^m \tag{10.55}$$

Some obvious and not so obvious choices for the correspondences are[1]

$$t^n\omega^m \rightarrow \mathcal{T}^n\mathcal{W}^m \qquad \text{[Normal]} \tag{10.56}$$

$$\rightarrow \mathcal{W}^m\mathcal{T}^n \qquad \text{[Antinormal]} \tag{10.57}$$

$$\rightarrow \tfrac{1}{2}\{\mathcal{T}^n\mathcal{W}^m + \mathcal{W}^m\mathcal{T}^n\} \qquad \text{[Symmetrization]} \tag{10.58}$$

$$\rightarrow \frac{1}{2^n}\sum_{\ell=0}^{n}\binom{n}{\ell}\mathcal{T}^{n-\ell}\mathcal{W}^m\mathcal{T}^{\ell} \tag{10.59}$$

$$= \frac{1}{2^m}\sum_{\ell=0}^{m}\binom{m}{\ell}\mathcal{W}^{m-\ell}\mathcal{T}^n\mathcal{W}^{\ell} \qquad \text{[Weyl[568]]} \tag{10.60}$$

[1] In manipulating these types of expressions the following relations are useful:

$$\mathcal{W}^m\mathcal{T}^n = \sum_{\ell=0}^{\min(m,n)} (-j)^\ell\, \ell! \binom{n}{\ell}\binom{m}{\ell} \mathcal{T}^{n-\ell}\,\mathcal{W}^{m-\ell}$$

$$\mathcal{T}^n\mathcal{W}^m = \sum_{\ell=0}^{\min(m,n)} j^\ell\, \ell! \binom{n}{\ell}\binom{m}{\ell} \mathcal{W}^{m-\ell}\,\mathcal{T}^{n-\ell}$$

They were given, in the quantum mechanical context, by Born and Jordan.[91].

$$\rightarrow \quad \frac{1}{n+1}\sum_{\ell=0}^{n} \mathcal{T}^{n-\ell}\mathcal{W}^{m}\mathcal{T}^{\ell} \tag{10.61}$$

$$= \frac{1}{m+1}\sum_{\ell=0}^{m} \mathcal{W}^{m-\ell}\mathcal{T}^{n}\mathcal{W}^{\ell} \qquad [\text{ Born-Jordan}^{[91]}] \tag{10.62}$$

$$\rightarrow \quad \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n \partial\tau^m}\,\phi(\theta,\tau)\,e^{j\theta\mathcal{T}+j\tau\mathcal{W}}\bigg|_{\theta,\tau=0} \qquad [\text{ Cohen}^{[125]}\text{ }] \tag{10.63}$$

Making a correspondence to the exponential, $e^{j\theta t+j\tau\omega}$, as we have done in Section 10.3, is equivalent to making a correspondence to the mixed moments. This is why we have given the same name to the correspondences although we have not yet proven that they are equivalent. To see this we start with

$$e^{j\theta t+j\tau\omega} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(j\theta)^n(j\tau)^m}{n!m!}\,t^n\omega^m \tag{10.64}$$

and take the correspondence of each side

$$e^{j\theta t+j\tau\omega} \rightarrow \mathcal{M} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(j\theta)^n(j\tau)^m}{n!m!}\,C_{nm}(\mathcal{T},\mathcal{W}) \tag{10.65}$$

and take average value

$$M(\theta,\tau) = \langle\mathcal{M}\rangle = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(j\theta)^n(j\tau)^m}{n!m!}\,\langle\, C_{nm}(\mathcal{T},\mathcal{W})\,\rangle \tag{10.66}$$

If we consider Eq. (10.65) as a Taylor series, the coefficients are

$$C_{nm}(\mathcal{T},\mathcal{W}) = \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n\partial\tau^m}\,\mathcal{M}(\theta,\tau)\bigg|_{\theta,\tau=0} \tag{10.67}$$

$$= \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n\partial\tau^m}\,e^{j\theta\mathcal{T}}\,e^{j\tau\mathcal{W}}\bigg|_{\theta,\tau=0} \tag{10.68}$$

However, we know that the most general characteristic function operator can be expressed in terms of the kernel, Eq. (10.36), and therefore

$$C_{nm}(\mathcal{T},\mathcal{W}) = \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n\partial\tau^m}\,\phi(\theta,\tau)\,e^{j\theta\mathcal{T}+j\tau\mathcal{W}}\bigg|_{\theta,\tau=0} \tag{10.69}$$

$$= \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n\partial\tau^m}\,\phi(\theta,\tau)\,e^{j\theta\tau/2}\,e^{j\theta\mathcal{T}}\,e^{j\tau\mathcal{W}}\bigg|_{\theta,\tau=0} \tag{10.70}$$

To see how the procedure indicated by Eq. (10.70) is carried out in practice we consider a few examples.

Example 10.3: Normal Ordering.

Take $\phi(\theta,\tau) = e^{-j\theta\tau/2}$, which gives

$$C_{nm}(\mathcal{T},\mathcal{W}) = \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n\partial\tau^m}\,e^{j\theta\mathcal{T}}\,e^{j\tau\mathcal{W}}\Big|_{\theta,\tau=0} \tag{10.71}$$

$$= \mathcal{T}^n\mathcal{W}^m \tag{10.72}$$

Example 10.4: Weyl Ordering.

We show that indeed Eq. (10.6) and Eq. (10.60) are equivalent. This was shown by McCoy.[380] For the Weyl rule the kernel is 1 and

$$C_{nm}(\mathcal{T},\mathcal{W}) = \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n\partial\tau^m}\,e^{j\theta\mathcal{T}+j\tau\mathcal{W}}\Big|_{\theta,\tau=0} \tag{10.73}$$

$$= \frac{1}{j^n j^m}\,\frac{\partial^{n+m}}{\partial\theta^n\partial\tau^m}\,e^{j\theta\tau/2}\,e^{j\theta\mathcal{T}}\,e^{j\tau\mathcal{W}}\Big|_{\theta,\tau=0} \tag{10.74}$$

$$= \frac{1}{2^m}\sum_{\ell=0}^{m}\binom{m}{\ell}\mathcal{W}^{m-\ell}\mathcal{T}^n\mathcal{W}^{\ell} \tag{10.75}$$

The step from Eq. (10.74) to Eq. (10.75) is not trivial but can be shown by using the relations in the footnote of Section 10.6.

Let us now consider an example where we calculate the characteristic function directly from the moments.

Example 10.5: Born-Jordan Distribution (Cohen[125]).

We consider the rule of Born-Jordan, Eq. (10.62), and calculate the characteristic function,

$$M(\theta,\tau) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(j\theta)^n(j\tau)^m}{n!m!}\langle C_{nm}(\mathcal{T},\mathcal{W})\rangle \tag{10.76}$$

$$= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{\ell=0}^{n}\frac{(j\theta)^n(j\tau)^m}{(n+1)n!m!}\int s^*(t)\,\mathcal{T}^{n-\ell}\mathcal{W}^m\mathcal{T}^{\ell}s(t)\,dt \tag{10.77}$$

$$= \sum_{n=0}^{\infty}\sum_{\ell=0}^{n}\frac{(j\theta)^n}{(n+1)n!}\int s^*(t)\,t^{n-\ell}e^{j\tau\mathcal{W}}t^{\ell}s(t)\,dt \tag{10.78}$$

$$= \sum_{n=0}^{\infty}\sum_{\ell=0}^{n}\frac{(j\theta)^n}{(n+1)n!}\int s^*(t)\,t^{n-\ell}\,(t+\tau)^{\ell}\,s(t+\tau)\,dt \tag{10.79}$$

where in the last step we have used the fact that $e^{j\tau\mathcal{W}}$ is the translation operator. Now

$$\sum_{\ell=0}^{n} t^{-\ell}(t+\tau)^{\ell} = \frac{(1+\tau/t)^{n+1}-1}{\tau/t} \tag{10.80}$$

and further

$$\sum_{n=0}^{\infty} \frac{t^n\,(j\theta)^n}{(n+1)n!}\,\frac{(1+\tau/t)^{n+1}-1}{\tau/t} = \frac{e^{j\theta t(1+\tau/t)}-e^{j\theta t}}{j\theta} \tag{10.81}$$

$$= e^{j\theta\tau/2+j\theta t}\,\frac{\sin\theta\tau/2}{\theta\tau/2} \tag{10.82}$$

Hence

$$M(\theta,\tau) = \frac{\sin\theta\tau/2}{\theta\tau/2}\,e^{j\theta\tau/2}\int s^*(t)\,e^{j\theta t}s(t+\tau)\,dt \tag{10.83}$$

$$= \frac{\sin\theta\tau/2}{\theta\tau/2}\int s^*(u-\tfrac{1}{2}\tau)\,e^{j\theta u}s(u+\tfrac{1}{2}\tau)\,du \tag{10.84}$$

Therefore the kernel is

$$\phi(\theta,\tau) = \frac{\sin\theta\tau/2}{\theta\tau/2} \tag{10.85}$$

For a further discussion of this distribution see Section 12.4.

Chapter 11

Kernel Design for Reduced Interference

11.1 INTRODUCTION

In our study of the Wigner distribution we saw that it sometimes gives values in the time-frequency plane that do not correspond to our intuitive sense of how the energy should be distributed. This is particularly the case when the signal is multicomponent. These spurious values are sometimes called interference terms because they happen whenever we have two or more components that "interfere." It must be understood that the word interference as used in optics is a real effect, but its usage here connotes something artificial and undesirable. A major development in the field occurred when the conditions on the kernel that minimized these spurious values were understood and implemented by Choi and Williams,[117] and Zhao, Atlas, and Marks,[626] Loughlin, Pitton and L. Atlas,[40] Jeong and Williams,[284] and Cunningham and Williams.[165] In this chapter we use a simple but general approach to see what those conditions are and also describe how this has led to the concept of kernel design.

11.2 REDUCED INTERFERENCE DISTRIBUTIONS

We now derive the conditions on the kernel so that the cross terms for multicomponent signals are small in some sense. These kernels produce "reduced interference distributions", a term coined by Williams. Williams, Jeong, and Cunningham introduced the concept of kernel design and formulated the general methodology to produce reduced interference distributions.

We shall approach the problem by first considering a simple case and subsequently show that the result is general. Consider the simplest multicomponent signal,

$$s(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \tag{11.1}$$

where A_1 and A_2 are constants. We calculate the general time-frequency distribution, leaving the kernel totally arbitrary. The generalized characteristic function, Eq. (9.4), expressed in terms of the self and cross terms is

$$M(\theta,\tau) = M_{11} + M_{22} + M_{12} + M_{21} \tag{11.2}$$

with

$$M_{11}(\theta,\tau) = 2\pi|A_1|^2\,\phi(\theta,\tau)\,e^{j\tau\omega_1}\,\delta(\theta) \tag{11.3}$$

$$M_{12}(\theta,\tau) = 2\pi A_1^* A_2\,\phi(\theta,\tau)e^{j\tau\bar{\omega}_{12}}\,\delta(\theta-\omega_1+\omega_2) \tag{11.4}$$

and

$$\bar{\omega}_{12} = \tfrac{1}{2}(\omega_1+\omega_2) \tag{11.5}$$

The terms M_{22} and M_{21} are obtained from the above by interchanging the subscripts 1 and 2. The distribution is

$$C(t,\omega) = \frac{1}{4\pi^2}\iint M(\theta,\tau)\,e^{-j\theta t-j\tau\omega}\,d\theta\,d\tau = C_{11}+C_{22}+C_{12}+C_{21} \tag{11.6}$$

where

$$C_{11} = \frac{|A_1|^2}{2\pi}\int \phi(0,\tau)\,e^{-j\tau(\omega-\omega_1)}\,d\tau \tag{11.7}$$

$$C_{12} = A_1^* A_2\,e^{-j(\omega_1-\omega_2)t}K(\omega) \tag{11.8}$$

with

$$K(\omega) = \frac{1}{2\pi}\int \phi(\omega_1-\omega_2,\tau)\;e^{-j\tau(\omega-\bar{\omega}_{12})}\,d\tau \tag{11.9}$$

To make the cross term small we must make K small. But K is a Fourier transform with respect to τ and to make a Fourier transform relatively flat the integrand must be a peaked function. Therefore, *as a function of* τ we must take $\phi(\theta,\tau)$ highly peaked.

Everything we have said for the sum of two sine waves applies to the case of a signal composed of two impulses,

$$s(t) = A_1\,\delta(t-t_1) + A_2\,\delta(t-t_2) \tag{11.10}$$

because there is total symmetry in the general equation with regard to the signal and spectrum. The effects discussed will be the same except that now we focus on the θ variable. Therefore if we don't want cross terms for the case of two impulses then the kernel must be peaked as a function of θ. For the general case, when we have two components that are not necessarily parallel to the time or frequency axis, the kernel must be peaked in both θ and τ. We also know that the characteristic function is given by $M(\theta,\tau) = \phi(\theta,\tau)A(\theta,\tau)$ and that the maximum is at the origin.

Hence the maximum of the kernel is also at the origin and we may take it to be one at the origin since that is the requirement for the preservation of total energy. If we also want the marginals to be satisfied, the kernel must be one along the θ and τ axes.

Therefore we seek kernels whose values away from either the θ or τ axis are small relative to the value at the θ and τ axes. A way to describe this region is to observe that the product $\theta\tau$ is large when we are away from either axis. We conclude that for cross term minimization

$$\phi(\theta,\tau) << 1 \qquad \text{for } \theta\tau >> 0 \tag{11.11}$$

Product Kernels. Product kernels are characterized by functions of one variable. If we let $x = \theta\tau$ then the condition for cross time minimization becomes

$$\phi(x) << 1 \qquad \text{for } x >> 0 \tag{11.12}$$

For this case

$$K(\omega) = \frac{1}{|\omega_2 - \omega_1|}\,\frac{1}{2\pi}\int \phi(x)\, e^{-jxy}\, dx \qquad ; \qquad y = \frac{\omega - \bar{\omega}_{12}}{|\omega_1 - \omega_2|} \tag{11.13}$$

Let us now see how some of the distributions we have considered thus far behave with regard to the cross term property derived above.

Example 11.1: Wigner Distribution.

For the Wigner distribution the kernel is one throughout the θ,τ plane and hence it does not satisfy the cross term minumization property.

Example 11.2: Spectrogram.

For the spectrogram the kernel depends on the window, $h(t)$. For a Gaussian window we have using Eq. (9.20)

$$\phi(\theta,\tau) = e^{-\alpha\tau^2/4 - \theta^2/(4\alpha)} \qquad [h(t) = (\alpha/\pi)^{1/4}\, e^{-\alpha t^2/2}] \tag{11.14}$$

We see that the kernel is peaked at the origin and furthermore for reasonable values of α it falls off appropriately. That is the one of the reasons the spectrogram sometimes has good cross term behavior. Note that the kernel is not one along the axes and therefore does not satisfy the marginals.

Example 11.3: Choi-Williams Distribution.

For the Choi-William distribution, which we will be studying in detail later, the kernel is $e^{-\theta^2\tau^2/\sigma}$. We can control the relative fall off by varying the parameter σ. If σ is very large, the function is relatively flat and cross term suppression is not achieved. For small values of σ the function is peaked at the origin, is one along the axis, and falls off rapidly away from the axis. Therefore this kernel satisfies the cross term minimization property.

Example 11.4: Further Generalization of Choi-Williams kernel.

Once the basic conditions for cross term minimization are understood it is easy to find kernels that satisfy the basic properties and give finer control for certain signals. One extension has been introduced by Diethorn,[175] and Papandreou and Boudreaux-Bartels,[420]

$$\phi(\theta,\tau) = e^{-(\theta/a)^n(\tau/b)^m} \tag{11.15}$$

This gives individual control in the θ, τ directions.

11.3 KERNEL DESIGN FOR PRODUCT KERNELS

An effective means for designing product kernels has been formulated by Williams et. all.[284] The constraints on product kernels transcribe themselves as very simple constraints on the Fourier transform of the kernel. In addition, since product kernels treat time and frequency on an equal footing, the two constraints, one for each domain, collapse into one. For example, to satisfy the time and frequency marginals for an arbitrary kernel we must have that $\phi(0,\tau) = 1$ and $\phi(\theta,0) = 1$. However, for product kernels both of these constraints are met by simply requiring that $\phi(0) = 1$.

The Fourier transform of a product kernel is defined by

$$h(t) = \frac{1}{2\pi}\int \phi(x)\, e^{jxt}\, dx \qquad ; \qquad \phi(\theta\tau) = \int h(t)\, e^{-j\theta\tau t}\, dt \tag{11.16}$$

In Table 11.1 we list the constraints for obtaining desirable properties on ϕ and $h(t)$.

Table 11.1. Conditions on the kernel and $h(t)$

Condition	$\phi(\theta,\tau)$	$h(t)$
Reality of distribution	ϕ : real	$h(t) = h(-t)$
Finite support	See Eq. (9.42)	$h(t) = 0$ for $\lvert t\rvert \geq \frac{1}{2}$
Cross term minimization	$\phi(x) << 1$ for $x >> 0$	Smooth tapering at $\pm\frac{1}{2}$
Marginals	$\phi(0) = 1$	$\int h(t)\, dt = 1$

The main advantage of formulating kernel design in $h(t)$ is that we have been able to collapse a number of diverse conditions into a few rather simple ones. In addition, many methods that have been developed for filter design can be used to produce such functions.

11.4 PROJECTION ONTO CONVEX SETS

Finding a kernel with all the properties we have enumerated may not be straightforward, indeed as the number of constraints increases the problem becomes that much harder. There is a method that automatically finds a kernel that satisfies a list of properties if those properties are "convex functions". A collection of functions is said to be a convex set if for any two functions, f and g, and a real number, a, a new function, h, constructed by way of

$$h = af + (1-a)g \quad \text{for } 0 \leq a \leq 1 \tag{11.17}$$

is also a member of the collection.

For the case of kernels, suppose that $\phi_1(\theta,\tau)$ and $\phi_2(\theta,\tau)$ are kernels that satisfy a particular constraint. Forming a new kernel by

$$\phi_3(\theta,\tau) = a\,\phi_1(\theta,\tau) + (1-a)\,\phi_2(\theta,\tau) \tag{11.18}$$

we ask whether, for any a between zero and one, $\phi_3(\theta,\tau)$ also satisfies that constraint? If it does, then the set of kernels satisfying this constraint form a convex set. For example, consider the time marginal constraint, which is $\phi(\theta,0) = 1$. For any two functions that satisfy this constraint, will ϕ_3 satisfy it? Consider

$$\phi_3(\theta,0) = a\,\phi_1(\theta,0) + (1-a)\,\phi_2(\theta,0) = a + (1-a) = 1 \tag{11.19}$$

and we see that $\phi_3(\theta,0)$ satisfies the time marginal constraint. Hence such functions form a convex set. The other constraints we have considered also form convex sets.

The method called "projection onto convex sets" automatically picks out the functions, if they exist, that satisfy all the conditions. Furthermore, if such a function does not exist then the method picks out the best function in the mean square sense. This method was devised by Oh, Marks II, Atlas, and Pitton.[408]

11.5 BARANIUK–JONES OPTIMAL KERNEL DESIGN

In Chapter 9 we mentioned that the kernel can be signal dependent and most of the constraints on the kernel we have derived apply to signal dependent kernels. Once we choose a signal dependent kernel we are no longer dealing with bilinear distributions. Signal dependent kernels are important for many reasons. We shall see in Section 14.2 how manifestly positive distributions may be obtained by choosing signal dependent kernels. In Section 11.2 we have seen how to choose kernels that minimize the cross terms, the criterion being that the kernel must be peaked near the origin. However, it must be peaked in such a way it that encompasses the self terms, which is clearly a signal dependent condition. A method to obtain kernels that do that effectively by adapting the kernel to the particular signal at hand has been devised by Baraniuk and Jones.[48] First the ambiguity function, $A(\theta,\tau)$, is calculated, and from it we form the generalized ambiguity function, $M(\theta,\tau)$, and

define a functional of the kernel by

$$f[\phi] = \iint |M(\theta,\tau)|^2 \, d\tau \, d\theta = \iint |A(\theta,\tau)\,\phi(\theta,\tau)|^2 \, d\tau \, d\theta \tag{11.20}$$

Since the self terms are concentrated around the origin we want this functional of the kernel to be maximized. The maximum of the characteristic function is at the origin and therefore one imposes in the maximization the constraint that the kernel never increases in any radial direction. Furthermore, we want to control how concentrated around the origin the kernel should be. This is achieved by controlling

$$\iint |\phi(\theta,\tau)|^2 \, d\tau \, d\theta \leq \alpha_c \tag{11.21}$$

One chooses α_c and constrains the maximization of the functional f with this constraint. The bigger α_c the more concentrated the kernel will be. In addition to the above, we may also want to constrain the solution to give the correct marginals and other quantities. An illustration is given in Fig. 11.1.

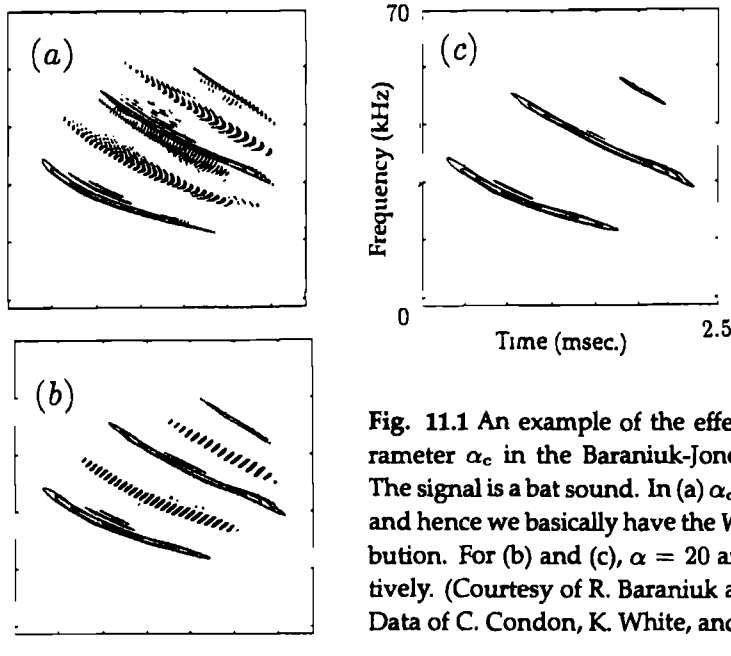

Fig. 11.1 An example of the effect of the parameter α_c in the Baraniuk-Jones approach. The signal is a bat sound. In (a) α_c is very large and hence we basically have the Wigner distribution. For (b) and (c), $\alpha = 20$ and 4, respectively. (Courtesy of R. Baraniuk and D. Jones. Data of C. Condon, K. White, and A. Feng.)

Chapter 12

Some Distributions

12.1 INTRODUCTION

We have already discussed in detail the Wigner distribution and the spectrogram. In this chapter we study additional distributions (some new, some old), and describe their properties and the physical and mathematical reasons for their introduction.

12.2 CHOI-WILLIAMS METHOD

As discussed in the last chapter, Choi, Jeong, Cunningham, and Williams[117, 284, 165] developed the theory of reduced interference distributions and the ideas that allow one to design kernels to accomplish that. Their first example was the distribution determined by the kernel

$$\phi(\theta,\tau) = e^{-\theta^2\tau^2/\sigma} \tag{12.1}$$

where σ is a parameter. We have already used this kernel in a number of examples for illustrative purposes. It is a product kernel and $\phi(0,\tau) = \phi(\theta,0) = 1$, which shows that both marginals are satisfied. Furthermore, it is straightforward to see that it satisfies the instantaneous frequency and group delay properties. If we take σ to be large then the Choi-Williams distribution approaches the Wigner distribution since the kernel then approaches one. For small σ it satisfies the reduced interference criteria, as discussed in the previous chapter. Substituting the kernel into the general class and integrating over θ, we obtain

$$P_{CW}(t,\omega) = \frac{1}{4\pi^{3/2}} \iint \frac{1}{\sqrt{\tau^2/\sigma}} \exp\left[-\frac{(u-t)^2}{4\tau^2/\sigma} - j\tau\omega\right] \tag{12.2}$$

$$\times \quad s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, du\, d\tau \tag{12.3}$$

To better understand the behavior of this distribution we consider a number of ex-

amples. For the sum of two sine waves

$$s(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \tag{12.4}$$

the distribution can be calculated exactly,

$$C_{CW}(t,\omega) = A_1^2\delta(\omega-\omega_1) + A_2^2\delta(\omega-\omega_2) + 2A_1A_2\cos[(\omega_2-\omega_1)t]\,\eta(\omega,\omega_1,\omega_2,\sigma) \tag{12.5}$$

where

$$\eta(\omega,\omega_1,\omega_2,\sigma) = \sqrt{\frac{1}{4\pi(\omega_1-\omega_2)^2/\sigma}}\ \exp\left[-\frac{(\omega-\frac{1}{2}(\omega_1+\omega_2)^2}{4(\omega_1-\omega_2)^2/\sigma}\right] \tag{12.6}$$

For large σ

$$\lim_{\sigma\to\infty}\eta(\omega,\omega_1,\omega_2,\sigma) = \delta(\omega-\tfrac{1}{2}(\omega_1+\omega_2)) \tag{12.7}$$

and therefore the distribution will be infinitely peaked at $\omega = \frac{1}{2}(\omega_1+\omega_2)$, which is precisely the Wigner distribution. As long as σ is kept finite, the cross terms will be finite and not concentrated at one point. They spread out with a lower maximum intensity. In Fig. 12.1 we illustrate the effect of different choices of σ where the delta functions are symbolically represented. Notice that the self terms are still delta functions along the frequencies ω_1 and ω_2. In general, this will not be true for other signals and it is not generally true of the Wigner distribution either. It is true for the Wigner distribution of a chirp only. For a chirp the Choi-Williams distribution is

$$C_{CW}(t,\omega) = \frac{1}{2\pi}\int e^{-\beta^2\tau^4/\sigma}\, e^{-j\tau(\omega-\beta t)}\,d\tau \qquad [s(t) = e^{\beta t^2/2)}] \tag{12.8}$$

The concentration is along the instantaneous frequency and the width depends on both β and σ. Figs. 12.1 – 12.6 illustrate the Choi-Williams distribution for a number of analytic and real signals and in some cases we compare it to the spectrogram and/or Wigner distribution.

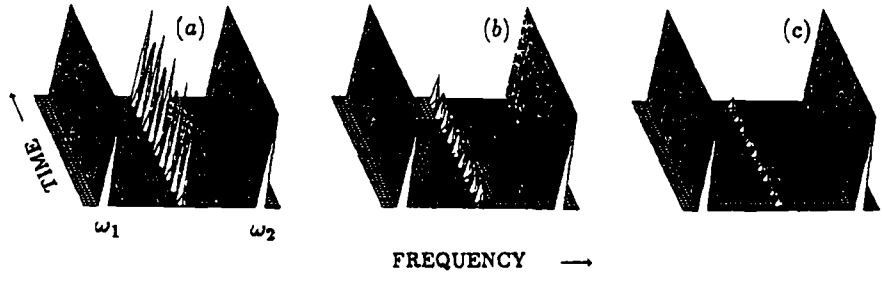

Fig. 12.1 The Wigner distribution (a) and Choi-Williams distribution (b) and (c), for the sum of two sine waves. Both distributions are infinitely peaked at the frequencies of the two sine waves. In the Wigner distribution the cross terms also go to infinity. In the Choi-Williams distribution their intensity is a function of the parameter σ. In (b) we have used a large σ, while in (c) a small σ is used. The Choi-Williams distribution becomes the Wigner distribution for $\sigma \to \infty$.

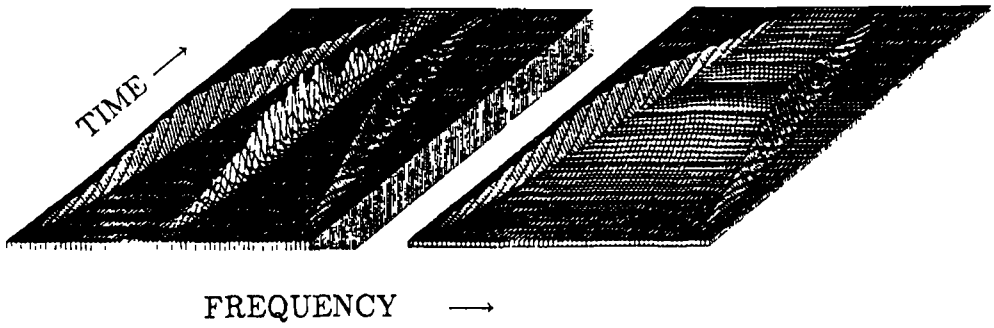

Fig. 12.2 A comparison of the Wigner (left) and Choi-Williams (right) distribution for the sum of two chirps. (Courtesy of W. Williams.)

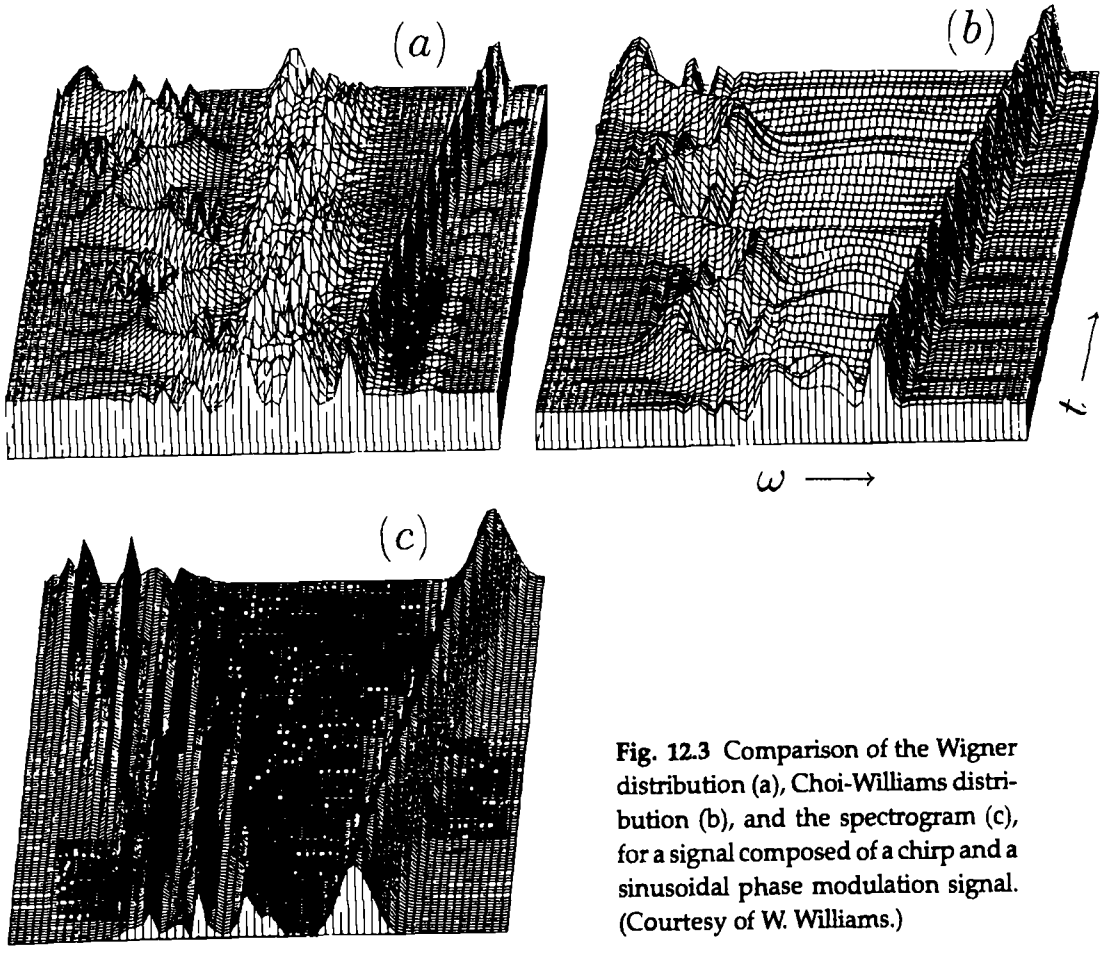

Fig. 12.3 Comparison of the Wigner distribution (a), Choi-Williams distribution (b), and the spectrogram (c), for a signal composed of a chirp and a sinusoidal phase modulation signal. (Courtesy of W. Williams.)

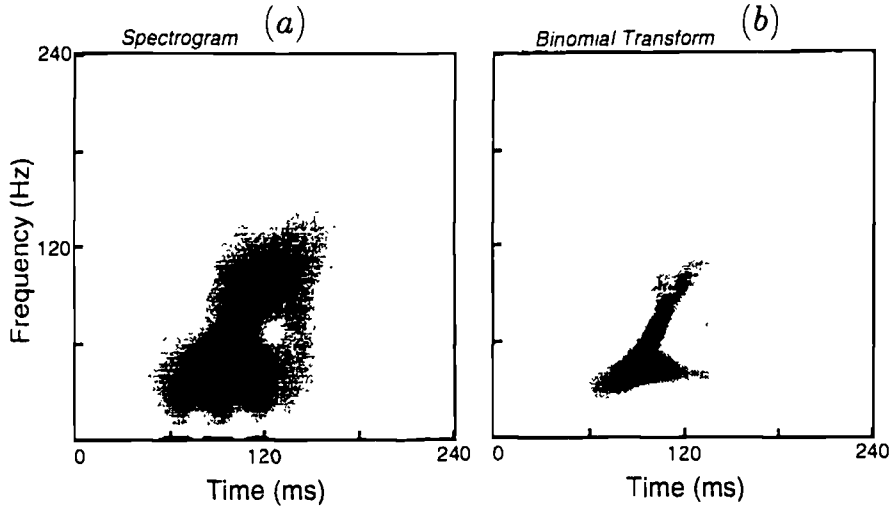

Fig. 12.4 Comparison of the spectrogram and the binomial distributions for the part of a heart sound due to the closing of the valve. The binomial distribution is a reduced interference distribution devised by Jeong and Williams. (Courtesy of J. Wood and D. T. Barry.)

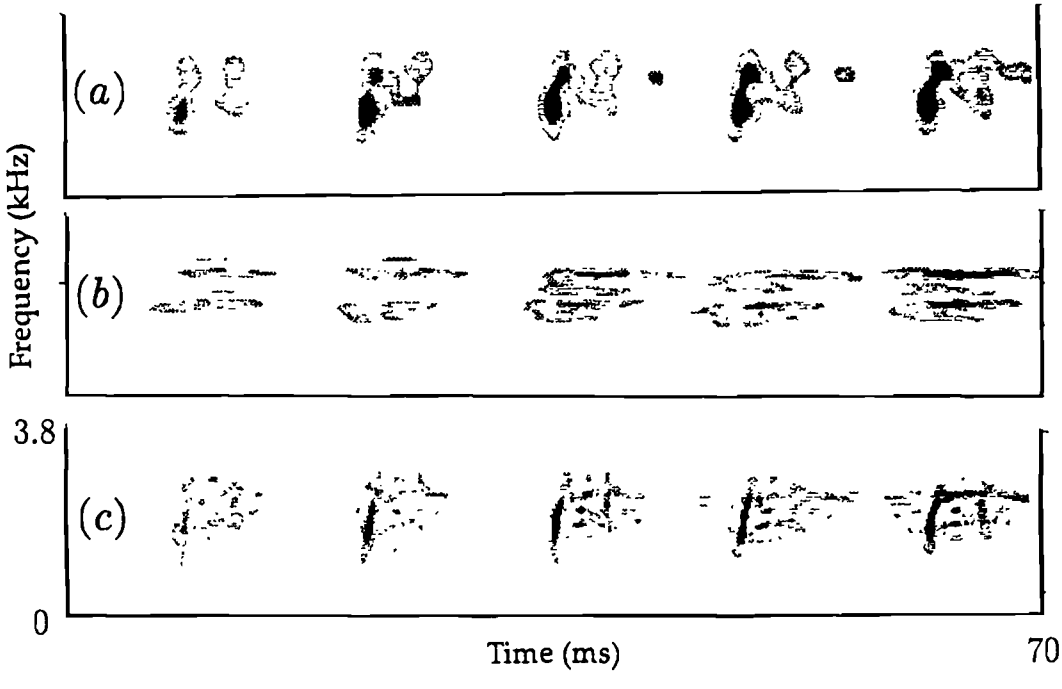

Fig. 12.5 A comparison of the narrowband spectrogram (a), wideband spectrogram (b), and Choi-Williams distribution (c) for a dolphin sound. (Courtesy of W. Williams and P. Tyack.)

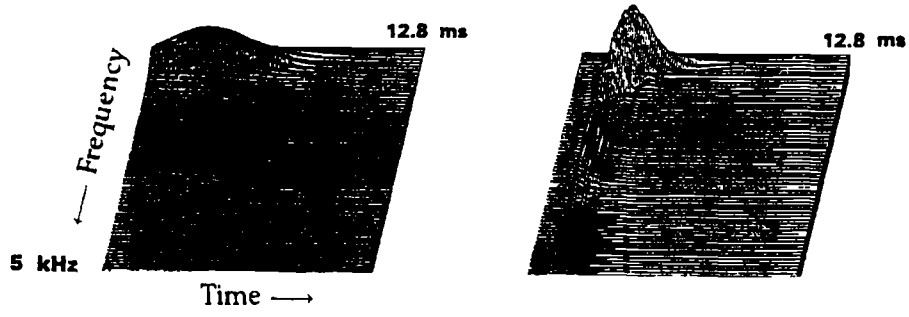

Fig. 12.6 In some individuals the jaw makes a click when speaking. These sounds are called TMJ sounds. Because of their very short duration the spectrogram (left) does not properly reveal the time-frequency structure. On the right is the Choi-Williams distribution. (Courtesy of S. E. Widmalm, W. J. Williams and C. Zheng.)

12.3 ZHAO-ATLAS-MARKS DISTRIBUTION[626]

In the Choi-Williams distribution the cross terms are spread out in the time frequency plane with low intensity. There is something else we can do with them. We can obscure them by placing them under the self terms, as shown by Loughlin, Pitton, and Atlas.[341]

Consider the case of the sum of two sine waves as in Section 11.2. The intensity of the cross terms are proportional to $K(\omega)$ as given by Eq. (11.9). For a kernel of the form

$$\phi(\theta,\tau) = f(\theta,\tau)\sin a\theta\tau \tag{12.9}$$

where a is a constant, we have

$$K(\omega) = \frac{1}{4\pi j}\int f(\omega_1-\omega_2,\tau) \left[e^{-j\tau[\omega-(a+\frac{1}{2})\omega_1+(a-\frac{1}{2})\omega_2])} - e^{-j\tau[\omega+(a-\frac{1}{2})\omega_1-(a+\frac{1}{2})\omega_2]}\right] d\tau \tag{12.10}$$

Therefore, the cross terms will be functions of $\omega = (a+\frac{1}{2})\omega_1 - (a-\frac{1}{2})\omega_2$ and $\omega = -(a-\frac{1}{2})\omega_1 + (a+\frac{1}{2})\omega_2$. Something unique happens at $a = \frac{1}{2}$. At that value we have

$$K(\omega) = \frac{1}{4\pi j}\int f(\omega_1-\omega_2,\tau)\left[e^{-j\tau(\omega-\omega_1)} - e^{-j\tau(\omega-\omega_2)}\right] d\tau \tag{12.11}$$

and we see that $K(\omega)$ is a function of $\omega-\omega_1$ and $\omega-\omega_2$, which are the location of the self terms. Loughlin, Pitton, and Atlas[341] have shown that for K to fall exactly on the self terms, the kernel $f(\theta,\tau)$ must produce a distribution whose cross terms lie exactly midway between the self terms.

Depending on the choice of $f(\theta,\tau)$ these kernels may or may not satisfy other desirable properties, such as the marginals. But nonetheless they can give a good

indication of the time-frequency structure of a signal. Note that for two impulses cross terms will be produced. However, for multicomponent signals that are mostly parallel in the frequency direction the cross terms will not be prominent because they will be hidden under the self terms.

A distribution with this property is the Zhao, Atlas, and Marks distribution,[626]

$$\phi_{ZAM}(\theta,\tau) = g(\tau)\,|\tau|\,\frac{\sin a\theta\tau}{a\theta\tau} \tag{12.12}$$

which gives

$$C(t,\omega) = \frac{1}{4\pi a}\int g(\tau)\,e^{-j\tau\omega}\int_{t-a|\tau|}^{t+a|\tau|} s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,du\,d\tau \tag{12.13}$$

for the distribution. In the original work of Zhao, Atlas, and Marks, g was taken to be equal to one and a to one half. This distribution has many interesting properties and a comprehensive analysis of it has been given by Oh and Marks.[408] In Figs. 12.7 and 12.8 we show two examples.

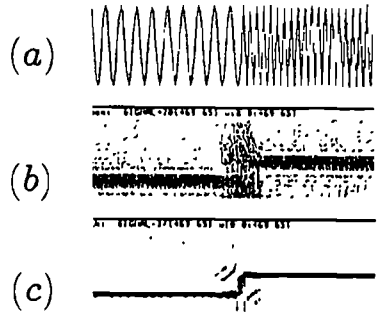

Fig. 12.7 A comparison of the spectrogram (b) and ZAM distribution (c) for a signal (a) that changes frequency abruptly. (Courtesy of Y. Zhao, L. Atlas, and R. Marks.)

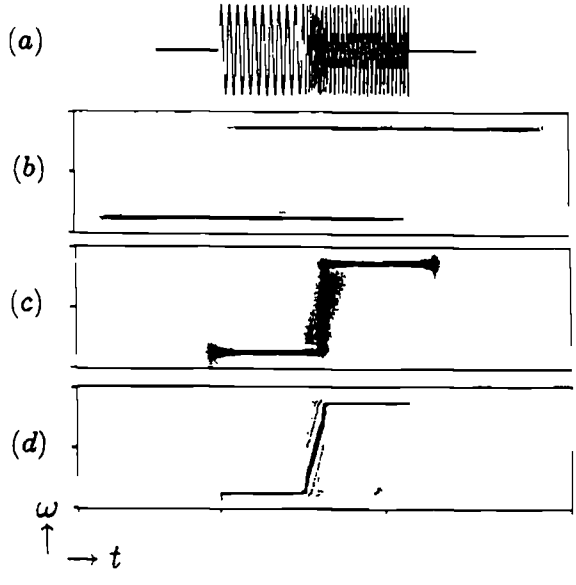

Fig. 12.8 The signal (a) changes from constant frequency to a higher frequency linearly in a finite time. In (b) is a narrowband spectrogram; in (c) we have a wide band spectrogram and in (d) we have the Zhao, Atlas, Marks distribution. (Courtesy of P. Loughlin, J. Pitton, and L. Atlas.)

12.4 BORN-JORDAN DISTRIBUTION[1]

In Section 10.6 we derived the sinc distribution using the correspondence rule of Born and Jordan. This distribution was first derived[125] in 1966, although its properties were not understood until the work of Jeong and Williams,[285] who studied its general properties and pointed out that it satisfies the properties of a reduced interference distributions. Loughlin, Pitton, and Atlas[40] and Cohen and Lee[134] have also considered its properties.

For the kernel take

$$\phi(\theta,\tau) = \frac{\sin(a\theta\tau)}{a\theta\tau} \tag{12.14}$$

and the distribution is

$$C(t,\omega) = \frac{1}{4\pi a}\int \frac{1}{|\tau|}\, e^{-j\tau\omega} \int_{t-a|\tau|}^{t+a|\tau|} s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, du\, d\tau \tag{12.15}$$

For the case of two sine waves as per Eq. (11.1)

$$K(\omega) = \frac{1}{2a}\frac{1}{\omega_2-\omega_1} \qquad \text{if} \quad (a+\tfrac{1}{2})\,\omega_1 - (a-\tfrac{1}{2})\,\omega_2 \le \omega \le -(a-\tfrac{1}{2})\,\omega_1 + (a+\tfrac{1}{2})\,\omega_2 \tag{12.16}$$

and zero otherwise. In deriving this expression we have assumed that $\omega_2 \ge \omega_1$. Thus the cross terms are uniformly spread in the region indicated. For the case $a = \frac{1}{2}$ we have

$$K(\omega) = \frac{1}{\omega_2-\omega_1} \qquad \text{if} \quad \omega_1 \le \omega \le \omega_2 \tag{12.17}$$

in which case the cross terms are totally restricted between the two frequencies.

12.5 COMPLEX ENERGY SPECTRUM

In Section 10.3 we derived the Margenau-Hill distribution by specifying a particular ordering of the characteristic function operator. Rihaczek[484] gave a plausibility argument based on physical grounds. Suppose we have a time dependent voltage and decompose it into its Fourier components,

$$V(t) = \frac{1}{\sqrt{2\pi}}\int V_\omega\, e^{j\omega t}\, d\omega \tag{12.18}$$

We can think of $V_\omega e^{j\omega t}$ as the voltage at a particular frequency. If we assume a unit resistance, the current for that frequency is $i_\omega = V_\omega e^{j\omega t}$. Hence the total current in the frequency band ω to $\omega+\Delta\omega$ is

$$i(t) = \int_\omega^{\omega+\Delta\omega} i_\omega(t)\, d\omega = \frac{1}{\sqrt{2\pi}}\int_\omega^{\omega+\Delta\omega} V_\omega\, e^{j\omega t}\, d\omega \tag{12.19}$$

[1] Born and Jordan did not consider joint distributions. This distribution was derived in reference [125] and called the Born-Jordan distribution because the derivation is based on the Born-Jordan rule. The derivation is repeated in Section 10.6.

Power is the product of voltage and current $V(t)\,i^*(t)$ and is the energy per unit time. Hence the energy in the time interval Δt and frequency interval $\Delta\omega$ is

$$E(t,\omega) = \int_t^{t+\Delta t} V(t)\,i^*(t)\,dt = \frac{1}{\sqrt{2\pi}} \int_t^{t+\Delta t} \int_\omega^{\omega+\Delta\omega} V_\omega^* V(t)\,e^{-j\omega t}\,d\omega\,dt \quad (12.20)$$

and therefore the energy density in time and frequency is

$$e(t,\omega) = \lim_{\Delta t,\Delta\omega\to 0} \frac{E(t,\omega)}{\Delta t\,\Delta\omega} = \frac{1}{\sqrt{2\pi}} V_\omega^* V(t)\,e^{-j\omega t} \quad (12.21)$$

$$= \frac{1}{\sqrt{2\pi}} s(t)\,S^*(\omega)\,e^{-j\omega t} \quad (12.22)$$

where in the last step we have identified the signal with the voltage, $s(t) = V(t)$, in which case $V_\omega = S(\omega)$. Equation (12.22) is the Margenau-Hill distribution derived in Section 10.3, also called the Rihaczek distribution and complex energy spectrum. Although it does satisfy the marginals it does not have many other desirable properties. In particular, it does not satisfy the instantaneous frequency condition, although it does satisfy the strong finite support property.

12.6 RUNNING SPECTRUM

The running spectrum, introduced by Page,[419] is a method to define a time frequency distribution. The Fourier transform of a signal considers the signal for all time. Suppose we consider it only up to a time t,

$$S_t^-(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} s(t')\,e^{-j\omega t'}\,dt' \quad (12.23)$$

This is the running Fourier transform defined by Page.

If the signal was indeed zero after time t, then the marginal in frequency is $|\,S_t^-(\omega)\,|^2$. If a distribution, $P^-(t,\omega)$, satisfies this marginal, then

$$\int_{-\infty}^{t} P^-(t',\omega)\,dt' = |\,S_t^-(\omega)\,|^2 \quad (12.24)$$

Differentiating with respect to time gives

$$P^-(t,\omega) = \frac{\partial}{\partial t}|\,S_t^-(\omega)\,|^2 = 2\,\mathrm{Re}\,\frac{1}{\sqrt{2\pi}} s^*(t)\,S_t^-(\omega)\,e^{j\omega t} \quad (12.25)$$

This is the Page distribution. Its kernel is $\phi(\theta,\tau) = e^{j\theta|\tau|/2}$. By inspection we see that it satisfies the marginals. The main characteristic of this distribution is that the future does not affect the past and hence the longer a frequency exists, the larger the intensity of that frequency as time increases. Once the particular frequency stops, the distribution at that frequency remains constant as time increases. The Page distribution satisfies the weak and strong finite support properties.

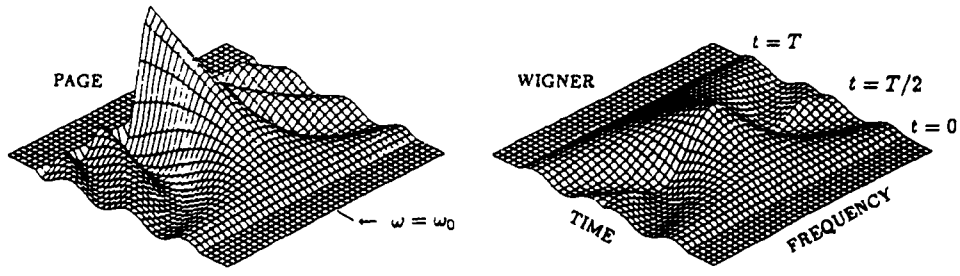

Fig. 12.9 A comparison of the Wigner distribution and Page distribution for a finite duration sine wave. The Page distribution does not go to zero at the end of the signal.

Example 12.1: Finite Sinusoid.

For the finite duration sinusoid,

$$s(t) = e^{j\omega_0 t} \qquad 0 \le t \le T \tag{12.26}$$

the running transform and distribution are

$$S_t^-(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{j\omega_0 t'} e^{-j\omega t'} \, dt' = j \frac{e^{-j(\omega-\omega_0)t} - 1}{(\omega - \omega_0)} \tag{12.27}$$

$$P^-(t,\omega) = \begin{cases} \dfrac{t}{\pi} \operatorname{sinc}(\omega - \omega_0)t & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases} \tag{12.28}$$

As time increases, the distribution becomes more and more peaked at ω_0. This is illustrated in Fig. 12.9.

Variations of the Page derivation were given by Turner,[543] Levin,[332] and others. Levin defined the future running transform by

$$S_t^+(\omega) = \frac{1}{\sqrt{2\pi}} \int_t^\infty s(t') \, e^{-j\omega t'} \, dt' \tag{12.29}$$

and using the same argument

$$\int_t^\infty P^+(t',\omega) \, dt' = |\, S_t^+(\omega) \,|^2 \tag{12.30}$$

and

$$P^+(t,\omega) = -\frac{\partial}{\partial t} |\, S_t^+(\omega) \,|^2 = 2 \operatorname{Re} \frac{1}{\sqrt{2\pi}} \, s^*(t) \, S_t^+(\omega) \, e^{j\omega t} \tag{12.31}$$

Note that if the two distributions are averaged we get the Rihaczek distribution.

Similar to the running Fourier transform, we can define the running signal transform for frequencies up to ω by

$$s_{\omega}^{-}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\omega} S(\omega') e^{j\omega' t} d\omega' \tag{12.32}$$

which gives the distribution

$$\hat{P}^{-}(t,\omega) = \frac{\partial}{\partial \omega} |s_{\omega}^{-}(t)|^2 = 2\,\mathrm{Re}\, \frac{1}{\sqrt{2\pi}} s_{\omega}^{-}(t) S^*(\omega) e^{-j\omega t} \tag{12.33}$$

Fig. 12.10 compares the Wigner, Page, and Margenau-Hill distributions for a finite duration signal that has been turned on and off as illustrated.

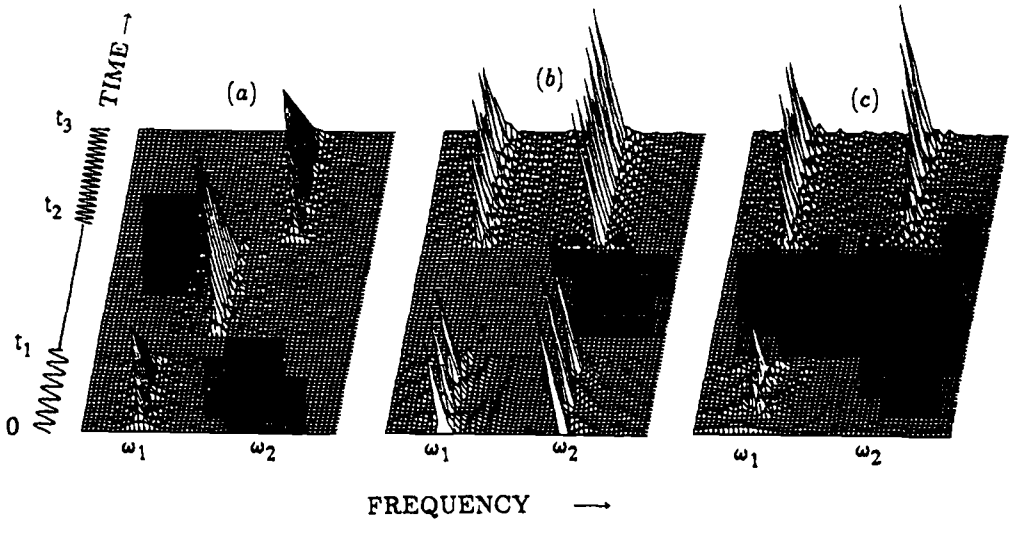

Fig. 12.10 A comparison of the Wigner (a), Margenau-Hill (b), and Page (c) distribution for a finite duration signal that is turned off and on as illustrated.

Chapter 13

Further Developments

13.1 INTRODUCTION

In this chapter we address a number of topics that are central to the description of signals in time and frequency.

13.2 INSTANTANEOUS BANDWIDTH[135, 145]

In previous chapters we have given a number of arguments to indicate that instantaneous frequency is the derivative of the phase. In Chapter 9 we showed that many time-frequency distributions give the derivative of the phase for the first conditional moment,

$$\langle \omega \rangle_t = \varphi'(t) \tag{13.1}$$

From this point of view, instantaneous frequency is an average, the average of the frequencies existing at a particular time. If instantaneous frequency is a conditional average it is natural to then ask for the conditional standard deviation about that average. Because it is the spread of frequencies at a particular time, we will call it the instantaneous bandwidth and use B_t or $\sigma_{\omega|t}$ to denote it. We now discuss a number of arguments that point to taking

$$B_t = \sigma_{\omega|t} = \left| \frac{A'(t)}{A(t)} \right| \tag{13.2}$$

as a plausible expression for instantaneous bandwidth. This expression does not depend on the phase, it depends only on the amplitude of the signal. This is reasonable since standard deviation is the spread about the mean and therefore the location of the mean is immaterial. We will present three arguments for its reasonableness.

Bandwidth Equation. In Section 1.5 we expressed the bandwidth of a signal in terms of its amplitude and phase,

$$B^2 = \int \left(\frac{A'(t)}{A(t)}\right)^2 A^2(t)\,dt + \int (\varphi'(t) - \langle\omega\rangle)^2 A^2(t)\,dt \qquad (13.3)$$

The second term in this expression averages all the deviations of the instantaneous frequency from the average frequency and certainly corresponds to our intuitive understanding of spread. But what is the meaning and origin of the first term? In Section 4.6 we showed that for any joint density, $P(x,y)$, the global spread, σ_y, is related to the conditional spread, $\sigma_{y\,x}$, in the following manner

$$\sigma_y^2 = \int \sigma_{y|x}^2\, P(x)\,dx + \int (\langle y\rangle_x - \langle y\rangle)^2\, P(x)\,dx \qquad (13.4)$$

The similarity of this general result with the bandwidth equation, Eq. (13.3), is suggestive. Associating x with time and y with frequency and comparing, we can immediately infer Eq. (13.2) for the conditional standard deviation. Of course, one cannot unambiguously equate quantities under the integral sign since we can always add a quantity that integrates to zero. However, if we added anything that integrates to zero, the expression for the standard deviation would not remain manifestly positive.

Joint Time-Frequency Representations. In Section 9.6 we showed that for distributions that satisfy the marginals and give the derivative of the phase for the instantaneous frequency, the standard deviation for a given time is

$$\sigma_{\omega|t}^2 = \tfrac{1}{2}[1 + 4\phi''(0)]\left(\frac{A'(t)}{A(t)}\right)^2 - \tfrac{1}{2}[1 - 4\phi''(0)]\frac{A''(t)}{A(t)} \qquad (13.5)$$

As mentioned in Section 9.6, the choice $\phi''(0) = \frac{1}{4}$ leads to a positive spread and in particular to Eq. (13.2).

There are an infinite number of densities that satisfy the condition $\phi''(0) = \frac{1}{4}$. Furthermore, it is not surprising that there are an infinite number of distributions that could produce a positive standard deviation. What is surprising is that they all produce the same result. All possible distributions that always give a positive result for the conditional standard deviation give the same result, namely, $|A'(t)/A(t)|$.

Instantaneous Bandwidth for the Spectrogram. In Chapter 7 we derived for the spectrogram the conditional first moment of frequency, Eq. (7.45). That expression depends on both the signal and window. We then showed that as we narrow the window, the conditional moment approaches the instantaneous frequency of the signal. We also obtained an exact expression for the instantaneous bandwidth for the spectrogram, Eq. (7.48), and discussed that it is reasonable that the expression goes to infinity as we narrow the window. The reason is that progressively narrowing the window is tantamount to progressively making shorter duration signals

with a corresponding increase in bandwidth. While the spectrogram can be used to estimate the instantaneous frequency, it appears that it cannot be used to estimate the instantaneous bandwidth. However, as the window gets narrower it can be shown[139] that the expression for the instantaneous bandwidth of the spectrogram, Eq. (7.48), breaks up into a number of terms, one of them window independent and the rest window dependent. In particular,

$$\sigma^2_{\omega|t} \to \left(\frac{A'(t)}{A(t)}\right)^2 + \text{window-dependent terms} \qquad [\text{spectrogram}] \qquad (13.6)$$

Thus we see that the window independent term is precisely the instantaneous bandwidth of the signal.

Purely Frequency Modulated Signals. The above three plausibility derivations all lead to $|A'(t)/A(t)|$ for the instantaneous bandwidth. An important consequence is that for purely frequency modulated signals, that is, signals whose amplitude is constant, the instantaneous bandwidth is zero,

$$B_t = \sigma_{\omega|t} = 0 \qquad \text{for} \qquad s(t) = A\,e^{j\varphi(t)} \qquad [\text{if } A = \text{constant}] \qquad (13.7)$$

Physically, what this means is that at each instant of time there is only one frequency, the derivative of the phase. This is in conformity with our intuition that for a purely frequency modulated signal the instantaneous frequency is known precisely. Notice also that for this case the global bandwidth is

$$B^2 = \int (\varphi'(t) - \langle\omega\rangle)^2\, dt \qquad (13.8)$$

For purely frequency modulated signals the spread in frequencies comes from the change in the instantaneous frequency only.

Constant Bandwidth. We now ask for what signals is the instantaneous bandwidth constant? Solving for $|A'(t)/A(t)| = p$ we immediately have

$$A(t) \sim a\,e^{-pt} \qquad (13.9)$$

Therefore signals whose amplitudes are decaying exponentials have a constant instantaneous bandwidth. For these signals the global bandwidth is

$$B^2 = p^2 + \int (\varphi'(t) - \langle\omega\rangle)^2\, dt \qquad (13.10)$$

Decay Rate Model and Formant Bandwidth.[149] For a decaying exponential the broadness is proportional to the decay constant. This fact can be used to obtain a sense of the broadness of an arbitrary signal at time t by fitting an exponential at that time,

$$s_F(t) = a\,e^{-pt}\cos(\omega t + \varphi_0) \qquad (13.11)$$

where $s_F(t)$ signifies the fitted curve around time t. The adjustable numbers to obtain the best fit are the three constants a, p, φ_0. The fitting is done locally, that is, we take a small piece of the function at that time and do the best possible fit by obtaining the constants a, p, and φ_0. Since we are trying to fit locally, we must decide on the interval of time around the time of interest. In the case of speech, 5 milliseconds is typically used. Once the curve is fitted, the broadness at that time is then given by p. For different times we have different sets of "constants".

This idea has been utilized in speech where the p's thus obtained are called formant bandwidth. However, speech is a multicomponent signal and the fitting is done by a sum of decaying exponentials. In this way we obtain the spread of each component at each time.

We now want to show the relation between this method of estimating a local bandwidth and Eq. (13.2). For simplicity we consider one component. Suppose the signal we are trying to fit is of the form

$$s(t) = A(t)\cos(\varphi(t)) \tag{13.12}$$

Expanding the amplitude and phase separately in a Taylor series about some time $t = t_0$, we have

$$s(t) = [\, A(t_0) + A'(t_0)(t-t_0)\cdots] \quad \cos(\, \varphi(t_0) + \varphi'(t_0)(t-t_0)\cdots) \tag{13.13}$$

Similarly, we expand the presumed fitted curve,

$$s_F \sim a\, e^{-p(t-t_0)} \cos(\omega(t-t_0)+\varphi_0) \sim a\,(1-p(t-t_0)\cdots)\cos(\omega(t-t_0)+\varphi_0) \tag{13.14}$$

In the limit of $t \to t_0$ we see that the best fit is when

$$a = A(t_0) \qquad ; \qquad -ap = A'(t_0) \tag{13.15}$$

which gives

$$|p| \sim \left| \frac{A'(t_0)}{A(t_0)} \right| \tag{13.16}$$

We see that this method of obtaining the spread in frequencies approximates the instantaneous bandwidth defined by Eq. (13.2).

Poletti Formulation. An interesting formulation of these concepts was devised by Poletti[446] who defined a new signal by

$$\beta(t) = \frac{d}{dt}\log s(t) = \frac{A'(t)}{A(t)} + j\varphi'(t) \tag{13.17}$$

This signal is called the dynamical signal. The relation between the dynamical signal and the real signal is $s'(t) = \beta(t)\, s(t)$. Now the instantaneous frequency and instantaneous bandwidth are simply given by

$$\omega_i = \langle \omega \rangle_t = \operatorname{Im}\beta(t) \qquad\qquad B_t = \operatorname{Re}\beta(t) \tag{13.18}$$

where Im and Re stand for the imaginary and real parts, respectively.

Group Delay and Its Spread. All the concepts and results developed for instantaneous frequency can be applied to group delay and its spread. The analogy is both physical and mathematical. We can think of group delay as the average time for a particular frequency. It is given by the negative of the derivative of the spectral phase. In complete mathematical analogy with our consideration of the spread about the instantaneous frequency we have the spread about the group delay,

$$\sigma_{t|\omega} = \left| \frac{B'(\omega)}{B(\omega)} \right| \tag{13.19}$$

where $B(\omega)$ is the spectral amplitude.

13.3 MULTICOMPONENT SIGNALS

We have frequently alluded to multicomponent signals and emphasized that a multicomponent signal is not just the sum of signals. We are now in a position to understand their nature and in particular to address the question of when a signal is multicomponent. The concept of instantaneous frequency and instantaneous bandwidth developed in the previous section will be central to the explanation. The origin of the concept of multicomponent signals arose with the observation that sometimes there are well delineated regions in the time-frequency plane. Perhaps the earliest observation of this was in the study of speech. The delineated regions are called formants. It is important to note that we get components with almost any distribution although the particular details may be different.

Before we address the case of multicomponent signals let us characterize a mono component signal. Generally speaking, a monocomponent signal will look like a single mountain ridge, as illustrated in Fig. 13.1. At each time the ridge is characterized by the peak. If it is well localized the peak is the instantaneous frequency. The width of the ridge is the conditional standard deviation (actually about twice that), which is the instantaneous bandwidth.

A typical multicomponent signal is illustrated in Fig. 13.2. It consists of two (or more) ridges, each characterized by its own instantaneous frequency and instantaneous bandwidth. Why do we have two ridges instead of one? Because the width of each one is small in comparison to the ridge separation. Therefore if we have a signal of the form

$$s(t) = s_1(t) + s_2(t) = A_1(t)\, e^{j\varphi_1(t)} + A_2(t)\, e^{j\varphi_2(t)} \tag{13.20}$$

we will have a multicomponent signal if the instantaneous bandwidths of each part are small in comparison to the separation between the ridges. But the separation between the ridges is given by the difference in instantaneous frequency and hence the condition for a multicomponent signal is that

$$\left| \frac{A_1'(t)}{A_1(t)} \right| \quad , \quad \left| \frac{A_2'(t)}{A_2(t)} \right| << \left| \varphi_2'(t) - \varphi_1'(t) \right| \tag{13.21}$$

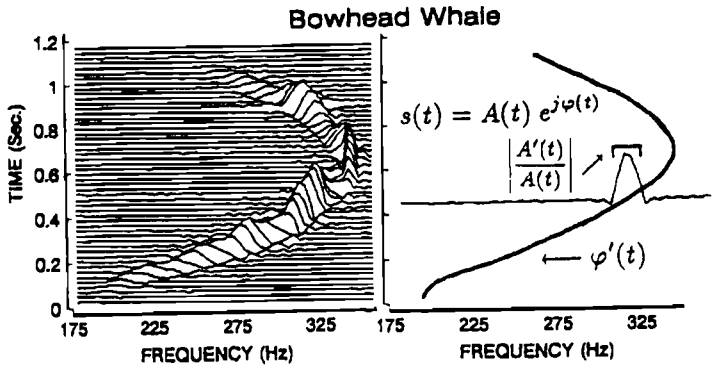

Fig. 13.1 The characteristics of a monocomponent signal.

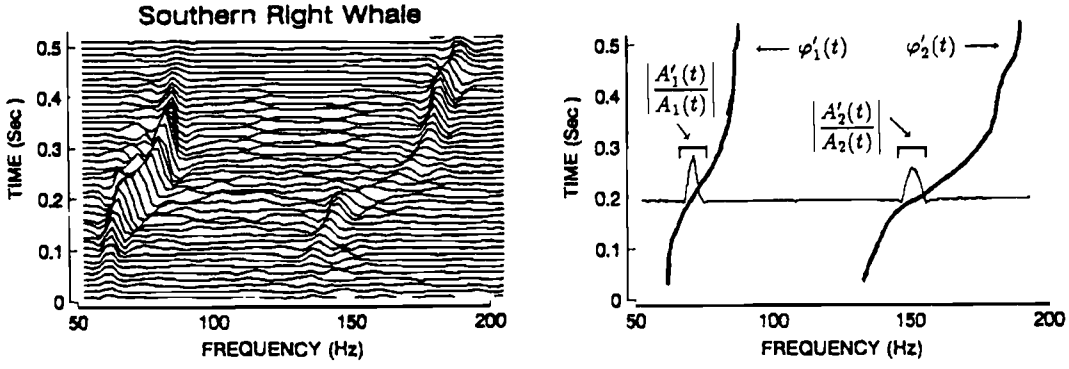

Fig. 13.2 A multicomponent signal is characterized by the narrowness of both parts in comparison to their separation.

Global or Local. There is a tendency to say that a signal is or is not multicomponent. The preceding discussion shows that the condition is a local one; it applies for one given time. Of course, it is possible that it will be the case for all time, but that does not have to be so. Hence, a signal may be multicomponent at some times and monocomponent at others.

Spectrum of a Multicomponent Signal. Generally, the spectrum gives no indication as to whether a signal is mono or multicomponent. This is reasonable since the spectrum is just the projection of the time-frequency plot on the frequency axis. An example is given in Fig. 13.3 (a) where the spectrum does not give any indication that we have a multicomponent signal. However, there are situations where the spectrum may indicate components and this is illustrated in Fig. 13.3 (b). The reason is the components are limited to mutually exclusive bands for all time.

Multicomponent in Time. The above discussion has used situations where the components have a narrow spread in the frequency direction. Of course, we can have a situation where components are narrow in the time direction or a combina-

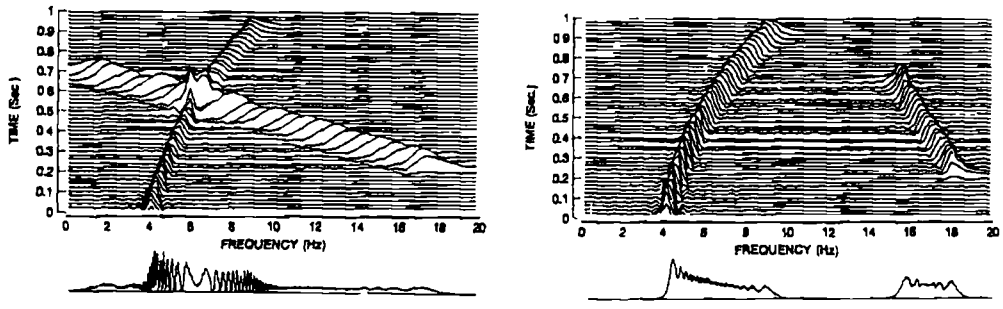

Fig. 13.3 Two examples of multicomponent signals and their energy density spectra. Generally the spectrum cannot be used to determine whether a signal is multicomponent, as in (a). Sometimes it can, as in (b).

tion of the two. The spread along the time dimension of each part must be small relative to the separation. The separation is the difference in the group delay and so the condition for a multicomponent signal in time is

$$\left| \frac{B_1'(\omega)}{B_1(\omega)} \right| \quad , \quad \left| \frac{B_2'(\omega)}{B_2(\omega)} \right| \quad << \quad |\psi_2'(\omega) - \psi_1'(\omega)| \tag{13.22}$$

There are many situations where the components may not be well delineated and this will depend on the amplitudes and phases. As with real mountain ridges there are many hazy situations in which separate ridges cannot be distinguished.

13.4 SPATIAL /SPATIAL-FREQUENCY DISTRIBUTIONS

The intuitive concept of frequency is how often something repeats. Thus far we have emphasized repetition in time, although we can have any other variable and ask for the repetition. For example, an ordinary two dimensional picture is an example of variation in density as a function of position in space. Now fix on a particular direction. The frequency is then the number of ups and downs in the density per unit distance. Therefore, instead of time-frequency we can sensibly speak of position/position-frequency or spatial/spatial-frequency distributions. Everything we have previously done is immediately applicable to this case with the proper transliteration. Signals are functions of space rather than time. Also, we can have two dimensional situations and have frequencies in each direction, the distributions will then be four dimensional. Just as time-frequency analysis gives us the local behavior in frequency, spatial/ spatial-analysis gives us the local behavior of spatial variations. The advantage and use of this type of distribution have been developed by a number of people and excellent discussions regarding them can be found in the articles by Jacobson and Wechsler[273] and in the review article by Cristobal, Gonzalo, and Bescos.[161]

13.5 DELTA FUNCTION DISTRIBUTION FOR FM SIGNALS

We have seen in Section 13.2 that for constant amplitude signals the instantaneous bandwidth is zero. Therefore for signals with constant amplitude one would like the joint time-frequency distribution to be totally concentrated along the instantaneous frequency,

$$P(t,\omega) = \delta(\omega - \varphi'(t)) \qquad \text{for} \quad s(t) = e^{j\varphi(t)} \tag{13.23}$$

Given a distribution we know how to obtain the kernel. It is Eq. (9.19). Applying that formula we have

$$\phi(\theta,\tau) = \frac{\int e^{j\theta t + j\tau\varphi'(t)}\, dt}{\int e^{j\theta t + j[\varphi(t+\tau/2)-\varphi(t-\tau/2)]}\, dt} \tag{13.24}$$

This kernel insures that we get Eq. (13.23) for signals that are purely frequency modulated. Note that it is a functional of the signal.

Example 13.1: Quadratic Phase.

Take $\varphi(t) = \beta t^2/2 + \omega_0 t$; then

$$\varphi(t+\tfrac{1}{2}\tau) - \varphi(t-\tfrac{1}{2}\tau) = \beta t\tau + \omega_0\tau \tag{13.25}$$

and since $\varphi'(t) = \beta t + \omega_0$ we have

$$\phi(\theta,\tau) = \frac{\int e^{j\theta t + j\tau\varphi'(t)}\, dt}{\int e^{j\theta t + j\tau(\beta t+\omega_0)}\, dt} = 1 \tag{13.26}$$

which is the kernel of the Wigner distribution.

Example 13.2: Cubic phase.

We take $\varphi(t) = \frac{1}{3}\gamma t^3$ and use Eq. (13.24) to obtain

$$\phi(\theta,\tau) = \frac{\int e^{j\theta t + j\tau\gamma t^2}\, dt}{\int e^{j\theta t + j\gamma[t^2\tau + \tau^3/12]}\, dt} = e^{-j\gamma\tau^2/12} \tag{13.27}$$

Marginals. The condition for the frequency marginal is $\phi(\theta,0) = 1$. Taking $\tau = 0$ in Eq. (13.24) we see that the time marginal is satisfied. Now consider the condition for the frequency marginals

$$\phi(0,\tau) = \frac{\int e^{j\tau\varphi'(t)}\, dt}{\int e^{j[\varphi(t+\frac{1}{2}\tau)-\varphi(t-\frac{1}{2}\tau)]}\, dt} \neq 1 \tag{13.28}$$

which shows that the marginal cannot be satisfied exactly for this kernel. However, it is approximately satisfied. To see that, expand $\varphi(t+\tau/2)$ and $\varphi(t+\tau/2)$ in a Taylor series

$$\varphi(t+\tfrac{1}{2}\tau) - \varphi(t-\tfrac{1}{2}\tau) \sim \tau\,\varphi'(t) \tag{13.29}$$

and

$$\phi(0,\tau) = \frac{\int e^{j\tau\varphi'(t)}\,dt}{\int e^{j[\varphi(t+\frac{1}{2}\tau)-\varphi(t-\frac{1}{2}\tau)]}\,dt} \sim \frac{\int e^{j\tau\varphi'(t)}\,dt}{\int e^{j\tau\varphi'(t)}\,dt} = 1 \tag{13.30}$$

These results are clearly unsatisfactory and argue that there should be another formulation of the general class that takes this difficulty into account in an exact way. However, that problem is unsolved as of this writing.

Concentration for Bilinear Distributions. In general we can not have a bilinear distribution satisfying the marginals that is a delta function along the instantaneous frequency. The reason is that the delta function is positive and we know that we cannot have positive bilinear distribution satisfying the marginals. The seeming exception of the Wigner distribution for the signal given by Eq. (8.43) is readily explained because for that signal the Wigner distribution can be expressed in a form which is a not bilinear, but of higher order.[280, 141] The fact that the bilinear distributions may go negative presents a problem in the definition of concentration, since for example standard deviations may go negative. Using the square of a distribution as a measure of concentration, Janssen[276] has considered this problem and shown that for distributions characterized by kernels of the form $\phi(\theta,\tau) = e^{ja\theta\tau}$, the Wigner distribution ($a = 0$) is generally the most concentrated. However the question for a general kernel remains open.

13.6 GABOR REPRESENTATION AND TIME-FREQUENCY DISTRIBUTIONS

When a one dimensional signal is expanded in a complete set of functions, the coefficients give an indication of the relative weight of the particular expansion function. For example, in the Fourier case, expanding a signal in the complex exponentials, the coefficients, that is, the Fourier spectrum, give an indication of the intensity or relative dominance of a frequency for the particular signal being expanded.

Gabor[210] conceived of the possibility of expanding a one dimensional signal in terms of two dimensional time-frequency functions. The time-frequency plane is discretized into a lattice where the coordinates are

$$t_i = nT \qquad \omega_i = m\Omega \qquad -\infty \le n, m \le \infty \tag{13.31}$$

and where T and Ω are the time and frequency lattice intervals. Gabor proposed that an arbitrary signal be expanded in the form

$$s(t) = \sum_{n,m} c_{n,m}\, h_{n,m}(t) \qquad h_{n,m}(t) = h(t-mT)\, e^{jn\Omega t} \quad n, m = -\infty, \infty \tag{13.32}$$

where $h(t)$ is a one dimensional function as yet to be specified. If such an expansion can be done the coefficients squared would give an indication of intensity at the time frequency point t_i, ω_i. Gabor suggested that an appropriate function for h is the Gaussian because it is the function that is most compact in the sense of the time-bandwidth product,

$$h(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2} \tag{13.33}$$

On the issue of whether such an expansion is possible in principle, it has been shown to be possible when $T\Omega \leq 1$. However, the expansion coefficients are not unique and for that reason a number of different approaches have been developed for their calculation. We shall not delve into their calculation here because it is not germane to our main point. Suffice it to say that they can be obtained for an arbitrary signal.

Qian and Morris[468] showed that this type of expansion can be effectively used to understand the cross terms in the Wigner distribution and in addition offers an easy way for systematically adding them to the self terms. The method was further developed by and Qian and Chen.[470] The approach can be applied to any of the bilinear distributions. Notice that what appears in all bilinear distributions is the factor $s^*(u - \frac{1}{2}\tau)s(u + \frac{1}{2}\tau)$, which, when expanded using Eq. (13.32), is

$$s^*(u - \tfrac{1}{2}\tau)s(u + \tfrac{1}{2}\tau) = \sum_{n',m'} \sum_{n,m} c^*_{n',m'}\, c_{n,m}\, h^*_{n',m'}(u - \tfrac{1}{2}\tau)\, h_{n,m}(u + \tfrac{1}{2}\tau) \tag{13.34}$$

We find the generalized distribution to be

$$C(t,\omega) = \sum_{n',m'} \sum_{n,m} c^*_{n',m'}\, c_{n,m}\, C_{n',m';n,m}(t,\omega) \tag{13.35}$$

where

$$C_{n',m';n,m} = \frac{1}{4\pi^2} \iiint h^*_{n',m'}(u - \tfrac{1}{2}\tau)\, h_{n,m}(u + \tfrac{1}{2}\tau)\, \phi(\theta,\tau)\, e^{-j\theta t - j\tau\omega + j\theta u}\, du\, d\tau\, d\theta \tag{13.36}$$

Let us now specialize to the Wigner distribution, in which case the $C_{n',m';n,m}$ can be done analytically since the h_{nm}'s are Gaussians. If we break up the summation as

$$W(t,\omega) = \sum_{n,m} |\, c_{n,m}\,|^2 W_{n,m;n,m}(t,\omega) + \sum_{n',m' \neq n,m} c^*_{n',m'}\, c_{n,m}\, W_{n',m';n,m}(t,\omega) \tag{13.37}$$

then the first summation is manifestly positive since the h_{nm} is a Gaussian. If all the terms of the second summation are added, then we get the exact Wigner distribution. An effective way to add the terms of the second summation is to first add nearest neighbor terms, then add second nearest neighbors, and so on. The motive is to systematically add the relative importance of the cross terms.

13.7 EXPANSION IN SPECTROGRAMS

From a calculation point of view, the spectrogram is particularly easy to compute since the only calculation involved is a single Fourier transform of the windowed signal. Furthermore, Fourier transforms are very efficiently calculated by way of the fast Fourier transform technique. We now show that an arbitrary real time-frequency distribution may be expressed as an infinite sum of spectrograms. This decomposition offers an efficient method for the calculation of a time-frequency distribution, because for many cases the time-frequency distribution can be approximated with a finite number of such terms. From a computational point of view there may still be a considerable saving. This method was developed into an effective calculation scheme by Cunningham and Williams.[167]

We write here Eq. (9.12), one of the ways to express the general time-frequency distribution,

$$C(t,\omega) = \frac{1}{2\pi}\iint r(t-\tfrac{1}{2}(x+x'),x-x')\,e^{-j\omega(x-x')}\,s^*(x')\,s(x)\,dx\,dx' \qquad (13.38)$$

Let us momentarily assume that we can express r in the following form,

$$r(t-\tfrac{1}{2}(x+x'),x-x') = \sum_{n=1}^{\infty}\frac{1}{\lambda_n}\,u_n(x-t)\,u_n^*(x'-t) \qquad (13.39)$$

where λ_n's and u_n are constants and functions as yet to be determined. Substituting this expansion in Eq. (13.38) we have

$$C(t,\omega) = \frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{1}{\lambda_n}\left|\frac{1}{\sqrt{2\pi}}\int s(x)\,u_n(x-t)\,e^{-j\omega x}\,dx\right|^2 \qquad (13.40)$$

which is a sum of spectrograms with windows u_n. All this depends on whether indeed we can decompose r as indicated by Eq. (13.39), which we now consider.

Hermitian Functions. If we have a two dimensional function, $K(z,z')$, that satisfies the Hermitian property

$$K(z,z') = K^*(z',z) \qquad (13.41)$$

then solving the integral equation

$$u(z) = \lambda\int K(z,z')\,u(z')\,dz' \qquad (13.42)$$

results in eigenvalues and eigenfunctions, λ_n's and u_n's, which form a complete set. For this to be possible the kernel must be square integrable. Such functions are called Hilbert-Schmidt kernels. The kernel, K, is then expressible in terms of the eigenfunctions and eigenvalues,

$$K(z,z') = \sum_{n=1}^{\infty}\frac{1}{\lambda_n}\,u_n(z)\,u_n^*(z') \qquad (13.43)$$

Now to specialize to our case. Let $z = x - t$ and $z' = x' - t$ and define

$$K(z, z') = r(-\tfrac{1}{2}(z + z'), z - z') \tag{13.44}$$

As a function of z and z', r is Hermitian if the distribution is real. That is,

$$K(z, z') = K^*(z', z) \qquad \text{if} \qquad r(u, \tau) = r^*(u, -\tau) \tag{13.45}$$

Hence the solution of

$$u(z) = \lambda \int r(-\tfrac{1}{2}(z + z'), z - z')\, u(z')\, dz \tag{13.46}$$

will result in eigenvalues and eigenfunctions so that r may be expanded as indicated by Eq. (13.39).

In practice, we usually have a signal in a discretized form and the eigenvalue problem is formulated in terms of a matrix equation for which there are standard routines for solving for the eigenvalues and eigenfunctions. This has to be done only once. Once the eigenvalues and eigenvectors have been obtained for a specific kernel, they may be used for any signal. The windows are not the usual windows used in calculating spectrograms, but that is of no consequence since the main motive is for the purpose of numerical computation.

Complex Distributions. If r does not satisfy the Hermitian property then it is possible to expand it in the following form,

$$r = \sum_{n=1}^{\infty} \frac{1}{\sigma_n}\, u_n(x - t)\, v_n^*(x' - t) \tag{13.47}$$

where now the u_n's and v_n's are different complete sets. Such a decomposition is achieved by a method known as singular value decomposition, which is discussed in Section 13.9. Substituting into Eq. (13.38) we have

$$C(t, \omega) = \frac{1}{2\pi} \sum_{n=1}^{\infty} \frac{1}{\sigma_n}$$

$$\left(\frac{1}{\sqrt{2\pi}} \int s(x)\, u_n(x - t)\, e^{-j\omega x}\, dx \right) \left(\frac{1}{\sqrt{2\pi}} \int s(x)\, v_n(x - t)\, e^{-j\omega x}\, dx \right)^* \tag{13.48}$$

which is a sum consisting of products of short-time Fourier transforms. This approach was developed by O'Hair and Suter.[411]

13.8 SPECTROGRAM IN TERMS OF OTHER DISTRIBUTIONS

The concept of expanding a spectrogram in terms of other distributions was considered by White[581] and Amin[19] who made a singular value decomposition of the kernel. We restrict ourselves to real distributions, which insures the kernel is Hermitian. White showed how the spectrogram can be expanded in terms of modified

Wigner distributions. We shall consider the general case where we expand any distribution in terms of modified distributions. Recall from Section 9.7 that any two distributions are related by

$$C_1(t,\omega) = \iint g_{12}(t'-t,\omega'-\omega)\, C_2(t',\omega')\, dt'\, d\omega' \tag{13.49}$$

with

$$g_{12}(t,\omega) = \frac{1}{4\pi^2} \iint \frac{\phi_1(\theta,\tau)}{\phi_2(\theta,\tau)}\, e^{j\theta t + j\tau\omega}\, d\theta\, d\tau \tag{13.50}$$

If the distributions are real, then $g_{12}(t,\omega)$ is Hermitian and

$$g_{12}(t,\omega) = \sum_{n=1}^{\infty} \frac{1}{\epsilon_n}\, \eta_n(t)\, \eta_n^*(\omega) \tag{13.51}$$

where the ϵ_n's and η_n's are obtained by solving the integral equation,

$$\eta_n(t) = \epsilon_n \int g_{12}(t,\omega)\, \eta_n(\omega)\, d\omega \tag{13.52}$$

Substituting this decomposition into Eq. (13.49) we have

$$C_1(t,\omega) = \sum_{n=1}^{\infty} \frac{1}{\epsilon_n}\, C_2^{(n)}(t,\omega) \tag{13.53}$$

where

$$C_2^{(n)}(t,\omega) = \iint \eta_n(t'-t)\, \eta_n(\omega'-\omega)\, C_2(t',\omega')\, dt'\, d\omega' \tag{13.54}$$

This shows that an arbitrary distribution can be expanded as a sum of modified distributions, $C_2^{(n)}(t,\omega)$. The modification indicated by Eq. (13.54) is that each term in the sum is the distribution smoothed independently in the time and frequency directions.

13.9 SINGULAR VALUE DECOMPOSITION OF DISTRIBUTIONS

If we have a function of two variables, $C(t,\omega)$, it is possible to expand it in terms of a sum of product functions,

$$C(t,\omega) = \sum_{n=1}^{\infty} \frac{1}{\sigma_n}\, u_n(t)\, v_n^*(\omega) \tag{13.55}$$

where the u_n's and $v'n$ are complete sets obtained by solving the coupled integral equations

$$u_n(t) = \sigma_n \int C(t,\omega)\, v_n(\omega)\, d\omega \quad ; \quad v_n(\omega) = \sigma_n \int C^*(\omega,t)\, u_n(t)\, dt \tag{13.56}$$

The solution of these equations results in common eigenvalues but two complete sets of functions, the u's and the v's. This decomposition of a two dimensional function is called a singular value decomposition. Note that for the case $C(t,\omega) = C^*(\omega,t)$ we have Hermiticity, in which case we can use the standard approach discussed in Section 13.9. Since in our case $C(t,\omega)$ will be a joint time-frequency distribution, it is in general not Hermitian.

In the general case we need in principle an infinite number of terms to represent the distribution exactly. However, in many situations the first four or five terms are sufficient to achieve a very high accuracy. The concept of decomposing a distribution in this manner was first done by Marinovich and Eichmann[363] for the Wigner distribution.

There are two basic reasons for seeking such a decomposition. First, suppose the signal is in a noisy environment. If we decompose the distribution and keep only the first few terms we will reduce the noise significantly because the signal is well represented by the first few terms but the noise is typically spread out over all the terms. Therefore, by truncating the series after four or five terms we retain most of the signal but lose most of the noise.

The second reason is for the purpose of classification. The basic idea is that the σ's contain unique characterizations of the time-frequency structure of a distribution and may be used for classification. Suppose, for example, we have ten signals and do a singular value decomposition of each one. This will result in ten sets of singular values and eigenfunctions, all different from each other. Suppose we keep just the first five terms. For the eigenvalues we will have ten sets of five numbers. Now suppose we have an unknown signal, one of the ten but in a noisy environment, and want to classify it. We do a singular value decomposition and compare the first five singular values to our sets. Since we are comparing numbers the comparison is easy and fast. Any reasonable measure of comparison can be used, such as the Euclidean distance. The closest distance is used to classify the signal.

13.10 SYNTHESIS

Suppose we want to design a signal that has a time-frequency structure of our choosing. A way to accomplish this is to form the time-frequency distribution and then calculate the signal that generates it. If the time-frequency distribution we have constructed is a legitimate one, that is, a representable one, then the simple inversion formula, Eq. (9.56), yields the signal. However, in general, the time-frequency distribution we construct will not be a representable one. For example, suppose we want a Wigner distribution to have certain time frequency structure. In general we will not be able to construct a representable Wigner distribution because we will not be able to properly construct the cross terms. We are simply not that capable. What is done is to seek a signal that reproduces the given time-frequency distribution as close as possible, for example, in the least squares sense. Obtaining a signal from a distribution is the synthesis problem. It was first considered by Boudreaux-Bartels and Parks[93, 95] for the Wigner distribution. For the

Choi-Williams distribution, where the cross terms are less of a problem, Jeong and Williams have devised an effective synthesis scheme.[287]

Another important application of the synthesis problem arises in the following circumstance. Suppose we have a time-frequency distribution of some signal that is multicomponent, and we want the signal that produces only one of the components. What we can do is to erase everything in the time-frequency plane but the component of interest. Having erased parts of a legitimate distribution, the resulting distribution will not usually be a representable one because, for example, we may have erased cross terms which are needed to have a representable distribution. Nonetheless, we want the signal that generates that component. The synthesis problem will again have to be done in terms of the best signal that generates a distribution close to the one we have at hand.

A further reason for the synthesis problem that is if we have a signal with noise and calculate the time-frequency distribution, the noise will typically be spread throughout the time-frequency plane. If we literally erase the noise but keep the signal, the result will be the signal and noise only around the signal in time and frequency. Finding the signal this produces this curtailed distribution will, hopefully, result in a signal with less noise than the original.

13.11 RANDOM SIGNALS

The main consideration in this book has been deterministic signals. To apply the ideas developed to random signals, one ensemble averages. Ensemble averaging means averaging over all possible realizations of the signal, each one having a certain probability of occurrence. For deterministic signals we have used averaging for many purposes, such as in the definition of global and local averages. These two uses of averages must be differentiated. We shall use an overline to denote ensemble averaging. Very often both averages are taken. For example, if we have a random signal then $\overline{\langle t \rangle}$ means that we are calculating the average time for a particular signal of the ensemble and then averaging over all possible signals. Sometimes the order of averaging matters and sometimes not, but we will not address this issue here.

We define a random time-frequency distribution, $\overline{C}(t,\omega)$, by ensemble averaging the deterministic joint distribution over all possible signals. Using the general time-frequency distribution, Eq. (9.1), we have

$$\overline{C}(t,\omega) = \frac{1}{4\pi^2} \iiint \overline{s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, \phi(\theta,\tau)}\, e^{-j\theta t - j\tau\omega + j\theta u}\, du\, d\tau\, d\theta \quad (13.57)$$

If we assume that the kernel is independent of the signal, then

$$\overline{C}(t,\omega) = \frac{1}{4\pi^2} \iiint \overline{s^*(u-\tfrac{1}{2}\tau)}\,\overline{s(u+\tfrac{1}{2}\tau)}\, \phi(\theta,\tau)\, e^{-j\theta t - j\tau\omega + j\theta u}\, du\, d\tau\, d\theta \quad (13.58)$$

$$= \frac{1}{4\pi^2} \iiint R(u-\tfrac{1}{2}\tau, u+\tfrac{1}{2}\tau)\, \phi(\theta,\tau)\, e^{-j\theta t - j\tau\omega + j\theta u}\, du\, d\tau\, d\theta \quad (13.59)$$

where $R(t,t')$ is the autocorrelation function of the random process

$$R(t,t') = \overline{s^*(t)\, s(t')} \tag{13.60}$$

If we want to work with the form given by Eq. (9.9), then

$$\overline{C}(t,\omega) = \frac{1}{2\pi}\iint r(t-u,\tau)R(u-\tfrac{1}{2}\tau,u+\tfrac{1}{2}\tau)\, e^{-j\omega\tau}\, du\, d\tau \tag{13.61}$$

If in addition we assume that the random process is stationary, the autocorrelation function is then a function of the difference of the times,

$$R(t,t') = R(t-t') \tag{13.62}$$

in which case we have

$$\overline{C}(t,\omega) = R(u-\tfrac{1}{2}\tau,u+\tfrac{1}{2}\tau) = R(\tau) \tag{13.63}$$

$$\overline{C}(t,\omega) = \frac{1}{2\pi}\iint r(t-u,\tau)R(\tau)\, e^{-j\omega\tau}\, du\, d\tau \tag{13.64}$$

For the Wigner distribution, Martin[376] and Martin and Flandrin[378] have developed the main ideas and Martin has coined the phrase "Wigner-Ville *spectrum*" to indicate that we have ensemble averaged the Wigner distribution. White[578] developed a comprehensive theory for the general case and devised specific methods for obtaining the important parameters of a random process and the errors involved in estimating the parameters. Amin's[20] early work on spectrum estimators is closely related to the concept of kernal choice for the random case. Considerable work has been done in this area recently by Amin,[26] Chaparro, El-Jaroudi and Kayhan,[110] Riedel and Sidorenko,[481] and Sayeed and Jones[501] who have addressed the fundamental problems and specific applications.

Kernel Choice. Most of the ideas regarding the choice of kernel carry over to the random case. For example, if we have a wide-sense stationary process, we expect the marginal in frequency to be white noise. The condition for that, as Posch[455] has shown, is that $\phi(\theta,0) = 1$, which is the condition to obtain the frequency marginal for the deterministic case. The issue of which is the best kernel to use for estimating the properties of the random process has been formulated by Amin,[26] who has shown that the random case necessitates unique constraints on the kernel. He has developed a comprehensive theory of kernel choice.

13.12 NUMERICAL COMPUTATION

The calculation of time-frequency distributions is fairly straightforward. However, as is usual with numerical methods there are many tricks and procedures that have been devised and that are best obtained from the original sources. We just give a broad outline of the basic ideas and discuss some of the fundamental issues that

have arisen. We assume that we have discrete samples and that the signal is sufficiently bandlimited so that the Nyquist sampling theorem applies: For a bandlimited signal, the signal can be reconstructed from discrete sampled values if the sampling is done at a sampling rate $\omega_s \geq 2\omega_{\max}$ where $\omega_{\max}$ is the highest frequency in the signal.

All distributions have at least one Fourier transform to be done. Historically there have been a number of methods proposed to calculate integrals involving sines and cosines, for example the method of Fillon. However, the advent of the Fast Fourier Transform has over shadowed these methods because indeed it is fast.

The simplest distribution is the spectrogram, Eq. (7.5) One chooses a specific time, t, calculates the modified signal, $s(\tau)\,h(\tau - t)$, as a function of τ, then takes the Fourier transform with respect to τ. This is repeated for each time desired. For the Wigner distribution $s^*(t - \frac{1}{2}\tau)\,s(t + \frac{1}{2}\tau)$ is calculated as a function of τ, for a fixed time t and the Fourier transform taken. That gives us the Wigner distribution of frequencies at time t. The procedure is repeated for any other time. Note that it appears because of the factor of one half we must have signal values at points in between the sampled values. If we do have them then of course there is no problem. But that requires that we over sample by a factor of two. If we do not, one can, for example, interpolate to get them. Because of this it has been believed that to construct the Wigner distribution from discrete samples one must sample the signal at twice the Nyquist rate; otherwise aliasing will occur. That can not be the case since in principle having a sampled signal at the Nyquist rate allows us to construct the signal for any time. Having constructed the signal for an arbitrary time one can then calculate the continuous Wigner distribution. Therefore we see that in principle we should not have to over sample. For the Wigner distribution, Poletti[442] and Nuttall[405] have shown that the Wigner distribution can be computed from a signal sampled at the Nyquist rate without interpolation of the sampled values or reconstitution of the continuous signal.

The same considerations apply to the calculation of any of other bilinear time-frequency distributions. The basic issue is how to define the discrete version of bilinear time-frequency distributions so that at the discrete time frequency points they have the same value as the continuous version, and where the calculation can be performed directly from the discrete version of the signal, sampled at the Nyquist rate. In addition, the discrete version should satisfy the same general properties of the continuous version, that is, the marginals and so forth. This is a fundamental problem that has recently been solved by Morris and Wu,[387] O'Hair and B. W. Suter,[410], Jeong and Williams,[286] and Cunningham and Williams.[167]

For the general bilinear case three Fourier transforms have to be performed but for many kernels one of the integrations can be done analytically resulting in the form given by Eq. (9.9). This is the case, for example, with the Choi-Williams distribution. For such distributions we have two integrals, only one of which is a Fourier transform. To calculate such a distribution the inner integral, that is, Eq. (9.10) is effectively done by the rectangular rule. Subsequently the Fourier transform is taken.

We also mention that decomposition of a time-frequency distribution as a sum of weighted spectrograms is a very effective calculational method. This is described

in Section 13.7. Also the computational approaches for calculating the positive distributions are described in Section 14.2 and the calculation method for obtaining optimum kernels is described in Section 11.5.

13.13 SIGNAL ANALYSIS AND QUANTUM MECHANICS

Historically, the origins of the mathematics and many of the ideas used in time-frequency analysis originated and were guided by corresponding developments in quantum mechanics. In fact, the original papers of Gabor and Ville continuously evoked the quantum analogy. This parallel development will undoubtedly continue because there is a very strong mathematical similarity between quantum mechanics and signal analysis, and results can be mathematically transposed. However, the transposition cannot be carried over, necessarily, to the interpretation of the ideas, because the physical interpretations are drastically different. The most fundamental difference between the two subjects is that quantum mechanics is an inherently probabilistic theory, while signal analysis is deterministic.[1] We emphasize that the fundamental idea of modern physics is that we can only predict probabilities for observables such as position and velocity and that this is not a reflection of human ignorance but rather the way nature is. The probabilities are predicted by solving Schrödinger's equation of motion.

The reason for the mathematical similarity is that in quantum mechanics the fundamental quantity is the wave function. In the position representation the wave function is $\psi(q)$, where q is position. In the momentum representation it is $\phi(p)$, where p is the momentum. The main point is that the two wave functions are Fourier transforms of each other:

$$\phi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int \psi(q)\, e^{-jqp/\hbar}\, dq \qquad ; \qquad \psi(q) = \frac{1}{\sqrt{2\pi\hbar}} \int \phi(p)\, e^{jqp/\hbar}\, dp \tag{13.65}$$

where $\hbar$ is Planck's constant divided by 2π. Moreover, the probability distribution, $P(q)$, for finding the particle at a certain position is the absolute square of the wave function and the probability of finding a particle with a certain momentum, $P(p)$, is the absolute square of the momentum wave function,

$$P(q) = |\,\psi(q)\,|^2 \qquad ; \qquad P(q) = |\,\phi(p)\,|^2 \tag{13.66}$$

Therefore, mathematically and mathematically only, we can associate the signal with the wave function, time with position, and frequency with momentum. The marginal conditions are formally the same, although the variables are different and

[1] It should be kept in mind that in quantum mechanics and signal analysis there is another layer that is probabilistic. In the case of quantum mechanics this comes about if we do not know the possible wave functions and assign probabilities for obtaining them. Hence in calculating averages for such a situation we have a double probability average, one due to the inherent probabilistic distribution, where the absolute square of the wave function is the probability, and the other an ensemble average over the possible wave functions. That aspect of the subject is called quantum statistical mechanics. In signal analysis we start with a deterministic theory and if the possible signals are probabilistically determined (random signal) then we have to ensemble average over the possible signals.

the interpretation is certainly different. In quantum mechanics the marginals are probability densities; in signal analysis they are deterministic intensities. In Table 13.1 we list some of the physical quantities in each field and the correspondence between them. We also point out that the operator method we describe in this book is fundamental in quantum mechanics.

We now come to the issue of interpretation and address some of the fundamental distinctions. One must be particularly cautious in transposing ideas because blind transpositions can lead to preposterous results. In quantum mechanics, physical quantities are represented by operators and it is the fundamental tenet that what can be measured for an observable are the eigenvalues of the operator. (That is the basis for the quantization of physical quantities.) From a classical point of view this produces bizarre results, which are nonetheless true and have been verified experimentally.

For a dramatic example of the difference, consider the sum of two continuous quantities. In quantum mechanics each quantity can be continuous and yet the sum is not necessarily continuous! Specifically, consider the position, q, and momentum, p, which are continuous in quantum mechanics because the eigenvalues of the position and momentum operators are continuous. Now consider the physical quantity made up of position and momentum, $q^2 + p^2$ (appropriately dimensioned). It is never continuous under any circumstances for any particle. It is always quantized, that is, it can have only certain values. The reason is that the eigenvalues of the operator, $q^2 + p^2$, are discrete. If we were to make a blind analogy, the corresponding statement in signal analysis would be that time and frequency are continuous, but that $t^2 + \omega^2$ (appropriately dimensioned) is never so and is always discrete. That would be a ludicrous statement to make in signal analysis.

We now address the issue of the uncertainty principle. We have already discussed some aspects of its interpretation in quantum mechanics and signal analysis in Chapter 3. The term uncertainty was coined in quantum mechanics, where it properly connotes the fact that quantum mechanics is an inherently probabilistic theory. In quantum mechanics the standard deviations involve the measurement of physical observables and are probabilistic statements. However, in nonprobabilistic contexts the uncertainty principle should be thought of as expressing the fact that a function and its Fourier transform cannot both be made arbitrarily narrow. It has nothing to do with "uncertainty" as used in quantum mechanics.

There is another important difference. In Chapter 2 we discussed that observable signals are real and that the energy density spectrum of real signals does not properly indicate the physical situation. For example, for real signals the average frequency is zero, which is not a reflection of the physical situation. This led to the desire to define a complex signal. In quantum mechanics wave functions are inherently complex, although they may be real. If a wave function is real, then the average momentum is zero. That is perfectly all right and does reflect the physical situation. It means that we have equal probability for the particle traveling to the right and the particle traveling to the left. Therefore there is no need to introduce the equivalent of an analytic signal. Momentum, unlike frequency, can be negative or positive.

Table 13.1. The relationship between quantum mechanics and signal analysis. The formal mathematical correspondence is (position, momentum) $\leftrightarrow$ (time, frequency). The wave function in quantum mechanics depends on time but this has no formal correspondence in signal analysis.

Quantum Mechanics (inherently probabilistic)		Signal Analysis (deterministic)	
Position:	q (random)	Time:	t
Momentum:	p (random)	Frequency:	ω
Time:	t	No correspondence	
Wave function:	$\psi(q,t)$	Signal:	$s(t)$
Momentum wave function:	$\phi(p,t) = \frac{1}{\sqrt{2\pi\hbar}} \int \psi(q,t)\, e^{-jqp/\hbar}\, dq$	Spectrum:	$S(\omega) = \frac{1}{\sqrt{2\pi}} \int s(t)\, e^{-j\omega t}\, dt$
Probability of position at time: t	$\lvert\psi(q,t)\rvert^2$	Energy density:	$\lvert s(t)\rvert^2$
Probability of momentum:	$\lvert\phi(p,t)\rvert^2$	Energy density spectrum:	$\lvert S(\omega)\rvert^2$
Expected value of position:	$\langle q \rangle = \int q\, \lvert\psi(q,t)\rvert^2\, dq$	Mean time:	$\langle t \rangle = \int t\, \lvert s(t)\rvert^2\, dt$
Expected value of momentum:	$\langle p \rangle = \int p\, \lvert\phi(p,t)\rvert^2\, dp$	Mean frequency:	$\langle \omega \rangle = \int \omega\, \lvert S(\omega)\rvert^2\, d\omega$
Standard deviation of position:	$\sigma_q = \sqrt{\langle q^2 \rangle - \langle q \rangle^2}$	Duration :	$T = \sqrt{\langle t^2 \rangle - \langle t \rangle^2}$
Standard deviation of momentum:	$\sigma_p = \sqrt{\langle p^2 \rangle - \langle p \rangle^2}$	Bandwidth :	$B = \sqrt{\langle \omega^2 \rangle - \langle \omega \rangle^2}$
Uncertainty principle:	$\sigma_q\, \sigma_p \geq \frac{1}{2}\hbar$	Time bandwidth relation:	$BT \geq \frac{1}{2}$
Current:	$\frac{d}{dq}$ phase of $\psi(q)$	Instantaneous frequency:	$\frac{d}{dt}$ phase of $s(t)$

Chapter 14

Positive Distributions Satisfying the Marginals

14.1 INTRODUCTION

Wigner showed that manifestly positive *bilinear* distributions satisfying the time and frequency marginals do not exist.[584, 585] The spectrogram, for example, is manifestly positive but does not satisfy the marginals while the Wigner distribution satisfies the marginals but is not manifestly positive. Historically there has been a sense expressed in the literature that manifestly positive distributions satisfying the marginals cannot exist. But they do. They are obtained by a simple procedure which ensures that the marginals are satisfied. These distributions are, of course, not bilinear in the signal.

14.2 POSITIVE DISTRIBUTIONS[127, 129]

Take

$$P(t,\omega) = |S(\omega)|^2 \, |s(t)|^2 \, \Omega(u,v) \tag{14.1}$$

where u, v are functions of t and ω

$$u(t) = \int_{-\infty}^{t} |s(t')|^2 \, dt' \qquad v(\omega) = \int_{-\infty}^{\omega} |S(\omega')|^2 \, d\omega' \tag{14.2}$$

and where Ω is any positive function satisfying

$$\int_0^1 \Omega(u,v)\, dv = 1 \quad ; \quad \int_0^1 \Omega(u,v)\, du = 1 \tag{14.3}$$

It is sufficient to define $\Omega(u,v)$ only for $0 \leq u,v \leq 1$. Note that u and v are the cumulative margins, that is, they are the sum of the densities up to given time and frequency values.

Marginals. First note that

$$dv = |S(\omega)|^2 \, d\omega \tag{14.4}$$

Now, integrate with respect to ω

$$\int P(t,\omega)\, d\omega = |s(t)|^2 \int |S(\omega)|^2 \, \Omega(u,v)\, d\omega \tag{14.5}$$

$$= |s(t)|^2 \int_0^1 \Omega(u,v)\, dv \tag{14.6}$$

$$= |s(t)|^2 \tag{14.7}$$

Similarly for the frequency marginal.

Positivity. The distributions given by Eq. (14.1) are positive as long as Ω is positive for the range $0 \leq u,v \leq 1$. Ω may or may not be a functional of the marginals or signal, but to generate all possible positive joint distributions we have to consider it a functional of the signal.[231] Examples of Ω's satisfying the above conditions are easily made up; for example,

$$\Omega(u,v) = 1 - (nu^{n-1} - 1)(mv^{m-1} - 1) \tag{14.8}$$

where n, m are any integers.

Relationship Between Kernels. We have introduced the form given by Eq. (14.1) because it makes the manifestly positive aspect clear. It is possible to obtain this form using the general class developed in Chapter 9. There exists a relationship between the kernels,[128] but generally it is easier to work directly with Eq. (14.1).

The Wigner Distribution for a Chirp. In Section 8.6 we pointed out that for one signal and one signal only, the Wigner distribution was manifestly positive. That case is given by Eq. (8.43). The reason it is positive is the it is not really bilinear. It can be obtained[280] from Eq. (14.1).

Weak and Strong Finite Support. For the distributions given by Eq. (14.1), the factor $|S(\omega)|^2 |s(t)|^2$ appears. Therefore, if the signal is zero at a certain time or the spectrum is zero at a frequency, the distribution will likewise be zero, assuming that Ω is not infinity at those values. Therefore these positive distributions satisfy the strong finite support property and hence also weak finite support.

Uncertainty Principle. Since these distributions satisfy the marginals, they satisfy the uncertainty principle.

Scale Invariance. We now show, following Loughlin,[344] that these distributions are scale invariant. The distribution of the scaled signal, $s_{sc}(t) = \sqrt{a}\, s(at)$, is

$$P_{sc}(t,\omega) = |S_{sc}(\omega)|^2 |s_{sc}(t)|^2 \Omega(u_{sc}(t), v_{sc}(\omega)) \tag{14.9}$$

$$= |S(\omega/a)|^2 |s(at)|^2 \Omega(u_{sc}(t), v_{sc}(\omega)) \tag{14.10}$$

But

$$u_{sc}(t) = \int_{-\infty}^{t} |s_{sc}(t')|^2 dt' = a \int_{-\infty}^{t} |s(at')|^2 dt' = \int_{-\infty}^{at} |s(t')|^2 dt' = u(at) \tag{14.11}$$

and similarly $v_{sc}(\omega) = v(\omega/a)$. Therefore we have that

$$P_{sc}(t,\omega) = |S(\omega/a)|^2 |s(at)|^2 \Omega(u(at), v(\omega/a)) \tag{14.12}$$

$$= P(at, \omega/a) \tag{14.13}$$

Time and Frequency Shifts. For a signal that is shifted by a time t_0 and frequency ω_0 the distribution is

$$P_{sh}(t,\omega) = |S_{sh}(\omega)|^2 |s_{sh}(t)|^2 \Omega(u_{sh}(t), v_{sh}(\omega)) \tag{14.14}$$

$$= |S(\omega - \omega_0)|^2 |s(t - t_0)|^2 \Omega(u_{sh}(t), v_{sh}(\omega)) \tag{14.15}$$

But

$$u_{sh}(t) = \int_{-\infty}^{t} |s_{sh}(t')|^2 dt' = \int_{-\infty}^{t} |s(t' - t_0)|^2 dt' = \int_{-\infty}^{t-t_0} |s(t')|^2 dt' = u(t - t_0) \tag{14.16}$$

and also $v_{sh}(\omega) = v(\omega - \omega_0)$, which shows that

$$P_{sh}(t,\omega) = P(t - t_0, \omega - \omega_0) \tag{14.17}$$

Conditional Distributions. The conditional distributions are

$$P(t|\omega) = \frac{P(t,\omega)}{|S(\omega)|^2} = |s(t)|^2 \Omega(u,v) \tag{14.18}$$

$$P(\omega|t) = \frac{P(t,\omega)}{|s(t)|^2} = |S(\omega)|^2 \Omega(u,v) \tag{14.19}$$

Consequences of Positivity. Our experience in everyday life is, of course, with positive densities. However, our experience in time-frequency analysis has dealt with either positive distributions that do not satisfy the marginals or non-manifestly positive ones that do satisfy the marginals. We now examine the nature of time-frequency distributions that are both manifestly positive and satisfy the marginals.

First let us note that if the marginal is zero at a certain point in time then the joint distribution must be zero at that time for all values of frequency; the strong finite support property is satisfied. This is so because to get the time marginal we add nonnegative pieces, which cannot add up to zero unless the pieces are zero themselves. This is no different than saying that if there are no people who are six feet tall, then there are no people at that height for any weight. Similarly, if the spectrum is zero for a certain value of frequency, the joint distribution will be zero for all times at that frequency.

Now consider the issue of instantaneous frequency and positive distributions. Can we obtain the derivative of the phase as the conditional average,

$$\langle \omega \rangle_t = \frac{1}{|\, s(t) \,|^2} \int_0^\infty \omega \, P(t,\omega) \, d\omega \stackrel{?}{=} \varphi'(t) \tag{14.20}$$

The answer is no in general, but yes sometimes. The general answer of no comes about by the following argument of Claasen and Mecklenbräuker.[120] Suppose we have an analytic signal and therefore the spectrum is zero for negative frequencies. By the arguments given above the joint distribution will also be zero for negative frequencies. If the distribution is positive the conditional average will be positive for an analytic signal. This is a satisfying result. However, we know from Section 2.7 that there are analytic signals that produce an instantaneous frequency, defined as the derivative of the phase, which may go negative. Hence, for those signals, Eq. (14.20) cannot be satisfied. But those are precisely the signals for which the derivative of the phase, as a representation of the concept of instantaneous frequency, apparently makes no sense. Therefore, a meaningful question is whether, for cases where the derivative of the phase meets our expectations of instantaneous frequency, positive distributions can satisfy Eq. (14.20). No general results regarding this question are known.

14.3 THE METHOD OF LOUGHLIN, PITTON, AND ATLAS

Construction of positive distributions has been achieved by Loughlin, Pitton, and Atlas.[343] They have formulated the problem in the following way. With our present knowledge we cannot fix the joint distribution. There are too many Ω's. The problem of finding a function for conditions that do not fix the function is a long standing one in many fields. One approach is the method of maximum entropy. The idea is that we find all the functions that satisfy the conditions and then choose among them the one that maximizes the entropy, the reason being that the maximum entropy solution is the one that is unbiased. In practice this is achieved by taking a

guess at the joint distribution, P_0, defining the cross entropy by

$$\Delta = -\iint P(t,\omega)\log\frac{P(t,\omega)}{P_0(t,\omega)}\,dt\,d\omega \tag{14.21}$$

and maximizing this expression with the constraints of the marginals, positivity and other possible constraints one may want to impose. In Figs. 14.1 – 14.3 we give a number of examples. Fonollosa and Nikias[202] have added the additional constraint that the distribution along a given axis be a specified function. In particular, if we have a two dimensional distribution and want to obtain the density of the variable $u = at + b\omega$ where a, b are the direction cosines specifying the axis, then the density of u is given by Eq. (4.97),

$$P(u) = \iint \delta(u-(at+b\omega))\,P(t,\omega)\,dt\,d\omega \tag{14.22}$$

$$= \int P(t,(u-at)/b)\,dt \tag{14.23}$$

In our case we do not know $P(t,\omega)$. Fonollosa and Nikias used the Wigner distribution to obtain $P(u)$ and impose the constraint for a finite number of axis.

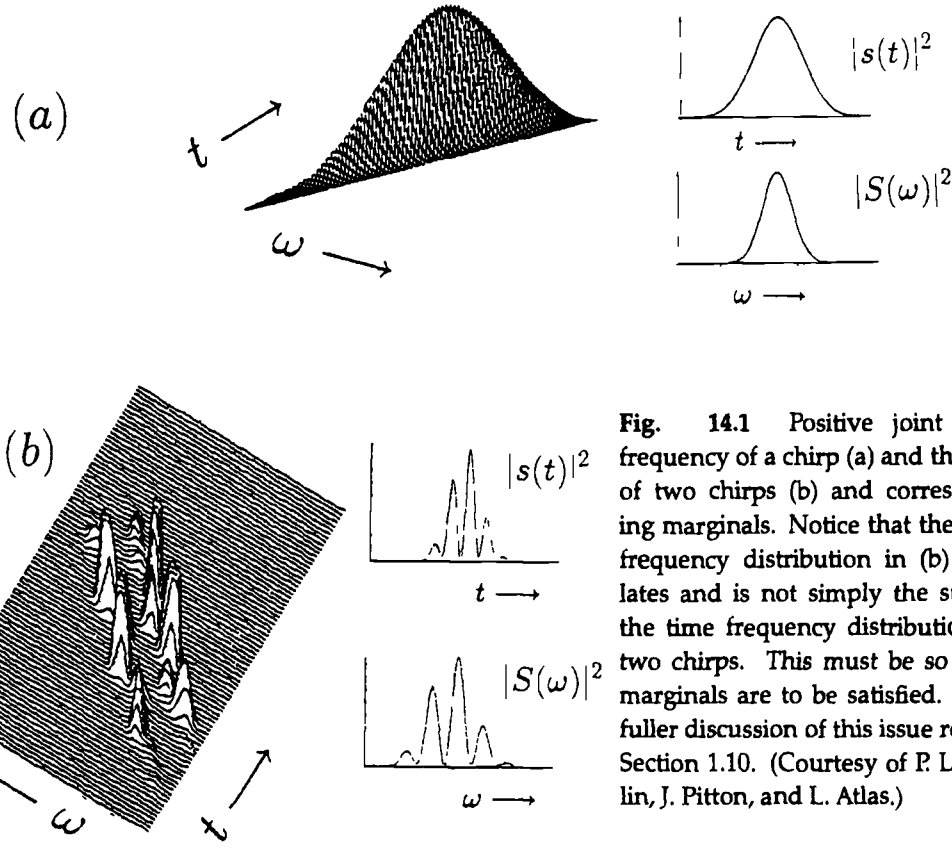

Fig. 14.1 Positive joint time-frequency of a chirp (a) and the sum of two chirps (b) and corresponding marginals. Notice that the time-frequency distribution in (b) oscillates and is not simply the sum of the time frequency distributions of two chirps. This must be so if the marginals are to be satisfied. For a fuller discussion of this issue refer to Section 1.10. (Courtesy of P. Loughlin, J. Pitton, and L. Atlas.)

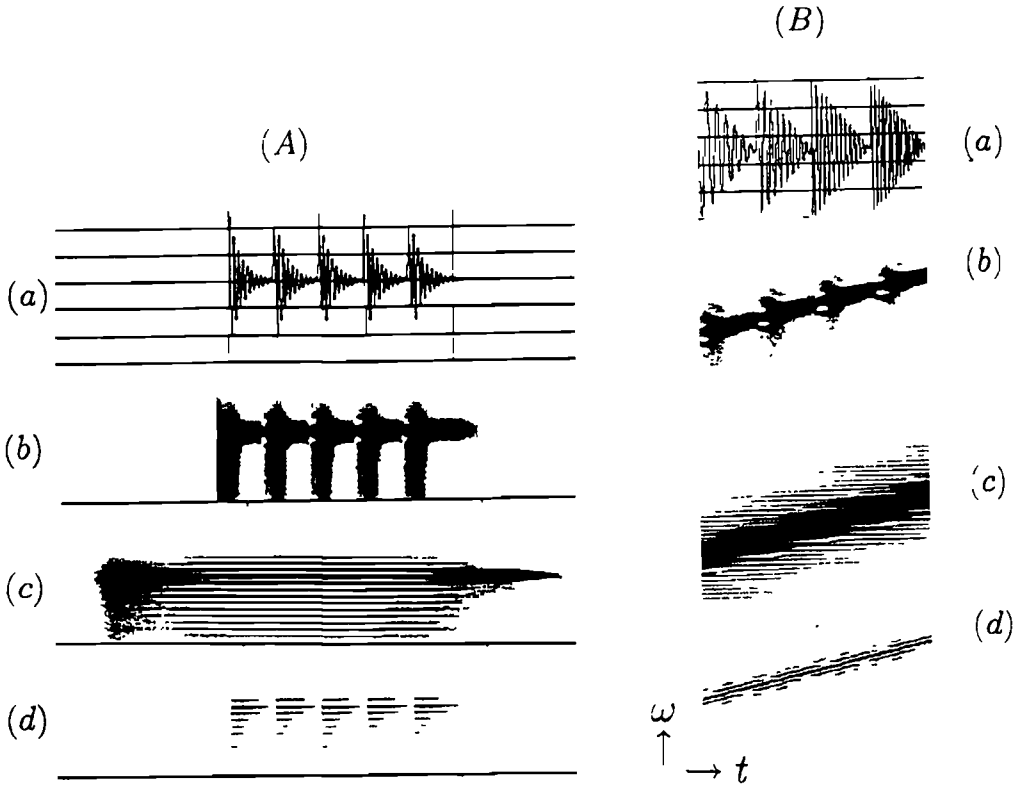

Fig. 14.2 For (A), the signal is a sequence of decaying exponentials, shown in (a). The frequency is constant. In (b) and (c) are the wide band and narrow band spectrograms. The manifestly positive distribution that satisfies the marginals is shown in (d). (B) is the same as (A) except that the frequency is continually increasing linearly. (Courtesy of P. Loughlin, J. Pitton and L. Atlas.)

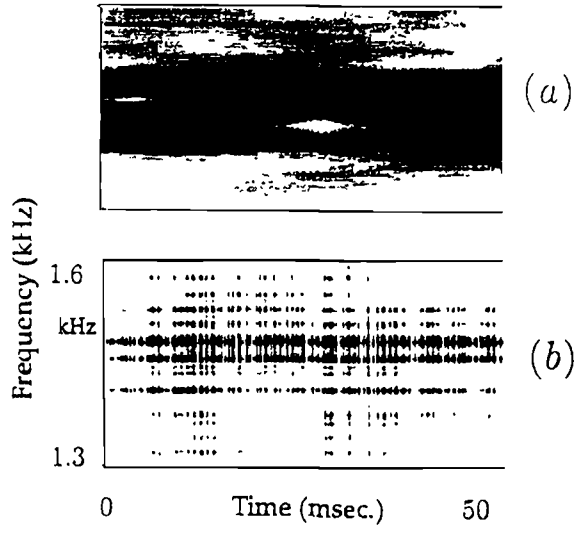

Fig. 14.3 The signal is the sound of a grinding machine. The spectrogram is (a) and the positive distribution is (b). The spectrograms gives no indication of the fine detail. The positive time-frequency distribution resolves the lines and moreover shows that some of the lines have periodic amplitude modulation. (Courtesy of P. Loughlin, J. Pitton, L. Atlas and G. Bernard. Data courtesy of Boeing Commerical Airplane Group.)

Chapter 15

The Representation of Signals

15.1 INTRODUCTION

In the previous chapters we developed the theory of joint representations for time and frequency. We now extend the methods to other physical variables. For a physical quantity, we want to obtain its density, average value, spread, and other results as was done for the case of frequency. This is accomplished by expanding the signal in the representation of the physical quantity in the same sense that the Fourier representation is the appropriate representation for the physical quantity "frequency". The basic idea of representing a signal in another representation is to write the signal as a linear combination of other functions, the expansion functions. They are obtained by solving the eigenvalue problem for an operator that represents, or is associated with, the physical quantity of interest. For example, the complex sinusoids are the eigenfunctions of the frequency operator, and the frequency operator is associated with or corresponds to the physical quantity "frequency". We will use "a" to signify the physical quantity we are studying and script capital $\mathcal{A}$ for the associated operator.

15.2 ORTHOGONAL EXPANSION OF SIGNALS

A signal is expanded in the form

$$s(t) = \int F(a)\, u(a,t)\, da \tag{15.1}$$

where $u(a,t)$ are the basis functions and $F(a)$ are the "expansion coefficients" or the "transform of the signal". As we will prove momentarily, $F(a)$ is given by

$$F(a) = \int s(t)\, u^*(a,t)\, dt \tag{15.2}$$

The basis functions, $u(a,t)$ are always functions of two variables, in this case time and a. The a's are the numerical values of the physical quantity and they may be continuous, discrete or both. Furthermore, the range of a's may be infinite or limited. The integration in Eq. (15.2) implies a specific region of integration. For example, for frequency the range is all possible values, while for scale we will see that the range is from zero to infinity. In this section we assume that the variable a is continuous and at the end of this section we address the discrete case.

The function $F(a)$ gives us an indication of how important a particular value of a is for the signal at hand. If $F(a)$ is relatively large only in a particular region we can then say that the signal is concentrated at those values of a.

The Expansion Functions, Operators, and the Eigenvalue Problem. Where do we get the expansion functions, the u's, that presumably are the natural functions for the physical quantity we are interested in? Also, how do we determine the possible numerical values, the a's, that the physical quantity may attain? Both are obtained by solving the eigenvalue problem for the operator that corresponds to the physical quantity. Operators are commands that change functions. For example, multiplication, squaring, differentiation, integration, and combinations of these are operators.

Generally, the eigenvalue problem is written as

$$\mathcal{A}u(a,t) = a\,u(a,t) \tag{15.3}$$

In the eigenvalue problem the operator is given and one seeks to find those functions, the u's, that when operated upon give back the same function multiplied by a number, in this case a. Generally, there are an infinite number of such functions each paired with an eigenvalue a. The $u(a,t)$'s are called the eigenfunctions and the a's are the eigenvalues. Solution of the eigenvalue problem means solving for both the u's and a's. For example, the operator $d/\,dx$, operating on e^{ax} returns $a\,e^{ax}$, hence e^{ax} is an eigenfunction with eigenvalue a. For this case there are an infinite number of eigenfunctions because we can take any number for a.

Linear Hermitian Operators. A linear operator is one that satisfies

$$\mathcal{A}(f+g) = \mathcal{A}f + \mathcal{A}g \tag{15.4}$$

For example, the operation of differentiation is linear because the derivative of the sum is the sum of the derivatives. However, the operation of squaring is not because the square of the sum is not the sum of the squares.

An operator is Hermitian or self adjoint if for any pair of functions, $f(t)$ and $g(t)$,

$$\int g^*(t)\,\mathcal{A}\,f(t)\,dt = \int f(t)\,\{\mathcal{A}\,g(t)\}^*\,dt \tag{15.5}$$

To prove that an operator is Hermitian one has to show that Eq. (15.5) does indeed hold for any pair of functions.

Importance of Hermiticity. The importance of Hermiticity is threefold. First, Hermiticity of the operator guarantees that the eigenfunctions are complete and orthogonal. This means that the eigenfunctions satisfy

$$\int u^*(a',t)\,u(a,t)\,dt = \delta(a-a') \tag{15.6}$$

$$\int u^*(a,t')\,u(a,t)\,da = \delta(t-t') \tag{15.7}$$

It is these properties of the expansion functions that allow us to transform between $s(t)$ and $F(a)$ as given by Eqs. (15.1) and (15.2). In particular, to see how to obtain $F(a)$, multiply Eq. (15.1) by $u^*(a',t)$ and integrate with respect to time,

$$\int s(t)\,u^*(a',t)\,dt = \iint F(a)\,u(a,t)\,u^*(a',t)\,da\,dt \tag{15.8}$$

$$= \int F(a)\,\delta(a-a')\,da \tag{15.9}$$

$$= F(a') \tag{15.10}$$

which proves the inverse relation, Eq. (15.2).

Second, if the operator is Hermitian the eigenvalues are guaranteed to be real. This is important because in nature measurable quantities are real. Hence, if the eigenvalues are to be measurable numerical values, the operator should be Hermitian to assure that this is so. That is not to say that non-Hermitian operators are not important but simply that they do not represent measurable physical quantities. For example, the translation operator that we studied in Section 1.4 is not Hermitian but is very useful.

Third, if an operator is Hermitian we can manipulate it in advantageous ways. We have already seen this in Section 1.4 and we will see it over and over. The proof that Hermitian operators do produce real eigenvalues and complete and orthogonal eigenfunctions can be found in any book on mathematical methods.

Range and Values of a. It is important to understand that the numerical values and range of the eigenvalues are obtained by solving the eigenvalue problem. It is not imposed or assumed. If the solution of the eigenvalue problem predicts certain values for the physical parameter a and we experimentally obtain a value that does not correspond to one of those values, then there are only three possibilities: the experiment is wrong, we have made an error in solving the eigenvalue problem, or we do not have the correct operator for that physical quantity.

Possible and Actual Values of the a's. One must be very clear about the following distinction. The solution of the eigenvalue problem gives us the possible values that the physical quantity may attain, that is, the a's. However, for a particular signal the actual values attainable are given by the function $F(a)$, which may or

may not range over all possible values of a. For example, any frequency is possible, but for a particular signal only certain frequencies may exist. Note that solving the eigenvalue problem has nothing to do with the signal at hand. The eigenvalue problem is solved once and for all, but $F(a)$ must be recalculated for each different signal.

Normalization. For linear operators, a constant times an eigenfunction is also a solution to the eigenvalue problem. To fix the constant the eigenfunctions are normalized so that Eqs. (15.6) and (15.7) are satisfied. In that case we say that we have normalized to a delta function.

The Usage of "spectrum." The set of eigenvalues obtained by solving the eigenvalue problem is often called the spectrum and the terms continuous and discrete spectrum are used to communicate whether the a's are continuous or discrete. The use of the word spectrum for the general case is unfortunate, because spectrum is associated with frequency which is only one particular case. But even more unfortunate, the word spectrum is often used to denote the Fourier transform, $S(\omega)$. However, the particular usage is usually clear from the context.

Example 15.1: Frequency.

From Chapter 1 we know that the frequency operator is

$$\mathcal{W} = \frac{1}{j}\frac{d}{dt} \tag{15.11}$$

First let us prove that it is Hermitian. For any two functions, f and g, we have, by integration by parts, that

$$\int g^*(t)\,\frac{1}{j}\frac{d}{dt}f(t)\,dt = \frac{1}{j}fg^*\Big|_{-\infty}^{\infty} - \frac{1}{j}\int f(t)\,\frac{d}{dt}g^*(t)\,dt \tag{15.12}$$

$$= \int f(t)\left(\frac{1}{j}\frac{d}{dt}g(t)\right)^* dt \tag{15.13}$$

and hence the frequency operator is Hermitian.

The eigenvalue problem is

$$\mathcal{W}u(\omega,t) = \omega\,u(\omega,t) \tag{15.14}$$

and the solutions are

$$u(\omega,t) = ce^{j\omega t} \tag{15.15}$$

Notice that all real values of ω are possible. Therefore we say that the range of frequencies are the real numbers from $-\infty$ to ∞. To obtain the normalization consider

$$\int u^*(\omega,t)\,u(\omega',t)\,dt = c^2\int e^{-j\omega t}e^{j\omega' t}\,dt = 2\pi c^2\delta(\omega-\omega') \tag{15.16}$$

Since we want to normalize to a delta function we must take $c^2 = 1/2\pi$ and hence the normalized frequency eigenfunctions are

$$u(\omega, t) = \frac{1}{\sqrt{2\pi}} e^{j\omega t} \tag{15.17}$$

Example 15.2: Time.

The time operator in the time representation is t, and the eigenvalue problem is

$$t\,u(t,t') = t'u(t,t') \tag{15.18}$$

where t' are the eigenvalues. This is the equation that led Dirac to invent the Dirac delta function. The solutions are

$$u(t,t') = \delta(t-t') \tag{15.19}$$

where t' can be any number. Therefore the eigenvalues are continuous. The eigenfunctions are complete and orthogonal.

Discrete Case. If the eigenvalues are discrete, then the notation a_n is used to denote them, where n is an integer index. By convention, whenever possible, the eigenvalues are arranged in order of increasing magnitude. The corresponding eigenfunctions are denoted by $u_n(t)$ and the eigenvalue problem is written as

$$\mathcal{A}\,u_n(t) = a_n\,u_n(t) \tag{15.20}$$

We emphasize that before we have solved the eigenvalue problem we do not know whether the solutions will be discrete or continuous or both. As in the continuous case, the eigenfunctions are orthogonal and complete,

$$\int u_m^*(t)\,u_n(t)\,dt = \delta_{nm} \tag{15.21}$$

$$\sum_n u_n^*(t)\,u_n(t') = \delta(t-t') \tag{15.22}$$

A signal can be expanded as

$$s(t) = \sum_n c_n\,u_n(t) \tag{15.23}$$

where the coefficients, c_n, are

$$c_n = \int u_n^*(t)\,s(t)\,dt \tag{15.24}$$

The proof of Eq. (15.24) is the same as for the continuous case which we gave in the previous section.

Signals in different representations: Terminology. One can uniquely go back and forth between $s(t)$ and $F(a)$. Nothing is lost. Therefore we say that $s(t)$ is the signal in the time representation and that $F(a)$ is the signal in the a representation. Similarly, for the discrete case we say that the set of $\{c_n\}'s$ is the signal in the u_n representation.

15.3 OPERATOR ALGEBRA

We now give an elementary exposition of the basic operator methods we will subsequently use. The two fundamental operators are the time and frequency operators and generally other operators will be functions of these. The primary idea is that for a physical quantity we will associate an operator.

Very often we will be dealing with two or more physical quantities and thus with two or more operators. If we have two operators, $\mathcal{A}$ and $\mathcal{B}$, then the operator $\mathcal{AB}$ means to operate first with $\mathcal{B}$ and then with $\mathcal{A}$. Generally, the order of operation is not interchangeable, as, for example, putting on socks and shoes. The order matters. If the order doesn't matter, as, for example, putting on a hat and shoes, then we say the operators commute. To determine if two operators commute, we operate with $\mathcal{AB}$ and $\mathcal{BA}$ on an arbitrary function to see whether the same answer is obtained. Equivalently, we can operate with $\mathcal{AB} - \mathcal{BA}$ to examine whether zero is obtained. The operator $\mathcal{AB} - \mathcal{BA}$ is called the commutator of $\mathcal{A}$ and $\mathcal{B}$ and is denoted by $[\mathcal{A}, \mathcal{B}]$,

$$[\mathcal{A}, \mathcal{B}] = \mathcal{AB} - \mathcal{BA} \qquad [\text{ Commutator of } \mathcal{A} \text{ and } \mathcal{B} \text{ }] \tag{15.25}$$

As we will see the commutator of two quantities plays a fundamental role. Some properties of the commutator are

$$[\mathcal{A}, \mathcal{B}] = -[\mathcal{B}, \mathcal{A}] \tag{15.26}$$

$$[c\mathcal{A}, \mathcal{B}] = c\,[\mathcal{A}, \mathcal{B}] \tag{15.27}$$

$$[\mathcal{A}, \mathcal{B} + \mathcal{C}] = [\mathcal{A}, \mathcal{B}] + [\mathcal{A}, \mathcal{C}] \tag{15.28}$$

It is also useful to define the anticommutator,

$$[\mathcal{A}, \mathcal{B}]_+ = \mathcal{AB} + \mathcal{BA} \qquad [\text{ Anticommutator of } \mathcal{A} \text{ and } \mathcal{B} \text{ }] \tag{15.29}$$

Note that $[\mathcal{A}, \mathcal{B}]_+$ and $[\mathcal{A}, \mathcal{B}]$ are respectively symmetric and antisymmetric with the interchange of the two operators.

Forming Hermitian Operators. One often forms a new operator from Hermitian operators and it is important to know whether the new operator is Hermitian. Assuming that $\mathcal{A}$ and $\mathcal{B}$ are Hermitian, the following are readily verified to be Hermi-

tian operators:

$$\left.\begin{array}{c} c\mathcal{A} \\ \mathcal{A}^n \\ \mathcal{A}+\mathcal{B} \\ [\mathcal{A},\mathcal{B}]_+ \\ [\mathcal{A},\mathcal{B}]/j \end{array}\right\} \text{ are Hermitian if } \mathcal{A} \text{ and } \mathcal{B} \text{ are Hermitian and } c \text{ is real} \qquad (15.30)$$

Operator Equations and Fundamental Commutation Relation. An operator equation means that when we operate with either side on an arbitrary function the same result is obtained. For example, the fundamental operator commutation rule between time and frequency

$$\mathcal{TW} - \mathcal{WT} = j \qquad (15.31)$$

means that operating with the left hand side on an arbitrary function is the same as multiplying that function by j. To prove this consider

$$(\mathcal{TW} - \mathcal{WT})\, s(t) = \left(t\frac{1}{j}\frac{d}{dt} - \frac{1}{j}\frac{d}{dt}t\right) s(t) \qquad (15.32)$$

$$= \frac{1}{j}\left(t\frac{ds}{dt} - t\frac{ds}{dt} - s\right) \qquad (15.33)$$

$$= js(t) \qquad (15.34)$$

which proves Eq. (15.31).

Operator equations are manipulated almost like algebraic equations but careful attention must be paid to the fact that operators may not commute and therefore can not be switched around an expression.

Functions of an Operator. By a function of an operator, $Q(\mathcal{A})$, we mean, by definition, that in the ordinary Taylor expansion of the function one substitutes the operator for the variable,

$$Q(\mathcal{A}) = \sum_n c_n \mathcal{A}^n \qquad \text{if} \qquad Q(x) = \sum_n c_n x^n \qquad (15.35)$$

When will $Q(\mathcal{A})$ thus defined be Hermitian? We already know that $\mathcal{A}^n$ is Hermitian and therefore $c_n\mathcal{A}^n$ is Hermitian if c_n is real. But having the c_n's real means that the function $Q(x)$ is real, hence we conclude that

$$Q(\mathcal{A}) \quad \text{is Hermitian if } \mathcal{A} \text{ is Hermitian and } Q(x) \text{ is a real function} \qquad (15.36)$$

A problem that often arises is the operation of $Q(\mathcal{A})$ on an arbitrary function. Generally, we just do it. However, a useful formula can be derived. Consider first the action of $Q(\mathcal{A})$ on an eigenfunction of $\mathcal{A}$,

$$Q(\mathcal{A})\,u(a,t) = \sum_n c_n \mathcal{A}^n\,u(a,t) = \sum_n c_n a^n\,u(a,t) \tag{15.37}$$

But the sum is recognized to be $Q(a)$ and therefore we have

$$Q(\mathcal{A})\,u(a,t) = Q(a)\,u(a,t) \tag{15.38}$$

Now consider the operation of $Q(\mathcal{A})$ on an arbitrary function $s(t)$,

$$Q(\mathcal{A})s(t) = Q(\mathcal{A}) \int F(a)\,u(a,t)\,da = \int F(a)Q(\mathcal{A})\,u(a,t)\,da \tag{15.39}$$

and therefore, using Eq. (15.38),

$$Q(\mathcal{A})\,s(t) = \int F(a)\,Q(a)\,u(a,t)\,da \tag{15.40}$$

The Inverse of an Operator. The inverse of $\mathcal{A}$, $\mathcal{A}^{-1}$, is defined so that

$$\mathcal{A}^{-1}\mathcal{A} = \mathcal{A}\mathcal{A}^{-1} = \mathcal{I} \tag{15.41}$$

where $\mathcal{I}$ is the unit operator. Typically we leave out the unit operator because it is understood to be there where appropriate. For example, we write $[\mathcal{T},\mathcal{W}] = j$ instead of $[\mathcal{T},\mathcal{W}] = j\mathcal{I}$.

The inverse of an operator that is the product of two operators is given by

$$(\mathcal{A}\mathcal{B})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1} \tag{15.42}$$

This is so because

$$(\mathcal{A}\mathcal{B})^{-1}(\mathcal{A}\mathcal{B}) = \mathcal{B}^{-1}\mathcal{A}^{-1}\mathcal{A}\mathcal{B} = \mathcal{B}^{-1}\mathcal{B} = \mathcal{I} \tag{15.43}$$

The Adjoint of an Operator. The adjoint is another operator, denoted by $\mathcal{A}^\dagger$, which forces the equality

$$\int g^*\mathcal{A}f\,dt = \int f\,\{\mathcal{A}^\dagger g\}^*\,dt \tag{15.44}$$

If the adjoint of an operator happens to equal the operator itself, then we see that Eq. (15.44) becomes the definition of a Hermitian operator, Eq. (15.5), and hence

$$\text{if} \quad \mathcal{A} = \mathcal{A}^\dagger \qquad \text{then } \mathcal{A} \text{ is Hermitian (self adjoint)} \tag{15.45}$$

As can be seen from the definition of adjoint, the adjoint of a constant is the complex conjugate of the constant,

$$\mathcal{A}^\dagger = c^* \qquad \text{if} \quad \mathcal{A} = c = \text{constant} \tag{15.46}$$

The adjoint of a product of operators is given by

$$(\mathcal{AB})^\dagger = \mathcal{B}^\dagger \mathcal{A}^\dagger \tag{15.47}$$

Another important property of the adjoint is that

$$\left.\begin{array}{c} \mathcal{A}+\mathcal{A}^\dagger \\ (\mathcal{A}-\mathcal{A}^\dagger)/j \end{array}\right\} \text{are Hermitian whether } \mathcal{A} \text{ is Hermitian or not.} \tag{15.48}$$

An arbitrary, possibly non Hermitian, operator can be written in the following way

$$\mathcal{A} = \tfrac{1}{2}(\mathcal{A}+\mathcal{A}^\dagger) + \tfrac{1}{2}j(\mathcal{A}-\mathcal{A}^\dagger)/j \tag{15.49}$$

Notice that we have multiplied and divided by j in the second term. The importance of this decomposition is twofold. First, it breaks up an operator into symmetric and antisymmetric parts. Second, it expresses an operator as a sum of a Hermitian operator plus j times a Hermitian operator. This is the operator analog of writing a complex number in terms of its real and imaginary part.

Product of Two Hermitian Operators. The product of two Hermitian operators is not necessarily Hermitian. However, using the above decomposition by letting $\mathcal{A} \to \mathcal{AB}$ it can be written in the following useful way,

$$\mathcal{AB} = \tfrac{1}{2}\left(\mathcal{AB}+(\mathcal{AB})^\dagger\right) + \tfrac{1}{2}\left(\mathcal{AB}-(\mathcal{AB})^\dagger\right) \tag{15.50}$$

$$= \tfrac{1}{2}(\mathcal{AB}+\mathcal{BA}) + \tfrac{1}{2}(\mathcal{AB}-\mathcal{BA}) \tag{15.51}$$

$$= \tfrac{1}{2}[\mathcal{A},\mathcal{B}]_+ + \tfrac{j}{2}[\mathcal{A},\mathcal{B}]/j \tag{15.52}$$

This expresses $\mathcal{AB}$ in terms of its commutator and anticommutator.

Unitary Operators. An operator $\mathcal{U}$ is said to be unitary if its adjoint is equal to its inverse,

$$\mathcal{U}^\dagger = \mathcal{U}^{-1} \qquad [\text{Unitary}] \tag{15.53}$$

The importance of unitary operators is that they preserve normalization when operating on a function. That is, $f(t)$ and $\mathcal{U}f(t)$ have the same normalization if $\mathcal{U}$ is unitary,

$$\int |f(t)|^2\,dt = \int |\mathcal{U}f(t)|^2\,dt \tag{15.54}$$

To see this consider

$$\int |\mathcal{U}f(t)|^2\,dt = \int \{\mathcal{U}f(t)\}^*\mathcal{U}f(t)\,dt = \int f^*(t)\mathcal{U}^\dagger\mathcal{U}f(t)\,dt = \int |f(t)|^2\,dt \tag{15.55}$$

In general, an operator of the form

$$\mathcal{U} = e^{j\mathcal{A}} \tag{15.56}$$

will be unitary if $\mathcal{A}$ is Hermitian. To see this, first note that if $\mathcal{A}$ is Hermitian then

$$\mathcal{U}^\dagger = e^{-j\mathcal{A}^\dagger} \tag{15.57}$$

This can be proven straightforwardly

$$\mathcal{U}^\dagger = (e^{j\mathcal{A}})^\dagger = \left(\sum_n \frac{j^n}{n!}\mathcal{A}^n\right)^\dagger = \sum_n \frac{(-j)^n}{n!}\mathcal{A}^{\dagger n} = e^{-j\mathcal{A}^\dagger} \tag{15.58}$$

To show that $\mathcal{U}$ is unitary if $\mathcal{A}$ is Hermitian, consider

$$\mathcal{U}\mathcal{U}^\dagger = e^{j\mathcal{A}}e^{-j\mathcal{A}^\dagger} = e^{j\mathcal{A}}e^{-j\mathcal{A}} = \mathcal{I} \tag{15.59}$$

Therefore, $\mathcal{U}^\dagger = \mathcal{U}^{-1}$, which is the definition of a unitary operator.

Unitary operators are not Hermitian but they do obey the simple manipulative rule

$$\int g^* \mathcal{U} f \, dt = \int f \, (\mathcal{U}^\dagger g)^* \, dt = \int f \, (\mathcal{U}^{-1} g)^* \, dt \tag{15.60}$$

This is the case since the middle step is the definition of adjoint, and the right side follows because of Eq. (15.53).

Example 15.3: Translation Operator.

The translation operator defined by $\mathcal{U} = e^{j\tau\mathcal{W}}$ is unitary. That follows because the frequency operator is Hermitian. Also, we can see that from the fact that $e^{j\tau\mathcal{W}} f(t) = f(t+\tau)$ and by noting that $f(t)$ and $f(t+\tau)$ have the same normalization for any τ. The inverse of the translation operator is $\mathcal{U}^{-1} = e^{-j\tau\mathcal{W}}$. This follows from Eq. (15.53) or can be verified directly,

$$e^{-j\tau\mathcal{W}} e^{j\tau\mathcal{W}} s(t) = e^{-j\tau\mathcal{W}} s(t+\tau) = s(t+\tau-\tau) = s(t) \tag{15.61}$$

15.4 AVERAGES

If the density of a is taken to be $|F(a)|^2$, then the average value of a is

$$\langle a \rangle = \int a \, |F(a)|^2 \, da \tag{15.62}$$

and more generally the average of any function, $g(a)$, is

$$\langle g \rangle = \int g(a) \, |F(a)|^2 \, da \tag{15.63}$$

In Section 1.4 we saw that we can calculate frequency averages directly from the signal, without calculating the Fourier transform. This is a special case of the more general result whereby the average of $g(a)$ can be calculated directly from the signal, and the calculation of the transform $F(a)$ can be avoided. This is a fundamental result of operator theory. In particular,

$$\langle g \rangle = \int g(a) \, | F(a) |^2 \, da = \int s^*(t) \, g(\mathcal{A}) \, s(t) \, dt \tag{15.64}$$

To prove this, consider

$$\int s^*(t) \, g(\mathcal{A}) \, s(t) \, dt = \iiint F^*(a') \, u^*(a', t) \, g(\mathcal{A}) F(a) \, u(a, t) \, dt \, da' da \tag{15.65}$$

Now, since $g(\mathcal{A}) \, u(a, t) = g(a) \, u(a, t)$, we have

$$\int s^*(t) \, g(\mathcal{A}) \, s(t) \, dt = \iiint F^*(a') \, u^*(a', t) \, g(a) F(a) \, u(a, t) \, dt \, da' da \tag{15.66}$$

$$= \iint F^*(a') \, g(a) F(a) \delta(a - a') \, da' da \tag{15.67}$$

$$= \int g(a) \, | F(a) |^2 \, da \tag{15.68}$$

Averages of Hermitian Operators Are Real. We use both $\langle \mathcal{A} \rangle$ and $\langle a \rangle$ to signify the average of a,

$$\langle a \rangle = \int s^*(t) \, \mathcal{A} \, s(t) \, dt \tag{15.69}$$

It is important to appreciate that the average defined by Eq. (15.69) will always be real provided the operator is Hermitian. The reason is that if the operator is Hermitian, then Eq. (15.64) holds and the middle term is real if g is a real function.

Example 15.4: Frequency.

The frequency operator is

$$\mathcal{W} = \frac{1}{j} \frac{d}{dt} \tag{15.70}$$

and therefore

$$\langle g(\omega) \rangle = \int | S(\omega) |^2 g(\omega) \, d\omega = \int s^*(t) \, g\left(\frac{1}{j} \frac{d}{dt} \right) s(t) \, dt \tag{15.71}$$

which is a relation we have repeatedly used in the previous chapters to simplify calculations.

Averages of Non-Hermitian Operators. If $\mathcal{A}$ is not Hermitian, then its average, calculated by way of Eq. (15.69), will be a complex number. The real and imaginary parts can be explicitly written down by using Eq. (15.49). Taking average values of both sides of that equation we have

$$\langle \mathcal{A} \rangle = \tfrac{1}{2}\langle \mathcal{A} + \mathcal{A}^\dagger \rangle + \tfrac{1}{2}j\langle (\mathcal{A} - \mathcal{A}^\dagger)/j \rangle \tag{15.72}$$

This is the average value of an operator in terms of its real and imaginary parts. If $\mathcal{A}$ is Hermitian then the second term, the imaginary part, is zero.

A particularly important case is when the operator is the product of two Hermitian operators. Taking average values of Eq. (15.51) we have

$$\langle \mathcal{AB} \rangle = \tfrac{1}{2}\langle [\mathcal{A},\mathcal{B}]_+ \rangle + \tfrac{1}{2}j\langle [\mathcal{A},\mathcal{B}]/j \rangle \tag{15.73}$$

which expresses the average value for the product of two operators in terms of the real and imaginary parts since both $\langle (\mathcal{AB} + \mathcal{BA}) \rangle$ and $\langle [\mathcal{A},\mathcal{B}]/j \rangle$ are real.

The Covariance. In Section 1.8 we defined the covariance between time and frequency as the average of $t\,\varphi(t)$, where φ is the phase of the signal. We now define the covariance of any two quantities by[87]

$$\begin{aligned} \mathrm{Cov}_{ab} &= \tfrac{1}{2}\langle \mathcal{AB} + \mathcal{BA} \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle && (15.74)\\ &= \tfrac{1}{2}\langle [\mathcal{A},\mathcal{B}]_+ \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle && (15.75)\\ &= \tfrac{1}{2}\langle [\mathcal{A} - \langle \mathcal{A} \rangle, \mathcal{B} - \langle \mathcal{B} \rangle]_+ \rangle && (15.76) \end{aligned}$$

The rational for this definition will be discussed in Section 17.10.

Example 15.5: Time-Frequency Covariance.

Using the commutation relation, $\mathcal{TW} - \mathcal{WT} = j$, we have

$$\langle \mathcal{TW} + \mathcal{WT} \rangle = \langle 2\mathcal{TW} \rangle - j = 2\int s^* t \mathcal{W} s\, dt - j = 2\int t\,\varphi'(t)\,|\,s(t)\,|^2\,dt = 2\langle t\,\varphi'(t) \rangle \tag{15.77}$$

and therefore

$$\mathrm{Cov}_{t\omega} = \langle t\,\varphi'(t) \rangle - \langle t \rangle \langle \omega \rangle \tag{15.78}$$

which is the same definition we used in Eq.(1.124). We mentioned in Section 1.8 that

$$\int t\,\varphi'(t)\,|\,s(t)\,|^2\,dt = -\int \omega\,\psi'(\omega)\,|\,S(\omega)\,|^2\,d\omega \tag{15.79}$$

A simple proof will now be given. Instead of evaluating $\langle \mathcal{TW} + \mathcal{WT} \rangle$ in the time representation we evaluate it in the spectral representation. Using the same steps as above,

$$\langle \mathcal{TW}+\mathcal{WT} \rangle = \langle 2\mathcal{WT} \rangle + j = 2\int S^* \omega \mathcal{T} S\, d\omega + j = -2\int \omega\psi'(\omega)|S(\omega)|^2\,d\omega = -2\langle \omega\,\psi' \rangle \tag{15.80}$$

Since the two methods of calculation must give the same result we have Eq. (15.79).

15.5 THE UNCERTAINTY PRINCIPLE FOR ARBITRARY VARIABLES

If we have two physical quantities then we will have an uncertainty principle between them if the operators do not commute. Specifically, the uncertainty principle for any two quantities represented by the operators $\mathcal{A}$ and $\mathcal{B}$ is

$$\sigma_a \, \sigma_b \geq \tfrac{1}{2} \, | \, \langle [\mathcal{A}, \mathcal{B}] \rangle \, | \tag{15.81}$$

where σ_a is the standard deviation defined in the usual way,

$$\sigma_a^2 = \langle \mathcal{A}^2 \rangle - \langle \mathcal{A} \rangle^2 \tag{15.82}$$

$$= \int s^*(t) \; (\mathcal{A} - \langle \mathcal{A} \rangle)^2 \, s(t) \, dt \tag{15.83}$$

$$= \int | \, (\mathcal{A} - \langle \mathcal{A} \rangle) \, s(t) \, |^2 \, dt \tag{15.84}$$

Similarly for $\mathcal{B}$. Furthermore, the signal that minimizes the uncertainty product is obtained by solving

$$(\mathcal{B} - \langle \mathcal{B} \rangle) \, s(t) = \lambda (\mathcal{A} - \langle \mathcal{A} \rangle) \, s(t) \tag{15.85}$$

where

$$\lambda = \frac{\langle [\mathcal{A}, \mathcal{B}] \rangle}{2\sigma_a^2} \tag{15.86}$$

Note that in general λ may be complex since $[\mathcal{A}, \mathcal{B}]$ is not Hermitian. As with the uncertainty principle for time and frequency, Eqs. (3.3) and (3.4), a more general result can be proven,

$$\sigma_a \, \sigma_b \geq \tfrac{1}{2} \sqrt{| \, \langle [\mathcal{A}, \mathcal{B}] \rangle \, |^2 + 4 \, \mathrm{Cov}_{ab}^2} \tag{15.87}$$

Proof of the Uncertainty Principle. First some preliminaries. For convenience define

$$\mathcal{A}_0 = \mathcal{A} - \langle \mathcal{A} \rangle \quad ; \quad \mathcal{B}_0 = \mathcal{B} - \langle \mathcal{B} \rangle \tag{15.88}$$

and note that the operators $\mathcal{A}_0$ and $\mathcal{B}_0$ are Hermitian and that their mean is zero. Furthermore, it is easy to verify that

$$[\mathcal{A}_0, \mathcal{B}_0] = [\mathcal{A}, \mathcal{B}] \tag{15.89}$$

and

$$[\mathcal{A}_0, \mathcal{B}_0]_+ = [\mathcal{A}, \mathcal{B}]_+ - 2\mathcal{A}\langle \mathcal{B} \rangle - 2\mathcal{B}\langle \mathcal{A} \rangle + 2\langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \tag{15.90}$$

Taking expectation values of both sides of Eq. (15.89) we have correspondingly

$$\langle [\mathcal{A}_0, \mathcal{B}_0] \rangle = \langle [\mathcal{A}, \mathcal{B}] \rangle \tag{15.91}$$

and

$$\langle [\mathcal{A}_0, \mathcal{B}_0]_+ \rangle = 2\,\mathrm{Cov}_{ab} \tag{15.92}$$

With these preliminaries we can now derive the uncertainty principle for arbitrary quantities by considering

$$\sigma_a^2 \sigma_b^2 = \int |\mathcal{A}_0\, s(t)|^2\, dt \times \int |\mathcal{B}_0\, s(t)|^2\, dt \tag{15.93}$$

$$\geq \left| \int \{\mathcal{A}_0\, s(t)\}^* \{\mathcal{B}_0 s(t)\}\, dt \right|^2 \tag{15.94}$$

$$= \left| \int s^*(t)\, \mathcal{A}_0\, \mathcal{B}_0 s(t)\, dt \right|^2 \tag{15.95}$$

$$= |\langle \mathcal{A}_0\, \mathcal{B}_0 \rangle|^2 \tag{15.96}$$

where in going from step (15.93) to step (15.94) we have used the Schwarz inequality, Eq. (3.10), and in going from (15.94) to (15.95) we have used the Hermiticity property of the operator. Therefore

$$\sigma_a^2 \sigma_b^2 \geq |\langle \mathcal{A}_0\, \mathcal{B}_0 \rangle|^2 \tag{15.97}$$

Using Eq. (15.92) and (15.91) we have

$$\langle \mathcal{A}_0\, \mathcal{B}_0 \rangle = \tfrac{1}{2}\langle [\mathcal{A}_0, \mathcal{B}_0]_+ \rangle + \tfrac{1}{2} j \langle [\mathcal{A}_0, \mathcal{B}_0]/j \rangle \tag{15.98}$$

$$= \mathrm{Cov}_{ab} + \tfrac{1}{2} j \langle [\mathcal{A}_0, \mathcal{B}_0]/j \rangle \tag{15.99}$$

and hence

$$\sigma_a^2 \sigma_b^2 \geq \left| \mathrm{Cov}_{ab} + \tfrac{1}{2} j \langle [\mathcal{A}_0, \mathcal{B}_0]/j \rangle \right|^2 \tag{15.100}$$

$$= \mathrm{Cov}_{ab}^2 + \tfrac{1}{4} |\langle [\mathcal{A}_0, \mathcal{B}_0] \rangle|^2 \tag{15.101}$$

$$= \mathrm{Cov}_{ab}^2 + \tfrac{1}{4} |\langle [\mathcal{A}, \mathcal{B}] \rangle|^2 \tag{15.102}$$

which is Eq. (15.87). Since Cov_{ab}^2 is positive it can be dropped to obtain the more standard uncertainty principle, Eq. (15.81).

The minimum uncertainty signal is obtained when the two functions are proportional, $\lambda\, \mathcal{A}_0 s = \mathcal{B}_0 s$ and $\mathrm{Cov}_{ab}^2 = 0$. That is,

$$\lambda\, (\mathcal{A} - \langle \mathcal{A} \rangle)\, s = (\mathcal{B} - \langle \mathcal{B} \rangle)\, s \tag{15.103}$$

$$\langle [\mathcal{A} - \langle \mathcal{A} \rangle, \mathcal{B} - \langle \mathcal{B} \rangle]_+ \rangle = 0 \tag{15.104}$$

The constant λ is thus far arbitrary. To fix it let us first write the above in terms of $\mathcal{A}_0$ and $\mathcal{B}_0$,

$$\lambda \mathcal{A}_0 s = \mathcal{B}_0 s \tag{15.105}$$

$$0 = \langle \mathcal{A}_0 \mathcal{B}_0 \rangle + \langle \mathcal{B}_0 \mathcal{A}_0 \rangle \tag{15.106}$$

Multiply the first equation by $\mathcal{A}_0$ and take average values; then multiply by $\mathcal{B}_0$ and take average values, to obtain

$$\lambda \langle \mathcal{A}_0^2 \rangle = \langle \mathcal{A}_0 \mathcal{B}_0 \rangle \quad ; \quad \lambda \langle \mathcal{B}_0 \mathcal{A}_0 \rangle = \langle \mathcal{B}_0^2 \rangle \tag{15.107}$$

Since $\langle \mathcal{A}_0^2 \rangle = \sigma_a^2$ and $\langle \mathcal{B}_0^2 \rangle = \sigma_b^2$, we may write this as

$$\lambda \sigma_a^2 = \langle \mathcal{A}_0 \mathcal{B}_0 \rangle \quad ; \quad \sigma_b^2/\lambda = \langle \mathcal{B}_0 \mathcal{A}_0 \rangle \tag{15.108}$$

Adding and subtracting these two equations, we have

$$\lambda \sigma_a^2 + \sigma_b^2/\lambda = 0 \tag{15.109}$$

$$\lambda \sigma_a^2 - \sigma_b^2/\lambda = \langle [\mathcal{A}_0, \mathcal{B}_0] \rangle = \langle [\mathcal{A}, \mathcal{B}] \rangle \tag{15.110}$$

Solving for λ we obtain

$$\lambda = \frac{\langle [\mathcal{A}, \mathcal{B}] \rangle}{2\sigma_a^2} \quad ; \quad \lambda^2 = -\frac{\sigma_b^2}{\sigma_a^2} \tag{15.111}$$

Note that λ is purely imaginary.

Example 15.6: Time and Frequency.

The commutator of the time and frequency operator is j, so

$$\sigma_t \, \sigma_\omega \geq \tfrac{1}{2} \, | \langle [T, W] \rangle | = \tfrac{1}{2} |j| = \tfrac{1}{2} \tag{15.112}$$

Chapter 16

Density of a Single Variable

16.1 INTRODUCTION

In this chapter we develop the basic methods to study the density of a single variable. The object is to generalize the ideas and methods that have been developed for the cases of frequency and time to arbitrary variables.

16.2 DENSITY OF A SINGLE VARIABLE

In the Fourier case the density of frequency is taken to be $|S(\omega)|^2$ and we can argue by analogy that for an arbitrary variable it should be $|F(a)|^2$, where $F(a)$ is the "a" transform of the signal, Eq. (15.2). We are now in a position to "derive" this by way of the characteristic function method. The word derive is in quotes because the general result for calculating averages, Eq. (15.64), has built into it that $|F(a)|^2$ is the density. The following derivation, using the characteristic function operator method, illustrates the simplicity and consistency of the method and will be the basis for obtaining joint densities of two variables. If $P(a)$ is the density of a then the characteristic function is

$$M(\alpha) = \langle e^{j\alpha a} \rangle = \int e^{j\alpha a} P(a)\, da \tag{16.1}$$

and the distribution is obtained from the characteristic function by

$$P(a) = \frac{1}{2\pi} \int M(\alpha)\, e^{-j\alpha a}\, d\alpha \tag{16.2}$$

As we have pointed out in previous chapters the characteristic function is an aver-

age and therefore it can be calculated directly from the signal by way of

$$M(\alpha) = \int s^*(t)\, e^{j\alpha\mathcal{A}} s(t)\, dt \tag{16.3}$$

Expanding the signal in terms of the transform we have

$$e^{j\alpha\mathcal{A}} s(t) = e^{j\alpha\mathcal{A}} \int F(a)\, u(a,t)\, da \tag{16.4}$$

$$= \int F(a)\, e^{j\alpha\mathcal{A}}\, u(a,t)\, da \tag{16.5}$$

$$= \int e^{j\alpha a} F(a)\, u(a,t)\, da \tag{16.6}$$

and therefore

$$M(\alpha) = \iiint F^*(a')\, u^*(a',t)\, e^{j\alpha a} F(a)\, u(a,t)\, da'\, da\, dt \tag{16.7}$$

$$= \iint F^*(a')\, e^{j\alpha a}\, \delta(a-a') F(a)\, da'\, da \tag{16.8}$$

$$= \int |F(a)|^2\, e^{j\alpha a}\, da \tag{16.9}$$

Comparing Eq. (16.9) with Eq. (16.1) we have

$$P(a) = |F(a)|^2 \tag{16.10}$$

Discrete Case. For the characteristic function we have

$$M(\alpha) = \langle e^{j\alpha\mathcal{A}} \rangle = \int s^*(t)\, e^{j\alpha\mathcal{A}} s(t)\, dt \tag{16.11}$$

$$= \int \sum_n \sum_k c_k^*\, u_k^*(t)\; e^{j\alpha\mathcal{A}} c_n\, u_n(t)\, dt \tag{16.12}$$

$$= \int \sum_n \sum_k c_k^*\, u_k^*(t)\, e^{j\alpha a_n}\, c_n\, u_n(t)\, dt \tag{16.13}$$

$$= \sum_n \sum_k c_k^* \delta_{kn} c_n\, e^{j\alpha a_n} \tag{16.14}$$

$$= \sum_n |c_n|^2\, e^{j\alpha a_n} \tag{16.15}$$

which gives

$$P(a) = \frac{1}{2\pi}\int M(\alpha)\, e^{-j\alpha a}\, d\alpha \tag{16.16}$$

$$= \frac{1}{2\pi}\int \sum_n |\, c_n\,|^2\, e^{j\alpha a_n}\, e^{-j\alpha a}\, d\alpha \tag{16.17}$$

$$= \sum_n |\, c_n\,|^2\, \delta(a_n - a) \tag{16.18}$$

The only values which are not zero are the a_n's and their density is

$$P(a_n) = |\, c_n\,|^2 \tag{16.19}$$

Example 16.1: Distribution of Frequency.

For frequency

$$M(\tau) = \langle e^{j\tau\mathcal{W}} \rangle = \int s^*(t)\, e^{j\tau\mathcal{W}} s(t) \tag{16.20}$$

The operator $e^{j\tau\mathcal{W}}$ is the translation operator

$$e^{j\tau\mathcal{W}} s(t) = s(t+\tau) \tag{16.21}$$

and therefore

$$M(\tau) = \int s^*(t)\, s(t+\tau)\, dt \tag{16.22}$$

The distribution is given by

$$P(\omega) = \frac{1}{2\pi}\int M(\tau)\, e^{-j\tau\omega}\, d\tau \tag{16.23}$$

$$= \frac{1}{2\pi}\iint s^*(t)\, s(t+\tau)\, e^{-j\tau\omega}\, d\tau\, dt \tag{16.24}$$

$$= |\, S(\omega)\,|^2 \tag{16.25}$$

as expected.

Example 16.2: Distribution of Inverse Frequency.

Method 1. We define inverse frequency by

$$\tau = \frac{\omega_0}{\omega} \tag{16.26}$$

where ω_0 is an arbitrarily chosen reference frequency. As we are making an ordinary transformation we can use the method developed in Section 4.7. Since

$$P(\omega) = |S(\omega)|^2 \tag{16.27}$$

the distribution of r is, according to Eq. (4.86),

$$P(r) = P(\omega)\left|\frac{d\omega}{dr}\right|_{\omega=\omega_0/r} = \frac{\omega_0}{r^2}\,|S(\omega_0/r)\,|^2 \tag{16.28}$$

Method 2. Alternatively we can define the inverse frequency operator

$$\mathcal{R} = \frac{\omega_0}{\mathcal{W}} \tag{16.29}$$

and use the operator methods developed. Solving the eigenvalue problem in the frequency representation

$$\frac{\omega_0}{\omega}u(r,\omega) = ru(r,\omega) \tag{16.30}$$

gives

$$u(r,\omega) = \frac{\sqrt{\omega_0}}{r}\,\delta(\omega-\omega_0/r) = \frac{\sqrt{\omega_0}}{\omega}\,\delta(r-\omega_0/\omega) \tag{16.31}$$

Calling $G(r)$ the inverse frequency transform we have

$$G(r) = \int S(\omega)\,u^*(r,\omega)\,d\omega = \int S(\omega)\sqrt{\frac{\omega_0}{r^2}}\,\delta(\omega-\omega_0/r)\,d\omega = \sqrt{\frac{\omega_0}{r^2}}\,S(\omega_0/r) \tag{16.32}$$

The distribution is

$$P(r) = |G(r)|^2 = \frac{\omega_0}{r^2}\,|\,S(\omega_0/r)\,|^2 \tag{16.33}$$

which is the same as Eq. (17.124).

16.3 MEAN VALUES

As we showed in the previous chapter, the average of a can be calculated directly from the signal by way of

$$\langle a\rangle = \int s^*(t)\,\mathcal{A}\,s(t)\,dt \tag{16.34}$$

We now obtain an interesting form analogous to the one we obtained for frequency, Eq. (1.90). Rewrite Eq. (16.34) as

$$\langle a\rangle = \int \left(\frac{\mathcal{A}s}{s}\right)|\,s(t)\,|^2\,dt \tag{16.35}$$

and break up $\mathcal{A}s/s$ into its real and imaginary parts

$$\frac{\mathcal{A}s}{s} = \left(\frac{\mathcal{A}s}{s}\right)_R + j\left(\frac{\mathcal{A}s}{s}\right)_I \tag{16.36}$$

to write

$$\langle a\rangle = \int\left[\left(\frac{\mathcal{A}s}{s}\right)_R + j\left(\frac{\mathcal{A}s}{s}\right)_I\right]|\,s(t)\,|^2\,dt \tag{16.37}$$

Since the mean value of a Hermitian operator has to be real we must have

$$\int \left(\frac{\mathcal{A}s}{s}\right)_I |\, s(t)\,|^2\, dt \;=\; 0 \tag{16.38}$$

which leaves

$$\langle a \rangle \;=\; \int \left(\frac{\mathcal{A}s}{s}\right)_R |\, s(t)\,|^2\, dt \tag{16.39}$$

Example 16.3: Spectral Average.

Taking $\mathcal{A}$ to be the frequency operator, $\mathcal{A} = \mathcal{W} = \frac{1}{j}\frac{d}{dt}$, we have

$$\frac{\mathcal{W}s}{s} \;=\; \frac{1}{s}\frac{1}{j}\frac{d}{dt}s \;=\; \frac{1}{j}\frac{A'(t)}{A(t)} + \varphi'(t) \tag{16.40}$$

The real part is $\varphi'(t)$, and using Eq. (16.39) we immediately have

$$\langle \omega \rangle \;=\; \int \varphi'(t)\, A^2(t)\, dt \tag{16.41}$$

16.4 BANDWIDTH

As with the average, the standard deviation can also be written directly in terms of the signal,

$$\sigma_a^2 \;=\; \int (a - \langle a \rangle)^2\, |\, F(a)\,|^2\, da \tag{16.42}$$

$$\;=\; \int s^*(t)\, (\mathcal{A} - \langle \mathcal{A} \rangle)^2\, s(t)\; dt \tag{16.43}$$

$$\;=\; \int \left| \left(\frac{\mathcal{A}s}{s} - \langle \mathcal{A} \rangle\right) s(t) \right|^2 dt \tag{16.44}$$

$$\;=\; \int \left| \left[\left(\frac{\mathcal{A}s}{s}\right)_R - \langle \mathcal{A} \rangle + j \left(\frac{\mathcal{A}s}{s}\right)_I \right] s(t) \right|^2 dt \tag{16.45}$$

The first two terms are real since $\langle \mathcal{A} \rangle$ is real, hence

$$\sigma_a^2 \;=\; \int \left(\frac{\mathcal{A}s}{s}\right)_I^2 |\, s(t)\,|^2\, dt \;+\; \int \left[\left(\frac{\mathcal{A}s}{s}\right)_R - \langle \mathcal{A} \rangle \right]^2 |\, s(t)\,|^2\; dt \tag{16.46}$$

Example 16.4: Bandwidth Equation.

From the previous example we have

$$\left(\frac{\mathcal{W}s}{s}\right)_I = -\frac{A'(t)}{A(t)} \qquad\qquad \left(\frac{\mathcal{W}s}{s}\right)_R = \varphi'(t) \tag{16.47}$$

and therefore

$$(\Delta\omega)^2 = \int \left(\frac{A'(t)}{A(t)}\right)^2 A^2(t)\,dt + \int \big(\varphi'(t) - \langle\omega\rangle\big)^2 A^2(t)\,dt \tag{16.48}$$

which is the bandwidth equation for frequency.

16.5 ARBITRARY STARTING REPRESENTATION

In the above considerations we have been transforming between the time representation and the a representation. However, there is no particular reason why we must start with the time representation. We can work between any two representations, say, a and b. Suppose we have a quantity, b, and its associated operator, $\mathcal{B}$, and we solve for the eigenfunctions, which we call $u(b,t)$. The signal in the b representation is then

$$F(b) = \int u^*(b,t)\,s(t)\,dt \tag{16.49}$$

The average and bandwidth of a can be calculated using $F(b)$ as the signal, instead of $s(t)$,

$$\langle\mathcal{A}\rangle = \int F^*(b)\,\mathcal{A}\,F(b)\,db \tag{16.50}$$

$$\sigma_a^2 = \int F^*(b)\,(\mathcal{A} - \langle\mathcal{A}\rangle)^2\,F(b)\,db \tag{16.51}$$

where now $\mathcal{A}$ has to be expressed in the variables of the b representation. The identical derivation of Section 16.3 leads to

$$\langle a\rangle = \int \left(\frac{\mathcal{A}F}{F}\right)_R |\,F(b)\,|^2\,db \tag{16.52}$$

and

$$\sigma_a^2 = \int \left(\frac{\mathcal{A}F}{F}\right)_I^2 |\,F(b)\,|^2\,db + \int \left[\left(\frac{\mathcal{A}F}{F}\right)_R - \langle\mathcal{A}\rangle\right]^2 |\,F(b)\,|^2\,db \tag{16.53}$$

Example 16.5: Mean Time and Duration in Terms of the Spectrum.

We write the spectrum in terms of its amplitude and phase $S(\omega) = B(\omega)\,e^{j\psi(\omega)}$ to obtain

$$\frac{\mathcal{T}S}{S} = -\frac{1}{S}\frac{1}{j}\frac{d}{d\omega}S = -\frac{1}{j}\frac{B'(\omega)}{B(\omega)} - \psi'(\omega) \tag{16.54}$$

Therefore

$$\left(\frac{\mathcal{T}S}{S}\right)_I = \frac{B'(\omega)}{B(\omega)} \qquad ; \qquad \left(\frac{\mathcal{T}S}{S}\right)_R = -\psi'(\omega) \tag{16.55}$$

and the mean time and duration expressed in terms of the spectrum are

$$\langle t\rangle = -\int \psi'(\omega)\,|\,S(\omega)\,|^2\,d\omega \tag{16.56}$$

$$\sigma_t^2 = \int \left(\frac{B'(\omega)}{B(\omega)}\right)^2 |\,B(\omega)\,|^2\,d\omega + \int \big(\psi'(\omega) + \langle\omega\rangle\big)^2\,|\,S(\omega)\,|^2\,d\omega \tag{16.57}$$

Chapter 17

Joint Representations for Arbitrary Variables

17.1 INTRODUCTION

We now generalize the methods we developed for time-frequency representations to obtain joint representations for arbitrary quantities. In generalizing we will use the time-frequency case for examples so that we may check by recovering the known results. In the next chapter we specialize to the study of scale.

17.2 MARGINALS

For two quantities, a and b, represented by the operators $\mathcal{A}$ and $\mathcal{B}$ we seek joint distributions, $P(a,b)$. In the previous chapter we saw that the densities of a and b are given by $|F(a)|^2$ and $|F(b)|^2$, where $F(a)$ and $F(b)$ are the a and b transforms, respectively. The joint density should therefore satisfy the marginal conditions

$$\int P(a,b)\,db = |F(a)|^2 \tag{17.1}$$

$$\int P(a,b)\,da = |F(b)|^2 \tag{17.2}$$

17.3 CHARACTERISTIC FUNCTION OPERATOR METHOD

As for the time-frequency case, we approach the problem of finding joint densities by the characteristic function method which was generalized for arbitrary operators by Scully and Cohen.[505] The characteristic function is

$$M(\alpha,\beta) = \langle e^{j\alpha a+j\beta b}\rangle = \iint e^{j\alpha a+j\beta b}P(a,b)da\,db \tag{17.3}$$

and the distribution is obtained from $M(\alpha,\beta)$ by Fourier inversion

$$P(a,b) = \frac{1}{4\pi^2}\iint M(\alpha,\beta)\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta \tag{17.4}$$

Since the characteristic function is the average of $e^{j\alpha a + j\beta b}$, we expect to calculate it by finding the average of a yet unspecified operator. We shall call that operator the characteristic function operator and denote it by $\mathcal{M}(\alpha,\beta)$. Accordingly we write

$$M(\alpha,\beta) = \langle\, \mathcal{M}(\alpha,\beta)\,\rangle = \int s^*(t)\, \mathcal{M}(\alpha,\beta)\, s(t)\, dt \tag{17.5}$$

There are many possibilities for $\mathcal{M}$. Among them are

$$\mathcal{M}(\alpha,\beta) \rightarrow e^{j\alpha\mathcal{A}+j\beta\mathcal{B}} \tag{17.6}$$

$$\rightarrow e^{j\alpha\mathcal{A}}\, e^{j\beta\mathcal{B}} \tag{17.7}$$

$$\rightarrow e^{j\beta\mathcal{B}}\, e^{j\alpha\mathcal{A}} \tag{17.8}$$

or more involved combinations like

$$\mathcal{M}(\alpha,\beta) = e^{j\alpha\mathcal{A}/2}\, e^{j\beta\mathcal{B}}\, e^{j\alpha\mathcal{A}/2} \tag{17.9}$$

Since there are an infinite number of possible orderings, there are an infinite number of distributions. It is this ordering ambiguity that gives rise to the general class of time-frequency distributions and we will see that it also gives rise to a general class for arbitrary variables.

The method for obtaining a distribution is as follows: Choose an ordering, that is a characteristic function operator, $\mathcal{M}(\alpha,\beta)$, calculate the characteristic function by way of Eq. (17.5), and then obtain the distribution using

$$P(a,b) = \frac{1}{4\pi^2}\iint \langle\, \mathcal{M}(\alpha,\beta)\,\rangle\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta \tag{17.10}$$

17.4 METHODS OF EVALUATION

We now consider the evaluation of the characteristic function and particularly of the expectation value of $e^{j\alpha\mathcal{A}+j\beta\mathcal{B}}$. Although this is only one form, it is an important one because very often other characteristic function operators may be expressed in terms of it. The evaluation of $\langle\, e^{j\alpha\mathcal{A}+j\beta\mathcal{B}}\,\rangle$ is a problem that appears in many fields and has a long history.[589] We discuss two methods for its simplification.

Method 1. Solve the eigenvalue problem,[505]

$$\{\, \alpha\mathcal{A} + \beta\mathcal{B}\,\}\; u(\lambda,t) = \lambda\; u(\lambda,t) \tag{17.11}$$

Since α and β are real numbers and $\mathcal{A}$ and $\mathcal{B}$ are Hermitian, the quantity $\alpha\mathcal{A} + \beta\mathcal{B}$ is also Hermitian. Therefore the solution to the eigenvalue problem gives rise to a complete set of eigenfunctions, the u's. Any signal can hence be expressed in terms of them by

$$s(t) = \int u(\lambda, t) F(\lambda)\, d\lambda \tag{17.12}$$

with the inverse transformation given by

$$F(\lambda) = \int u^*(\lambda, t)\, s(t)\, dt \tag{17.13}$$

Now consider

$$e^{j\alpha\mathcal{A}+j\beta\mathcal{B}} s(t) = e^{j\alpha\mathcal{A}+j\beta\mathcal{B}} \int u(\lambda, t) F(\lambda)\, d\lambda = \int e^{j\lambda}\, u(\lambda, t) F(\lambda)\, d\lambda \tag{17.14}$$

Substituting for $F(\lambda)$ we have

$$e^{j\alpha\mathcal{A}+j\beta\mathcal{B}}\, s(t) = \int G(t, t')\, s(t')\, dt' \tag{17.15}$$

where

$$G(t, t') = \int e^{j\lambda}\, u^*(\lambda, t')\, u(\lambda, t)\, d\lambda \tag{17.16}$$

This method works quite well in many cases but has the disadvantage that the eigenvalue problem must be solved.

Example 17.1: Time and Frequency.

The eigenvalue problem

$$(\theta\mathcal{T} + \tau\mathcal{W})\, u(\lambda, t) = \lambda\, u(\lambda, t) \tag{17.17}$$

in the time representation is

$$(\theta t - j\tau \frac{d}{dt})\, u(\lambda, t) = \lambda\, u(\lambda, t) \tag{17.18}$$

and the solutions, normalized to a delta function, are

$$u(\lambda, t) = \frac{1}{\sqrt{2\pi\tau}}\, e^{j(\lambda t - \theta t^2/2)/\tau} \tag{17.19}$$

Calculating G of Eq. (17.16)

$$G(t, t') = \delta(\tau + t - t')\, e^{j\theta\tau/2}\, e^{j\theta t} \tag{17.20}$$

and substituting in Eq. (17.15), we have

$$e^{j\theta\mathcal{T}+j\tau\mathcal{W}} s(t) = \int G(t, t')\, s(t')\, dt' = e^{j\theta\tau/2}\, e^{j\theta t} s(t + \tau) \tag{17.21}$$

And so the characteristic function is

$$M(\theta,\tau) = \int s^*(t)\, e^{j\theta\tau/2}\, e^{j\theta t} s(t+\tau)\, dt \tag{17.22}$$

$$= \int s^*(t-\tfrac{1}{2}\tau)\, e^{j\theta t} s(t+\tfrac{1}{2}\tau)\, dt \tag{17.23}$$

which is the characteristic function of the Wigner distribution.

Method 2. The second method seeks to directly simplify the operator $e^{j\alpha\mathcal{A}+j\beta\mathcal{B}}$. This, however, is one of the most difficult problems in mathematical analysis and simplification has thus far been achieved only for special cases. One such case is when both operators commute with their commutator,

$$[[\mathcal{A},\mathcal{B}],\mathcal{A}] = [[\mathcal{A},\mathcal{B}],\mathcal{B}] = 0 \tag{17.24}$$

in which case[589]

$$e^{j\alpha\mathcal{A}+j\beta\mathcal{B}} = e^{-\alpha\beta[\mathcal{A},\mathcal{B}]/2}\, e^{j\beta\mathcal{B}}\, e^{j\alpha\mathcal{A}} \tag{17.25}$$

$$= e^{\alpha\beta[\mathcal{A},\mathcal{B}]/2}\, e^{j\alpha\mathcal{A}}\, e^{j\beta\mathcal{B}} \tag{17.26}$$

This relation holds for time and frequency since their commutator is a number and hence commutes both with the time and frequency operator. We have already used this method in Chapter 10.

Another special case is when

$$[\mathcal{A},\mathcal{B}] = c_1 + c_2\,\mathcal{A} \tag{17.27}$$

in which case[143]

$$e^{j\alpha\mathcal{A}+j\beta\mathcal{B}} = e^{j\mu\alpha c_1/c_2}\, e^{j\alpha\mu\mathcal{A}}\, e^{j\beta\mathcal{B}}\, e^{j\alpha\mathcal{A}} \tag{17.28}$$

where

$$\mu = \frac{1}{j\beta c_2}\left[1-(1+j\beta c_2)\, e^{-j\beta c_2}\right] \tag{17.29}$$

This case will arise when we study the scale operator. These relations are proved in the appendix. Note that for $c_2 = 0$, Eq. (17.24) holds and we have

$$\mu = 0 \quad ; \quad \mu/c_2 = j\beta \qquad \text{for } c_2 \to 0 \tag{17.30}$$

and hence Eq. (17.28) becomes Eq. (17.25).

17.5 GENERAL CLASS FOR ARBITRARY VARIABLES

As with time-frequency case, instead of dealing with all possible orderings we can use the kernel method to obtain all possible representations. Choose a specific ordering to produce a specific characteristic function, $M(\alpha,\beta)$. Using it, form a new characteristic function by

$$M_{\text{new}}(\alpha,\beta) = \phi(\alpha,\beta)\, M(\alpha,\beta) \tag{17.31}$$

where ϕ is the kernel. The new distribution is

$$P_{\text{new}}(a,b) = \frac{1}{4\pi^2}\iint M_{\text{new}}(\alpha,\beta)\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta \tag{17.32}$$

$$= \frac{1}{4\pi^2}\iint \phi(\alpha,\beta)\, M(\alpha,\beta)\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta \tag{17.33}$$

and it will satisfy the marginals of the original distribution if the kernel satisfies

$$\phi(0,\beta) = \phi(\alpha,0) = 1 \tag{17.34}$$

Equation (17.33) is the general class for arbitrary variables. It can be written in the triple integral form by using Eq. (17.5) for M,

$$P(a,b) = \frac{1}{4\pi^2}\iint \langle\, \mathcal{M}(\alpha,\beta)\rangle\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta \tag{17.35}$$

$$= \frac{1}{4\pi^2}\iiint \phi(\alpha,\beta)\, s^*(t)\, \mathcal{M}(\alpha,\beta)\, s(t)\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta\, dt \tag{17.36}$$

17.6 TRANSFORMATION BETWEEN DISTRIBUTIONS

Suppose we have two distributions, P_1 and P_2, with corresponding kernels, ϕ_1, ϕ_2, and corresponding characteristic functions, M_1 and M_2. The respective characteristic functions are

$$M_1(\alpha,\beta) = \phi_1(\alpha,\beta)\, M(\alpha,\beta) \tag{17.37}$$

$$M_2(\alpha,\beta) = \phi_2(\alpha,\beta)\, M(\alpha,\beta) \tag{17.38}$$

which gives

$$M_1(\alpha,\beta) = \frac{\phi_1(\alpha,\beta)}{\phi_2(\alpha,\beta)}\, M_2(\alpha,\beta) \tag{17.39}$$

Now

$$P_1(a,b) = \frac{1}{4\pi^2}\iint M_1(\alpha,\beta)\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta \tag{17.40}$$

$$= \frac{1}{4\pi^2}\iint \frac{\phi_1(\alpha,\beta)}{\phi_2(\alpha,\beta)}\, M_2(\alpha,\beta)\, e^{-j\alpha a - j\beta b}\, d\alpha\, d\beta \tag{17.41}$$

and since

$$M_2(\alpha,\beta) = \iint e^{j\alpha a + j\beta b} P_2(a,b)\, da\, db \tag{17.42}$$

we have

$$P_1(a,b) = \frac{1}{4\pi^2} \iiiint \frac{\phi_1(\alpha,\beta)}{\phi_2(\alpha,\beta)} e^{j\alpha(a'-a)+j\beta(b'-b)} P_2(a',b')\, d\alpha\, d\beta\, da\, db' \tag{17.43}$$

This can be written as

$$P_1(a,b) = \iint g(a'-a, b'-b) P_2(a',b')\, da'\, db' \tag{17.44}$$

with

$$g(a,b) = \frac{1}{4\pi^2} \iint e^{j\alpha a + j\beta b} \frac{\phi_1(\alpha,\beta)}{\phi_2(\alpha,\beta)}\, d\alpha\, d\beta \tag{17.45}$$

17.7 LOCAL AUTOCORRELATION

In studying time-frequency distributions we saw that a fruitful approach was to generalize the relationship between the power spectrum and the autocorrelation function by defining a local autocorrelation function. A similar development holds for arbitrary variables. Consider the variable b. We write its density as

$$P(b) = \frac{1}{2\pi} \int R(\beta)\, e^{-j\beta b}\, d\beta \tag{17.46}$$

where $R(b)$ is the characteristic function of b (which we write as R instead of M as is the convention). We generalize by taking the joint distribution of a, b to be the local autocorrelation of b around some value a and write

$$P(a,b) = \frac{1}{2\pi} \int R_a(\beta)\, e^{-j\beta b}\, d\beta \tag{17.47}$$

Comparing Eq. (17.47) with Eq. (17.33), we find

$$R_a(\beta) = \frac{1}{2\pi} \int \phi(\alpha,\beta)\, M(\alpha,\beta)\, e^{-j\alpha a}\, d\alpha \tag{17.48}$$

In terms of the characteristic function operator this may be written as

$$R_a(\beta) = \frac{1}{2\pi} \iint \phi(\alpha,\beta)\, s^*(t)\, \mathcal{M}(\alpha,\beta)\, s(t)\, e^{-j\alpha a}\, d\alpha\, dt \tag{17.49}$$

17.8 INSTANTANEOUS VALUES

We rewrite Eq. (16.46) here, which expresses the bandwidth directly in terms of the signal,

$$\sigma_a^2 = \int \left(\frac{\mathcal{A}s}{s}\right)_I^2 |\, s(t)\,|^2\, dt + \int \left[\, \left(\frac{\mathcal{A}s}{s}\right)_R - \langle\, \mathcal{A}\,\rangle \right]^2 |\, s(t)\,|^2\, dt \qquad (17.50)$$

Comparing it to the general result that connects the global and local standard deviation, Eq. (4.71),

$$\sigma_a^2 = \int \sigma_{a|b}^2\, P(b)\, db + \int (\langle\, a\,\rangle_b - \langle\, a\,\rangle)^2\, P(b)\, db \qquad (17.51)$$

we conclude that the instantaneous value, a_t, and its conditional standard deviation, $\sigma_{a|t}$, are given by

$$a_t = \langle\, a\,\rangle_t = \left(\frac{\mathcal{A}s}{s}\right)_R \qquad ; \qquad \sigma_{a|t}^2 = \left(\frac{\mathcal{A}s}{s}\right)_I^2 \qquad (17.52)$$

In an eigenstate we expect the value of the physical quantity to be the fixed for all time. For an eigenstate, $\mathcal{A}\,u(a,t) = a\,u(a,t)$ and therefore

$$\frac{\mathcal{A}\,u(a,t)}{u(a,t)} = a \qquad (17.53)$$

Because the operator is Hermitian, the eigenvalue, a, is real, and Eq. (17.52) gives

$$a_t = a \qquad ; \qquad \sigma_{a|t}^2 = 0 \qquad (17.54)$$

We see that in an eigenstate the instantaneous value is equal to the global average and the local standard deviation is zero for all time.

Relation Between Global and Instantaneous Quantities. We expect the average value to be the average of the local value,

$$\langle\, \mathcal{A}\,\rangle = \int \langle\, \mathcal{A}\,\rangle_t\, |\, s(t)\,|^2\, dt = \int \left(\frac{\mathcal{A}s}{s}\right)_R |\, s(t)\,|^2\, dt \qquad (17.55)$$

This is seen to be the case since

$$\int s(t)^*\, \mathcal{A}s(t)\, dt = \int \frac{\mathcal{A}s}{s}\, |\, s(t)\,|^2\, dt \qquad (17.56)$$

$$= \int \left[\, \left(\frac{\mathcal{A}s}{s}\right)_R + j \left(\frac{\mathcal{A}s}{s}\right)_I \right] |\, s(t)\,|^2\, dt \qquad (17.57)$$

$$= \int \left(\frac{\mathcal{A}s}{s}\right)_R |\, s(t)\,|^2\, dt \qquad (17.58)$$

where the second term in Eq. (17.57) is zero because we know that the average value of a Hermitian operator is real.

17.9 LOCAL VALUES FOR ARBITRARY VARIABLE PAIRS

We now consider the more general question of the conditional value of a quantity for a fixed value of another arbitrary quantity. We shall use the notation $\langle a \rangle_b$ or a_b for the conditional value of a for a given value of b. In Section 16.4 we showed that the bandwidth of a can be written in the following way:

$$\sigma_a^2 = \int \left(\frac{\mathcal{A}F}{F}\right)_I^2 |F(b)|^2\, db + \int \left[\left(\frac{\mathcal{A}F}{F}\right)_R - \langle \mathcal{A} \rangle\right]^2 |F(b)|^2\, db \qquad (17.59)$$

where $F(b)$ is the signal in the b representation. Comparing with Eq. (17.51) gives

$$a_b = \langle a \rangle_b = \left(\frac{\mathcal{A}(b)F(b)}{F(b)}\right)_R \qquad (17.60)$$

$$\sigma_{a|b}^2 = \left(\frac{\mathcal{A}(b)F(b)}{F(b)}\right)_I^2 \qquad (17.61)$$

Again, we expect that the global average be the average of the conditional value,

$$\langle a \rangle = \int \langle \mathcal{A} \rangle_b |F(b)|^2\, db = \int \left(\frac{\mathcal{A}F}{F}\right)_R |F(b)|^2\, db \qquad (17.62)$$

The proof is identical to that given in the previous section where b was time.

Example 17.2: Group Delay.

For a we take time and for b we take frequency. Hence $F(b)$ is the spectrum $S(\omega)$ Writing the spectrum as

$$S(\omega) = B(\omega)\, e^{j\psi(\omega)} \qquad (17.63)$$

we have

$$\frac{\mathcal{T}S}{S} = -\frac{1}{S}\frac{1}{j}\frac{d}{d\omega}S = -\frac{1}{j}\frac{B'(\omega)}{B(\omega)} - \psi'(\omega) \qquad (17.64)$$

Therefore

$$\langle t \rangle_\omega = -\psi'(\omega) \qquad (17.65)$$

which is the group delay. Its standard deviation is, according to Eq. (17.61),

$$\sigma_{t|\omega}^2 = \left(\frac{B'(\omega)}{B(\omega)}\right)^2 \qquad (17.66)$$

17.10 THE COVARIANCE

In Chapter 1 we showed how the covariance can be defined for the case of time and frequency. In Section 15.5 we gave an expression for the covariance for arbitrary variables but did not justify that definition. We do so now. Let us place ourselves in the b representation. The joint moment is then $\langle ba_b \rangle$ since a_b is the instantaneous value of a for a given b. The covariance is

$$\text{Cov}_{ab} = \langle ba_b \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \tag{17.67}$$

$$= \left\langle b \left(\frac{\mathcal{A}(b)F(b)}{F(b)} \right)_R \right\rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \tag{17.68}$$

We could have equally well put ourselves in the a representation, in which case

$$\text{Cov}_{ab} = \langle ab_a \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \tag{17.69}$$

$$= \left\langle a \left(\frac{\mathcal{B}(a)F(a)}{F(a)} \right)_R \right\rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \tag{17.70}$$

These two definitions are identical because

$$\left\langle b \left(\frac{\mathcal{A}(b)F(b)}{F(b)} \right)_R \right\rangle = \left\langle a \left(\frac{\mathcal{B}(a)F(a)}{F(a)} \right)_R \right\rangle \tag{17.71}$$

We prove this by showing first that the definition given in Section 15.5 for the covariance is the same as given above, namely

$$\text{Cov}_{ab} = \tfrac{1}{2} \langle (\mathcal{AB} + \mathcal{BA}) \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \tag{17.72}$$

Consider

$$\left\langle a \left(\frac{\mathcal{B}F(a)}{F(a)} \right)_R \right\rangle = \frac{1}{2} \left\langle a \left(\frac{\mathcal{B}F(a)}{F(a)} \right) \right\rangle + \frac{1}{2} \left\langle a \left(\frac{\mathcal{B}F(a)}{F(a)} \right)^* \right\rangle \tag{17.73}$$

$$= \frac{1}{2} \left[\int F^* \mathcal{A} \mathcal{B} F + \int a F \mathcal{B}^\dagger F^* \right] \tag{17.74}$$

$$= \tfrac{1}{2} \langle \mathcal{AB} + \mathcal{BA} \rangle \tag{17.75}$$

If we start with the expression given by the left hand side of Eq. (17.71), the identical derivation leads to the same result. Hence we conclude that the two expressions for the covariance, Eqs. (17.68) and (17.70), are identical.

17.11 GENERALIZATION OF THE SHORT-TIME FOURIER TRANSFORM

We now generalize the short-time Fourier transform to arbitrary variables. First consider the case where the two variables are time and b and subsequently generalize to two arbitrary variables.

The Short-Time b Transform. We focus in on a particular time by the usual procedure of defining a modified signal according to

$$s_t(\tau) = s(\tau)\, h(\tau - t) \tag{17.76}$$

where t is the fixed time of interest, τ is the running time, and $h(t)$ is the windowing function. Instead of taking the Fourier transform of this modified signal, we take the b transform,

$$F_t(b) = \int s(\tau)\, h(\tau - t)\, u^*(b, \tau)\, d\tau \tag{17.77}$$

In analogy with the short-time Fourier transform this shall be called the short-time b transform. The joint time – b density is then

$$P(t, b) = |\, F_t(b)\,|^2 = \left| \int s(\tau)\, h(\tau - t)\, u^*(b, \tau)\, d\tau \right|^2 \tag{17.78}$$

The Short–a b Transform. We first recall that the signals in the a and b representations, are given by

$$F(a) = \int u^*(a, t)\, s(t)\, dt \tag{17.79}$$

$$F(b) = \int u^*(b, t)\, s(t)\, dt \tag{17.80}$$

These two equations are the transformation from the time representation to the a and b representations. However, we can transform directly from the a to the b representation,

$$F(b) = \int \alpha(b, a)\, F(a)\, da \tag{17.81}$$

where $\alpha(b, a)$ is called the transformation matrix. It is given by

$$\alpha(b, a) = \int u^*(b, t)\, u(a, t)\, dt \tag{17.82}$$

which can be verified by substituting it into Eq. (17.81). Now we window the signal in the a representation to obtain a modified signal,

$$F_a(a') = F(a')\, h(a' - a) \tag{17.83}$$

where a' is the running variable, and a is the fixed value of interest. Taking the b transform of the modified signal, we have

$$F_a(b) = \int \alpha(b,a')\, h(a'-a) F(a')\, da' \tag{17.84}$$

and it is reasonable to call this the short-a b transform. The joint density is therefore

$$P(a,b) = |\, F_a(b)\,|^2 = \left| \int \alpha(b,a')\, h(a'-a) F(a')\, da' \right|^2 \tag{17.85}$$

which is a joint distribution for the variables a and b analogous to the spectrogram.

17.12 UNITARY TRANSFORMATION

Another method for obtaining joint distributions for variables other than time and frequency is to start with the general class of time-frequency distributions,

$$P(t,\omega) = \frac{1}{4\pi^2} \iiint s^*(u)\, e^{j\alpha\mathcal{T}+j\beta\mathcal{W}} s(u)\, \phi(\theta,\tau)\, e^{-j\alpha t - j\beta\omega}\, du\, d\alpha\, d\beta \tag{17.86}$$

and for the signal, $s(t)$, use a transformed signal. The transformed signal is obtained by a unitary transformation, so chosen to be in the representation of one of the variables, a, we are interested (see section 15.3). The other variable, b, is as yet unspecified. In the general time-frequency representation, Eq. (9.1), replace $s(t)$ by $\mathcal{U}s(t)$ and replace the variables t, ω by a, b:

$$P(a,b) = \frac{1}{4\pi^2} \iiint \mathcal{U}s^*(u)\, e^{j\alpha\mathcal{T}+j\beta\mathcal{W}} \mathcal{U}s(u)\, \phi(\theta,\tau)\, e^{-j\alpha a - j\beta b}\, du\, d\alpha\, d\beta \tag{17.87}$$

This approach was developed by Baraniuk and Jones.[49]

It can be directly verified that the marginals for this distribution are

$$\int P(a,b)\, db = |\, \mathcal{U}s(a)\,|^2 \tag{17.88}$$

$$\int P(a,b)\, da = |\, S_{\mathcal{U}s}(b)\,|^2 \tag{17.89}$$

where $S_{\mathcal{U}s}(b)$ is the Fourier transform of the transformed signal, $\mathcal{U}s(a)$. Let us suppose our interest was to have a distribution where one of the marginals is $|\,\mathcal{U}s(a)\,|^2$. We are not free to choose the other – it is forced and not at our disposal. The advantage of the Baraniuk-Jones procedure is that we do not have to calculate new distributions, but can use the already calculated time-frequency distribution. It cannot be applied to two arbitrary variables because both variables are not under our control. We now ask, what are the two variables or operators that generate Eq. (17.87) for the joint distribution?[52]

Similarity and Unitary Transformation. To see the relation between the form given by Eq. (17.87) and the general procedure developed in the previous sections we first have to define the concept of a similarity and unitary transformation of operators. If we have an operator and form a new operator, $\mathcal{A}'$, according to

$$\mathcal{A}' = \mathcal{U}^{-1}\mathcal{A}\mathcal{U} \qquad [\text{ similarity transformation }] \tag{17.90}$$

then that is called a similarity transformation. If, in addition, the operator is unitary, then $\mathcal{U}^{-1} = \mathcal{U}^{\dagger}$, in which case

$$\mathcal{A}' = \mathcal{U}^{\dagger}\mathcal{A}\mathcal{U} \qquad [\text{ unitary transformation }] \tag{17.91}$$

and the transformation is called a unitary transformation. The reason for defining a unitary operator transformation in this way is that by doing so we obtain the important result that averages retain their form in both representations. In particular, if we call s' the transformed signal

$$s' = \mathcal{U}^{\dagger}s = \mathcal{U}^{-1}s \tag{17.92}$$

then

$$\langle a' \rangle = \int s'^{*}(t)\,\mathcal{A}'\,s'(t)\,dt \tag{17.93}$$

$$= \int \{\mathcal{U}^{-1}s(t)\}^{*}\,\mathcal{U}^{\dagger}\,\mathcal{A}\mathcal{U}\mathcal{U}^{-1}s(t)\,dt \tag{17.94}$$

$$= \int s^{*}(t)\,\mathcal{A}\,s(t)\,dt \tag{17.95}$$

$$= \langle a \rangle \tag{17.96}$$

An important property of a unitary transformation is that the eigenvalues of operators remain the same. Also, Hermiticity is retained; that is, $\mathcal{A}'$ is Hermitian if $\mathcal{A}$ is. Another important property of a unitary transformation is that it leaves algebraic equations functionally the same. In particular, it leaves the commutator invariant. If

$$[\mathcal{A},\mathcal{B}] = \mathcal{C} \tag{17.97}$$

then

$$[\mathcal{A}',\mathcal{B}'] = \mathcal{C}' \tag{17.98}$$

To see this, multiply Eq. (17.97) on the left by $\mathcal{U}^{\dagger}$ and on the right by $\mathcal{U}$

$$\mathcal{U}^{\dagger}[\mathcal{A},\mathcal{B}]\mathcal{U} = \mathcal{U}^{\dagger}\mathcal{C}\mathcal{U} = \mathcal{C}' \tag{17.99}$$

But

$$\mathcal{U}^{\dagger}[\mathcal{A},\mathcal{B}]\mathcal{U} = \mathcal{U}^{\dagger}(\mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A})\mathcal{U} \tag{17.100}$$

$$= \mathcal{U}^{\dagger}(\mathcal{A}\mathcal{U}\mathcal{U}^{-1}\mathcal{B} - \mathcal{B}\mathcal{U}\mathcal{U}^{-1}\mathcal{A})\mathcal{U} \tag{17.101}$$

$$= \mathcal{U}^{\dagger}\mathcal{A}\mathcal{U}\mathcal{U}^{-1}\mathcal{B}\mathcal{U} - \mathcal{U}^{\dagger}\mathcal{B}\mathcal{U}\mathcal{U}^{-1}\mathcal{A}\mathcal{U} \tag{17.102}$$

$$= [\mathcal{A}',\mathcal{B}'] \tag{17.103}$$

and hence Eq. (17.98) follows.

We are now in a position to see which two operators generate the joint distribution, Eq. (17.87). They are operators that are related to the time and frequency operator by a unitary transformation,[52]

$$\mathcal{A} = \mathcal{U}^\dagger \mathcal{T} \mathcal{U} \quad ; \quad \mathcal{B} = \mathcal{U}^\dagger \mathcal{W} \mathcal{U} \tag{17.104}$$

To prove this we calculate the characteristic function and distribution for $\mathcal{A}$ and $\mathcal{B}$ as per the general procedure given by Eqs. (17.3) and (17.4),

$$M(\alpha,\beta) = \phi(\alpha,\beta) \int s^*(t)\, e^{j\beta\mathcal{A}+j\beta\mathcal{B}}\, s(t)\, dt \tag{17.105}$$

$$= \phi(\alpha,\beta) \int s^*(t)\, e^{j\alpha\beta/2}\, e^{j\beta\mathcal{A}}\, e^{j\beta\mathcal{B}}\, s(t)\, dt \tag{17.106}$$

$$= \phi(\alpha,\beta) \int s^*(t)\, e^{j\alpha\beta/2}\, (\mathcal{U}^{-1}\, e^{j\alpha\mathcal{T}}\mathcal{U})\, (\mathcal{U}^{-1}\, e^{j\beta\mathcal{W}}\mathcal{U})\, s(t)\, dt \tag{17.107}$$

$$= \phi(\alpha,\beta) \int \mathcal{U}s^*(t)\, e^{j\alpha\mathcal{T}+j\beta\mathcal{W}}\, \mathcal{U}s(t)\, dt \tag{17.108}$$

where in going from step (17.107) to (17.108) we have used the fact that

$$e^{j\alpha\mathcal{A}} = \mathcal{U}^\dagger\, e^{j\alpha\mathcal{T}}\, \mathcal{U} \quad ; \quad e^{j\beta\mathcal{B}} = \mathcal{U}^\dagger\, e^{j\beta\mathcal{W}}\, \mathcal{U} \tag{17.109}$$

which can be readily proven. The distribution is the double Fourier transform of M, which gives Eq. (17.87).

Therefore for the case where the two new variables are connected to time and frequency by a similarity transformation, the general class of distributions for a and b is obtained from the general class of time-frequency distributions by substituting the transformed signal for $s(t)$. This simplification can be used whenever applicable.

Generalization.[52] This result can be further generalized. Suppose we have two arbitrary variables and use the procedure we have developed to obtain the general class, that is, Eq. (17.10), which we repeat here,

$$P(a,b) = \frac{1}{4\pi^2} \iiint s^*(t)\, \mathcal{M}(\alpha,\beta)\, s(t) e^{-j\alpha a' - j\beta b'}\, dt\, d\alpha\, d\beta \tag{17.110}$$

Now suppose that we have two new variables which are related to $\mathcal{A}, \mathcal{B}$ by way of

$$\mathcal{A}' = \mathcal{U}^\dagger\, \mathcal{A}\, \mathcal{U} \quad ; \quad \mathcal{B}' = \mathcal{U}^\dagger\, \mathcal{B}\, \mathcal{U} \tag{17.111}$$

Then we do not have to recalculate their general class. It can be obtained from Eq. (17.110) by substituting the transformed signal for $s(t)$,

$$P(a',b') = \frac{1}{4\pi^2} \iiint \mathcal{U}s^*(t)\, \mathcal{M}(\alpha,\beta)\, \mathcal{U}\, s(t)\, e^{-j\alpha a' - j\beta b'}\, dt\, d\alpha\, d\beta \qquad (17.112)$$

The proof is the same as in the time-frequency case. This shows that variables connected by a unitary transformation have distributions that are functionally identical.

17.13 INVERSE FREQUENCY

It is sometimes of interest to obtain a distribution of time and inverse frequency.

Method 1. The straightforward way to obtain such a distribution is to use the general class of time-frequency distributions and the method of Section 4.7 to transform distributions. The distribution in time and inverse frequency, $P(t,r)$, is given by

$$P(t,r)\ dt\, dr = P(t,\omega)\ dt\, d\omega \qquad (17.113)$$

where $P(t,\omega)$ is a time-frequency distribution.

Using

$$r = \omega_0/\omega \qquad ; \qquad dr = -\omega_0/\omega^2\, d\omega = -\frac{r^2}{\omega_0}\, d\omega \qquad (17.114)$$

we obtain

$$P(t,r) = \frac{\omega_0}{r^2}\, P(t,\omega_0/r) \qquad (17.115)$$

where the P on the right hand side is the time-frequency distribution. Any time-frequency distribution can be used for $P(t,\omega)$. If we use the Wigner distribution

$$P(t,r) = \frac{\omega_0}{r^2}\, W(t,\omega_0/r) \qquad (17.116)$$

$$= \frac{1}{2\pi}\frac{\omega_0}{r^2} \int s^*(t-\tfrac{1}{2}\tau)\, e^{-j\tau\omega_0/r}\, s(t+\tfrac{1}{2}\tau)\, d\tau \qquad (17.117)$$

$$= \frac{1}{2\pi}\frac{\omega_0}{r^2} \int S^*(\omega_0/r-\theta/2)\, e^{-jt\theta}\, S(\omega_0/r-\theta/2)\, d\theta \qquad (17.118)$$

To obtain the general class of time-inverse frequency distributions we can use the general class of time-frequency distributions, Eq. (9.1), together with Eq. (17.115). Doing so yields

$$P(t,r) = \frac{1}{4\pi^2}\frac{\omega_0}{r^2} \iiint e^{-j\theta t - j\tau\omega_0/r + j\theta u}\, \phi(\theta,\tau)\, s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, du\, d\tau\, d\theta \qquad (17.119)$$

In terms of the spectrum it is

$$P(t,r) = \frac{1}{4\pi^2}\frac{\omega_0}{r^2}\iiint e^{-j\theta t - j\tau\omega_0/r + j\tau u}\,\phi(\theta,\tau)S^*(u+\tfrac{1}{2}\theta)S(u-\tfrac{1}{2}\theta)\,du\,d\tau\,d\theta \tag{17.120}$$

The marginals of these distributions are the marginals of time and inverse frequency

$$\int P(t,r)\,dr = |\,s(t)\,|^2 \qquad ; \qquad \int P(t,r)\,dt = \frac{\omega_0}{r^2}\,|\,S(\omega_0/r)\,|^2 \tag{17.121}$$

Method 2. Alternatively, although in this case much harder, we can use operator methods. Here we do not need to, since we are making a transformation of ordinary functions; nevertheless, it is of some interest to do so. Take

$$M(\theta,\kappa) = \int S^*(\omega)\,e^{j\theta\mathcal{T}+j\kappa\mathcal{R}}\,S(\omega)\,d\omega \tag{17.122}$$

for the characteristic function. We first address the simplification of $e^{j\theta\mathcal{T}+j\kappa\mathcal{R}}S(\omega)$. Using the method given by Eqs. (17.11)-(17.16) we solve the eigenvalue problem

$$[\,\theta\mathcal{T}+\kappa\mathcal{R}\,]\,u(\lambda,\omega) = \lambda\,u(\lambda,\omega) \tag{17.123}$$

in the frequency representation

$$\left[\,j\theta\frac{d}{d\omega}+\kappa\frac{\omega_0}{\omega}\,\right]u(\lambda,\omega) = \lambda\,u(\lambda,\omega) \tag{17.124}$$

The solution is

$$u(\lambda,\omega) = \frac{1}{\sqrt{2\pi\theta}}\,e^{-j[\,\lambda\omega-\omega_0\kappa\ln|\omega|\,]/\theta} \tag{17.125}$$

which gives for the G of Eq. (17.16)

$$G(\omega,\omega') = \int e^{j\lambda}u(\lambda,\omega)\,u^*(\lambda,\omega')\,d\lambda = e^{j\omega_0\kappa\ln(\omega/\omega')/\theta}\delta(\theta+\omega'-\omega) \tag{17.126}$$

and therefore

$$e^{j\theta\mathcal{T}+j\kappa\mathcal{R}}S(\omega) = \int e^{j\omega_0\kappa\ln(\omega/\omega')/\theta}\,\delta(\theta+\omega'-\omega)\,S(\omega')\,d\omega' \tag{17.127}$$

$$= S(\omega-\theta)\,\exp\left[j\frac{\omega_0\kappa}{\theta}\ln\frac{\omega}{\omega-\theta}\right] \tag{17.128}$$

Using this result we calculate

$$M(\theta,\kappa) = \int S^*(\omega)\,S(\omega-\theta)\,\exp\left[j\frac{\omega_0\kappa}{\theta}\ln\frac{\omega}{\omega-\theta}\right]d\omega \tag{17.129}$$

$$= \int S^*(\omega+\tfrac{1}{2}\theta)\,S(\omega-\tfrac{1}{2}\theta)\,\exp\left[j\frac{\omega_0\kappa}{\theta}\ln\frac{\omega+\frac{1}{2}\theta}{\omega-\frac{1}{2}\theta}\right]d\omega \tag{17.130}$$

The distribution is therefore

$$P(t,\tau) = \frac{1}{4\pi^2}\iiint S^*(\omega+\tfrac{1}{2}\theta)\,S(\omega-\tfrac{1}{2}\theta)\,\exp\left[j\frac{\omega_0\kappa}{\theta}\ln\frac{\omega+\frac{1}{2}\theta}{\omega-\frac{1}{2}\theta}\right]\,e^{-j\theta t-j\kappa\tau}\,d\theta\,d\kappa\,d\omega \tag{17.131}$$

which after considerable algebra leads to

$$P(t,\tau) = \frac{1}{8\pi\tau^3}\int\left(\frac{\omega_0 u}{\sinh u/2}\right)^2 e^{-j\omega_0 u t/\tau}S^*\left(\frac{\omega_0 u}{2\tau}\frac{e^{u/2}}{\sinh u/2}\right)S\left(\frac{\omega_0 u}{2\tau}\frac{e^{-u/2}}{\sinh u/2}\right)du \tag{17.132}$$

This distribution satisfies the marginals of time and inverse frequency, Eqs. (17.121).

17.14 APPENDIX

The simplification of $e^{\mathcal{A}+\mathcal{B}}$ is generally difficult but for the case where[1]

$$[\mathcal{A},\mathcal{B}] = \beta + \alpha\mathcal{A} \tag{17.133}$$

simplification is possible. This encompasses the time-frequency and time-scale case.

If one tries to expand $e^{\mathcal{A}+\mathcal{B}}$ and attempts to use the commutation relations to rearrange each term, a very significant mess ensues. One of the well known tricks to simplify this type of expression is to introduce a parameter λ and write an equation similar to the one we are trying to simplify. In particular, we write

$$f(\lambda) = e^{\lambda\mathcal{B}}\,e^{\lambda\mathcal{A}}\,e^{-\lambda(\mathcal{A}+\mathcal{B})} \tag{17.134}$$

The idea is to obtain a differential equation for $f(\lambda)$ and solve it. The solution, hopefully, will be a simpler form than what we started out with.

Before proceeding we establish some operator relations. Taking the commutator of $\mathcal{A}$ with $[\mathcal{A},\mathcal{B}]$ we have

$$[\mathcal{A},[\mathcal{A},\mathcal{B}]] = [\mathcal{A},\alpha\mathcal{A}] = 0 \tag{17.135}$$

Also,

$$[\mathcal{B},[\mathcal{A},\mathcal{B}]] = \alpha[\mathcal{B},\mathcal{A}] = -\alpha\beta - \alpha^2\mathcal{A} \tag{17.136}$$

There is a well known relation[589] that is very useful for simplification of operator identities,

$$e^{\lambda\mathcal{A}}\mathcal{B}e^{-\lambda\mathcal{A}} = \mathcal{B} + \lambda[\mathcal{A},\mathcal{B}] + \frac{\lambda^2}{2!}[\mathcal{A},[\mathcal{A},\mathcal{B}]] + \frac{\lambda^3}{3!}[\mathcal{A},[\mathcal{A},[\mathcal{A},\mathcal{B}]]] + \cdots \tag{17.137}$$

[1] For the sake of neatness we use here the notation $[\mathcal{A},\mathcal{B}] = \beta + \alpha\mathcal{A}$ rather than $[\mathcal{A},\mathcal{B}] = c_1 + c_2\mathcal{A}$ used in the text. The α, β used in this appendix should not be confused with their usage in the text.

Using Eqs. (17.135) and (17.136) we have

$$e^{\lambda A} B e^{-\lambda A} = B + \lambda\beta + \lambda\alpha A \tag{17.138}$$

$$e^{\lambda B} A e^{-\lambda B} = e^{-\lambda\alpha} A + \frac{\beta}{\alpha}(e^{-\lambda\alpha} - 1) \tag{17.139}$$

Differentiating Eq. (17.134) with respect to λ and using the relations just given we establish that

$$\frac{df(\lambda)}{d\lambda} = -\lambda(\beta + \alpha A)\, e^{-\lambda\alpha} f \tag{17.140}$$

Solving, we have

$$f(\lambda) = \exp[e^{-\lambda\alpha}(\lambda + 1/\alpha)(\beta + \alpha A)/\alpha + k] \tag{17.141}$$

where k is a constant of integration which is found by imposing the requirement that $f(0) = 1$. This gives

$$k = -\frac{(\alpha A + \beta)}{\alpha^2} \tag{17.142}$$

and

$$f(\lambda) = e^{-\mu\lambda\beta/\alpha}\, e^{-\mu\lambda A} \quad ; \quad \mu = \frac{1}{\lambda\alpha}\{1 - (1 + \lambda\alpha)\, e^{-\lambda\alpha}\} \tag{17.143}$$

Hence

$$e^{\lambda B}\, e^{\lambda A}\, e^{-\lambda(A+B)} = e^{-\mu\lambda\beta/\alpha}\, e^{-\mu\lambda A} \tag{17.144}$$

or

$$e^{\lambda(A+B)} = e^{\mu\lambda\beta/\alpha}\, e^{\mu\lambda A}\, e^{\lambda B}\, e^{\lambda A} \tag{17.145}$$

Taking $\lambda = 1$ gives

$$e^{A+B} = e^{\mu\beta/\alpha}\, e^{\mu A}\, e^{B}\, e^{A} \quad ; \quad \mu = \frac{1}{\alpha}\{1 - (1 + \alpha)\, e^{-\alpha}\} \tag{17.146}$$

Chapter 18

Scale

18.1 INTRODUCTION

In this chapter we develop, using the general methods of the previous chapters, the basic properties of scale. Scale is considered to be a physical attribute just like frequency. The frequency transform, the Fourier transform, allows us to ascertain the frequency content of a signal. Similarly, we will develop the scale transform, which allows the determination of the scale content. The first objective is to obtain the operator that represents scale. The scale transform is obtained by solving the eigenvalue problem for the operator. Having the scale operator and transform allows us to obtain average scale, scale bandwidth, instantaneous scale, and other properties in complete analogy to frequency. Scaling can be applied to any variable. For concreteness we develop scale for time functions and subsequently consider the scaling of other variables, such as frequency. In the next chapter we develop joint representations of scale and time and frequency.

18.2 THE SCALE AND COMPRESSION OPERATOR

For the scale operator, C, we take the Hermitian operator

$$C = \tfrac{1}{2}(TW + WT) = \tfrac{1}{2}[T, W]_+ \tag{18.1}$$

That this operator gives results that conform to our intuitive sense of scaling will be seen as we develop its properties. In the time representation C is given by

$$C = \frac{1}{2j}\left(t\frac{d}{dt} + \frac{d}{dt}t\right) \tag{18.2}$$

Using the commutation relation for time and frequency, $(\mathcal{TW} - \mathcal{WT}) = j$, it can be written in the following alternative forms,

$$\mathcal{C} = \mathcal{TW} - \tfrac{1}{2}j = \mathcal{WT} + \tfrac{1}{2}j \tag{18.3}$$

Compression Operator. In Chapter 1 we showed that from the frequency operator, $\mathcal{W}$, we obtain the translation operator by forming $e^{j\tau\mathcal{W}}$,

$$e^{j\tau\mathcal{W}}\, s(t) = s(t+\tau) \tag{18.4}$$

We expect a similar result for the scale operator in that it will define a compression operator. In fact,

$$e^{j\sigma\mathcal{C}}\, s(t) = e^{\sigma/2}\, s(e^{\sigma/2}t) \tag{18.5}$$

That is, $e^{j\sigma\mathcal{C}}$ compresses functions with the factor $e^{\sigma/2}$. If we take $\ln\sigma$ for σ we have

$$e^{j\ln\sigma\mathcal{C}}\, s(t) = \sqrt{\sigma}\, s(\sigma t) \tag{18.6}$$

The significance of the factors, $e^{\sigma/2}$ or $\sqrt{\sigma}$, is that they preserve normalization. They entered in an automatic way because the operator $e^{j\sigma\mathcal{C}}$ is unitary.

Equation. (18.6) is basic and we give two proofs. The first proof relies strictly on the algebraic properties of the operator; the second depends on the eigenfunctions of scale and is presented after we obtain them. To show Eq. (18.6) consider first the action of $\mathcal{C}$ on t^n,

$$\mathcal{C}t^n = \left(\mathcal{TW} - \tfrac{1}{2}j\right)t^n = -j\left(t\frac{d}{dt} + \tfrac{1}{2}j\right)t^n = -jtnt^{n-1} - \tfrac{1}{2}jt^n \tag{18.7}$$

and hence

$$\mathcal{C}t^n = -j(n+\tfrac{1}{2})t^n \tag{18.8}$$

By repeated action we have

$$\mathcal{C}^k t^n = (-j)^k(n+\tfrac{1}{2})^k t^n \tag{18.9}$$

Now

$$e^{j\sigma\mathcal{C}}t^n = \sum_{k=0}^{\infty}\frac{(j\sigma)^k}{k!}\mathcal{C}^k t^n = \sum_{k=0}^{\infty}\frac{(j\sigma)^k}{k!}(-j)^k(n+\tfrac{1}{2})^k t^n = e^{\sigma(n+1/2)}t^n \tag{18.10}$$

To obtain the action of $e^{j\sigma\mathcal{C}}$ on an arbitrary function, $s(t)$, we expand the function in a power series

$$e^{j\sigma\mathcal{C}}\, s(t) = e^{j\sigma\mathcal{C}}\sum_{n=0}^{\infty}a_n t^n = e^{\sigma/2}\sum_{n=0}^{\infty}a_n\, e^{\sigma n}t^n = e^{\sigma/2}s(e^{\sigma}t) \tag{18.11}$$

which is the desired result.

Basic Commutation Relations. Unlike the commutator for time and frequency the commutator of scale with time or frequency does not produce a number. The commutation relations are

$$[\mathcal{T},\mathcal{C}] = j\mathcal{T} \quad ; \quad [\mathcal{W},\mathcal{C}] = -j\mathcal{W} \tag{18.12}$$

which can be proven directly by using the definition of $\mathcal{C}$ and $[\mathcal{T},\mathcal{W}] = j$. These relations are fundamental and determine the uncertainty principle for scale, as we discuss in Section 18.9. Furthermore,

$$[\mathcal{T},[\mathcal{T},\mathcal{C}]] = 0 \quad ; \quad [\mathcal{W},[\mathcal{W},\mathcal{C}]] = 0 \tag{18.13}$$

$$[\mathcal{C},[\mathcal{T},\mathcal{C}]] = \mathcal{T} \quad ; \quad [\mathcal{C},[\mathcal{W},\mathcal{C}]] = \mathcal{W} \tag{18.14}$$

All these relations are important for the simplification of expressions involving the scale operator.

18.3 THE SCALE EIGENFUNCTIONS

To obtain the scale transform we solve the eigenvalue problem for the scale operator. We shall use c and $\gamma(c,t)$ to indicate the eigenvalues and eigenfunctions, respectively. The eigenvalue problem is

$$\mathcal{C}\,\gamma(c,t) = c\,\gamma(c,t) \tag{18.15}$$

Explicitly, using Eq. (18.3),

$$-jt\frac{d\gamma(c,t)}{dt} - \tfrac{1}{2}j\,\gamma(c,t) = c\,\gamma(c,t) \tag{18.16}$$

The solutions, normalized to a delta function, are

$$\gamma(c,t) = \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln t}}{\sqrt{t}} \quad , \quad t \geq 0 \tag{18.17}$$

It is possible to obtain this solution only if time is positive, and thus for scaling we must have an origin in time. This is reasonable since to scale is to enlarge or reduce. That means we must multiply time t by a number from zero to one for enlargement and a number bigger than one for reduction. At no time do we multiply by a negative number. Also, the appearance of the logarithm is reasonable because it has the effect of putting on an equal footing the range of zero to one (enlargement) with the range of one to infinity (reduction).

Completeness and Orthogonality. Since the scale operator is Hermitian we expect that the eigenfunctions are complete and orthogonal,

$$\int_0^\infty \gamma^*(c',t)\,\gamma(c,t)\,dt = \delta(c-c') \tag{18.18}$$

$$\int \gamma^*(c,t')\,\gamma(c,t)\,dc = \delta(t-t') \qquad t,t' \geq 0 \tag{18.19}$$

These relations are proven in the appendix.

Properties of the Scale Eigenfunctions. In Table 18.1 we list the main algebraic properties of the eigenfunctions and contrast them to the frequency eigenfunctions,

$$u(\omega,t) = \frac{1}{\sqrt{2\pi}}\,e^{j\omega t} \tag{18.20}$$

Notice that with respect to the time variable, multiplication/division in the scale case corresponds to addition/subtraction in the frequency case.

Table 18.1. Properties of the scale kernel and frequency kernel

Scale kernel: $\gamma(c,t)$	Frequency kernel: $u(\omega,t)$
$\gamma(c,tt') = \gamma(c,t)\,\gamma(c,t')$	$u(\omega,t+t') = u(\omega,t)\,u(\omega,t')$
$\gamma(c,t/t') = \gamma(c,t)\,\gamma^*(c,t')$	$u(\omega,t-t') = u(\omega,t)\,u^*(\omega,t')$
$\gamma(c+c',t) = \gamma(c,t)\,\gamma(c',t)$	$u(\omega+\omega',t) = u(\omega,t)\,u(\omega',t)$
$\frac{\partial}{\partial c}\gamma(c,t) = j\ln t\,\gamma(c,t)$	$\frac{\partial}{\partial\omega}u(\omega,t) = jt\,u(\omega,t)$
$\frac{\partial}{\partial t}\gamma(c,t) = \frac{jc-1/2}{t}\,\gamma(c,t)$	$\frac{\partial}{\partial t}u(\omega,t) = j\omega\,u(\omega,t)$
$e^{j\sigma c}\gamma(c,t) = e^{\sigma/2}\gamma(c,te^{\sigma})$	$e^{j\sigma\omega}\,u(\omega,t) = u(\omega+\sigma,t)$

18.4 THE SCALE TRANSFORM

Any function can be expanded in terms of the scale eigenfunctions,

$$s(t) = \int D(c)\,\gamma(c,t)\,dc \tag{18.21}$$

$$= \frac{1}{\sqrt{2\pi}}\int D(c)\,\frac{e^{jc\ln t}}{\sqrt{t}}\,dc \;;\quad t \geq 0 \tag{18.22}$$

and the inverse transformation is

$$D(c) = \int_0^\infty s(t)\,\gamma^*(c,t)\,dt \tag{18.23}$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^\infty s(t)\,\frac{e^{-jc\ln t}}{\sqrt{t}}\,dt \tag{18.24}$$

We call $D(c)$ the scale transform. It can also be written as

$$D(c) = \frac{1}{\sqrt{2\pi}} \int_0^\infty s(t)\,t^{-jc-1/2}\,dt \tag{18.25}$$

which shows that it is the Mellin transform of the signal with the complex argument $-jc+\frac{1}{2}$. We know from the last chapter that the density of a quantity is the absolute square of its transform and therefore the density of scale is

$$P(c) = |\,D(c)\,|^2 \qquad \text{[Intensity or density of scale]} \tag{18.26}$$

Invariance of the Scale Transform to Compression. The fundamental property of the Fourier transform, $S(\omega)$, is that the energy density function of a translated function is identical to the energy density of the original function. That is, the Fourier transform of the signal, $s(t)$, and the translated signal, $s_{tr} = s(t+t_0)$, differ only by a phase factor:

$$s(t+t_0) \Longleftrightarrow S_{tr}(\omega) = e^{j\omega t_0}\,S(\omega) \qquad \text{if } s(t) \Longleftrightarrow S(\omega) \tag{18.27}$$

Hence they have identical energy density spectra,

$$|\,S_{tr}(\omega)\,|^2 = |\,S(\omega)\,|^2 \tag{18.28}$$

A similar idea holds for scale in that the scale energy density of a scaled function is equal to the scale energy density of the original function. If we compare the scale transform of $s(t)$ and $s_{sc} = \sqrt{a}\,s(at)$, they will differ only by a constant phase factor,

$$\sqrt{a}\,s(at) \Longleftrightarrow D_{sc}(c) = e^{jc\ln a}\,D(c) \qquad \text{if } s(t) \Longleftrightarrow D(c) \tag{18.29}$$

Therefore

$$|\,D_{sc}(c)\,|^2 = |\,D(c)\,|^2 \tag{18.30}$$

This a basic result which shows that the scaling of any time function leaves the energy density scale spectrum unchanged.

Alternative Proof of the Compression Operator. Using the fact that the scale eigenfunctions satisfy $e^{j\sigma c}\gamma(c,t) = e^{\sigma/2}\gamma(c,\,e^{\sigma}t)$, we have

$$e^{j\sigma C}\,s(t) = e^{j\sigma C}\int D(c)\,\gamma(c,t)\,dc = \int D(c)\,e^{j\sigma c}\,\gamma(c,t)\,dc \tag{18.31}$$

$$= \int D(c)\,e^{\sigma/2}\,\gamma(c,\,e^{\sigma}t)\,dc = e^{\sigma/2}\,s(\,e^{\sigma}t) \tag{18.32}$$

which is identical to Eq. (18.5).

The Scale Transform of the Sum of Scaled Functions. Let us first consider a signal composed of the sum of translated functions,

$$s(t) = \sum_{k=1}^{n} c_k f(t + t_k) \tag{18.33}$$

The Fourier transform of $s(t)$ is the sum of the Fourier transform of each term. But the Fourier transform of $f(t - t_k)$ is $e^{j\omega t_k} F(\omega)$, where F is the Fourier transform of $f(t)$. Therefore we have

$$S(\omega) = \sum_{k=1}^{n} c_k e^{j\omega t_k} F(\omega) = F(\omega) \sum_{k=1}^{n} c_k e^{j\omega t_k} \tag{18.34}$$

We now obtain the analog for the sum of scaled signals,

$$s(t) = \sum_{k=1}^{n} c_k \sqrt{a_k}\, f(a_k t) \tag{18.35}$$

The scale transform of each term is $c_k\, e^{jc \ln a_k} E(c)$, where $E(c)$ is the scale transform of $f(t)$. Therefore

$$D(c) = \sum_{k=1}^{n} c_k\, e^{jc \ln a_k} E(c) = E(c) \sum_{k=1}^{n} c_k\, e^{jc \ln a_k} \tag{18.36}$$

If we consider the situation where the amplitudes are arbitrary

$$s(t) = \sum_{k=1}^{n} A_k f(a_k t) \tag{18.37}$$

then

$$D(c) = E(c) \sum_{k=1}^{n} \frac{A_k}{\sqrt{a_k}}\, e^{jc \ln a_k} \tag{18.38}$$

Relation with the Fourier Transform. An interesting viewpoint of the scale transform is obtained by considering a new function constructed by replacing time with the logarithm of time,

$$f_\ell(t) = \frac{1}{\sqrt{t}}\, s(\ln t) \tag{18.39}$$

The factor $1/\sqrt{t}$ is inserted to preserve normalization. Now consider the scale transform of f_ℓ,

$$D_\ell(c) = \frac{1}{\sqrt{2\pi}} \int_0^\infty f_\ell(t)\, \frac{e^{-jc \ln t}}{\sqrt{t}}\, dt \tag{18.40}$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^\infty s(\ln t)\, \frac{e^{-jc\ln t}}{t}\, dt \tag{18.41}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty s(\tau)\, e^{-jc\tau}\, d\tau \tag{18.42}$$

which is the Fourier transform of $s(t)$,

$$D_\ell(c) = S(c) \tag{18.43}$$

Inversely we can consider the scale transform to be the Fourier transform of the function $s(e^t)\, e^{t/2}$. If we define

$$f(t) = s(e^t)\, e^{t/2} \tag{18.44}$$

then the Fourier transform of $f(t)$ is

$$F(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty s(e^t)\, e^{t/2}\, e^{-jct}\, dt \tag{18.45}$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^\infty s(t)\, \frac{e^{-jc\ln t}}{\sqrt{t}}\, dt \tag{18.46}$$

$$= D(c) \tag{18.47}$$

18.5 SIGNALS WITH HIGH SCALE CONTENT

What kind of signals have high scale content, that is, a high concentration of scale around a particular value? Signals with the highest scale content are the scale eigenfunctions, because their transform is a delta function of scale. As far as other functions are concerned a way to develop our intuition is to use the crutch of Fourier transforms developed in the above section. For a function $s(t)$ consider the function $e^{t/2}s(e^t)$ and ask if that new function has a high frequency content. If it does then the original function will have a high scale content, because its scale transform will be functionally the same as the Fourier transform of $e^{t/2}s(e^t)$.

Example 18.1: The Highest and the Lowest Scale Content Signals.

The eigenfunctions have the highest scale content because

$$D(c) = \delta(c - c') \quad \text{if } s(t) = \frac{1}{\sqrt{2\pi}}\, \frac{e^{jc'\ln t}}{\sqrt{t}} \tag{18.48}$$

What functions have the lowest or flattest scale content? We approach this by asking what functions have the lowest frequency content. They are functions that have a flat

spectrum. Signals that give rise to flat spectra are impulses, $\delta(t-t_0)$. Therefore, using our argument above, the functions that have the lowest scale content are

$$D(c) = \frac{1}{\sqrt{2\pi}} \frac{e^{-jc\ln t_0}}{\sqrt{t_0}} \tag{18.49}$$

$$\text{for } s(t) = \frac{\delta(\ln t - \ln t_0)}{\sqrt{t_0}} = \frac{\delta(\ln(t/t_0))}{\sqrt{t_0}} = \sqrt{t_0}\,\delta(t-t_0) \tag{18.50}$$

where for the last step in Eq. (18.50) we have used Eq. (18.93). We see therefore that impulses have the lowest frequency and scale content.

18.6 SCALE CHARACTERISTIC FUNCTION

The characteristic function for the density of scale is

$$M(\sigma) = \int e^{j\sigma c} \mid D(c) \mid^2 dc \tag{18.51}$$

$$= \int_0^\infty s^*(t)\, e^{j\sigma C}\, s(t)\, dt \tag{18.52}$$

$$= \int_0^\infty s^*(t)\, e^{\sigma/2}\, s(e^{\sigma} t)\, dt \tag{18.53}$$

or

$$M(\sigma) = \int_0^\infty s^*(e^{-\sigma/2}t)\, s(e^{\sigma/2}t)\, dt \tag{18.54}$$

As usual, the distribution is given by

$$P(c) = \frac{1}{2\pi} \int M(\sigma)\, e^{-j\sigma c}\, d\sigma \tag{18.55}$$

$$= \frac{1}{2\pi} \int \int_0^\infty s^*(e^{-\sigma/2}t)\, s(e^{\sigma/2}t)\, e^{-j\sigma c}\, d\sigma\, dt \tag{18.56}$$

Making the transformation

$$x = e^{-\sigma/2}\, t \quad ; \quad y = e^{\sigma/2}\, t \tag{18.57}$$

$$\sigma = \ln(y/x) \quad ; \quad t = \sqrt{xy} \quad ; \quad t\, d\sigma\, dt = dx\, dy \quad ; \quad d\sigma\, dt = \frac{dx\, dy}{\sqrt{xy}} \tag{18.58}$$

we have

$$P(c) = \frac{1}{2\pi} \int_0^\infty \int_0^\infty s^*(x)\, s(y)\, \frac{e^{-jc\,(\ln y - \ln x)}}{\sqrt{xy}}\, dx\, dy \tag{18.59}$$

$$= \mid D(c) \mid^2 \tag{18.60}$$

Total Energy Conservation. The total energy, E, of a signal is obtained by integrating over all time the energy density, $|s(t)|^2$. It should also be obtainable by integrating the scale density over all scale. That is indeed the case since

$$\int_0^\infty |s(t)|^2\,dt = \int |D(c)|^2\,dc \tag{18.61}$$

This shows that the total energy is preserved in the transformation. This is the analog to Parceval's theorem for time and frequency. It can be proven by direct substitution, although it follows from general considerations since the compression operator is unitary.

18.7 MEAN SCALE AND BANDWIDTH

There are three equivalent ways we can calculate the mean value for scale. The first is by direct use of the scale density, the second is to use the general formula for calculating the average value of any operator, and the third is to use Eq. (17.58),

$$\langle c\rangle = \int c\,|D(c)|^2\,dc \tag{18.62}$$

$$= \int_0^\infty s^*(t)\,C\,s(t)\,dt \tag{18.63}$$

$$= \int_0^\infty \left(\frac{Cs}{s}\right)_R |s(t)|^2\,dt \tag{18.64}$$

All three formulas will lead to the same result, but the last is the easiest to use. Calculating

$$\frac{Cs}{s} = t\,\varphi'(t) + \frac{1}{j}\left(t\,\frac{A'(t)}{A(t)} + \frac{1}{2}\right) \tag{18.65}$$

we see that the real part is $t\,\varphi'(t)$ and therefore

$$\langle c\rangle = \int_0^\infty t\,\varphi'(t)\,A^2(t)\,dt \tag{18.66}$$

The Scale Bandwidth. Similarly, the scale bandwidth can also be calculated by using any one of the three alternative formulas,

$$\sigma_c^2 = \int (c - \langle c\rangle)^2\,|D(c)|^2\,dc \tag{18.67}$$

$$= \int_0^\infty s^*(t)\,(C - \langle c\rangle)^2\,s(t)\,dt = \int |(C - \langle c\rangle)\,s(t)|^2\,dt \tag{18.68}$$

$$= \int \left(\frac{Cs}{s}\right)_I^2 |s(t)|^2\,dt + \int \left[\left(\frac{Cs}{s}\right)_R - \langle C\rangle\right]^2 |s(t)|^2\,dt \tag{18.69}$$

where, the last equation is obtained from Eq. (16.46). Again, the last equation is easiest since we have just calculated the real and imaginary parts of $\mathcal{C}s/s$, Eq. (18.65). We immediately have

$$\sigma_c^2 = \int_0^\infty \left(t\frac{A'(t)}{A(t)} + \frac{1}{2} \right)^2 A^2(t)\,dt + \int_0^\infty \left(t\varphi'(t) - \langle c \rangle \right)^2 A^2(t)\,dt \qquad (18.70)$$

18.8 INSTANTANEOUS SCALE

In Chapter 16 we showed that the instantaneous value of a quantity is given by the real part of $\mathcal{A}s/s$, where $\mathcal{A}$ is an arbitrary operator. Therefore, instantaneous scale is

$$c_t = \left(\frac{\mathcal{C}s}{s} \right)_R = t\varphi'(t) = t\omega_i \qquad (18.71)$$

That is, instantaneous scale is time multiplied by instantaneous frequency. The general relation of a conditional mean to the global mean is that the global mean is the average of the conditional mean. For scale we expect

$$\langle c \rangle = \int_0^\infty c_t \,|\, s(t) \,|^2\, dt \qquad (18.72)$$

and that is the case, as can be seen by Eq. (18.64).

18.9 UNCERTAINTY PRINCIPLE FOR SCALE

In Section 15.5 we showed that there is always an uncertainty principle for two operators if they do not commute. Specializing to scale and time we have

$$\sigma_t^2 \sigma_c^2 \geq \tfrac{1}{2} \,|\, \langle [\mathcal{T}, \mathcal{C}] \rangle \,| \qquad (18.73)$$

From Eq. (18.12) we know that $[\mathcal{T}, \mathcal{C}] = j\mathcal{T}$ and therefore we have the following uncertainty principle for time and scale,[307, 151]

$$\sigma_t^2 \sigma_c^2 \geq \tfrac{1}{2} \,|\, \langle t \rangle \,| \qquad (18.74)$$

where $\langle t \rangle$ is the mean time.

Minimum Uncertainty Product Signal. Using Eq. (15.103), the minimum uncertainty product signal is

$$\left(\frac{1}{j} \left[t\frac{d}{dt} + \frac{1}{2} \right] - \langle c \rangle \right) s(t) = \lambda \left(t - \langle t \rangle \right) s(t) \qquad (18.75)$$

where

$$\lambda = \frac{\langle [T, C] \rangle}{2\sigma_t^2} = j\frac{\langle t \rangle}{2\sigma_t^2} \tag{18.76}$$

and the solution is[151]

$$s(t) = k\, t^{\alpha_2}\, e^{-\alpha_1 t + j\langle c \rangle \ln(t/\langle t \rangle)} \tag{18.77}$$

where k is a normalizing constant and

$$\alpha_1 = \frac{\langle t \rangle}{2\sigma_t^2} \quad ; \quad \alpha_2 = \frac{1}{2}\left(\frac{\langle t \rangle^2}{\sigma_t^2} - 1\right) \tag{18.78}$$

18.10 FREQUENCY AND OTHER SCALING

In the preceding sections we have considered the scaling of time. We can equally consider the scaling of frequency or the scaling of any other variable. Let us first consider frequency scaling. From a structural point of view it is clear that to obtain the frequency scaling operator, C_ω, we just replace t by ω in Eq. (18.2),

$$C_\omega = \frac{1}{2j}\left(\frac{d}{d\omega}\omega + \omega\frac{d}{d\omega}\right) \tag{18.79}$$

Everything we have derived for time scaling can be immediately transliterated for frequency by substituting ω for t appropriately. In particular, the frequency scale eigenfunctions are

$$\gamma_\omega(c, \omega) = \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln\omega}}{\sqrt{\omega}} \tag{18.80}$$

and the transformations between the frequency domain and the scale domain are

$$S(\omega) = \frac{1}{\sqrt{2\pi}}\int D_\omega(c)\,\frac{e^{jc\ln\omega}}{\sqrt{\omega}}\,dc \quad , \quad \omega \geq 0 \tag{18.81}$$

$$D_\omega(c) = \frac{1}{\sqrt{2\pi}}\int_0^\infty S(\omega)\,\frac{e^{-jc\ln\omega}}{\sqrt{\omega}}\,d\omega \tag{18.82}$$

We have used $D_\omega(c)$ to emphasize that we are defining a different transform, the frequency scale transform. For the frequency scaling representation only the positive half of the spectrum is considered, which is equivalent to considering analytic signals. For frequency functions we have

$$e^{j\sigma C_\omega}\, S(\omega) = e^{\sigma/2} S(e^\sigma \omega) \quad ; \quad e^{j\ln\sigma C_\omega}\, S(\omega) = \sqrt{\sigma}\, S(\sigma\omega) \tag{18.83}$$

For mean frequency-scale and bandwidth we have

$$\langle c \rangle = \int_0^\infty \omega\, \psi'(\omega)\, B^2(\omega)\, d\omega \tag{18.84}$$

$$\sigma_c^2 = \int_0^\infty \left(t\frac{B'(\omega)}{B(\omega)} + \frac{1}{2} \right)^2 B^2(\omega)\, d\omega + \int_0^\infty \left(\omega\psi'(\omega) - \langle c \rangle \right)^2 B^2(\omega)\, d\omega \quad (18.85)$$

where $B(\omega)$ and $\psi(\omega)$ are the spectral amplitude and phase. Similarly, instantaneous scale in the frequency domain is given by

$$c_\omega = \omega\psi'(\omega) \quad (18.86)$$

Other Domains. For an arbitrary domain, say the a domain, the scale operator is

$$C_a = \frac{1}{2j}\left(\frac{d}{da}a + a\frac{d}{da} \right) \quad (18.87)$$

and in general we have

$$e^{j\sigma C_a} F(a) = e^{\sigma/2} F(e^\sigma a) \quad ; \quad e^{j \ln \sigma C_a} F(a) = \sqrt{\sigma}\, F(\sigma a) \quad (18.88)$$

To convert the equations we have derived for time scaling to scaling in the a domain, all we have to do is substitute a for t and $s(a)$ for $s(t)$ in any of the formulas. The above application for frequency was an example of this.

18.11 APPENDIX

We prove the completeness relations Eq. (18.18) and (18.19). Consider the left hand side of Eq. (18.18)

$$\int_0^\infty \gamma^*(c',t)\,\gamma(c,t)\, dt = \frac{1}{2\pi}\int_0^\infty \frac{e^{-jc' \ln t}}{\sqrt{t}} \frac{e^{jc \ln t}}{\sqrt{t}} \quad (18.89)$$

$$= \frac{1}{2\pi}\int_0^\infty \frac{e^{j(c-c')\ln t}}{t}\, dt \quad (18.90)$$

$$= \frac{1}{2\pi}\int_{-\infty}^\infty e^{j(c-c')x}\, dx \quad (18.91)$$

$$= \delta(c-c') \quad (18.92)$$

where in going from Eq. (18.89) to Eq. (18.91) we have made the transformation $x = \ln t$. To prove Eq. (18.19) we first establish the following delta function identity

$$\delta(\ln(x/a)) = a\,\delta(x-a) \quad (18.93)$$

with a positive. To prove it consider

$$\int_0^\infty \delta(\ln(x/a))\, f(x) dx = a\int_{-\infty}^\infty \delta(y)\, f(a\,e^y)\, e^y dy = a\, f(a) \quad (18.94)$$

which is the same answer that would be obtained if we integrate the right hand side of Eq. (18.93) with $f(x)$. Now consider

$$\int \gamma^*(c,t')\,\gamma(c,t)\,dc = \frac{1}{2\pi}\int \frac{e^{-jc\ln t'}}{\sqrt{t'}}\,\frac{e^{jc\ln t}}{\sqrt{t}}\,dc \tag{18.95}$$

$$= \frac{1}{\sqrt{tt'}}\,\delta(\ln t - \ln t') \tag{18.96}$$

$$= \frac{1}{\sqrt{tt'}}\,\delta(\ln(t/t')) \tag{18.97}$$

$$= \sqrt{\frac{t'}{t}}\,\delta(t-t') \tag{18.98}$$

$$= \delta(t-t') \tag{18.99}$$

which proves Eq. (18.19).

Chapter 19

Joint Scale Representations

19.1 INTRODUCTION

In this chapter we obtain joint representations where scale is one of the variables and the other variable is time or frequency. We also consider joint representations of the three variables, time, frequency, and scale. The general approach we use is the characteristic function operator method developed in Chapters 10 and 17.

19.2 JOINT TIME-SCALE REPRESENTATIONS

We use $P(t,c)$ for the time-scale distribution and $M(\theta,\sigma)$ for the corresponding characteristic function. As usual they are connected by

$$M(\theta,\sigma) = \iint e^{j\theta t+j\sigma c}\, P(t,c)\, dt\, dc = \langle e^{j\theta t+j\sigma c} \rangle \tag{19.1}$$

$$P(t,c) = \frac{1}{4\pi^2} \iint M(\theta,\sigma)\; e^{-j\theta t-j\sigma c}\, d\theta\, d\sigma \tag{19.2}$$

Marginals. The energy densities of time and scale are $|\, s(t)\,|^2$ and $|\, D(c)\,|^2$, respectively. Therefore, we would ideally like

$$P(c) = \int P(t,c)\, dt = |\, D(c)\,|^2 \tag{19.3}$$

$$P(t) = \int P(t,c)\, dc = |\, s(t)\,|^2 \tag{19.4}$$

However, as in the time-frequency case, there are joint representations that do not satisfy the marginals exactly but are nonetheless useful.

Joint Characteristic Function. For the characteristic function operator we use the notation $M(\theta,\sigma;\mathcal{T},\mathcal{C})$. The average of the characteristic function operator is the characteristic function,

$$M(\theta,\sigma) = \langle M(\theta,\sigma;\mathcal{T},\mathcal{C}) \rangle = \int_0^\infty s^*(t)\, M(\theta,\sigma;\mathcal{T},\mathcal{C})\, s(t)\, dt \tag{19.5}$$

As we discussed in Chapter 17, there are many possible orderings. In Table 19.1 we list some of the possibilities and the corresponding characteristic functions and joint distributions. In the appendix we work out the first two choices in detail because they utilize and illustrate the main mathematical techniques. Marinovich,[179, 363] Bertrand and Bertrand[68, 71] and Altes[12] introduced the concept of joint scale representations. They obtained, by other methods, the first two representations of Table 19.1. Further development of the theory has been made by Rioul,[488] Posch,[454] Rioul and Flandrin,[490] Papandreou, Hlawatsch and Boudreaux-Bartels,[421, 423] Baraniuk and Jones,[49, 53] Shenoy and Parks,[510, 511] and Cohen.[143, 151]

It is straightforward to verify, by direct integration, that the time-scale distributions of Table 19.1 satisfy the marginals of time and scale, Eqs. (19.3) and (19.4). However, the easiest way to ascertain that is by noting that each of those characteristic functions satisfy the characteristic functions of the marginals. That is,

$$M(\theta,0;\mathcal{T},\mathcal{C}) = e^{j\theta\mathcal{T}} \quad ; \quad M(0,\sigma;\mathcal{T},\mathcal{C}) = e^{j\sigma\mathcal{C}} \tag{19.6}$$

Of course, checking the characteristic function does not check whether the algebra to obtain the distribution was correctly done.

19.3 GENERAL CLASS OF TIME-SCALE DISTRIBUTIONS

To obtain the general class of joint time-scale representations we use the method developed in Section 17.5. We start with any characteristic function and define a new one by

$$M_{\text{new}}(\theta,\sigma) = \phi(\theta,\sigma)\, M(\theta,\sigma) \tag{19.7}$$

where ϕ is the kernel function whose role is identical to that in the time-frequency case. The general class of time-scale functions is then

$$P(t,c) = \frac{1}{4\pi^2} \iint \phi(\theta,\sigma)\, M(\theta,\sigma)\, e^{-j\theta t - j\sigma c}\, d\theta\, d\sigma \tag{19.8}$$

Suppose we choose the characteristic function given by ordering one of Table 19.1. Then

$$P(t,c) = \frac{1}{4\pi^2} \iiint s^*(e^{-\sigma/2}u)\, e^{-j\theta t - j\sigma c + j\theta u}\, \phi(\theta,\sigma)\, s(e^{\sigma/2}u)\, d\theta\, du\, d\sigma \tag{19.9}$$

Any other starting characteristic function would serve equally well.

Table 19.1. Time-scale representations

#	Ordering	Characteristic Function: $M(\theta,\sigma)$	Distribution: $P(t,c)$
1	$e^{j\sigma C/2}e^{j\theta T}e^{j\sigma C/2}$	$\int_0^\infty e^{j\theta t}\, s^*(e^{-\sigma/2}t)s(e^{\sigma/2}t)\,dt$	$\frac{1}{2\pi}\int s^*(e^{-\sigma/2}t)e^{-j\sigma c}s(e^{\sigma/2}t)\,d\sigma$
2	$e^{j\theta T+j\sigma C}$	$\int_0^\infty e^{2j\theta t\,\mathrm{sinh}(\sigma/2)/\sigma}\, s^*(e^{-\sigma/2}t)s(e^{\sigma/2}t)\,dt$	$\frac{1}{2\pi}\int \frac{\sigma}{2\sinh(\sigma/2)}e^{-j\sigma c}$ $s^*(e^{-\sigma/2}\frac{\sigma t}{2\sinh(\sigma/2)})\,s(e^{\sigma/2}\frac{\sigma t}{2\sinh(\sigma/2)})\,d\sigma$
3	$e^{j\theta T/2}e^{j\sigma C}e^{j\theta T/2}$	$\int_0^\infty \exp[j\theta t\cosh(\sigma/2)]\, s^*(e^{-\sigma/2}t)s(e^{\sigma/2}t)\,dt$	$\frac{1}{2\pi}\int \frac{1}{\cosh(\sigma/2)}e^{-j\sigma c}$ $s^*(e^{-\sigma/2}\frac{t}{\cosh(\sigma/2)})s(e^{\sigma/2}\frac{t}{\cosh(\sigma/2)})\,d\sigma$
4	$e^{j\sigma C}e^{j\theta T}$	$\int_0^\infty \exp[j\theta e^{\sigma/2}t]\, s^*(e^{-\sigma/2}t)s(e^{\sigma/2}t)$	$s(t)\gamma^*(c,t)D^*(c)$
5	$e^{j\theta T}e^{j\sigma C}$	$\int_0^\infty \exp[j\theta e^{-\sigma/2}t]\, s^*(e^{-\sigma/2}t)s(e^{\sigma/2}t)$	$s^*(t)\gamma(c,t)D(c)$
6	$\frac{1}{2}(\{4\}+\{5\})$	$\frac{1}{2}(\{4\}+\{5\})$	Real part $\{\, s^*(t)\gamma(c,t)D(c)\,\}$

Relation Between Representations. Suppose we have two time-scale distributions, P_1 and P_2, with corresponding kernels, ϕ_1 and ϕ_2. Their characteristic functions are

$$M_1(\theta,\sigma) = \phi_1(\theta,\sigma)\,M(\theta,\sigma) \quad ; \quad M_2(\theta,\sigma) = \phi_2(\theta,\sigma)\,M(\theta,\sigma) \tag{19.10}$$

Hence

$$M_1(\theta,\sigma) = \frac{\phi_1(\theta,\sigma)}{\phi_2(\theta,\sigma)}\,M_2(\theta,\sigma) \tag{19.11}$$

By taking the Fourier transform of both sides we obtain

$$P_1(t,c) = \iint g(t'-t,c'-c)P_2(t',c')\,dt'\,dc' \tag{19.12}$$

with

$$g(t,c) = \frac{1}{4\pi^2}\iint e^{j\theta t+j\sigma c}\frac{\phi_1(\theta,\sigma)}{\phi_2(\theta,\sigma)}\,d\theta\,d\sigma \tag{19.13}$$

Local autocorrelation method. The general time-scale distribution can be written in the form

$$P(t,c) = \frac{1}{2\pi}\int R_t(\sigma)\, e^{-jc\sigma}\, d\sigma \tag{19.14}$$

with

$$R_t(\sigma) = \frac{1}{2\pi}\iint \phi(\theta,\sigma)\, M(\theta,\sigma)\, e^{-j\theta t}\, d\theta \tag{19.15}$$

We shall call R_t the generalized local scale autocorrelation function in analogy with the local autocorrelation function for frequency, Eq. (9.7). Using Eq. (19.9) for the general class, $R_t(\sigma)$ is explicitly given by

$$R_t(\sigma) = \frac{1}{2\pi}\iint e^{j\theta(u-t)}\, \phi(\theta,\sigma)\, s^*(e^{-\sigma/2}u)\, s(e^{\sigma/2}u) du\, d\theta \tag{19.16}$$

19.4 JOINT FREQUENCY- SCALE REPRESENTATIONS

Because the scale operator has the identical form in any domain we can immediately write down joint distributions of frequency and frequency scaling. Using $M(\tau,\sigma)$ and $P(\omega,c)$ for the frequency-scale characteristic function and distribution, respectively, and $\mathcal{C}_\omega$ for the frequency scale operator, we calculate the characteristic function by

$$M(\tau,\sigma) = \langle\, \mathcal{M}(\tau,\sigma;\mathcal{W},\mathcal{C})\,\rangle \tag{19.17}$$

which is achieved by way of

$$M(\tau,\sigma) = \int_0^\infty S^*(\omega)\, \mathcal{M}(\tau,\sigma;\mathcal{W},\mathcal{C})\, S(\omega)\, d\omega \tag{19.18}$$

This is identical in structure to the time-frequency case. Therefore, all we have to do is substitute τ for θ and S for s. However, it must be understood that now c stands for scaling in the frequency domain, that is, scaling of frequency functions. The marginals are

$$\int P(\omega,c)\, d\omega = |\, D_\omega(c)\,|^2 \tag{19.19}$$

$$\int P(\omega,c)\, dc = |\, S(\omega)\,|^2 \tag{19.20}$$

where $D_\omega(c)$ is the transform defined by Eq. (18.24).

19.5 JOINT REPRESENTATION OF TIME, FREQUENCY, AND SCALE

We now consider joint representations of the three variables, time, frequency, and scale. By scale we mean time-scaling, although with a small modification the procedure can be used to obtain distributions of time, frequency, and frequency-scaling.

We use the notation $P(t,\omega,c)$ to signify the joint distribution and $M(\theta,\tau,\sigma)$ to signify the characteristic function. In Chapter 4 we discussed multivariate joint distributions but used the two dimensional case for the sake of clarity. We are now dealing with a three dimensional situation in which case the characteristic function and distribution are related by

$$M(\theta,\tau,\sigma) = \iiint P(t,\omega,c)\, e^{j\theta t+j\tau\omega+j\sigma c}\, dt\, d\omega\, dc \tag{19.21}$$

$$P(t,\omega,c) = \frac{1}{8\pi^3}\iiint M(\theta,\tau,\sigma)\, e^{-j\theta t-j\tau\omega-j\sigma c}\, d\theta\, d\sigma\, d\tau \tag{19.22}$$

The marginals we have to satisfy are

$$\int P(t,\omega,c)\, dc = P(t,\omega) \tag{19.23}$$

$$\int_0^\infty P(t,\omega,c)\, dt = P(\omega,c) \tag{19.24}$$

$$\int P(t,\omega,c)\, d\omega = P(t,c) \tag{19.25}$$

where the right hand sides are the two dimensional densities of the corresponding variables.

Characteristic Function. Many orderings are possible but it is sufficient to consider one ordering and write all other possibilities by using the kernel method. We take the ordering

$$\mathcal{M}(\theta,\tau,\sigma) = e^{j\sigma\mathcal{W}/2}\, e^{j\theta\mathcal{T}+j\tau\mathcal{W}}\, e^{j\sigma\mathcal{W}/2} \tag{19.26}$$

The characteristic function is

$$M(\theta,\tau,\sigma) = \int_0^\infty s^*(t)\, e^{j\sigma\mathcal{W}/2}\, e^{j\theta\mathcal{T}+j\tau\mathcal{W}}\, e^{j\sigma\mathcal{W}/2}\, s(t)\, dt \tag{19.27}$$

which evaluates to

$$M(\theta,\tau,\sigma) = \int_0^\infty e^{j\theta t}\, s^*(e^{-\sigma/2}(t-\tfrac{1}{2}\tau))\, s(e^{\sigma/2}(t+\tfrac{1}{2}\tau))\, dt \tag{19.28}$$

From Eq. (19.22), the distribution is

$$P(t,\omega,c) = \frac{1}{4\pi^2}\iint s^*(\,e^{-\sigma/2}(t-\tfrac{1}{2}\tau))\, e^{-j\tau\omega-j\sigma c}\, s(\,e^{\sigma/2}(t+\tfrac{1}{2}\tau))\, d\tau\, d\sigma \tag{19.29}$$

To obtain the general class of time-frequency-scale distributions we define a general characteristic function by

$$M_{\text{new}}(\theta,\tau,\sigma) = \phi(\theta,\tau,\sigma)\, M(\theta,\tau,\sigma) \tag{19.30}$$

where $\phi(\theta,\tau,\sigma)$ is a kernel of three variables. Therefore

$$M_{\text{new}}(\theta,\tau,\sigma) = \phi(\theta,\tau,\sigma)\int e^{j\theta t}\, s^*(\,e^{-\sigma/2}(t-\tfrac{1}{2}\tau))\; s(\,e^{\sigma/2}(t+\tfrac{1}{2}\tau))\,dt \qquad (19.31)$$

and the general class is

$$P(t,\omega,c) = \frac{1}{8\pi^3}\iiint M_{\text{new}}(\theta,\tau,\sigma)\, e^{-j\theta t-j\tau\omega-j\sigma c}\,d\theta\,d\sigma\,d\tau \qquad (19.32)$$

Explicitly

$$\begin{aligned} P(t,\omega,c) &= \Big(\frac{1}{2\pi}\Big)^5 \iiiint \phi(\theta,\tau,\sigma)\, s^*(\,e^{-\sigma/2}(u-\tfrac{1}{2}\tau))\, s(\,e^{\sigma/2}(u+\tfrac{1}{2}\tau)) \\ &\quad\times\quad e^{-j\theta(t-u)-j\tau\omega-j\sigma c}\,d\theta\,d\sigma\,d\tau\,du \end{aligned} \qquad (19.33)$$

The marginals of this general class are the general classes of the marginals. For example, consider integrating out scale. Integrating over c gives a $2\pi\delta(\sigma)$ and therefore

$$P(t,\omega) = \int P(t,\omega,c)\,dc \qquad (19.34)$$

$$= \frac{1}{4\pi^2}\iiint \phi(\theta,\tau,0)\, s^*(u-\tfrac{1}{2}\tau)\, s(u+\tfrac{1}{2}\tau)\, e^{-j\theta(t-u)-j\tau\omega}\,d\theta\,d\tau\,du \qquad (19.35)$$

This is precisely the general class of time-frequency distributions, Eq. (9.1), with the general kernel $\phi(\theta,\tau,0)$.

19.6 APPENDIX

We derive the distributions corresponding to orderings 1 and 2 of Table 19.1.

Ordering 1: $\mathcal{M}(\theta,\sigma) = e^{j\sigma C/2}\, e^{j\theta T}\, e^{j\sigma C/2}$. We have

$$M(\theta,\sigma) = \langle e^{j\sigma C/2}\, e^{j\theta T}\, e^{j\sigma C/2}\rangle \qquad (19.36)$$

$$= \int s^*(t)\, e^{j\sigma C/2}\, e^{j\theta T}\, e^{j\sigma C/2} s(t)\,dt \qquad (19.37)$$

$$= \int s^*(t)\, e^{j\sigma C/2}\, e^{j\theta t}\; e^{\sigma/4} s(e^{\sigma/2}t)\,dt \qquad (19.38)$$

$$= \int s^*(t)\, \exp(j\theta e^{\sigma/2}t)\; e^{\sigma/2} s(e^{\sigma}t)\,dt \qquad (19.39)$$

or

$$M(\theta,\sigma) = \int s^*(e^{-\sigma/2}t)\, e^{j\theta t} s(e^{\sigma/2}t)\,dt \qquad (19.40)$$

The distribution is

$$P(t,c) = \frac{1}{4\pi^2}\iint M(\theta,\sigma)\, e^{-j\theta t - j\sigma c}\, d\theta\, d\sigma \tag{19.41}$$

$$= \frac{1}{4\pi^2}\iiint s^*(e^{-\sigma/2}t')\, e^{j\theta t'}\, s(e^{\sigma/2}t')\, e^{-j\theta t - j\sigma c}\, d\theta\, d\sigma\, dt' \tag{19.42}$$

$$= \frac{1}{2\pi}\iint s^*(e^{-\sigma/2}t')\delta(t-t')s(e^{\sigma/2}t')\, e^{-j\sigma c}\, d\sigma\, dt' \tag{19.43}$$

Hence

$$P(t,c) = \frac{1}{2\pi}\int s^*(e^{-\sigma/2}t)\, e^{-j\sigma c} s(e^{\sigma/2}t)\, d\sigma \tag{19.44}$$

Ordering 2: $\mathcal{M}(\theta,\sigma) = e^{j\theta\mathcal{T}+j\sigma\mathcal{C}}$. Since the scale operator satisfies

$$[\mathcal{T},\mathcal{C}] = j\mathcal{T} \tag{19.45}$$

we can use Eq. (17.28) to obtain

$$e^{j\theta\mathcal{T}+j\sigma\mathcal{C}} = e^{j\theta\eta\mathcal{T}}\, e^{j\sigma\mathcal{C}}\, e^{j\theta\mathcal{T}} \tag{19.46}$$

with

$$\eta = -\frac{1}{\sigma}\left\{1-(1-\sigma)\, e^{\sigma}\right\} \tag{19.47}$$

The characteristic function is therefore

$$M(\theta,\sigma) = \int s^*(t)\, e^{j\theta\mathcal{T}+j\sigma\mathcal{C}} s(t)\, dt = \int s^*(t)\, e^{j\theta\eta\mathcal{T}}\, e^{j\sigma\mathcal{C}}\, e^{j\theta\mathcal{T}} s(t)\, dt \tag{19.48}$$

Now we can follow essentially the same steps as in ordering 1 to obtain

$$M(\theta,\sigma) = \int s^*(e^{-\sigma/2}t)\, e^{2j\theta t \sinh(\sigma/2)/\sigma} s(e^{\sigma/2}t)\, dt \tag{19.49}$$

In one of the steps leading to Eq. (19.49) use is made of the fact that

$$e^{\sigma/2} + \eta e^{-\sigma/2} = \frac{e^{\sigma/2}-e^{-\sigma/2}}{\sigma} = \frac{2}{\sigma}\sinh(\sigma/2) \tag{19.50}$$

The distribution is

$$P(t,c) = \frac{1}{4\pi^2}\iiint s^*(e^{-\sigma/2}t')\, e^{2j\theta t' \sinh(\sigma/2)/\sigma} s(e^{\sigma/2}t')\, e^{-j\theta t - j\sigma c}\, d\theta\, d\sigma\, dt' \tag{19.51}$$

which reduces to

$$P(t,c) = \frac{1}{2\pi}\int \frac{\sigma}{2\sinh(\sigma/2)}\, e^{-j\sigma c} s^*\left(e^{-\sigma/2}\frac{\sigma t}{2\sinh(\sigma/2)}\right)\, s\left(e^{\sigma/2}\frac{\sigma t}{2\sinh(\sigma/2)}\right)\, d\sigma \tag{19.52}$$

Bibliography

[1] R. M. S. S. Abeysekera and B. Boashash, "Time frequency domain features of ECG signals: Their application in P wave detection using the cross Wigner–Ville distribution," *Proc. IEEE ICASSP–89*, pp. 1524–1527, 1989.

[2] S. Abrahamsson, B. Brusmark, C. G. Gaunaurd and H. C. Strifors, "Target identification by means of impulse radar," *Proc. SPIE*, vol. 1471, pp. 130–141, 1991.

[3] S. Abrahamsson, B. Brusmark, H. C. Strifors, and C. G. Gaunaurd, "Extraction of target features in the combined time–frequency domain by means of impulse radar," *Proc. SPIE*, vol. 1700, pp. 102–114, 1992.

[4] M. H. Ackroyd, "Instantaneous and time–varying spectra – An introduction," *Radio Electron. Eng.*, vol. 239, pp. 145–152, 1970.

[5] Ackroyd, M. H., "Short–time spectra and time–frequency energy distributions," *J. Acoust. Soc. Am.*, vol. 50, pp. 1229–1231, 1970.

[6] Adamopoulos, P. G. and Hammond, J. K., "The use of the Wigner–Ville spectrum as a method of identifying/characterizing nonlinearities in systems," *Proc. IEEE ICASSP–87*, vol. 3, pp. 1541–1544, 1987.

[7] Adamopoulos, P. G., Hammond, J. K., and Lee, J. S., "The Wigner distribution, multi–equilibria nonlinear systems and chaotic behavior," *Proc. IEEE ICASSP–88*, pp. 2216–2219, 1988.

[8] D. V. Ageyev, "The active band of the frequency spectrum of a function of time," *Trudy Gor'kovskogo Politekhnicheskogo In–ta*, vol. 11, pp. 5, 1955.

[9] J. B. Allen and L. R. Rabiner, "A Unified Approach to STFT Analysis and Synthesis," *Proceedings of the IEEE*, vol. 65, pp. 1558–1564, 1977.

[10] J. B. Allen, "Short–Time Spectral Analysis, Synthesis and Modification by Discrete Fourier Transform." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, pp. 235–238, 1977.

[11] R. A. Altes, "Detection, estimation and classification with spectrograms," *J. Acoust. Soc. Am.*, vol. 67, pp. 1232–1246, 1980.

[12] R. A. Altes, "Wideband, proportional–bandwidth Wigner–Ville analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1005–1012, 1990.

[13] P. O. Amblard and J. L. Lacoume, "Construction of Fourth Order Cohen's Class: A Deductive Approach," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 257–260, 1992.

[14] M. G. Amin, "The application of the GLPW technique to time–varying spectrum estimation," in: *Applied Signal Processing*, ed. M. H. Hamza, pp. 298–302, Acta Press, 1985.

[15] M. G. Amin, "Time–varying spectrum estimation of a general class of nonstationary processes," *Proceedings of the IEEE*, vol. 74, pp. 1800–1802, 1986.

[16] M. G. Amin, "On the application of the single pole filter in recursive power spectrum estimation," *Proceedings of the IEEE*, vol. 75, pp. 729– 731, 1987.

[17] M. G. Amin, "Time and lag window selection in Wigner–Ville distribution," *Proc. IEEE ICASSP–87*, vol. 3, pp. 1529–1532, 1987.

[18] M. G. Amin, "Computationally lag invariant recursive spectrum estimators," *IEEE Trans. Acoust , Speech, Signal Processing*, vol. 35, pp. 1713–1724, 1987.

[19] M. G. Amin, "Recursion in Wigner distributions ," *Proc. SPIE*, vol. 975, pp. 221–231, 1988

[20] M. G Amin, "Mixed time–based spectrum estimators," *IEEE Trans Acoust., Speech, Signal Processing*, vol. 36, pp. 739–750, 1988.

[21] M G. Amin and M. T. Shiavoni, "Time–varying spectrum estimation via multidimensional filter representation," *Proc SPIE*, vol. 1152, pp. 437–448, 1989.

[22] M. G. Amin and F. Allen, "Separation of signals and cross–terms in Wigner Distribution using adaptive filtering," *Proceedings of the 32nd Midwest Symposium on Circuits and Systems*, vol. 2, pp. 853–856, 1989

[23] M. G. Amin and F Allen, "A novel approach to cross–terms elimination in Wigner distribution using adaptive filtering in frequency domain," *Proc SPIE*, vol. 1152, pp. 496–505, 1989.

[24] M. G. Amin, "High temporal resolution estimators through reduced rank periodograms," *Proc. IEEE ICASSP– 90*, vol. 4, pp. 2471–2474, 1990.

[25] M G. Amin, "Statistical Trade Offs in Modern Time–Frequency Kernel Design," *Proc of Twenty–Fifth Annual Asilomar Conference on Signals, Systems and Computers*, pp. 359–363, 1991.

[26] M. G. Amin, "Time–frequency spectrum analysis and estimation for non–stationary random processes," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 208–232, Longman Cheshire, 1992.

[27] M. G. Amin, "Spectral Decomposition of time–frequency distribution kernels," *Proceedings of the Asilomar Conference on Circuits, Systems, and Computers*, vol. 1, pp. 488–492, 1992.

[28] M. G. Amin and W. J Williams. "High Spectral Resolution Time–Frequency Distribution Kernels," *Proc. SPIE*, vol. 1770, pp. 302–314, 1992.

[29] M. G. Amin, P. Davis, and F. Allen ,"An Application of the LMS Algorithm in smoothing Pseudo Wigner Distribution," *IEEE Transactions on Signal Processing*, vol. 41, pp. 930–934, 1993.

[30] M. G. Amin and W. J. Williams, "Deterministic exponential modeling techniques for high spectral resolution time–frequency distributions," *Proc. IEEE ICASSP– 93*, vol. 3, pp. 253–256, 1993.

[31] M. G. Amin and J. F. Carroll, "A linearly constrained minimization approach for time–frequency kernel design," *Proc. 27th Asilomar Conf. on Sigs., Syst. and Comps.*, pp. 476–480, 1993.

[32] M. G. Amin, "Introducing the spectral diversity," *IEEE Transactions on Signal Processing*, vol. 41, pp. 185–193, 1993

[33] M. G. Amin, "The running time–frequency distributions," *Journal of Circuits, Systems, and Signal Processing*, to appear, 1995.

[34] M. G. Amin, "Spectral decomposition of the time–frequency distribution kernels," *IEEE Transactions on Signal Processing* vol. 42, pp. 1156–1166, 1994.

[35] J. C. Andrieux, M. R. Feix, G. Mourgues, P. Bertrand, B. Izrar, and V. T. Nguyen, "Optimum smoothing of the Wigner–Ville distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 764–769, 1987.

[36] R. A. Athale and J. N. Lee and E. L. Robinson and H. H. Szu, "Acousto–optic processors for real–time generation of time–frequency representations," *Opt. Lett.*, vol. 8, pp. 166–168, 1983

[37] L E Atlas, Y. Zhao, R. J. Marks II, "Application of the generalized time–frequency representation to speech signal analysis," *IEEE Proc. Pacific Rim Conf on Comm., Comp. and Sig. Proc.*, pp. 517–520, 1987.

[38] L. E. Atlas and W. Kooiman and P. J Loughlin and R. Cole, "New non–stationary techniques for the analysis and display of speech transients," *Proc. IEEE ICASSP– 90*, pp. 385–388, 1990.

[39] L. E. Atlas, P. Loughlin, J. Pitton and W. Fox, "Applications of cone–shaped kernel time–frequency representations to speech and sonar analysis," *2nd Int. Symp. on Sig. Proc. and its Applications*, pp. 882–885, 1990.

[40] L. E. Atlas and P. J. Loughlin and J. W. Pitton, "Signal analysis with cone kernel time–frequency representations and their application to speech," in. *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 375–388, Longman Cheshire, 1992

[41] L. E. Atlas, J. Fang, P. Loughlin, and W. Music, "Resolution advantages of quadratic signal processing," *Proc. SPIE* , vol. 1566, pp. 134–143, 1991.

[42] L. E. Atlas, P. Loughlin, J. Pitton, "Truly nonstationary techniques for the analysis and display of voiced speech," *Proc. IEEE ICASSP-91*, pp. 433–436, 1991.

[43] L. E. Atlas and J. Fang, "Quadratic Detectors for General Nonlinear Analysis of Speech," *Proc. IEEE ICASSP-92*, pp. II–9–II–12, 1992.

[44] R. Bamler and H. Glunder, "The Wigner distribution function of two–dimensional signals; coherent optical generation and display," *Opt. Acta.*, vol. 30, pp. 1789–1803, 1983.

[45] R. G. Baraniuk and D. L. Jones, "Optimal kernels for time–frequency analysis," *Proc. SPIE*, vol. 1348, pp. 181–187, 1990.

[46] R. G. Baraniuk, D. L. Jones, T. Brotherton, and S. L. Marple, "Applications of adaptive time–frequency representations to underwater acoustic signal processing," in: *Proceedings of the 25th Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 1109-1113, 1991.

[47] R. G. Baraniuk and D. L. Jones, "Signal–dependent time–frequency analysis using a radially Gaussian kernel," *Signal Processing*, vol. 32, pp. 263–284, 1993.

[48] R. G. Baraniuk and D. L. Jones, "A signal–dependent time–frequency representation: optimal kernel design," *IEEE Trans. on Signal Processing*, vol. 41, pp. 1589–1602, 1993.

[49] R. G. Baraniuk and D. L. Jones, "Warped wavelet bases: unitary equivalence and signal processing," *Proc. IEEE ICASSP-93*, pp. 320–323, 1993.

[50] R. G. Baraniuk, "A signal transform covariant to scale changes," *Electronics Letters*, vol. 29, pp. 1675–1676, 17, 1993.

[51] R. G. Baraniuk and D. L. Jones, "A signal–dependent time–frequency representation: Fast algorithm for optimal kernel design," *IEEE Trans. Sig. Proc.*, vol. 42, pp. 134–146, 1994.

[52] R. G. Baraniuk and L. Cohen, "On Joint Distributions for Arbitrary Variables," *IEEE Signal Processing Letters*, submitted, 1994.

[53] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: a new twist on signal processing," *IEEE Trans. on Signal Processing*, submitted, 1993.

[54] N. F. Barber and F. Ursell, "The response of a resonant system to a gliding tone," *Phil. Mag.*, vol. 39, pp. 345–361, 1948.

[55] D. T. Barry, N. M. Cole and J. C. Wood, "Time–frequency analysis of resonating biological systems," *in: ISSPA 90 – Signal Processing*, vol. 2, pp. 930–933, 1990.

[56] D. T. Barry, N. M. Cole and J. C. Wood, "Muscle sounds occur at the resonant frequency of skeletal muscle," *IEEE Biomed. Engr.*, vol. 37, pp. 525–531, 1990.

[57] D. T. Barry and J. C. Wood, "Time–frequency analysis of the human first heart sound," *Proceedings of the IEEE Society of Engineering in Medicine and Biology*, pp. 2097–2098, 1991.

[58] D. T. Barry, "Fast Calculation of the Choi–Williams Distribution," *IEEE Trans. on Signal Processing*, vol. 40, pp. 450–455, 1992.

[59] H. O. Bartelt, K. H. Brenner and W. Lohmann, "The Wigner distribution function and its optical production," *Opt. Commun.*, vol. 32, pp. 32–38, 1980.

[60] M. J. Bastiaans, "The Wigner distribution applied to optical signals and systems," *Opt. Commun.*, vol. 25, pp. 26–30, 1978.

[61] M. J. Bastiaans, "Wigner distribution function and its application to first order optics" *Jour. Opt. Soc. Am.*, vol. 69, pp. 1710–1716, 1979.

[62] M. J. Bastiaans, "Signal description by means of a local frequency spectrum," *Proc. SPIE*, vol. 373, pp. 49–62, 1981.

[63] M. J. Bastiaans, "On the sliding window representation in digital signal processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 868–873, 1985.

[64] E. Bedrosian, "A product theorem for Hilbert transforms," *Proceedings of the IEEE*, vol. 51, pp. 868–869, 1963.

[65] S. C. Bekir, "A contribution to the unaliased discrete–time Wigner distribution," *J. Acoust. Soc Am.*, vol. 93, pp. 363–371, 1993.

[66] L. R. Berriel–Valdos, C. Gonzalo and J. Bescos, "Computation of the Wigner distribution function by the Hartley transform," *Opt. Commun.*, vol. 68, pp. 339–344, 1988.

[67] M. V. Berry, "Semiclassical Mechanics in Phase Space: A Study of Wigner's Function," Phil. Trans. Roy. Soc., vol. A-287, pp. 237–271, 1977.

[68] J. Bertrand and P. Bertrand "Representation temps–frequence des signoux," *C. R. Acad. Sc.*, vol. 299 (serie 1), pp. 635–638, 1984

[69] J. Bertrand and P. Bertrand, "Time–frequency representations of broad– band signals" in:*Wavelets, time–frequency methods and phase space*, edited by J. M. Coombes, A. Grossmann and Ph. Tchamitchian, Springer–Velag, 1989.

[70] J. Bertrand, P. Bertrand, and J. P. Ovarlez, "Discrete Mellin transform for signal analysis," *Proc. IEEE ICASSP– 90*, pp. 1603–1606, 1990.

[71] J. Bertrand and P. Bertrand "Affine Time–Frequency Distributions," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 118–140, Longman Cheshire, 1992.

[72] J. Bertrand and P. Bertrand, "A class of affine Wigner functions with extended covariance properties," *Jour. Math. Phys.*, vol. 33, pp. 2515–2527, 1992.

[73] P. Bertrand, J. P. Doremus, B. Izrar, V. T. Nguyen and M. R. Feix, "Obtaining non–negative quantum distribution functions" *Phys. Lett.*, vol. 94, pp. 415–417, 1983.

[74] P. Bertrand and J. Bertrand "Time–Frequency Representations of Broad Band Signals," in: *The Physics of Phase Space*, edited by Y. S. Kim and W. W. Zachary, Springer Verlag, pp. 250–252, 1987.

[75] T. K. Bhattacharya, and S. Haykin, "Application of Time–Frequency Distributions for Improved Radar Target Detection," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 509–512, 1992

[76] A. Blanc–Lapierre and B. Picinbono, "Remarques sur la Notion de Spectre instantane de puissance," *Public. Sci. Univ. d'Alger*, Serie B1, pp. 2–32, 1955.

[77] B. Boashash, L. B. White, and J. Imberger, "Wigner–Ville Analysis of Non–Stationary Random Signals with Application to Temperature Gradient Microstructure," *Proc. IEEE ICASSP– 86*, pp 2323–2326, 1986.

[78] B. Boashash, B. Lovell, and L. B. White, "Time–frequency analysis and pattern recognition using singular value decomposition of the Wigner–Ville distribution," *Proc. SPIE*, vol. 826, pp. 104–114, 1987.

[79] B. Boashash, "Note on the use of the Wigner distribution for time–frequency signal analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1518–1521, 1988.

[80] B. Boashash, G. Jones and P. O'Shea, "Instantaneous frequency of signals, estimation techniques and applications," *Proc. SPIE*, vol. 1152, pp. 382–400, 1989.

[81] B. Boashash and P. O'Shea, "A Methodology for Detection and Classification of Some Underwater Acoustic Signals Using Time–Frequency Analysis Techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp 1829–1841, 1990

[82] B. Boashash and L. B. White, "Instantaneous frequency estimation and automatic time–varying filtering," *Proc. IEEE ICASSP– 90*, pp. 1221–1224, 1990.

[83] B. Boashash, "Time–frequency signal analysis," in: *Advances in Spectrum Estimation and Array Processing*, Ed. S. Haykin, Prentice Hall, pp. 418–517, 1990.

[84] B. Boashash and A. P. Reilly, "Algorithms for time–frequency signal analysis," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 163–181, Longman Cheshire, 1992.

[85] B. Boashash and G. Jones, "Instantaneous frequency and time–frequency distributions," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 43–73, Longman Cheshire, 1992.

[86] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal – Part I: Fundamentals," *Proceedings of the IEEE*, vol. 80, pp. 520–538, 1992.

[87] D. Bohm, *Quantum Mechanics*, Prentice–Hall, 1951.

[88] P. Boles and B. Boashash, "The cross Wigner–Ville distribution – A two dimensional analysis method for the processing of vibrosis seismic signals," *Proc. IEEE ICASSP– 88*, pp. 904–907, 1988.

[89] P. Boles and B. Boashash, "Application of the cross Wigner–Ville distribution to seismic surveying," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 445–466, Longman Cheshire, 1992

[90] F. Bopp, "La Mechanique Quantique Est–Elle Une Mechanique Statistique Classique Particuliere?," *Ann. L'Inst. H. Poincare*, vol. 15, pp. 81–112, 1956.

[91] M. Born and P. Jordan, "Zur Quantenmechanik," *Zeit. f. Phys.*, vol. 34, pp. 858–888, 1925.

[92] G. F. Boudreaux–Bartels and T. W. Parks, "Reducing Aliasing in the Wigner Distribution Using Implicit Spline Approximation," *Proc. IEEE ICASSP– 83*, pp. 1438–1441, 1983.

[93] G. F. Boudreaux–Bartels and T. W. Parks, "Signal estimation using modified Wigner distributions," *Proc. IEEE ICASSP– 84*, pp. 22.3.1–22.3.4. 1984.

[94] G. F. Boudreaux–Bartels, "Time–varying signals processing using the Wigner distribution time–frequency signal representation," in: *Advances in Geophysical Data Processing*, Edited by M. Simaan, JAI Press, pp. 33–79, 1985.

[95] G. F. Boudreaux–Bartels and T. W. Parks, "Time–Varying Filtering and Signal Estimation Using Wigner Distribution Synthesis Techniques," *IEEE Trans. on Acoust , Speech, Signal Processing*, vol. 34, pp. 442–451, 1986.

[96] G. F. Boudreaux–Bartels, "Wigner distribution analysis of acoustic well logs," *Proc. IEEE ICASSP–87*, vol. 4, pp. 2237–2240, 1987.

[97] G. F. Boudreaux–Bartels and A. Papandreou, "A Generalization of the Choi–Williams Exponential Distribution," *1991 Asilomar Conference on Signals and Systems*, pp. 364–369, 1991.

[98] A. C. Bovik, P. Maragos, and T. F. Quatieri, "Measuring Amplitude and Frequency Modulations in Noise Using Multiband Energy Operators," *Proc IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 4–6, 1992.

[99] A. C. Bovik, P. Maragos, and T. F. Quatieri, "AM–FM Energy Detection and Separation in Noise Using Multiband Energy Operators," *IEEE Trans Signal Processing*, vol. 41, pp 3245–3265, 1993.

[100] H. Brand, R. Graham and A. Schenzle, "Wigner distribution of bistable steady states in optical subharmonic generation," *Opt Commun.*, vol. 32, pp. 359–363, 1980.

[101] B. R. Breed and T. E. Posch, "A range and azimuth estimator based on forming the spatial Wigner distribution," *Proc. IEEE ICASSP– 84*, vol. 3, pp. 41B.9.1–2, 1984.

[102] K. H. Brenner and A. W. Lohmann, "Wigner distribution function display of complex 1D signals," *Opt. Commun.*, vol. 42, pp. 310–314, 1982.

[103] T. W. Brotherton, T. Pollard, R. Barton, A. Krieger, and S. L. Marple, "Application of Time–Frequency and Time–Scale Analysis to Underwater Acoustic Transients," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 513–516, 1992.

[104] T. Brotherton, T. Pollard, and D. L. Jones, "Applications of time–frequency and time–scale representations to fault detection and classification," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 95–98, 1992.

[105] N. Bukhrain, V. Y. Petrunkin, S. A. Rogov, S. V. Rozov, V. G. Samsonov, and A. G. Tkachenko, "Acoustoelectronic system for generating a modified Wigner distribution," *Optoelectronics, Instrumentation, and Data Processing*, vol. 6, pp. 51–54, 1987.

[106] R. J. Bulgrin, B. J. Rubal, C. R. Thompson, and J. M. Moody, "Comparison of Short–Time Fourier, Wavelet and Time–Domain Analysis of Intracardiac Sounds," *Procedings of the 30th Ann. Rocky Mountain Bionengineering Symposium*, vol. 29, pp. 465–472, 1993.

[107] J. R. Carson and T. C. Fry, "Variable frequency electric circuit theory with application to the theory of frequency–modulation," *Bell System Tech. J.*, vol. 16, pp. 513–540, 1937.

[108] N. Cartwright, "Correlations without joint distributions in quantum mechanics," *Found. of Phys.*, vol. 4, pp. 127–136, 1974.

[109] N. Cartwright, "A non–negative quantum mechanical distribution function," *Physica*, vol. 83A, pp. 210–212, 1976.

[110] L. F. Chaparro, A. El–Jaroudi, and S. A. Kayhan, "Evolutionary Spectral Estimation on Non–Stationary Signals", in: *Fundamentals of Discrete–Time Systems*, Editors: M. Jamshidi, M. Mansour, B. Anderson, and N. Bose, TSI Press, pp. 219–226, 1993.

[111] H. Chen and G. F Boudreaux–Bartels, "On the Application of Two–Dimensional AR Modeling of the Wigner Distribution for Efficient Coding, Filtering, and Pattern Recognition," *Proc. IEEE ICASSP– 90*, pp. 2463–2466, 1990.

[112] S–C. Chen, and X. Yang, "Speech recognition with high recognition rate by smoothed spaced pseudo Wigner–Ville distribution (SSPWD) and overlap slide window spectrum methods," *Proc. IEEE ICASSP– 88*, vol. 1, pp. 191–194, 1988.

[113] D. B. Chester, F. J. Taylor and M. Doyle, "Application of the Wigner distribution to speech processing," *IEEE Acoust., Speech and Signal Processing Spectrum Estimation Workshop II*, pp. 98–102, 1983

[114] D. B. Chester, F. J. Taylor and M. Doyle, "On the Wigner distribution," *Proc. IEEE ICASSP–83*, vol. 2, pp. 491–494, 1983.

[115] D. B. Chester, F. J. Taylor and M. Doyle, "The Wigner distribution in speech processing applications," *Jour. Franklin Inst.*, vol. 318, pp. 415–430, 1984.

[116] P. R. Chitrapu and D. J. Waldo, "Time–Frequency Analysis and Linear Prediction of Cardiac Late Potentials," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 227–230, 1992.

[117] H. I. Choi and W. J. Williams, "Improved time–frequency representation of multicomponent signals using exponential kernels," *IEEE Trans on Acoust., Speech, Signal Processing*, vol. 37, pp. 862–871, 1989.

[118] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution – A tool for time–frequency signal analysis – Part I: continuous–time signals," *Philips Jour. of Research*, vol. 35, pp. 217–250, 1980.

[119] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution – a tool for time–frequency signal analysis – Part II: Discrete time signals," *Philips Jour. Research.*, vol. 35, pp. 276–300, 1980.

[120] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution – a tool for time–frequency signal analysis – Part III: Relations with other time–frequency signal transformations," *Philips Jour. Research.*, vol. 35, pp. 372–389, 1980

[121] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "Time–frequency signal analysis by means of the Wigner distribution," *Proc. IEEE ICASSP–81*, vol. 1, pp. 69–72, 1981.

[122] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The aliasing problem in discrete time Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1067–1072, 1983.

[123] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "On the time–frequency discrimination of energy distributions: can they look sharper than Heisenberg?," *Proc. IEEE ICASSP–84*, pp. 41B.7.1–41B.7.4, 1984.

[124] F. S. Cohen, G. F. Boudreaux–Bartels and S. Kadambe, "Tracking of unknown non stationary chirp signals using unsupervised clustering in the Wigner distribution space," *Proc. IEEE ICASSP–88*, pp. 2180–2183, 1988.

[125] L. Cohen, "Generalized phase–space distribution functions," *Jour. Math. Phys.*, vol. 7, pp. 781–786, 1966.

[126] L. Cohen, "Quantization Problem and Variational Principle in the Phase Space Formulation of Quantum Mechanics," *Jour. Math. Phys*, vol. 17, pp. 1863–1866, 1976.

[127] L. Cohen and Y. Zaparovanny, "Positive Quantum Joint Distributions," *J. Math. Phys.*, vol. 21, pp. 794–796, 1980.

[128] L. Cohen, and T. Posch, "Generalized Ambiguity Functions," *Proc. IEEE ICASSP–85*, pp. 1033–1036, 1985.

[129] L. Cohen and T. Posch, "Positive time–frequency distribution functions," *IEEE Trans. ASSP*, vol. 33, pp. 31–38, 1985.

[130] L. Cohen, "Positive and Negative Joint Distributions," in: *Non–Equilibrium Quantum Statistical Physics*, eds. G. Moore and M. O. Scully, Plenum Publishing, 1986.

[131] L. Cohen, "Joint Quantum Probabilities and The Uncertainty Principle," in. *New Techniques and Ideas in Quantum Measurement theory*, edited by D. Greenberger, New York Academy of Sciences, 1986

[132] L. Cohen, "A critical review of the fundamental ideas of joint time–frequency distributions," *IEEE Proc. Int. Symp. on Circuits and Systems*, vol. 1, pp 28–33, 1986.

[133] L. Cohen, "Wigner distribution for finite duration or band limited signals and limiting cases," *IEEE Tran. Acoust., Speech, Signal Processing*, vol. 35, pp. 796–806, 1987.

[134] L. Cohen, "On a fundamental property of the Wigner distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 559–561, 1987.

[135] L. Cohen and C. Lee, "Instantaneous Frequency, Its Standard Deviation and Multicomponent Signals," *Proc. SPIE*, vol. 975, pp. 186–208, 1988

[136] L. Cohen and C. Lee, "Instantaneous Mean Quantities in Time–Frequency Analysis," *Proc IEEE ICASSP– 88*, 2188–2191, 1988.

[137] L. Cohen, "Time–Frequency Distributions – A Review," *Proc. of the IEEE*, vol. 77, pp. 941–981, 1989.

[138] L. Cohen and C. Lee, "Standard deviation of instantaneous frequency," *Proc. IEEE ICASSP– 89*, pp. 2238–2241, 1989.

[139] L. Cohen and C. Lee, "Local bandwidth and optimal windows for the short time Fourier transform," *Proc. SPIE*, vol. 1152, pp. 401–425, 1989.

[140] L. Cohen and C. Lee, "Optimal Windows For The Short Time Fourier Transform," *ISSPA 90*, vol. 2, pp. 912–915, 1990.

[141] L. Cohen, "Distributions Concentrated Along The Instantaneous Frequency," *Proc. SPIE*, vol. 1348, pp. 149–157, 1990.

[142] L Cohen and C. Lee, "Instantaneous bandwidth for signals and spectrogram," *Proc. IEEE ICASSP–90*, pp. 2451–2454, 1990.

[143] L. Cohen, "A General Approach For Obtaining Joint Representations in Signal Analysis And An Application to Scale," *Proc. SPIE*, vol. 1566, pp. 109–133, 1991.

[144] L. Cohen, "A Primer on Time Frequency Distributions," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 3–42, Longman Cheshire, 1992.

[145] L. Cohen and C. Lee, "Instantaneous bandwidth," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 98–117, Longman Cheshire, 1992.

[146] L. Cohen, "Scale and Inverse Frequency Representations," in: *The Role of Wavelets in Signal Processing Applications*, edited by: B W. Suter, M. E Oxley, and G. T Warhola, AFIT/EN–TR–92–3, pp. 118–129, 1992.

[147] L. Cohen, "What is a Multicomponent Signal?," *Proc. IEEE ICASSP– 92*, vol. 5, pp. 113–116, 1992.

[148] L. Cohen, "A General Approach For Obtaining Joint Representations in Signal Analysis," *Proceedings of the IEEE*, submitted, May 1992.

[149] L. Cohen, K. Assaleh, and A. Fineberg, "Instantaneous Bandwidth and Formant Bandwidth" in: *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 13–17, 1992.

[150] L. Cohen, "Instantaneous Anything," *Proc. IEEE ICASSP– 93*, vol. 5, pp. 105–108, 1993.

[151] L. Cohen, "The scale representation," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3275–3293, 1993.

[152] N. M. Cole and D. T Barry "Muscle sounds frequencies of the frog are modulated by skeletal muscle tension," *Biophys. J.*, vol. 66, pp. 1104–114, 1994.

[153] T D. Cole and A. S. El–Dinary, "Estimation of Target Vibration form Laser Radar Backscatter Using Time–Frequency Distributions," in: *Proc. SPIE*, vol. 1936, 1993.

[154] M. Conner and L. Yao, "Optical generation of the Wigner distribution of 2–D real signals," *Appl. Opt.*, vol. 24, pp. 3825–3829, 1985.

[155] C. E. Cook and M. Bernfield, *Radar Signals*, Academic Press, 1967.

[156] A H. Costa and G. F. Boudreaux–Bartels, "Optimal Time–Frequency Representations Using a Multiform, Tiltable Exponential Kernel," *IEEE Trans. Signal Processing*, submitted, 1993

[157] G. Cristobal, J. Bescos and J. Santamaria, "Application of the Wigner distribution for image representation and analysis," *Proc. IEEE Int. Assoc Pattern Recognition; Eighth International Conference on Pattern Recognition*, pp. 998–1000, 1986.

[158] G. Cristobal, J Bescos, J. Santamaria and J. Montes, "Wigner distribution representation of digital images," *Patt. Rec. Lett.*, vol. 5, pp. 215–221, 1987.

[159] G. Cristobal, J. Bescos and J. Santamaria, "Local image filtering and image classification through the Wigner distribution," *Proc. SPIE*, vol. 813, pp. 377–378, 1987.

[160] G. Cristobal, J. Bescos and J. Santamaria, "Image analysis through the Wigner distribution," *Appl Opt.*, vol. 28, pp. 262–271, 1989.

[161] G. Cristobal, C. Gonzalo and J. Bescos, "Image filtering and analysis through the Wigner distribution," in: *Advances in Electronics and Electron Physics*, vol. 80, pp. 309–397, Academic Press, 1991.

[162] G. S. Cunningham and W. J. Williams. "Fast Implementation of Time–Frequency Distributions," *Proc. IEEE-SP Int Sympos. Time–Frequency and Time–Scale Analysis*, pp. 241–244, 1992.

[163] G. S. Cunningham and W. J Williams. "Fast Computation of the Wigner Distribution for Finite–Length Signals," *Proc. IEEE ICASSP– 92*, pp. 23–26, 1992.

[164] G. S. Cunningham and W J Williams. "Decomposition of Time–Frequency Distribution Kernels," *Proc. SPIE*, vol. 1770, pp. 351–357, 1992.

[165] G. S Cunningham and W. J. Williams "High–Resolution Signal Synthesis for Time–Frequency Distributions," *Proc. IEEE ICASSP– 93*, vol. 4, pp. 400–403, 1993.

[166] G. S. Cunningham and W J Williams, "Kernel Decomposition of Time–Frequency Distributions," *IEEE Trans. on Signal Processing*, vol. 42, pp. 1425–1442, 1994.

[167] G. S. Cunningham and W. J. Williams, "Fast Implementations of Generalized Discrete Time–Frequency Distributions," *IEEE Trans. on Signal Processing*, vol. 42, p. 1496–1508, 1994.

[168] I. Daubechies, "Time – Frequency Localization Operators: A Geometric Phase Space Approach," *IEEE Trans. Inform. Theory*, vol. 34, 1988, pp. 605–612, 1988.

[169] I. Daubechies, "The wavelet transform, time–frequency localization and signal analysis," *IEEE Trans on Information Theory*, vol. 36, pp. 961–1005, 1990.

[170] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.

[171] P. Davis and M.G. Amin, "Frequency tracking: adaptive filtering vs. time–frequency distributions," *Proc. IEEE ICASSP– 91*, pp. 3225–3228, 1991.

[172] C. Detka, A. El–Jaroudi and L. F. Chaparro, "Relating the Bilinear Distributions and the Evolutionary Spectrum," *Proc. IEEE ICASSP– 93*, vol. 4, pp. 496–499, 1993.

[173] C. S Detka and A. El–Jaroudi, "The Transitory Evolutionary Spectrum," in: *Proc. IEEE ICASSP– 94*, 1994.

[174] C. Detka, P. Loughlin and A. El–Jaroudi, "On combining evolutionary spectral estimates," *Seventh SP Workshop on Statistical Signal and Array Processing*, pp. 26–29, 1994

[175] E. J. Diethorn, "The Generalized Exponential Time–Frequency Distribution," *AT&T Bell Laboratories Tech. Memo.*, 1992.

[176] E. J. Diethorn, "The Generalized Exponential Time–Frequency Distribution," *IEEE Trans. on Signal Processing*, vol. 42, p. 1028–1037, 1994.

[177] J. H. Eberly and K Wodkiewicz, "The time–dependent physical spectrum of light," *J. Opt. Soc. Am.*, vol. 67, pp 1152–1261, 1977.

[178] G. Eichmann and B. Z. Dong, "Two–dimensional optical filtering of 1–D signals," *Appl. Opt.*, vol. 21, pp. 3152–3156, 1982.

[179] G. Eichmann and N. M. Marinovich, "Scale–invariant Wigner distribution and ambiguity functions," *Proc. SPIE*, vol. 519, pp. 18–24, 1985.

[180] A. S. El–Dinary and T. D. Cole, "Non–Recursive FM Demodulation of Laser Radar Backscatter Using Time–Frequency Distributions," *Proc. IEEE ICASSP– 94*, vol. 4, pp. 353–356, 1994.

[181] J. Fang, L. E. Atlas, and G. Bernard, "Advantages of Cascaded Quadratic Detectors for Analysis of Manufacturing Sensor Data," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 345–348, 1992.

[182] R. M. Fano, "Short–time autocorrelation functions and power spectra," *J. Acoust. Soc. Am.*, vol. 22, pp. 546–550, 1950.

[183] P. D. Finch and R. Groblicki, "Bivariate probability densities with given margins," *Found. of Phys.*, vol. 14, pp. 549–552, 1984.

[184] L. M. Fink, "Relations Between The Spectrum And Instantaneous Frequency Of A Signal," Translated in: *Problems Inform Transm.*, vol. 2, pp. 11–21, 1966.

[185] P. Flandrin, "Some features of time–frequency representations of multicomponent signals," *Proc. IEEE ICASSP– 84*, pp. 41. B. 4. 1–41. B. 4. 4, 1984.

[186] P. Flandrin, Sageloli, J., Sessarego, J. P., and Zakharia, M., "Application of time–frequency analysis to the characterization of surface waves on elastic targets," *Acoustics Letters*, vol. 10, pp. 23–28, 1986

[187] P. Flandrin, "When is the Wigner–Ville spectrum non–negative?," *Signal Processing III: Theories and Applications*, vol. 1, pp. 239–242, Elsevier, 1986.

[188] P. Flandrin, "On detection–estimation procedures in the time–frequency plane," *Proc. IEEE ICASSP– 86*, vol. 4, pp. 2331–2334, 1986.

[189] P. Flandrin, "On the positivity of the Wigner–Ville spectrum," *Signal Processing*, vol. 11, pp. 187–189, 1987.

[190] P. Flandrin, "Time–frequency receivers for locally optimum detection, *Proc. IEEE ICASSP– 88*, vol. 5, pp. 2725–2728, 1988.

[191] P. Flandrin, "A time–frequency formulation of optimum detection," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 36, pp. 1377–1384, 1988.

[192] P. Flandrin, "Maximum signal energy concentration in a time–frequency domain," *Proc. IEEE ICASSP– 88*, vol. 4, pp. 2176–2179, 1988.

[193] P. Flandrin, "Time–frequency and time–scale," *Proceedings of the Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling*, pp. 77–80, 1988.

[194] P. Flandrin, "Non–destructive evaluation in the time–frequency domain by means of the Wigner–Ville distribution," *Signal Processing and Pattern Recognition in Nondestructive Evaluation of Materials*, ed. C. H. Chen, Springer–Verlag, pp. 109–116, 1988.

[195] P. Flandrin, "Some aspects of non–stationary signal processing with emphasis on time–frequency and time–scale methods," *Wavelets: Time–Frequency Methods and Phase Space*, J. M. Combes, A. Grossman, and P. Tchamitchian, eds., pp. 68–98, Springer Verlag, 1989.

[196] P. Flandrin, "Scale–Invariant Wigner Spectra and Self–Similarity" in: *Signal Processing V – Theories and Applications*, L. Torres et al. eds., pp. 149–152, Elsevier, 1990.

[197] P. Flandrin and O. Rioul, "Affine smoothing of the Wigner–Ville distribution," *Proc. IEEE ICASSP–90*, pp. 2455–2458, 1990.

[198] P. Flandrin, "On the Time–Scale Analysis of Self Similar Processes," *The Role of Wavelets in Signal Processing Applications*, edited by: B. W. Suter, M. E. Oxley, and G. T. Warhola, AFIT/EN–TR–92–3, pp. 137–150, 1992.

[199] J. R. Fonollosa and C. L. Nikias, "Wigner Polyspectra: Higher–Order Spectra in Time Varying Signal Processing," *Proc. IEEE ICASSP– 91*, pp. 3085–3088, 1991.

[200] J. R. Fonollosa and C. L. Nikias, "Wigner Higher Order Moment Spectra: A review," *Proc. SPIE*, vol. 1770, pp. 280–289, 1992.

[201] J. R. Fonollosa and C. L. Nikias, "Wigner Higher–Order Moment Spectra: Definition, Properties, Computation and Application to Transient Signal Analysis," *IEEE Trans. Signal Processing*, vol. 41, pp. 245–266, 1993.

[202] J. R. Fonollosa and C. L. Nikias, "A new positive time–frequency distribution," *Proc. IEEE ICASSP–94*, vol. 4, pp. 301–304, 1994.

[203] B. D. Forrester, "Use of the Wigner–Ville distribution in helicopter fault detection," *Australian Symp. on Signal Processing and Its Applications – 89*, pp. 78–82, 1989.

[204] B. D. Forrester, "Time–frequency analysis in machine fault detection," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 406–423, Longman Cheshire, 1992.

[205] M. L. Fowler, L. H. Sibul, "A Unified Formulation for Detection Using Time–Frequency and Time–Scale Methods," *Asilomar Conf. on Sig., Syst. and Comp.*, pp. 637–642, 1991.

[206] W. L. J. Fox, J. C. Luby, J. W. Pitton, P. J. Laughlin, and L. E. Atlas, "Sonar and Radar Range Doppler Processing Using Cone–Shaped Kernel Time–Frequency Representation," *24th Asilomar Conference on Signals, Systems, and Computers*, pp. 1079–1083, 1990.

[207] B. Friedlander and B. Porat, "Detection of transient signals by the Gabor representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 169–180, 1989.

[208] D. H. Friedman, "Instantaneous–frequency distribution vs. time: an interpretation of the phase structure of speech," *Proc. IEEE ICASSP– 85*, vol. 3, pp. 1121–1124, 1985.

[209] C. Frishberg, "Wigner distribution formulation of the crystallography problem," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 488–497, Longman Cheshire, 1992.

[210] D. Gabor, "Theory of communication," *Journal of the IEE*, vol. 93, pp. 429–457, 1946.

[211] W. A. Gardner, *Statistical Spectral Analysis – A Non–Probabilistic Theory*, Prentice–Hall, Englewood Cliffs, 1988.

[212] H. Garudadri, M. P. Beddoes, J. H. V. Gilbert and A. P. Benguerel, "Identification of invariant acoustic cues in stop consonants using the Wigner distribution," *Proc. Int. Symp. on Applied Signal Processing and Digital Filtering IASTED 85*, ed. M. H. Hamza, Acta Press, pp. 196–200, 1985.

[213] H. Garudadri, M. P. Beddoes, A. P. Benguerel and J. H. V. Gilbert, "On computing the smoothed Wigner distribution," *Proc. IEEE ICASSP-87*, pp. 1521–1524, 1987

[214] C. G. Gaunaurd, and H. C. Strifors "Frequency and time domain analysis of transient resonance scattering resulting from the interaction of a sound pulse with submerged elastic shells," *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 40, pp. 313–324, 1993.

[215] C G. Gaunaurd, and H. C. Strifors, "Signal processing of ideal acoustic echoes from submerged shells in the time–frequency domain, via Wigner–type distributions," *Proc. Intern. Conference on Theoretical and Computational Acoustics*, D. Lee et. all., editors, World Scientific, 1994.

[216] N. L. Gerr, "Introducing a Third–Order Wigner Distribution," *Proc. of the IEEE*, vol. 76, pp. 290–292, 1988.

[217] G. B. Giannakis and A. V. Dandawate, "Consistent Kth–Order Time–Frequency Representations for (Almost) Cyclostationary Signals," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 123–126, 1992.

[218] C Gonzalo, J. Bescos, L. R. Berriel–Valdos and J. Santamaria, "Space–variant digital filtering through the Wigner distribution," *Signal Processing IV Theories and Applications*, Eds. J. L. Lacoume, A. Chehikian, N. Martin and J. Malbos, Elsevier, vol. 2, pp. 639–642, 1988.

[219] C. Gonzalo, J. Bescos, L. R. Berriel–Valdos and J. Santamaria, "Space–variant filtering through the Wigner distribution function," *Applied Optics*, vol. 28, pp. 730–736, 1989

[220] C. Gonzalo and J. Bescos, "Digital implementation of the 4D Wigner distribution function: Application to space variant processing of real images," *Signal Processing V. Theories and Applications*, Eds. L. Torres, E. Masgrau and M. A. Lagunas, Elsevier, vol. 1, pp. 717–720, 1990.

[221] O. D. Grace, "Instantaneous power spectra," *J. Acoust. Soc Am.*, vol. 69, pp. 191–198, 1981.

[222] J E. Gray, T. R. Rice, J. E. Conte, and G. Watson, "A Method for Determining the Non–linear Time Dependence of Windowed Data Gathered from Scattered Signals," *Proc. IEEE-SP Int. Sympos Time–Frequency and Time–Scale Analysis*, pp. 525–528, 1992.

[223] Y. Grenier, "Parametric Time–Frequency Representations," in: *Signal Processing*, Eds. J. L. Lacoume, T. S. Durrani and R. Stora, Elsevier, pp. 339–397, 1987.

[224] C. Griffin, P. Rao, and F. J. Taylor, "Roundoff error analysis of the discrete Wigner distribution using fixed–point arithmetic," *Proc. IEEE ICASSP-89*,

[225] C. Griffin and A. H. Nuttall, "Comparison of two Kernels for the Modified Wigner Distribution Function," *Proc. SPIE*, vol. 1566, pp. 439–451, 1991.

[226] D. W. Griffin and J. S. Lim, "Signal estimation from modified short–time Fourier transform," *Proc IEEE ICASSP-83*, pp. 804–807, 1983.

[227] M. F. Griffin, A. M. Finn, and S. Tulpule, "Diagnostic Signal Analysis Using Rank Reduction and The Wigner Distribution," in: *Proc. 24th Asilomar Conference on Signals, Systems, and Computers*, vol. 2, 1068-107, 1990

[228] M. F. Griffin and S. Tulpule, "Time–Frequency Distributions for Propulsion–System Diagnostics," *Proc. SPIE*, vol. 1566, pp. 179–189, 1991.

[229] M. F. Griffin, A. M. Finn, and S. Tulpule, "Signal Enhancement for Diagnostic Signal Analysis," *Proc IEEE ICASSP-91,*, 3281–3284, 1991

[230] L. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 23, pp. 207–222, 1975.

[231] R. Groblicki, "A generalization of Cohen's theorem," *Jour. Math. Phys.*, vol. 24, pp. 841–843, 1983.

[232] D. Groutage, J. Schemp and L. Cohen, "Characterization and Analysis of Marine Sounds Using Time–Frequency and Time–Prony Techniques," in: *OCEANS 94 OSATES*, to appear, 1994

[233] D. Groutage and D. Bennink, "Analysis of Resonant Structures Using Advanced Time–Prony Techniques," *ASME Journal of Acoustics and Vibration*, submitted, 1994

[234] Z. Guo, L.–G. Durand and H. C. Lee, "Comparison of Time–Frequency Distribution Techniques for Analysis of Simulated Doppler Ultrasound Signals of the Femoral Artery" *IEEE Trans. on Biomed Engr.*, vol. 41, 332–342, 1994.

[235] Z. Guo, L.–G. Durand and H. C. Lee, "The Time–Frequency Distributions of Nonstationary Signals Based on a Bessel Kernel," *IEEE Trans. Signal Processing*, vol. 42, pp. 1700–1707, 1994

[236] M. S. Gupta, "Definitions of instantaneous frequency and frequency measurability," *Amer. Jour Phys.*, vol. 43, pp. 1087–1088, 1975.

[237] A. K. Gupta, T. Iwai, and T. Asakura, "Parallel generation of the sectional Wigner distribution function for two–dimensional objects," *Jour. of Optics*, vol. 16, pp. 10–19, 1987.

[238] J. K. Hammond, "Evolutionary spectra in random vibrations," *Jour. Roy. Statist. Soc.*, vol. B 35, pp. 167–188, 1983.

[239] J. K. Hammond, Y. H. Tsao and R. F. Harrison, "Evolutionary spectral density models for random processes having a frequency modulated structure," *Proc. IEEE ICASSP– 83*, pp. 261–264, 1983.

[240] J. K. Hammond, J. S. Lee and R. F. Harrison, "The relationship between Wigner–Ville and evolutionary spectra for frequency modulated random processes," *Proc. IEEE ICASSP– 86*, pp. 2327–2330, 1986.

[241] F J. Harris and A. Abu Salem, "Performance comparison of Wigner–Ville based techniques to standard FM–discriminators for estimating instantaneous frequency of rapidly slewing FM sinusoid in the presence of noise," *Proc. SPIE*, vol. 975, pp. 232–244, 1989.

[242] F. J. Harris, "Techniques and limitations of spectrum analysis with the discrete Fourier transform," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 184–207, Longman Cheshire, 1992.

[243] J. D. Harris and D. Nelson, "Glottal Pulse Alignment in Voiced Speech for Pitch Determination," *Proc. IEEE ICASSP– 93*, pp. 519–522, 1993.

[244] S. Hearon and M. G. Amin, "Statistical trade offs in modern t–f distributions," *Proceedings of Asilomar Conference on Circuits, Systems, and Computers*, vol. 2, pp. 359–363, 1991.

[245] S. Hearon and M. Amin, "Minimum variance time–frequency distribution kernels," *IEEE Transactions on Signal Processing*, to appear, 1994.

[246] C. W. Helstrom, "An expansion of a signal in Gaussian elementary signals," *IEEE Trans. on Information Theory*, vol. 12, pp. 81–82, 1966.

[247] M. Hillery, R. F. O'Connell, M. O. Scully and E. P. Wigner, "Distribution functions in physics: Fundamentals," *Physics Reports*, vol. 106, pp. 121–167, 1984.

[248] R. Hippenstiel and P. De Oliveira, "Time–varying spectral estimation using the instantaneous power spectrum (IPS)," *IEEE Trans. Sig. Proc.*, vol. 38, pp. 1752–1759, 1990.

[249] F. Hlawatsch, "Interference terms in the Wigner distribution," in: *Digital Signal Processing–84*, Eds. V. Cappellini and A. G. Constantinides, Elsevier, 1984.

[250] F. Hlawatsch, "Transformation, Inversion and Conversion of Bilinear Signal Representations," *Proc. IEEE ICASSP– 85*, pp. 1029–1032, 1985

[251] F. Hlawatsch and W. Krattenthaler, "Signal Synthesis from Unitary Time–Frequency Signal Representations," In *Signal Processing III: Theories and Applications*, eds. I. T. Young et al., pp. 37–40, Elsevier, 1986.

[252] F. Hlawatsch, "Unitary Time–Frequency Signal Representations," In *Signal Processing III. Theories and Applications* , eds. I. T. Young et al., pp. 33–36, Elsevier, 1986.

[253] a) F. Hlawatsch, "A note on 'Wigner distribution for finite duration or band–limited signals and limiting cases'," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 927–929, 1988. b) Author's reply, ibid.

[254] F. Hlawatsch, W. Kozek and W. Krattenthaler, "Time–frequency subspaces and their application to time–varying filtering," *Proc. IEEE ICASSP– 90*, pp. 1607–1610, 1990.

[255] F. Hlawatsch, "Duality and Classification of Bilinear Time–Frequency Signal Representations," *IEEE Trans. Signal Processing*, vol. 39, pp. 1564–1574, 1991.

[256] F. Hlawatsch, A. Papandreou, and G. F. Boudreaux–Bartels, "The Hyperbolic Class of Quadratic Time–Frequency Representations Part II: Regularity and Unitarity, Subclasses, and the Altes–Marinovich Distribution," *IEEE Trans. Signal Processing*, submitted, 1992.

[257] F. Hlawatsch and G. F. Boudreaux–Bartels, "Linear and Quadratic Time–Frequency Signal Representations," *IEEE Signal Processing Magazine*, vol. 9, pp. 21–67, 1992.

[258] F. Hlawatsch, "Regularity and Unitarity of Bilinear Time–Frequency Signal Representations," *IEEE Trans. Information Theory*, vol. 38, pp. 82–94, 1992.

[259] F. Hlawatsch, "Bilinear Signal Synthesis," *IEEE Trans. Signal Processing* vol. 40, pp. 352–363, 1992.

[260] F. Hlawatsch and W. Kozek, "Time–Frequency Weighting and Displacement Effects in Linear, Time–Varying Systems," *Proc IEEE ISCAS–92*, pp. 1455–1458, 1992

[261] F. Hlawatsch, A. Papandreou, and G. F. Boudreaux–Bartels, "The Power Classes of Quadratic Time–Frequency Representations: A Generalization of the Affine and Hyperbolic Classes," *Proc. IEEE ICASSP– 93*, pp. 1265–1270, 1993

[262] F. Hlawatsch, "Time–Frequency Design and Processing of Signals via Smoothed Wigner Distributions," *IEEE Trans. Signal Processing*, vol. 41, pp. 278–287, 1993.

[263] F Hlawatsch, A Papandreou, and G. F Boudreaux–Bartels, "Regularity and Unitarity of Affine and Hyperbolic Time–Frequency Representations," *Proc IEEE ICASSP– 93*, vol. 3, pp. 245–248, 1993.

[264] F. Hlawatsch and R. L. Urbanke, "Bilinear Time–Frequency Representations of Signals: The Shift–Scale Invariant Class," *IEEE Trans Signal Processing*, vol. 42, pp. 357–366, 1994

[265] F. Hlawatsch and W. Kozek, "Second–Order Time–Frequency Synthesis of Nonstationary Random Processes," *IEEE Trans. Info Theory*, to appear, 1994

[266] R. L. Hudson, "When is the Wigner quasi–probability density non–negative?," *Rep. Math. Phys.*, vol. 6, pp. 249–252, 1974

[267] D. H. Hughes, "Backscattering of Sound by Spherical Shells in Water," Ph. D. dissertation, Washington State University, 1992.

[268] D. H Hughes and P L. Marston, "Local temporal variance of Wigner's distribution function as a spectroscopic observable. Lamb wave resonances of a spherical shell," *J Acoust. Soc. Am.*, vol 94, p. 499, 1993.

[269] D. H. Hughes, C. F. Gaumond and L. R. Dragonette, "Synthesized wave packet basis for monostatic scattering from a randomly ribbed, finite cylindrical shell," *J Acoust. Soc. Am.*, submitted, 1994

[270] J. Imberger and B. Boashash, "Application of the Wigner–Ville distribution to temperature gradient microstructure: A new technique to study small–scale variations," *Jour of Physical Oceanography*, vol. 16, pp. 1997–2012, 1986.

[271] M. Jacob and P. Fortier, "An Interpretation of the Spectrogram for Discrete–Time Signals," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 281–284, 1992.

[272] L. D. Jacobson and H. Wechsler, "The composite pseudo Wigner distribution (CPWD): A computable and versatile approximation to the Wigner distribution (WD)," *Proc IEEE ICASSP– 83*, pp. 254–256, 1983.

[273] L. D. Jacobson and H. Wechsler, "Joint spatial/spatial–frequency representations," *Signal Processing*, vol. 14, pp. 37–68, 1988.

[274] Janse, C. P., and Kaizer, A. J. M., "Time–Frequency Distributions of Loudspeakers. The Application of the Wigner Distribution," *J. Audio Eng. Soc.*, vol. 31, pp. 198– 223, 1983.

[275] A. J. E. M. Janssen, "Positivity of weighted Wigner distributions," *SIAM Journal of Mathematical Analysis*, vol. 12, pp. 752–758, 1981.

[276] A. J. E. M. Janssen, "On the locus and spread of pseudo–density functions in the time–frequency plane," *Philips Journal of Research*, vol. 37, pp. 79–110, 1982

[277] A. J. E M. Janssen, "Positivity properties of phase–plane distribution functions," *Jour Math. Phys*, vol. 25, pp. 2240–2252, 1984.

[278] A. J. E. M. Janssen and T. A. C. M. Claasen, "On positivity of time–frequency distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 1029–1032, 1985.

[279] A. J. E. M. Janssen, "Bilinear phase–plane distribution functions and positivity," *Jour Math. Phys*, vol. 26, pp. 1986–1994, 1985.

[280] A. J. E. M. Janssen, "A note on 'Positive time–frequency distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 701–703, 1987. See also: Author' reply, ibid.

[281] A. J. E. M. Janssen, "Positivity of time–frequency distribution functions," *Signal Processing*, vol. 14, pp. 243–252, 1988

[282] J. Jeong and W. J Williams, "Variable Windowed Spectrograms: Connecting Cohen's Class and the Wavelet Transform," *ASSP Workshop on Spectrum Estimation and Modeling*, pp. 270–273, 1990

[283] J. Jeong, G. S. Cunningham and W. J. Williams. "Instantaneous Frequency and Kernel Requirements for Discrete Time–Frequency Distributions," *Proc SPIE*, vol. 1348, pp. 170–180, 1990.

[284] J. Jeong and W. J. Williams, "A new formulation of generalized discrete–time time–frequency distributions," *Proc IEEE ICASSP– 91*, pp. 3189–3192, 1991.

[285] J. Jeong and W. Williams, "Kernel design for reduced interference distributions," *IEEE Trans Sig. Proc.*, vol. 40, pp. 402–412, 1992.

[286] J. Jeong and W. J. Williams. "Alias–Free Generalized Discrete–Time Time–Frequency Distributions," *IEEE Trans. on Signal Processing*, vol. 40, pp. 2757–2765, 1992.

[287] J. Jeong and W. J. Williams. "Time–Varying Filtering and Signal Synthesis," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 389–405, Longman Cheshire, 1992.

[288] J. Jeong, W. J. Williams. "A Unified Fast Recursive Algorithm for Data Shuffling in Various Orders," *IEEE Trans on Signal Processing*, vol. 40, pp. 1091–1095, 1992.

[289] J. Jeong and W. J. Williams. "Mechanism of the Cross–Terms in Spectrograms," *IEEE Trans. on Signal Processing*, vol. 40, pp. 2608–2613, 1992.

[290] D. L. Jones and T. W. Parks, "A resolution comparison of several time–frequency representations," *IEEE Trans. on Signal Processing*, vol. 40, pp. 413–420, 1992.

[291] D. L. Jones and R. G. Baraniuk, "An adaptive optimal kernel time–frequency representation," *Proc. IEEE ICASSP– 93*, vol. 4, 109–112, 1993.

[292] G. Jones and B. Boashash, "Instantaneous frequency, time delay, instantaneous bandwidth and the analysis of multicomponent signals," *Proc. IEEE ICASSP– 90*, pp. 2467–2470, 1990.

[293] S. Kadambe, G. F. Boudreaux–Bartels, P. Duvaut, "Window Length Selection for Smoothing Wigner Distributions by Applying an Adaptive Filter," *Proc. IEEE ICASSP– 89*, pp. 52.6.1–4, 1989.

[294] S. Kadambe and G. F. Boudreaux–Bartels, "A Comparison of the Existence of 'Cross Terms' in the Wigner Distribution and the Squared Magnitude of the Wavelet Transform and the Short Time Fourier Transform," *IEEE Trans on Signal Processing*, vol. 40, pp. 2498–2517, 1992.

[295] J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," *Proc. IEEE ICASSP– 90*, pp. 381–384, 1990.

[296] J. F. Kaiser, "Some Useful Properties of Teager's Energy Operators," *Proc. IEEE ICASSP– 93*, vol. 3, pp. 149–152, 1993.

[297] S. Kay and G. F. Boudreaux–Bartels, "On the optimality of the Wigner distribution for detection," *Proc. IEEE ICASSP– 85*, pp. 1017–1020, 1985.

[298] S. M. Kay, *Modern Spectral Estimation*, Prentice–Hall, Englewood Cliffs, 1987.

[299] S. A. Kayhan, L. F. Chaparro, and A. El–Jaroudi, "Wold–Cramer Evolutionary Spectral Estimators," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 115–118, 1992.

[300] A. S. Kayhan, A. El–Jaroudi, L. F. Chaparro, "The Evolutionary Periodogram for Non–Stationary Signals," *IEEE Trans. on Signal Processing*, vol. 42, pp. 1527–1536, 1994.

[301] A. S. Kayhan, A. El–Jaroudi, and L. F. Chaparro, "Data–Adaptive Evolutionary Spectral Estimation," *IEEE Trans. on Signal Processing*, to appear, 1994.

[302] E. Kelly and R. P. Wishner, "Matched–Filter Theory Velocity, Accelerating Targets," *IEEE Trans on Mil. Electr.*, vol. 9, pp. 56–69, 1965.

[303] E. L. Key, E. M. Fowle and R. D. Haggarty, "A method of designing signals of large time–bandwidth product," *IRE Intern. Convent. Rec.*, vol. 4, pp. 146–154, 1961.

[304] A. Khinchin, Bull. Univ. Moscow, 1, 1937.

[305] J. G. Kirkwood, "Quantum statistics of almost classical ensembles," *Phys. Rev.*, vol. 44, pp. 31–37, 1933.

[306] J. R. Klauder, A. C. Price, S. Darlington and W. J. Albershein, "The theory and design of chirp radars," *Bell System Tech. J.*, vol. 39, pp. 745–808, 1960.

[307] J. R. Klauder, "Path Integrals For Affine Variables," in: *Functional Integration: Theory and Applications*, J. P Antoine and E. Tirapogui, eds. Plenum Press, New York, 1980.

[308] Kobayashi, F., and Suzuki, H., "Time–varying signal analysis using squared analytic signals," *Proc. IEEE ICASSP– 87*, vol. 3, pp. 1525–1528, 1987.

[309] T. E. Koczwara and D. L. Jones, "On mask selection for time–varying filtering using the Wigner distribution," *Proc. IEEE ICASSP– 90*, pp. 2487–2490, 1990.

[310] K. Kodera, C. de Villedary and R. Gendrin, "A new method for the numerical analysis of non–stationary signals," *Physics of the Earth and Planetary Interiors*, vol. 12, pp. 142–150, 1976.

[311] K. Kodera, R. Gendrin and C. de Villedary, "Analysis of time–varying signals with small BT values," *IEEE Trans. Acoust., Speech, Signal Processing*, vol 26, pp. 64–76, 1978.

[312] R. Koenig, H. K Dunn and L. Y. Lacy, "The sound spectrograph," *J. Acoust. Soc. Am.*, vol. 18, pp. 19–49, 1946.

[313] W. Kolodziej and D. H. Park, "Detection and Diagnosis of Parameters Change in Linear Systems Using Time–Frequency Transformation," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 369–372, 1992

[314] W. Kozek, "Time–Frequency Signal Processing Based on the Wigner–Weyl Framework," *Signal Processing*, vol. 29, pp. 77–92, 1992.

[315] W. Kozek, "On the Generalized Weyl correspondence and its Application to Time–Frequency Analysis of Linear Systems," *Proc. IEEE-SP Int Sympos. Time–Frequency and Time–Scale Analysis*, pp. 167–170, 1992.

[316] W Krattenthaler and F Hlawatsch, "Improved Signal Synthesis from Pseudo Wigner Distribution," *IEEE Trans. Signal Processing*, vol. 39, pp. 506–509, 1991.

[317] W. Krattenthaler and F. Hlawatsch, "Time–frequency design and processing of signals via smoothed Wigner distributions," *IEEE Trans on Acous., Speech, Signal Processing*, vol. 41, pp 278–287, 1993.

[318] J. G. Kruger and A. Poffyn, "Quantum mechanics in phase space," *Physica*, vol. 85, pp. 84–100, 1976.

[319] B. V. K. V. Kumar and W. C. Carroll, "Performance of Wigner distribution function based detection methods," *Optical Engineering*, vol 23, pp. 732–737, 1984.

[320] B. V. K. V. Kumar, C. P. Neuman and K. J. DeVos, "Discrete Wigner synthesis," *Signal Processing*, vol. 11, pp. 277–304, 1986.

[321] B. V. K. V. Kumar and K. J. DeVos, "Linear system description using Wigner distribution functions," *Proc. SPIE*, vol. 826, pp. 115–124, 1987.

[322] R. Kumaresan, A. G. Sadasiv, C. S. Ramalingam, and J. F. Kaiser, "Instantaneous Non–linear Operators for Tracking Multicomponent Signal Parameters," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 404–407, 1992.

[323] R. Kumaresan and C. S. Ramalingam, "On separating Voiced–Speech into its Components" in: *Proc. of Twenty–Seventh Annual Asilomar Conference on Signals, Systems and Computers*, vol. 2, 1041–1046, 1993.

[324] V. V. Kuryshkin, "Some problems of quantum mechanics possessing a non–negative phase–space distribution functions," *International Journal of Theoretical Physics*, vol. 7, pp. 451–466, 1973.

[325] V. V. Kuryshkin, "Uncertainty principle and the problems of joint coordinate–momentum probability density in quantum mechanics," in: *The uncertainty Principle and the Foundations of Quantum Mechanics*, pp. 61–83, Wiley, 1977.

[326] S. Kwong W. Gang, and O. Y. Zheng, "Fundamental Frequency Estimation Based on Adaptive Time–Averaging Wigner– Ville Distribution," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 413–416, 1992.

[327] U. K. Laine, "Famlet, to be or not to be a wavelet?," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 335–338, 1992.

[328] D. G. Lampard, "Generalization of the Wiener–Khinchin theorem to non–stationary processes," *Jour. Applied Physics*, vol. 25, pp. 802, 1954.

[329] H. Landau, D. Slepian and H. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty," *Bell System Tech. J.*, vol. 40, pp. 43–63, 65–84, 1961.

[330] F. Lari and A. Zakhor, "Automatic Classification of Active Sonar Data Using Time–Frequency Transforms," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 21–24, 1992.

[331] R. M. Lerner, "Representation of signals," in: *Lectures on Communication System Theory*, E. J. Baghdady, ed. , pp. 203–242, McGraw–Hill, 1961

[332] M. J. Levin, "Instantaneous Spectra and Ambiguity Functions," *IEEE Trans. Inform. Theory*, vol. 10, pp. 95–97, 1964.

[333] A. L. Levshin, V. F. Pisarenko and G. A. Pogrebinsky, "On a frequency–time analysis of oscillations," *Ann. Gcophys.*, vol. 28, pp. 211–218, 1972.

[334] Y. Li, G. Eichmann and M. Conner, "Optical Wigner distribution and ambiguity function for complex signals and images," *Optics Communications*, vol. 67, pp. 177–179, 1988.

[335] Z.–b Lin, "Wideband ambiguity function of broadband signals," *J Acoust. Soc. Am.*, vol. 83, 2108–2116, 1988.

[336] Z. Y. Lin and J. De Z. Chen, "Time - Frequency Representation of the Electrogastrogram - Application of the Exponential Distribution," *IEEE Trans. on Biomed Engr.*, vol. 41, pp. 267–275, 1994.

[337] S. C. Liu, "Time–varying spectra and linear transformation," *Bell Systems Tech. J.*, vol. 50, pp 2365–2374, 1971.

[338] P. Loughlin, J. Pitton, L. E. Atlas, "New properties to alleviate interference in time–frequency representations," *Proc. IEEE ICASSP– 91*, pp. 3205–3208, 1991.

[339] P. Loughlin, J. Pitton and L. E. Atlas, "An information–theoretic approach to positive time–frequency distributions," *Proc. IEEE ICASSP– 92*, vol. 5, pp. 125–128, 1992.

[340] P. Loughlin, J Pitton and L. E. Atlas, "Proper time–frequency energy distributions and the Heisenberg uncertainty principle," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 151–154, 1992.

[341] P. Loughlin, L. E. Atlas and J. Pitton, "Advanced time–frequency representations for speech processing," in: *Visual Representations of Speech Signals*, M. Cooke, S. Beete and M. Crawford (eds.), John Wiley & Sons, 1993.

[342] P. Loughlin, J. Pitton and L. E. Atlas, "Bilinear time–frequency representations: new insights and properties," *IEEE Trans Sig. Proc.*, vol. 41, pp. 750–767, 1993.

[343] P. Loughlin, J. Pitton and L. E. Atlas, "Construction of positive time–frequency distributions," *IEEE Trans. Sig. Proc.*, to appear, 1994.

[344] P. Loughlin, J. Pitton and B. Hannaford, "Approximating time–frequency density functions via optimal combinations of spectrograms, *IEEE Sig. Proc. Letters*, submitted, 1994.

[345] P. Loughlin, "Comments on scale invariance of time–frequency distributions," *Signal Processing Letters*, submitted, 1994.

[346] B. Lovell and R. Williamson, "The statistical performance of some instantaneous frequency estimators," *IEEE Trans. Sig. Proc.*, vol. 40, pp. 1708–1723, 1992.

[347] B. Lovell, R. Williamson and B. Boashash, "The relationship between instantaneous frequency and time–frequency representations," *IEEE Trans. Sig. Proc.*, vol. 41, pp. 1458–1461, 1993.

[348] Loynes, R. M. "On the concept of spectrum for nonstationary processes" *J. Roy. Stat. Soc. Ser. B*, vol. 30, pp. 1–30, 1968.

[349] D. B. Malkoff and E. J. Smythe, "An integrated neural network–time–frequency method for detection and classification," in: *Proceedings of SPIE 1990 Technical Symposium on Optical Engineering and Phonetics in Aerospace Sensing, Orlando, Florida*, 1990.

[350] D. B. Malkoff, "Detection and classification by neural networks and time distributions," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 324–348, Longman Cheshire, 1992.

[351] S. G. Mallat and Z. Zhang, "Matching Pursuits with Time–Frequency Dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, p. 3397–3415, 1993.

[352] S. G. Mallat and Z. Zhang, "Adaptive Time–Frequency Decomposition Using Matching Pursuits," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 7–10, 1992.

[353] L. Mandel, "Interpretation of instantaneous frequency," *Amer. Jour. Phys.*, vol. 42, pp. 840–846, 1974.

[354] S. Mann and S. Haykin, "Adaptive "chirplet" transform: An adaptive generalization of the wavelet transform," *Optical Engineering*, vol. 31, pp. 1243–1256, 1992.

[355] S. Mann and S. Haykin, "Time–frequency perspectives: the chirplet transform," *Proc. IEEE ICASSP– 92*, vol. 3, pp. 417–420, 1992.

[356] S. Mann and S. Haykin, "'Chirplets' and 'Warblets': Novel Time–Frequency Methods," *Electronics Letters*, vol. 28, pp. 114–118, 1992,

[357] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On Amplitude and Frequency Demodulation Using Energy Operators," *IEEE Trans Signal Processing*, vol. 41, pp. 1532–1550, 1993.

[358] H. Margenau and R. N. Hill, "Correlation between measurements in quantum theory," *Prog. Theoret Phys.*, vol. 26, pp. 722–738, 1961.

[359] H. Margenau and L. Cohen, "Probabilities in quantum mechanics," in: *Quantum Theory and Reality*, edited by M. Bunge, Springer–Verlag, 1967.

[360] N. M. Marinovich and W. A. Smith, "Use of the Wigner distribution to analyze the time–frequency response of ultrasonic transducers," in: *Proc. IEEE 1984 Ultrasonics Symposium*, 1984.

[361] N. M. Marinovich and W. A. Smith, "Suppression of noise to extract the intrinsic frequency variation from an ultrasonic echo," *Proc. IEEE 1985 Ultrasonics Symposium*, pp. 841–846, 1985.

[362] N. M. Marinovich and G. Eichmann, "Feature extraction and pattern classification in space–spatial frequency domain," *Proc. SPIE*, vol. 579, pp. 15–20, 1985.

[363] N. M. Marinovich and G. Eichmann, "An expansion of the Wigner distribution and its applications," *Proc. IEEE ICASSP–85*, vol. 3, pp. 1021–1024, 1985.

[364] N. M. Marinovich, "The Wigner distribution and the ambiguity function: Generalizations, enhancement, compression and some applications," Ph. D. Dissertation, The City University of New York, NY, 1986.

[365] N. M. Marinovich, and W. A. Smith, "Application of joint time–frequency distributions to ultrasonic transducers," *1986 IEEE Information Symposium on Circuits and Systems*, pp. 50–54, 1986.

[366] N. M. Marinovich, "VLSI architecture of a real–time Wigner distribution processor for acoustic signals," *Proc. IEEE ICASSP–88*, vol. 4, pp. 2112–2115, 1988.

[367] N. M. Marinovitch, "Detection of transients in stationary colored noise by time frequency distributions," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 305–323, Longman Cheshire, 1992.

[368] W. D. Mark, "Spectral analysis of the convolution and filtering of non–stationary stochastic processes," *Jour. Sound Vib.*, vol. 11, pp. 19–63, 1970.

[369] W. D. Mark, "Power spectrum representation for nonstationary random vibration," in: *Random Vibration – Status and recent developments*, edited by D. H. Ielishakoff, Elsevier, 1986.

[370] W. D. Mark, Effects of intensity modulations on the power spectra of random processes," *Jour. Sound and Vibr.*, vol. 119, pp. 451–485, 1987

[371] R. J. Marks II, J. F. Walkup, and T. F. Krile, "Ambiguity function display: an improved coherent processor," *Appl. Opt.*, 16, 746–750, 1977.

[372] R. J. Marks II, *Introduction to Shannon Sampling and Interpolation Theory*", Springer–Verlag, 1991.

[373] S. L. Marple, *Digital Spectral Analysis*, Prentice–Hall, 1987.

[374] L. Marple, T. Brotherton, R. Barton, et al., "Travels through the time–frequency zone: advanced Doppler ultrasound processing techniques," *Proc. 27th Asilomar Conf. on Sigs., Syst. and Comps.*, vol. 2, pp. 1469–1473, 1993.

[375] P. L. Marston, L. Zhang, N. Sun, G. Kaduchak, and D. H. Hughes, "Ray representations of the backscattering of tone bursts by shells in water: Calculations and related experiments," in: *Second International Congress on Recent Developments in Air and Structure–Borne Sound and Vibration*, M. J. Crocker and P. K. Raju, editors, vol 3, 1992.

[376] W. Martin, "Time–frequency analysis of random signals," *Proc IEEE ICASSP–82*, pp. 1325–13281 1982.

[377] W. Martin, "Measuring the degree of non–stationarity by using the Wigner–Ville spectrum," *Proc IEEE ICASSP–84*, pp. 41B.3.1–4, 1984.

[378] W. Martin and P. Flandrin, "Wigner–Ville spectral analysis of non–stationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 1461–1470, 1985.

[379] W. Martin, and P. Flandrin, "Detection of changes in signal structure by using the Wigner–Ville spectrum," *Signal Processing*, vol. 8, pp. 215–233, 1985.

[380] N. H. McCoy, "On the Function in Quantum Mechanics Which Corresponds to a Given Function in Classical Mechanics," *Proc. Natl Acad. Sci. U.S.A.*, vol. 18, pp. 674–676, 1932.

[381] T. J. McHale and G. F. Boudreaux-Bartels, "An Algorithm for Synthesizing Signals from Partial Time-Frequency Models Using the Cross Wigner Distribution," *IEEE Trans. Signal Processing*, vol. 41, pp. 1986–1990, 1993.

[382] W. F. G. Mecklenbräuker, "A tutorial on non-parametric bilinear time-frequency signal representations," in: *Signal Processing*, Eds. J. L. Lacoume, T. S. Durrani and R. Stora, Elsevier, 277–338, 1985.

[383] F. Molinaro and F. Castanie, "A comparison of time-frequency methods," *Signal Processing V. – Theories and Applications*, Eds. L. Torres, E. Masgrau and M. A. Lagunas, Elsevier, vol. 1, pp. 145–149, 1990.

[384] P. M. Moore and Q. T. Tran, "Signal-Dependent Kernel for Reduction of Time-Domain Cross Terms in Time-Frequency Distributions," *Proc. IEEE-SP Int. Sympos. Time-Frequency and Time-Scale Analysis*, pp. 245–248, 1992.

[385] J. M. Morris and S. Qian, "A Cross-Term Free Joint Time-Frequency Representation via Wigner-Ville Distribution Decomposition," *Proc. 29th Annual Allerton Conf. Commun., Contr., & Computing*, pp. 553–554, 1991.

[386] J. M. Morris and D-S Wu, "Some Results on Joint Time-Frequency Representations via Gabor Expansions," *Proc. 1992 Conf. Inform. Sci. & Sys.*, vol. I, pp. 6–10, 1992.

[387] J. M. Morris and D-S. Wu, "On Alias-Free Formulations of Discrete-Time Cohen's Class of Distributions," *IEEE Trans. on Signal Processing*, submitted, 1994.

[388] J. M. Morris and D-S. Wu, "On the Resolution of the Crossterm-Deleted Wigner Representation," preprint, 1994.

[389] G. Mourgues, M. R. Feix, J. C. Andrieux and P. Bertrand, "Not necessary but sufficient conditions for the positivity of generalized Wigner functions," *Jour Math. Phys.*, vol. 26, pp. 2554–2555, 1985.

[390] J. E. Moyal, "Quantum mechanics as a statistical theory," *Proc. Camb. Phil. Soc.*, vol. 45, pp. 99–124, 1949.

[391] M. Mugur-Schachter, "A study of Wigner's theorem on joint probabilities," *Found. of Phys.*, vol. 9, pp. 389–404, 1979.

[392] S. K. Mullick and V. A. Topkar, "A Wigner distribution based receiver," *Signal Processing*, vol. 14, pp. 185–196, 1988.

[393] W. Music, W. Fox, P. Loughlin, L. E. Atlas and J. Pitton, "Shift-keyed signal identification using time-frequency processing," in: *Proc. 25th Asilomar Conf. on Signals, Systems and Computers*, vol. 2, 846–850, 1991.

[394] F. J. Narkowich, and R. F. O'Connell, "Necessary and sufficient conditions for a phase-space function to be a Wigner distribution," *Phys. Rev. A*, vol. 34, pp. 1–6, 1986.

[395] S. N. Nawab, T. F. Quatieri and J. S. Lim, "Signal reconstruction from short-time Fourier transform magnitude," *Proc. IEEE ICASSP- 82*, pp. 1046–1048, 1982.

[396] S. H. Nawab, T. F. Quatieri and J. S. Lim, "Algorithms for signal reconstruction from short-time Fourier transform magnitude," *Proc. IEEE ICASSP- 83*, pp. 800–803, 1983.

[397] S. N. Nawab and T. F. Quatieri "Short-Time Fourier Transform," *Advanced Topics in Signal Processing*, ed., J. S. Lim and A. V. Oppenheim, pp. 289–337, Prentice-Hall, 1988.

[398] S. N. Nawab, D. Beyerbach and E. Dorken, "Principal decomposition of time-frequency distribution," *IEEE Trans. Sig. Proc.*, vol. 41, pp. 3182–3186, 1993.

[399] D. Nelson, D. Barbour, and M. Pierce, "Speech Analysis by Alternate Spectral Analysis Methods," preprint, 1994.

[400] D. Nelson, "The Wavelet Mellin Transform," preprint, 1994.

[401] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectral Analysis: a Nonlinear Signal Processing Framework*, Prentice-Hall, 1993.

[402] A. H. Nuttall, "On the quadrature approximation to the Hilbert transform of modulated signals," *Proc. IEEE*, vol. 54, pp. 1458, 1966; and Reply by Bedrosian.

[403] A. H. Nuttall, "Wigner distribution function: Relation to short-term spectral estimation, smoothing, and performance in noise," *Naval Underwater Systems Center*, Technical Report 8225, 1988.

[404] A. H. Nuttall, "The Wigner distribution function with minimum spread," *Naval Underwater Systems Center*, Technical Report 8317, 1988.

[405] A. H. Nuttall, "Alias-free Wigner distribution function and complex ambiguity function for discrete signals," *Naval Underwater Systems Center*, Technical Report 8533, 1989.

[406] R. F. O'Connell and E. P. Wigner, "Quantum-mechanical distribution functions: Conditions for uniqueness," *Phys. Lett.*, vol. 83A, pp. 145–148, 1981.

[407] R. F. O'Connell, "The Wigner distribution function – 50th birthday," *Found. of Phys.*, vol 13, pp. 83–92, 1983.

[408] S. Oh, R. J. Marks II, L. E. Atlas and J. A. Pitton, "Kernel Synthesis for Generalized Time-Frequency Distributions Using the Method of Projection Onto Convex Sets," *Proc. SPIE*, vol. 1348, pp. 197–207, 1990.

[409] S. Oh and R. J. Marks II, "Some properties of the generalized time-frequency representation with cone shaped kernel," *IEEE Trans Sig. Proc.*, vol. 40, pp. 1735–1745, 1992

[410] J. R. O'Hair and B. W. Suter , "Kernel Design Techniques for Alias-Free Time-Frequency Distributions," *IEEE Trans. on Signal Processing*, submitted, 1994.

[411] J. R. O'Hair and B. W. Suter , "The Zak Transform and Decimated Time-Frequency Distributions," *IEEE Trans. on Signal Processing*, submitted, 1993.

[412] J. R. O'Hair and B. W. Suter , "Multirate. A New Computational Paradigm," *IEEE Trans. on Signal Processing*, submitted, 1994.

[413] J. R. O'Hair and B. W. Suter , "Kernel Design Techniques for Alias-Free Time-Frequency Distributions," *Proc. IEEE ICASSP- 94*, vol. pp. 333–336, 1994.

[414] A. V. Oppenheim, "Speech spectrograms using the fast Fourier transform," *IEEE Spectrum*, vol. 7, pp. 57–62, 1970.

[415] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 65, pp. 529–541, 1981.

[416] R. S. Orr, J. M. Morris, and S. Qian, "Use of the Gabor Representation for Wigner Distribution Crossterm Suppression," *Proc. IEEE ICASSP- 92*, vol. 5, pp. 29–31, 1992.

[417] J. P. Ovarlez, J. Bertrand, and P. Bertrand, "Computation of affine time-frequency distributions using the fast Mellin transform," *Proc. IEEE ICASSP- 92*, vol. 5, pp. 117–120, 1992.

[418] A. Pacut, W. J. Kolodieji and A. Said, "Discrete domain Wigner distributions – a comparison and an implementation," *Proc. IEEE Int. Symposium on Circuits and Systems*, pp. 1264–1267, 1989.

[419] C. H. Page, "Instantaneous power spectra," *Jour. Appl. Phys*, vol. 23, pp. 103–106, 1952

[420] A. Papandreou and G. F. Boudreaux-Bartels, "Distributions for Time-Frequency Analysis: a Generalization of Choi-Williams and the Butterworth Distribution," *Proc. IEEE ICASSP- 92*, vol. 5, pp. 181–184, Mar. 1992.

[421] A. Papandraou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "Quadratic time-frequency distributions: the new hyperbolic class and its intersection with the affine class," *Sixth SSAP Workshop on Statistical Signal and Array Processing*, pp. 26–29, 1992.

[422] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "A unified framework for the Bertrand distribution and the Altes distribution· the new hyperbolic class of quadratic time-frequency distributions," *Proc. IEEE-SP Int. Sympos. Time-Frequency and Time-Scale Analysis*, pp 27–30, 1992.

[423] A. Papandreou, F. Hlawatsch, and G F. Boudreaux-Bartels, "The Hyperbolic Class of Quadratic Time-Frequency Representations – Part I: Constant-Q Warping, the Hyperbolic Paradigm, Properties, and Members," *IEEE Trans. Signal Processing*, vol. 41, pp. 3425–3444, Dec. 1993.

[424] A. Papandreou and G. F. Boudreaux-Bartels, "A Generalization of the Choi-Williams and the Butterworth Time-Frequency Distributions," *IEEE Trans. Signal Processing*, vol. 41, pp. 463–472, 1993

[425] J. L. Park and H. Margenau, "Simultaneous measurability in quantum theory," *Intern Jour. Theoret Phys.*, vol. 1, pp. 211–283, 1968.

[426] T. W. Park and R. O Shenoy, "Time-frequency concentrated basis functions," *Proc IEEE ICASSP-90*, pp. 2459–2462, 1990.

[427] S. C. Pei, and T. Y. Wand, "The Wigner Distribution of Linear Time–variant Systems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 1681–1684, 1988.

[428] S. C. Pei and E. J. Tsai, "Cross–terms analysis in the modified instantaneous power spectrum," *IEEE Trans. Sig. Proc.*, vol. 41, pp. 477–480, 1993.

[429] S. Peleg and B. Friedlander, "A Technuque for Estimating the Parameters of Multiple Polynomial Phase Signals," *Proc. IEEE-SP Int. Sympos Time–Frequency and Time–Scale Analysis*, pp. 119–122, 1992.

[430] M. R. Perry, "A Correlation Approach to Time–Frequency Representations," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 249–252, 1992.

[431] A. P. Petropulu, " Detection of Multiple Chirp Signals Based on a Slice of the Instantaneous Higher Order Moments," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 261–264, 1992.

[432] F. Peyrin and R. Prost, "A unified definition for the discrete–time discrete–frequency, and discrete–time/frequency Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 858–867, 1986.

[433] F. Peyrin, B. Karoubi, D. Morlet, P. Rubel, P. Desseigne, and P. Touboul, "Time–Frequency and Time–Scale Analysis of high Resolution ECGS, *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 77–80, 1992.

[434] C. A. Pickover, and L. Cohen, "A Comparison of Joint Time–Frequency Distributions for Speech, Signals," *IEEE Int. Symp. Circuits and Systems*, pp. 42–45, 1986.

[435] W. J. Pielemeier and G. H. Wakefield, "Time–frequency and time–scale analysis for musical transcription," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 421–424, 1992.

[436] W. J. Pielemeier and G. H. Wakefield, "A high resolution time–frequency representation for musical instrument signals," *J. Acoust. Soc. Am.*, to appear, 1994.

[437] W. J. Pielemeier, "A Time–Frequency Distribution with Parameter Estimators for a Musical Signal Model," University of Michigan, Ph. D. Thesis, Ann Arbor, Michigan, 1994

[438] J. Pitton, W. Fox, L. E. Atlas, J. Luby and P. Loughlin, "Range–Doppler processing with the cone–kernel time–frequency representation," *IEEE Proc. Pac. Rim Conf. on Comm., Comps. and Sig. Proc.*, pp. 799–802, 1991.

[439] J. Pitton, P. Loughlin and L. E. Atlas, "Positive time–frequency distributions via maximum entropy deconvolution of the evolutionary spectrum," *Proc. IEEE ICASSP– 93*, vol. 4, pp. 436–439, 1993.

[440] J. Pitton, L. E. Atlas and P. Loughlin, "Deconvolution for positive time–frequency distributions," *Proc. 27th Asilomar Conf. on Sigs., Syst. and Comps.*, vol. 2, pp. 1450–1454, 1993.

[441] J. Pitton, L. E. Atlas and P. Loughlin, "Applications of positive time–frequency distributions to speech processing," *IEEE Trans. Speech and Audio Proc.*, to appear, 1994.

[442] M. Poletti, "Linearly swept frequency measurements, time delay spectrometry and the Wigner distribution," *J. Audio Eng. Soc.*, vol. 36, pp. 457–468, 1988

[443] M. Poletti, "The development of a discrete transform for the Wigner distribution and ambiguity function," *J. Acoust. Soc. Am.*, vol. 84, pp. 238–252, 1988

[444] M. Poletti, "The application of linearly swept frequency measurements," *J. Acoust. Soc. Am.*, vol. 84, pp. 599–610, 1988

[445] M. Poletti, "Application of the Wigner–Ville distribution to time delay spectrometry," *2nd IASTED International symposium on signal processing and its applications*, pp. 900–903, 1990.

[446] M. Poletti, "Instantaneous frequency and conditional moments in the time–frequency plane," *IEEE Trans. Signal Processing*, vol. 39., pp. 755–756, 1991.

[447] M. Poletti, "The homomorphic analytic signal," *3rd IASTED International symposium on signal processing and its applications*, pp. 49–52, 1992.

[448] M. Poletti, "Linearly swept frequency measurements and the Wigner–Ville distribution," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 424–444, Longman Cheshire, 1992.

[449] M. Poletti, "The development of instantaneous bandwidth via local signal expansion," *Signal Processing*, vol. 31, pp. 273–281, 1993.

[450] M. Poletti, "A new description of instantaneous spectral moments," *IEEE Trans Signal Processing*, submitted, 1994.

[451] M. Poletti, "The Homomorphic Analytic Signal," *IEEE Trans Signal Processing*, submitted, 1994.

[452] Portnoff, M. R., "Representation of Digital Signals and Systems Based on the Short-Time Fourier Transform," *IEEE Trans. Acoust., Speech, Sig Proc.*, vol. 28, pp. 55–69, 1980.

[453] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans on Acoustics, Speech, Signal Processing*, vol. 28, pp. 55–69, 1980.

[454] T. E. Posch, "Wavelet Transform and Time-Frequency Distributions" *Proc. SPIE*, vol. 1152, pp. 477–482, 1988.

[455] T. E. Posch, "Time-frequency distributions for a wide sense stationary random signal," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 725–727, 1990.

[456] T. E. Posch and B R. Breed, "Space-wave number distributions and matched field processing," in. *Time-Frequency Signal Analysis*, edited by B. Boashash, pp. 467–487, Longman Cheshire, 1992.

[457] T. E. Posch, "The Wave Packet Transform (WPT) as Applied to Signal Processing," *Proc IEEE-SP Int Sympos. Time-Frequency and Time-Scale Analysis*, pp. 143–146, 1992.

[458] R. K. Potter, G. Kopp and H. C Green, *Visible Speech*, Van Nostrand, 1947.

[459] D. Preis, F. Hlawatsch, P. J. Bloom, and J. A. Deer, "Wigner distribution analysis of filters with perceptible phase distortion," *Journal of the Audio Engineering Society*, vol. 35, pp. 1004–1012, 1987.

[460] R. Priebe and G. Wilson, "Application of matched wavelets to identification of metallic transients," *Proc. IEEE-SP Int. Sympos. Time-Frequency and Time-Scale Analysis*, pp. 349–352, 1992

[461] M. B. Priestley, "Evolutionary spectra and nonstationary processes," *J. Roy. Statist. Soc. Ser. B*, vol. 27, 204–237, 1965.

[462] M. B. Priestley, "Evolutionary spectra and non-stationary processes," *J. Roy. Statist. Soc., Ser. B*, vol 30, pp. 1–30, 1968

[463] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press, 1981,

[464] M. B. Priestley, *Non-Linear and Non-Stationary Time Series Analysis*, Academic Press, 1988.

[465] S. Puechmorel and B. Lacaze, "Optimal transforms related to paths in the time-frequency plane," *Proc. IEEE-SP Int. Sympos Time-Frequency and Time-Scale Analysis*, pp. 34–37, 1992.

[466] S. Qian, and J. M. Morris, "A Fast Algorithm for Real Joint Time-Frequency Transformations of Time-Varying Signals," *Electronic Letters*, vol. 26, pp. 537–539, 1990.

[467] S. Qian, and J. M. Morris, "A New Approach for Signal Synthesis from Smoothed Versions of the Wigner Distribution," *Proc IEEE ICASSP- 91*, pp. 3217–3219, 1991.

[468] S. Qian and J. Morris, "Wigner-Ville Distribution and Cross-Term Deleted Representation," *Signal Processing*, vol. 25, pp. 125–144, 1992.

[469] S. Qian and D. Chen, "Signal Representation in Adaptive Gaussian Functions and Adaptive Spectrogram," *Proceeding of 27h Annual Conference on Information Science and Systems*, pp. 59–65, 1993.

[470] S. Qian and D. Chen, "Decomposition of the Wigner-Ville Distribution," *IEEE Trans. on Signal Processing*, to appear, 1994.

[471] A A. Rakes, C. Hester, R. Inguva, and B. V. K. V. Kumar, "Wigner distribution function / Ambiguity function processor," *Proc. SPIE*, vol. 936, pp 260–269, 1988.

[472] P A. Ramamoorthy, V. K. Iyer, Y. and Ploysongsang, "Autoregressive modeling of the Wigner spectrum," *Proc IEEE ICASSP- 87*, vol. 3, pp. 1509–1512, 1987.

[473] P. Rao, C. Griffin, and F. J. Taylor, "Time-delay estimation using the Wigner distribution," in. *Proceedings of the 22nd Asilomar Conference on Signals and Systems*, 1988.

[474] P. Rao and F. J Taylor, "Estimation of instantaneous frequency using the Wigner distribution," *Electronics Letters*, vol. 26, pp. 246–248, 1990.

[475] S. Raz, "Synthesis of signals from Wigner distributions· Representation on biorthogonal bases," *Signal Processing*, vol. 20, pp. 303–314, 1990.

[476] T Reed, and H Wechsler, "Texture segmentation via non-stationarity detection using joint spatial/spatial-frequency representations Digital Signal Processing," *Proceedings of the 5th International Conference on Digital Signal Processing*, ed. V. Capellini and A. G Constantinides, pp. 896–900, Elsevier, 1987

[477] T. Reed and H. Wechsler, "Tracking on nonstationarities for texture fields," *Signal Processing*, vol. 14, pp. 95–102, 1988.

[478] K. S. Riedel, "Optimal Databased Complex Demodulation for Evolutionary Spectra," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 273–276, 1992.

[479] K. S. Riedel, "Data–based kernel estimation of evolutionary spectra," *IEEE Trans. on Signal Processing* vol. 41, pp. 2439–2447, 1993.

[480] K. S. Riedel and A. Sidorenko, "Data Adaptive Kernel Smoothers," *Computers in Physics*, to appear, 1994.

[481] K. S. Riedel and A Sidorenko, "Minimum bias multiple taper spectral estimation," *IEEE Trans. on Signal Processing*, to appear, 1994.

[482] K. S. Riedel, "Kernel Estimation of the Instantaneous Frequency," *IEEE Trans. on Signal Processing*, to appear, 1994.

[483] A. W. Rihaczek, "Hilbert transforms and the complex representation of real signals," *Proceedings of the IEEE*, vol. 54, pp. 434–435, 1966; and Reply by Bedrosian.

[484] A. W. Rihaczek, "Signal energy distribution in time and frequency," *IEEE Trans. Info. Theory* vol. 14, pp. 369–374, 1968.

[485] A. W. Rihaczek, *Principles of High Resolution Radar*, McGraw–Hill, 1968.

[486] M. D. Riley, "Beyond quasi–stationarity: Designing time–frequency representations for speech signals," *Proc. IEEE ICASSP– 87*, pp. 657–660, 1987

[487] M. D. Riley, *Speech time–frequency Representations*, Kluwer, 1989.

[488] O. Rioul, "Wigner–Ville Representations of Signal Adapted to Shifts and Dilations," *AT&T Bell Laboratories Tech. Memo.*, 1988.

[489] O. Rioul and M. Vetterli, "Wavelets and Signal Processing," *IEEE Signal Processing Mag.*, vol. 39, 1564–1574, 1991.

[490] O. Rioul and P. Flandrin, "Time–scale energy distributions: a general class extending wavelet transforms," *IEEE Trans. on Signal Processing*, vol. 40, pp. 1746–1757, 1992.

[491] G. Rizzoni and X. C. Chen, "Detection of Internal Combustion Engine Knock Using Time–Frequency Distributions," *36th. Midwest Symposium on Circuits and Systems*, vol. 1 pp. 360–363, 1993.

[492] R Rohrbaugh, "Application of time–frequency analysis to machinery condition assessment, *Proc. 27th Asilomar Conf. on Sigs., Syst. and Comps.*, vol. 2, pp. 1455–1458, 1993.

[493] R. Rohrbaugh, "Advanced Time-Frequency Analyis: Theory and Application to Machinery Condition Assessment," *Naval Surface Warfare Center*, Technical report SADP-U93/00039-715, 1993.

[494] C. E. Rosenthal and L. Cohen, "Time–Frequency Analysis of Acoustic Scattering," *Proc. 24th Asilomar Conference on Signals, Systems, and Computers*, pp. 1061–1067, 1991.

[495] A. Royer, "Wigner Function as Expectation Value of a Parity Operator," *Phys. Rev. A*, vol. 15, pp. 449–450, 1977.

[496] G J. Ruggeri, "On phase space description of quantum mechanics," *Prog. Theoret. Phys* vol. 46, pp. 1703–1712, 1971.

[497] B. E. A. Saleh and N. S. Subotic, "Time–variant filtering of signals in the mixed time frequency domain," *IEEE Trans Acoust., Speech, Signal Processing*, vol. 33, pp. 1479–1485, 1985.

[498] B. E. A. Saleh and M. K. Asi, "Filtering in the time–frequency domain," *Proc SPIE*, vol. 1152, pp. 426–436, 1989.

[499] B Samimy and G. Rizzoni, "Time–Frequency Analysis for Improved Detection of Internal Combustion Engine Knock," *Proc IEEE–SP Int Sympos. Time–Frequency and Time–Scale Analysis*, to appear, 1994.

[500] B. Samimy and G. Rizzoni, "Experimental Studies of Internal Combustion Engine Knock Detection," *SAE 1995 International Congress and Exposition*, to appear, 1995.

[501] A. M Sayeed and D. L. Jones, "Optimal Kernels For Non–Stationary Spectral Estimation," *Trans. on Signal Processing*, to appear, 1994.

[502] M. R. Schroeder and B. S. Atal, "Generalized short–time power spectra and autocorrelation functions," *J. Acoust. Soc. Am.*, vol. 34, pp. 1679–1683, 1962.

[503] B. Schweizer and A. Sklar, "Probability distributions with given margins: Note on a paper by Finch and Groblicki," *Found of Phys.*, vol. 16, pp. 1061–1064, 1986.

[504] R. J. Sclabassi, M. Sun, D. N. Krieger, P. Jasiukaitis and M. S. Scher, "Time–Frequency Domain Problems in the Neurosciences," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 498–519, Longman Cheshire, 1992.

[505] M. Scully and L. Cohen, "Quasi–Probability Distributions for Arbitrary Operators" in: *The Physics of Phase Space*, edited by Y. S. Kim and W. W. Zachary, Springer Verlag, 1987.

[506] J. P. Sessarego, J. Sageloli, P. Flandrin, and M. Zakharia, "Time–Frequency Analysis of Signals Related to Scattering Problems in Acoustics," in *Wavelets: Time–Frequency Methods and Phase Space*, J. M. Combes, A. Grossman, and Ph. Tchamitchian, Eds., pp. 147–153, Springer Verlag, 1989.

[507] S. I. Shah, L. F. Chaparro and A. S. Kayhan, "Evolutionary Maximum Entropy Spectral Analysis," in: *Proc. IEEE ICASSP– 94*, 1994.

[508] J. Shekel, "Instantaneous frequency," *Proc. IRE.*, vol. 41, pp. 548, 1953.

[509] R. G Shenoy and T. W. Parks, "Affine Wigner distributions," *Proc. IEEE ICASSP– 92*, vol. 5, pp 185–188, 1992.

[510] R. G. Shenoy and T. W. Parks, "Wideband ambiguity Functions and the Affine Wigner Distributions," *Signal Processing*, to appear, 1993.

[511] R. G. Shenoy and T. W. Parks, "The Weyl Correspondence and time–frequency analysis," *IEEE Trans on Signal Processing*, vol. 42, p. 318–331, 1994.

[512] H. Shevrin, W. J. Williams, R. E. Marshall, R. K. Hertel, J. A. Bond and L. A. Brakel, "Event–Related Potential Indicators of the Dynamic Unconscious," *Consciousness and Cognition*, vol. 1, pp. 340–366, 1992.

[513] Y. S. Shin, J. J. Jeon, and S. G. Spooner, "Pseudo Wignerr–Ville Distribution and its Application to Machinery Condition Monitoring," Proceedings of the 47th Meeting of the Mechanical Failures Prevention Group," pp. 235–250, 1993

[514] M. I. Skolnik, *Introduction to Radar Systems*, Mcgraw–Hill Book Co. 1980.

[515] D. Slepian and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty," *Bell System Technical Journal*, vol. 40, pp. 43–64, 1961.

[516] R. Smits, "Accuracy of quasy–stationary analysis of highly dynamic speech signals," *J. Acoust. Soc. Am.*, to appear, 1994.

[517] F. Soto and P. Claverie, "Some properties of the smoothed Wigner function," *Physica*, vol. 109A, pp. 193–207, 1981.

[518] F. Soto and P. Claverie, "When is the Wigner distribution function of multidimensional systems nonnegative?," *Jour. Math. Phys.*, vol. 24, pp. 97–100, 1983.

[519] M. D. Srinivas and E. Wolf, "Some nonclassical features of phase–space representations of quantum mechanics," *Phys. Rev. D*, vol. 11, pp 1477–1485, 1975.

[520] L. Stankovic and S. Stankovic, "Wigner distribution of noisy signals," *IEEE Trans. Sig. Proc.*, vol. 41, pp. 956–960, 1993.

[521] L. Stankovic and S. Stankovic, "An analysis of instantaneous frequency representation using time frequency distributions–generalized Wigner distribution," *IEEE Trans. Sig. Proc.*, to appear, 1994.

[522] L. Stankovic, "A method for time–frequency analysis," *IEEE Trans. Sig Proc.*, vol. 42, pp. 225–229, 1994.

[523] S. Stein and H. H. Szu, "Two dimensional optical processing of one–dimensional acoustic data," *Opt. Eng.*, vol. 21, pp. 804–813, 1982.

[524] L. R. O. Storey, "An investigation of whistling atmospherics," *Phil. Trans. Roy Soc.* vol A246, pp. 113–141, 1953.

[525] H. C Strifors, C. G. Gaunaurd, B. Brusmark, and S. Abrahamsson, "Aspect dependence of impulse radar extracted signature features in the combined time–frequency domain," *Proc. SPIE*, vol. 2234, 1994.

[526] H. C. Strifors and C. G. Gaunaurd "Scattering of short, ultra–wideband electromagnetic pulses by simple targets having reduced radar cross section" in: *Proceed. 2nd Confer. on Ultra–wideband, short–pulse, electromagnetic*, Plenum, 1994.

[527] H. C. Strifors S. Abrahamsson, and C. G. Gaunaurd "Transient scattering of EM–pulses by simple and complex targets in combined time–frequency domain using impulse radar," *Proceed., EURO–EM '94 conference*, 1994.

[528] H. C. Strifors, C. G. Gaunaurd, B. Brusmark, and S. Abrahamsson, "Transient interactions of an EM pulse with a dielectric spherical shell," *IEEE Trans. on Antennas and Propagation*, vol. 42, pp. 453–462, 1994

[529] N. S. Subotic and B. E. A. Saleh, "Generation of the Wigner distribution function of two–dimensional signals by a parallel optical processor," *Optics Letters*, vol. 9, pp. 471–473, 1984.

[530] N. S. Subotic and B. E. A. Saleh, "Optical time–variant processing of signals in the mixed time–frequency domain," *Opt. Commun.*, vol. 52, pp. 259–264, 1984.

[531] M. Sun, C. C. Li, L. N. Sekhar and R. J. Sclabassi, "A Wigner Frequency Analyzer for Nonstationary Signals," *IEEE Trans. on Instrumentation and Measurement*, vol. 38, pp. 961–966, 1989.

[532] M. Sun, C. C. Li, L. N. Sekhar and R. J. Sclabassi, "Elimination of Cross–Components of the Discrete Pseudo Wigner Distribution via Image Processing," *Proc. IEEE ICASSP– 89*, pp. 2230–2233, 1989.

[533] M. Sun and C. C. Li and L. N. Sekhar and R. J. Sclabassi, "Efficient computation of the discrete pseudo–Wigner distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1735–1742, 1989.

[534] M. Sun and R. J. Sclabassi, "Discrete Instantaneous Frequency and its Computation," *IEEE Trans. on Signal Processing*, vol. 41, pp. 1867–1880, 1993.

[535] A. L. Swindlehurst and T. Kailath, "Near–field source parameter estimation using a spatial Wigner distribution approach," *Proc. SPIE*, vol. 975, pp. 86–92, 1989.

[536] H. H. Szu, and J. A. Blodgett, "Wigner distribution and ambiguity function," in: *Optics in Four Dimensions*, ed. L. M. Narducci, pp. 355–381, American Institute of Physics, New York, 1981.

[537] H. H. Szu, "Two dimensional optical processing of one–dimensional acoustic data," *Opt. Eng.*, vol. 21, pp. 804–813, 1982.

[538] H. H. Szu and H. J. Caulfield, "The mutual time–frequency content of two signals," *Proceedings of the IEEE*, vol. 72, pp. 902–908, 1984.

[539] H. H. Szu, "Signal processing using bilinear and nonlinear time–frequency joint–distributions," in: *The Physics of Phase Space*, eds. Y. S. Kim and W. W. Zachary, Springer Verlag, pp. 179–199, 1987.

[540] F. M. Taylor, M. Doyle and D. Chester, "On the Wigner distribution," *Proc. IEEE ICASSP– 83*, pp. 491–494, 1983.

[541] Ya. P. Terletsky, Z. Eksp. Teor. Fiz., vol. 7, p. 1290, 1937.

[542] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, pp. 1055–1096, 1982.

[543] C. H. M. Turner, "On the concept of an instantaneous spectrum and its relationship to the auto-correlation function," *Jour. Appl. Phys.*, vol. 25, pp. 1347–1351, 1954.

[544] P. L. Tyack, W. J. Williams and G. S. Cunningham. "Time–Frequency Fine Structure of Dolphin Whistles," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 17–20, 1992.

[545] S. Usui and H. Akaike, "Wigner distribution analysis of BSPM for optimal sampling," *IEEE Engin. in Medicine and Biology*, vol. 9, pp. 29–32. 1990.

[546] V. G. Vaidya and R. M. Haralick, "The Use of GTFR with Cone Shaped Kernel for Motion Estimation," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 535–538, 1992.

[547] D. E. Vakman, *Sophisticated Signals and the Uncertainly Principle in Radar*, Springer–Verlag, 1968.

[548] D. E. Vakman, "On the Definition of Concepts of Amplitude, Phase and Instantaneous Frequency of a Signal," *Trans. in Radio Eng. and Electron. Phy*, pp. 754–759, 1972.

[549] D. E. Vakman, "Do we know what are the instantaneous frequency and instantaneous amplitude of a signal?," *Trans. in Radio Eng. and Electron. Phy*, vol. 21, pp. 95–100, 1976.

[550] D. E. Vakman, and L. A. Vainshtein, "Amplitude, phase, frequency–fundamental concepts of oscillation theory," *Sov. Phys. Usp.*, vol. 20, 1002–1016, 1978.

[551] D. E. Vakman, "Measuring the Frequency of an Analytic Signal," *Trans. in Radio Eng. and Electron Phy*, pp. 63–69, 1980.

[552] D. E. Vakman and L. A Vainshtein, *Frequency Separation in Theory of Oscilations and Waves*, Nauka Press, Moscow, USSR, 1983.

[553] D. E. Vakman, "Computer Measuring of Frequency Stability and the Analytic Signal," *IEEE Trans Instrumentaion and Measurments*, vol. 43, pp. 668–671, 1994.

[554] D. E. Vakman, "Dynamic Flicker Frequency Modualtion and Noise Instabilty of Oscilators," *IEEE Trans. Circuits and Systems*, vol. 41, pp. 321–325, 1994.

[555] B. Van der Pol, "The fundamental principles of frequency modulation," *Jour. IEE (London)*, vol. 93, pp. 153–158, 1946.

[556] E. F. Velez and R. G. Absher, "Transient analysis of speech signals using the Wigner time–frequency representation," *Proc. IEEE ICASSP- 89*, pp. 2242–2245, 1989.

[557] E. F. Velez and R. G. Absher, "Segmentation and classification of phonemes in continuous speech based on their time–frequency representations," *Proc. SPIE*, vol. 1348, pp. 188–196, 1990.

[558] E. F. Velez and R. G. Absher, "Spectral estimation based on the Wigner–Ville representation," *Signal Processing*, vol. 20, pp. 325–346, 1990.

[559] E. F. Velez, "Speech analysis using the smoothed Wigner–Ville distribution," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 351–374, Longman Cheshire, 1992.

[560] M. Vetterli, "Filter Banks Allowing Perfect Reconstruction," *Signal Proc.*, vol. 10, pp. 219–244, 1986.

[561] J. Ville, "Theorie et applications de la notion de signal analytique," *Cables et Transmissions*, vol. 2A, pp. 61–74, 1948. Translated from French by I. Selin, "Theory and applications of the notion of complex signal," RAND Corporation Technic al Report T–92, Santa Monica, CA, 1958.

[562] D. J. Waldo, and P. R. Chitrapu, "Crossterm Reduction by a Nonlinear Combination of Time–Frequency Signal Representations," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 253–256, 1992.

[563] W. J. Wang, and P. D. McFadden, "Early detection of Gear Failure by Vibration Analysis – I. Calculation of the Time Frequency Distribution", *Mechanical Systems and Signal Processing* vol. 7, pp. 193–203, 1993.

[564] W. A. Watkins, "The Hormonic Interval: Fact or Artifact in Spectral Analyis of Pulse Trains," *Marine Bio–acoustics*, vol. 2, pp. 15–43, 1966.

[565] J. B. Weaver, J. Lu, and D. M. Healy, Jr. "Signal Recovery and Wavelet Reproducing Kernels," *Proc IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 177–180, 1992.

[566] L. G. Weiss and R. K. Young, "Wideband, Nearfield Array Processing and Wavelet Transforms," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*; pp. 521–524, 1992.

[567] L. G. Weiss, "Wavelets and Widband Correlation Processing," *IEEE Signal Processing Magazine*, vol. 11, pp.13–32, 1994.

[568] Weyl, H., *The Theory of Groups and Quantum Mechanics*, E. P. Dutton and Co., 1931. (reprinted by Dover Publications, 1950.)

[569] A. D. Whalen, *Detection of Signals in Noise*, Academic Press, 1971

[570] L. B. White, B. Boashash, "Time–Frequency Analysis of Non–Stationary Signals," IREECON 87, Sydney 1987, Institution of Engineers Press.

[571] L. B. White, B. Boashash, "Pattern Classification in the Time–Frequency Domain," *Conference on Computer Systems and Information Technology*, Brisbane, 1987, Institution of Engineers Press, 1987.

[572] L. B. White, B. Boashash, "Time–Frequency Coherence – A Theoretical Basis for the Cross Spectral Analysis of Non–Stationary Signals," *Signal Processing : Theories, Implementations and Applications*, pp. 18–23, Ed. B. Boashash, Institution of Engineers Press, 1987.

[573] L. B. White and B. Boashash, "A design methodology for the optimal estimation of the spectral characteristics of non–stationary random signals," *ASSP Workshop on Spectral Estimation and Modeling*, pp. 65–70, 1988.

[574] L. B. White and B. Boashash, "On estimating the instantaneous frequency of a Gaussian random signal by the use of the Wigner–Ville distribution," *IEEE Trans. on Acoustics, Speech, Signal Processing*, vol. 36, pp. 417–420, 1988.

[575] L. B. White, "A Fast Algorithm for Pattern Classification in the Time–Frequency Domain," in: *Australian Symposium on Signal Processing and Applications*, 1989.

[576] L. B. White, "Resolution Enhancement in Time–Frequency Signal Processing using Inverse Methods," *Proc. IEEE ICASSP–89*, pp. 1622–1625, 1989.

[577] L. B. White, "Time–Frequency Filtering from Convex Projections," *Proc. SPIE*, vol. 1348, pp. 158–169, 1990.

[578] L. B. White, B. Boashash, "Cross Spectral Analysis of Nonstationary Processes," *IEEE Trans Infor Theory*, vol. 36, pp. 830–835, 1990.

[579] L. B. White, "Estimation of the instantaneous frequency of a noisy signal," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 233–249, Longman Cheshire, 1992.

[580] L. B. White, "Signal Synthesis from Cohen's Class of Bilinear Time–Frequency Distributions using Convex Projections," *Proc. IEEE ICASSP–91*, 1991.

[581] L. B. White, "Transition Kernels for Bilinear Time–Frequency Signal Representations," *IEEE Trans. Signal Processing*, vol. 39, pp 542–544, 1991.

[582] L. B. White, "The Wideband Ambiguity Function and Altes' Q–distribution Constrained Synthesis and Time–Scale Filtering," *IEEE Trans. Information Theory*, vol. 38, pp. 886–892, 1992.

[583] S. E. Widmalm, W. J. Williams and C. Zheng. "Time Frequency Distributions of TMJ Sounds," *Journal of Oral Rehabilitation*, vol. 18, pp. 403–412, 1991.

[584] E. P. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, vol. 40, pp. 749–759, 1932.

[585] E. P. Wigner, "Quantum–mechanical distribution functions revisited," *Perspectives in Quantum Theory*, W. Yourgrau and A. van der Merwe, eds., pp. 25–36, MIT Press, 1971.

[586] J. E. Wilbur, and F. J. Taylor, "High–speed Wigner processing based on a single modulus quadratic residue numbering system," *Proc. IEEE ICASSP–87*, pp. 1031–1034, 1987.

[587] J. E. Wilbur, and F. J. Taylor, "An acoustooptic Wigner processor for time–varying signal analysis," *Proceedings of the IEEE*, vol. 75, pp. 427–428, 1987.

[588] J. E. Wilbur, and F. J. Taylor, "Consistent speaker identification via Wigner smoothing techniques," *Proc. IEEE ICASSP–88*, vol. 1, pp. 591–594, 1988.

[589] R. M. Wilcox, "Exponential Operators and Parameter Differentiation in Quantum Physics," *Jour. Math. Phys.*, vol. 8, pp. 962–982, 1967.

[590] W. J. Williams, H. Shevrin and R. E. Marshall, "Information Modeling and Analysis of Event Related Potentials," *IEEE Trans. on Biomedical Engr.*, vol. 34, pp. 928–937, 1987.

[591] W. J. Williams and J. Jeong. "New Time–Frequency Distributions for the Analysis of Multicomponent Signals," *Proc. SPIE*, vol. 1152, pp. 483–495, 1989.

[592] W. J. Williams, J. Jeong, M. Driscol, and S. Krishnamachari, "Applications and Interpretations of Reduced Interference Time–Frequency Distributions," in: *24th Asilomar Conf. on Signals, Systems, & Computers*, vol. 2, 1049-1054, 1990.

[593] W. J. Williams, M. L. Brown, and A. O. Hero III. "Uncertainty, Information and Time–Frequency Distributions," *Proc. SPIE*, vol. 1566, pp. 144–156, 1991.

[594] W. Williams and J. Jeong, "Reduced interference time–frequency distributions," in: *Time–Frequency Signal Analysis*, edited by B. Boashash, pp. 74–97, Longman Cheshire, 1992.

[595] W. J. Williams. "Time–Frequency Analysis of Biological Signals" *Proc. IEEE ICASSP–93*, vol. 1, pp. 83–96, 1993.

[596] P. J. Wiseman and G. F. Boudreaux–Bartels, "Wigner Distribution Analysis of Acoustic Well Logs," *Proc. IEEE ICASSP–87*, pp. 2237–2240, 1987.

[597] K. M. Wong and Q. Jin, "Estimation of the time–varying, frequency of a signal: The Cramer–Rao bound and the application of Wigner distribution," *IEEE Trans. on Acoustics, Speech Signal Processing*, vol. 38, pp. 519–536, 1990.

[598] J. C. Wood, and D. T. Barry, "Radon transformation of time–frequency distributions for analysis of multicomponent signals," *Proc. IEEE ICASSP–92*, vol. 4, pp. 257–261, 1992.

[599] J. C. Wood, and D. T. Barry, "Radon transform of the Wigner Distribution," *Proc SPIE*, vol 1770, pp. 358–375, 1992.

[600] J. C. Wood, A. J. Buda, and D. T. Barry, "Time–frequency transforms. A new approach to first heart sound frequency analysis," *IEEE Trans in Biomedical Engineering*, vol. 39, pp 728–739, 1992.

[601] J C. Wood, and D. T. Barry, "The variation of first heart sound frequency dynamics across the human chest wall," *Computers in Engineering in Medicine and Biology*, in press, 1994.

[602] J. C. Wood, M. P. Festen, M. J. Lim, J. Buda, and D. T. Barry, "Differential effects of myocardial ischemia on regional first heart sound frequency," *Journal of Applied Physiology*, to appear, 1994.

[603] J. C. Wood, and D. T. Barry, "Radon transformation of time–frequency distributions for analysis of multicomponent signals," *IEEE Trans. of Acoustics, Speech, and Signal Processing*, in press, 1994.

[604] J. C. Wood, and D. T. Barry, "Time–frequency analysis of the first heart sound: techniques and applications," *Engineering in Medicine and Biology Magazine*, submitted, 1994.

[605] J. C. Wood, and D. T. Barry, "Analysis of multicomponent, nonstationary biological signals" *Medical and Life Sciences in Engineering*, in press, 1994.

[606] J. C. Wood, and D. T. Barry, "Linear signal synthesis from the Radon–Wigner domain," *IEEE Trans. of Acoustics, Speech, and Signal Processing*, vol. 42, 2094–2104, 1994.

[607] J. C. Wood, and D. T. Barry, "Tomographic reconstruction of the time–frequency plane," *IEEE Trans. of Acoustics, Speech, and Signal Processing*, vol. 42, 2105–2111, 1994

[608] J. C. Wood, "Frequency dynamics of the first heart sound," Ph. D. thesis, University of Michigan, Ann Arbor, MI, 1994.

[609] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, Pergamon Press, 1953.

[610] H. Xie and J. M. Morris, "Design of Orthogonal Wavelets with Better Time–Frequency Resolution," *Proc. SPIE*, vol. 2242, pp. 878–887, 1994.

[611] N. Yen and L. R. Dragonette, "Inverse acoustic scattering problem: time and frequency approach," *Review of progress in quantitative nondestructive evaluation*, Ed. by D. O. Thompson and D. E. Chimenti, vol. 10A, pp. 89–96, Plenum, 1981

[612] N. Yen, "Time and frequency representation of acoustic signals by means of the Wigner distribution function," *J. Acoust. Soc. Am.*, vol. 81, pp. 1841–1850, 1987.

[613] N. Yen, L. R. Dragonette and S. K. Numrich, "Graphical representations of the spherical–shell–forin–function based on Wigner distribution algorithm," *J. Acoust. Soc. Am.*, vol. 84, suppl. 1, pp. S185, Fall 1988.

[614] N. Yen, L. R. Dragonette, and S. K. Numrich, "Time–frequency analysis of acoustic scattering from elastic objects," *J. Acoust. Soc. Am.*, vol. 87, pp. 2359–2370, 1990.

[615] N. Yen and L. R. Drogonette, "Application of time–frequency analysis to acoustic response from an insonified object," in: *IASTED, Signal Processing and Digital Filtering*, Ed. by M. H. Hamza, ACTA press, 1990.

[616] N. Yen, "Wave packet decomposition," *J. Acoust. Soc. Am.*, vol 95, pp. 889–896, 1994.

[617] R. K. Young, *Wavelet Theory and Its Applications*, Kluwer Academic Publishers, 1993.

[618] K. B. Yu and S. Cheng, "Signal synthesis from Wigner distribution," *Proc IEEE ICASSP– 85*, pp. 1037–1040, 1985.

[619] K. B. Yu and S. Cheng, "Signal synthesis from pseudo–Wigner distribution and applications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1289–1302, 1987

[620] K. B. Yu, "Signal representation and processing in the mixed time–frequency domain," *Proc. IEEE ICASSP– 87*, pp. 1513–1516, 1987.

[621] L. A. Zadeh, "Frequency analysis of variable networks," *Proc. IRE.*, vol. 38, pp. 291, 1950.

[622] E. J. Zalubas and M. G. Amin, "The design of signal dependent time– frequency kernels by the McCllelan Transformation," *Proc. IEEE-SP Int Sympos. Time–Frequency and Time–Scale Analysis*, pp. 91–94, 1992.

[623] H. P. Zaveri, W. J. Williams, L. D. Iasemidis and J. C. Sackellares. "Time–Frequency Representation of Electrocorticograms in Temporal Lobe Epilepsy." *IEEE Trans on Biomed. Engr.*, vol. 39, pp. 502–509, 1992

[624] B. Zhang and S. Sato, "A New Time–Frequency Distribution and Its Application to Speech, Signal Processing," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 417–420, 1992.

[625] B. Zhang and S. Sato, "A time–frequency distribution of Cohen's class with a compound kernel and its application to speech signal processing," *IEEE Trans Sig Proc*, vol. 42, pp. 54–64, 1994.

[626] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone–shaped kernels for generalized time–frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1084–1091, 1990.

[627] C. Zhen, S. E. Widmalm and W. J. Williams, "New time–frequency analyses of EMG and TMJ sound signals," *Proc. IEEE Int. Conf. Engineering in Medicine and Biology*, pp. 741–742, 1989.

[628] C. S. Zheng, W. J. Williams and J. C. Sackellares, "RID Time–Frequency Analysis of Median Filters and Lowpass Filters in Reducing EMG Artifacts in EEG Recording," *IEEE Intl. Conference on Engineering in Medicine and Biology*, vol. 1, pp. 350–351, 1993.

[629] Y. M. Zhu, F. Peyrin and R. Goutte, "Equivalence between two–dimensional analytic and real signal Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1631–1634, 1989.

[630] Y. M. Zhu, F. Peyrin and R. Goutte, "The use of a two dimensional Hilbert transform for Wigner analysis of two–dimensional real signals," *Signal Processing*, vol. 19, pp. 205–220, 1990.

[631] Y. M. Zhu, Y. Gao, and R. Goutte, "An Investigation of the Wigner Model for Texture Discrimination," *Proc. IEEE-SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, pp. 539–542, 1992.

[632] T. P. Zielenski, "New detection method based on the cross–terms mechanism of the Wigner–Ville transform," *Signal Processing V. – Theories and Applications*, Eds. L. Torres, E. Masgrau and M. A. Lagunas, Elsevier, vol. 1, pp. 157–160, 1990.

Index

C

D

E

F

G

L

M

N

O

T