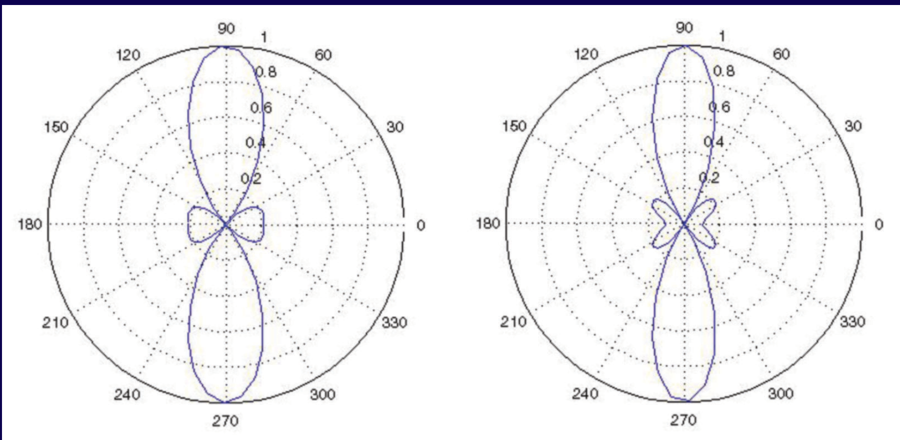
MONOGRAPHS AND RESEARCH NOTES IN MATHEMATICS

# Signal Processing

## A Mathematical Approach

*Second Edition*




Charles L. Byrne

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

# Signal Processing

A Mathematical Approach

*Second Edition*

# MONOGRAPHS AND RESEARCH NOTES IN MATHEMATICS



## Published Titles

*Iterative Optimization in Inverse Problems*, Charles L. Byrne

*Signal Processing: A Mathematical Approach*, *Second Edition*, Charles L. Byrne

*Modeling and Inverse Problems in the Presence of Uncertainty*, H. T. Banks, Shuhua Hu, and W. Clayton Thompson

*Sinusoids: Theory and Technological Applications*, Prem K. Kythe

*Blow-up Patterns for Higher-Order: Nonlinear Parabolic, Hyperbolic Dispersion and Schrödinger Equations*, Victor A. Galaktionov, Enzo L. Mitidieri, and Stanislav Pohozaev

*Set Theoretical Aspects of Real Analysis*, Alexander B. Kharazishvili

*Special Integrals of Gradshetyn and Ryzhik: the Proofs – Volume I*, Victor H. Moll

## Forthcoming Titles

*Stochastic Cauchy Problems in Infinite Dimensions: Generalized and Regularized Solutions*, Irina V. Melnikova and Alexei Filinkov

*Monomial Algebra, Second Edition*, Rafael Villarreal

*Groups, Designs, and Linear Algebra*, Donald L. Kreher

*Geometric Modeling and Mesh Generation from Scanned Images*, Yongjie Zhang

*Difference Equations: Theory, Applications and Advanced Topics, Third Edition*, Ronald E. Mickens

*Method of Moments in Electromagnetics, Second Edition*, Walton C. Gibson

*The Separable Galois Theory of Commutative Rings, Second Edition*, Andy R. Magid

*Dictionary of Inequalities, Second Edition*, Peter Bullen

*Actions and Invariants of Algebraic Groups, Second Edition*, Walter Ferrer Santos and Alvaro Rittatore

*Practical Guide to Geometric Regulation for Distributed Parameter Systems*, Eugenio Aulisa and David S. Gilliam

*Analytical Methods for Kolmogorov Equations, Second Edition*, Luca Lorenzi

*Handbook of the Tutte Polynomial*, Joanna Anthony Ellis-Monaghan and Iain Moffat

*Application of Fuzzy Logic to Social Choice Theory*, John N. Mordeson, Davendar Malik and Terry D. Clark

*Microlocal Analysis on Rˆn and on NonCompact Manifolds*, Sandro Coriasco

## Forthcoming Titles (continued)

*Cremona Groups and Icosahedron*, Ivan Cheltsov and Constantin Shramov

*Special Integrals of Gradshetyn and Ryzhik: the Proofs – Volume II*, Victor H. Moll

*Symmetry and Quantum Mechanics*, Scott Corry

*Lineability and Spaceability in Mathematics*, Juan B. Seoane Sepulveda, Richard W. Aron, Luis Bernal-Gonzalez, and Daniel M. Pellegrinao

*Line Integral Methods and Their Applications*, Luigi Brugnano and Felice Iaverno

*Reconstructions from the Data of Integrals*, Victor Palamodov

*Lineability: The Search for Linearity in Mathematics*, Juan B. Seoane Sepulveda

*Partial Differential Equations with Variable Exponents: Variational Methods and Quantitative Analysis*, Vicentiu Radulescu

*Complex Analysis: Conformal Inequalities and the Bierbach Conjecture*, Prem K. Kythe

MONOGRAPHS AND RESEARCH NOTES IN MATHEMATICS

# Signal Processing

## A Mathematical Approach

*Second Edition*

Charles L. Byrne

University of Massachusetts Lowell

Lowell, Massachusetts, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

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

First edition published in 2005 by A K Peters, Ltd.

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



Printed on acid-free paper
Version Date: 20140730

International Standard Book Number-13: 978-1-4822-4184-6 (Hardback)



**Library of Congress Cataloging-in-Publication Data**

Byrne, Charles L., 1947-
Signal processing : a mathematical approach / Charles L. Byrne, Department of Mathematical Sciences, University of Massachusetts Lowell. -- Second edition.
pages cm. -- (Monographs and research notes in mathematics)
Includes bibliographical references and index.
ISBN 978-1-4822-4184-6
1. Signal processing--Mathematics. I. Title.

TK5102.9.B96 2015
621.382'20151--dc23 2014028555

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

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

*I dedicate this book to Eileen,*
*my wife for forty-four wonderful years.*

*My thanks to my graduate student*
*Jessica Barker, who read most of this book*
*and made many helpful suggestions.*

# *Contents*

# Preface

In graduate school, and for the first few years as an assistant professor, my research was in pure mathematics, mainly topology and functional analysis. Around 1979 I was drawn, largely by accident, into signal processing, collaborating with friends at the Naval Research Laboratory who were working on sonar. Initially, I felt that the intersection of the mathematics that I knew and that they knew was nearly empty. After a while, I began to realize that the basic tools of signal processing are subjects with which I was already somewhat familiar, including Fourier series, matrices, and probability and statistics. Much of the jargon and notation seemed foreign to me, and I did not know much about the particular applications everyone else was working on. For a while it seemed that everyone else was speaking a foreign language. However, my knowledge of the basic mathematical tools helped me gradually to understand what was going on and, eventually, to make a contribution.

Signal processing is, in a sense, applied Fourier analysis, applied linear algebra, and some probability and statistics. I had studied Fourier series and linear algebra as an undergraduate, and had taught linear algebra several times. I had picked up some probability and statistics as a professor, although I had never had a course in that subject. Now I was beginning to see these tools in a new light; Fourier coefficients arise as measured data in array processing and tomography, eigenvectors and eigenvalues are used to locate sonar and radar targets, matrices become images and the singular-value decomposition provides data compression. For the first time, I saw Fourier series, matrices and probability and statistics used all at once, in the analysis of the sampled cross-sensor correlation matrices and the estimation of power spectra.

In my effort to learn signal processing, I consulted a wide variety of texts. Each one helped me somewhat, but I found no text that spoke directly to people in my situation. The texts I read were either too hard, too elementary, or written in what seemed to me to be a foreign language. Some texts in signal processing are written by engineers for engineering students, and necessarily rely only on those mathematical notions their students have encountered previously. In texts such as [116] basic Fourier series and transforms are employed, but there is little discussion of matrices and no mention of probability and statistics, hence no random models.

I found the book [121] by Papoulis helpful, although most of the examples deal with issues of interest primarily to electrical engineers. The books written by mathematicians tend to treat signal processing as a part of harmonic analysis or of stochastic processes. Books about Fourier analysis focus on its use in partial differential equations, or explore rigorously the mathematical aspects of the subject. I was looking for something different. It would have helped me a great deal if there had been a book addressed to people like me, people with a decent mathematical background who were trying to learn signal processing. My hope is that this book serves that purpose.

There are many opportunities for mathematically trained people to make a contribution in signal and image processing, and yet few mathematics departments offer courses in these subjects to their students, preferring to leave it to the engineering departments. One reason, I imagine, is that few mathematics professors feel qualified to teach the subject. My message here is that they probably already know a good deal of signal processing, but do not realize that they know it. This book is designed to help them come to that realization and to encourage them to include signal processing as a course for their undergraduates.

The situations of interest that serve to motivate much of what is discussed in this book can be summarized as follows: We have obtained data through some form of sensing; physical models, often simplified, describe how the data we have obtained relates to the information we seek; there usually isn't enough data and what we have is corrupted by noise, modeling errors, and other distortions. Although applications differ from one another in their details, they often make use of a common core of mathematical ideas. For example, the Fourier transform and its variants play an important role in remote sensing, and therefore in many areas of signal and image processing, as do the language and theory of matrix analysis, iterative optimization and approximation techniques, and the basics of probability and statistics. This common core provides the subject matter for this text. Applications of the core material to tomographic medical imaging, optical imaging, and acoustic signal processing are included in this book.

The term *signal processing* is used here in a somewhat restrictive sense to describe the extraction of information from measured data. I believe that to get information out we must put information in. How to use the mathematical tools to achieve this is one of the main topics of the book.

This text is designed to provide a bridge to help those with a solid mathematical background to understand and employ signal processing techniques in an applied environment. The emphasis is on a small number of fundamental problems and essential tools, as well as on applications. Certain topics that are commonly included in textbooks are touched on only briefly or in exercises or not mentioned at all. Other topics not usually considered to be part of signal processing, but which are becoming increasingly important, such as iterative optimization methods, are included. The book, then, is a rather personal view of the subject and reflects the author's interests.

The term *signal* is not meant to imply a restriction to functions of a single variable; indeed, most of what we discuss in this text applies equally to functions of one and several variables and therefore to image processing. However, there are special problems that arise in image processing, such as edge detection, and special techniques to deal with such problems; we shall not consider such techniques in this text. Topics discussed include the following: Fourier series and transforms in one and several variables; applications to acoustic and electro-magnetic propagation models, transmission and emission tomography, and image reconstruction; sampling and the limited data problem; matrix methods, singular value decomposition, and data compression; optimization techniques in signal and image reconstruction from projections; autocorrelations and power spectra; high-resolution methods; detection and optimal filtering; eigenvector-based methods for array processing and statistical filtering, time-frequency analysis, and wavelets.

The ordering of the first eighteen chapters of the book is not random; these main chapters should be read in the order of their appearance. The remaining chapters are ordered randomly and are meant to supplement the main chapters.

Reprints of my journal articles referenced here are available in pdf format at my website, http://faculty.uml.edu/cbyrne/cbyrne.html.

# Chapter 1

# Introduction

## 1.1 Chapter Summary

We begin with an overview of applications of signal processing and the variety of sensing modalities that are employed. It is typical of remote-sensing problems that what we want is not what we can measure directly, and we must obtain our information by indirect means. To illustrate that point without becoming entangled in the details of any particular application, we present a marbles-in-bowls model of remote sensing that, although simple, still manages to capture the dominate aspects of many real-world problems.

## 1.2 Aims and Topics

The term *signal processing* has broad meaning and covers a wide variety of applications. In this course we focus on those applications of signal processing that can loosely be called *remote sensing*, although the mathematics we shall study is fundamental to all areas of signal processing.

In a course in signal processing it is easy to get lost in the details and lose sight of the big picture. My main objectives here are to present the most important ideas, techniques, and methods, to describe how they relate to one another, and to illustrate their uses in several applications. For signal processing, the most important mathematical tools are Fourier series and related notions, matrices, and probability and statistics. Most students with a solid mathematical background have probably encountered each of these topics in previous courses, and therefore already know some signal processing, without realizing it.

Our discussion here will involve primarily functions of a single real variable, although most of the concepts will have multi-dimensional versions. It is not our objective to treat each topic with the utmost mathematical rigor, and we shall seek to avoid issues that are primarily of mathematical concern.

### 1.2.1 The Emphasis in This Book

This text is designed to provide the necessary mathematical background to understand and employ signal processing techniques in an applied environment. The emphasis is on a small number of fundamental problems and essential tools, as well as on applications. Certain topics that are commonly included in textbooks are touched on only briefly or in exercises or not mentioned at all. Other topics not usually considered to be part of signal processing, but which are becoming increasingly important, such as matrix theory and linear algebra, are included.

The term *signal* is not meant to imply a specific context or a restriction to functions of time, or even to functions of a single variable; indeed, most of what we discuss in this text applies equally to functions of one and several variables and therefore to image processing. However, this is in no sense an introduction to image processing. There are special problems that arise in image processing, such as edge detection, and special techniques to deal with such problems; we shall not consider such techniques in this text.

### 1.2.2 Topics Covered

Topics discussed in this text include the following: Fourier series and transforms in one and several variables; applications to acoustic and EM propagation models, transmission and emission tomography, and image reconstruction; sampling and the limited data problem; matrix methods, singular value decomposition, and data compression; optimization techniques in signal and image reconstruction from projections; autocorrelations and power spectra; high-resolution methods; detection and optimal filtering; eigenvector-based methods for array processing and statistical filtering; time-frequency analysis; and wavelets.

### 1.2.3 Limited Data

As we shall see, it is often the case that the data we measure is not sufficient to provide a single unique answer to our problem. There may be many, often quite different, answers that are consistent with what we have measured. In the absence of prior information about what the answer should look like, we do not know how to select one solution from the many possibilities. For that reason, I believe that to get information out we must put information in. How to do this is one of the main topics of the course. The example at the end of this chapter will illustrate this point.

---

## 1.3 Examples and Modalities

There are a wide variety of problems in which what we want to know about is not directly available to us and we need to obtain information by more indirect methods. In this section we present several examples of remote sensing. The term "modality" refers to the manner in which the desired information is obtained. Although the sensing of acoustic and electromagnetic signals is perhaps the most commonly used method, remote sensing involves a wide variety of modalities: electromagnetic waves (light, x-ray, microwave, radio); sound (sonar, ultrasound); radioactivity (positron and single-photon emission); magnetic resonance (MRI); seismic waves; and a number of others.

### 1.3.1 X-ray Crystallography

The patterns produced by the scattering of x-rays passing through various materials can be used to reveal their molecular structure.

### 1.3.2 Transmission Tomography

In transmission tomography x-rays are transmitted along line segments through the object and the drop in intensity along each line is recorded.

### 1.3.3 Emission Tomography

In emission tomography radioactive material is injected into the body of the living subject and the photons resulting from the radioactive decay are detected and recorded outside the body.

### 1.3.4 Back-Scatter Detectors

There is considerable debate at the moment about the use of so-called *full-body scanners* at airports. These are not scanners in the sense of a CAT scan; indeed, if the images were skeletons there would probably be less controversy. These are images created by the returns, or *backscatter*, of millimeter-wavelength (MMW) radio-frequency waves, or sometimes low-energy x-rays, that penetrate only the clothing and then reflect back to the machine.

The controversies are not really about safety to the passenger being imaged. The MMW imaging devices use about $10,000$ times less energy than a cell phone, and the x-ray exposure is equivalent to two minutes of flying in an airplane. At present, the images are fuzzy and faces are intentionally blurred, but there is some concern that the images will get sharper, will be permanently stored, and eventually end up on the net. Given what is already available on the net, the market for these images will almost certainly be non-existent.

### 1.3.5 Cosmic-Ray Tomography

Because of their ability to penetrate granite, cosmic rays are being used to obtain transmission-tomographic three-dimensional images of the interiors of active volcanos. Where magma has replaced granite there is less attenuation of the rays, so the image can reveal the size and shape of the magma column. It is hoped that this will help to predict the size and occurrence of eruptions.

In addition to mapping the interior of volcanos, cosmic rays can also be used to detect the presence of shielding around nuclear material in a cargo container. The shielding can be sensed by the characteristic scattering by it of muons from cosmic rays; here neither we nor the objects of interest are the sources of the probing. This is about as "remote" as sensing can be.

### 1.3.6 Ocean-Acoustic Tomography

The speed of sound in the ocean varies with the temperature, among other things. By transmitting sound from known locations to known receivers and measuring the travel times we can obtain line integrals of the temperature function. Using the reconstruction methods from transmission tomography, we can estimate the temperature function. Knowledge of the temperature distribution may then be used to improve detection of sources of acoustic energy in unknown locations.

### 1.3.7 Spectral Analysis

In our detailed discussion of transmission and remote sensing we shall, for simplicity, concentrate on signals consisting of a single frequency. Nevertheless, there are many important applications of signal processing in which the signal being studied has a *broad spectrum*, indicative of the presence of many different frequencies. The purpose of the processing is often to determine which frequencies are present, or not present, and to determine their relative strengths. The hotter inner body of the sun emits radiation consisting of a continuum of frequencies. The cooler outer layer absorbs the radiation whose frequencies correspond to the elements present in that outer layer. Processing these signals reveals a spectrum with a number of missing frequencies, the so-called *Fraunhofer lines*, and provides information about the makeup of the sun's outer layers. This sort of *spectral analysis* can be used to identify the components of different materials, making it an important tool in many applications, from astronomy to forensics.

### 1.3.8 Seismic Exploration

Oil companies want to know if it is worth their while drilling in a particular place. If they go ahead and drill, they will find out, but they would like to know what is the chance of finding oil without actually drilling. Instead, they set off explosions and analyze the signals produced by the seismic waves, which will tell them something about the materials the waves encountered. Explosive charges create waves that travel through the ground and are picked up by sensors. The waves travel at different speeds through different materials. Information about the location of different materials in the ground is then extracted from the received signals.

### 1.3.9 Astronomy

Astronomers know that there are radio waves, visible-light waves, and other forms of electro-magnetic radiation coming from the sun and distant regions of space, and they would like to know precisely what is coming from which regions. They cannot go there to find out, so they set up large telescopes and antenna arrays and process the signals that they are able to measure.

### 1.3.10 Radar

Those who predict the weather use radar to help them see what is going on in the atmosphere. Radio waves are sent out and the returns are analyzed and turned into images. The location of airplanes is also determined by radar. The radar returns from different materials are different from one another and can be analyzed to determine what materials are present. Synthetic-aperture radar is used to obtain high-resolution images of regions of the earth's surface. The radar returns from different geometric shapes also differ in strength; by avoiding right angles in airplane design *stealth* technology attempts to make the plane invisible to radar.

### 1.3.11 Sonar

Features on the bottom of the ocean are imaged with sonar, in which sound waves are sent down to the bottom and the returning waves are analyzed. Sometimes near or distant objects of interest in the ocean emit their own sound, which is measured by sensors. The signals received by the sensors are processed to determine the nature and location of the objects. Even changes in the temperature at different places in the ocean can be determined by sending sound waves through the region of interest and measuring the travel times.

### 1.3.12 Gravity Maps

The pull of gravity varies with the density of the material. Features on the surface of the earth, such as craters from ancient asteroid impacts, can be imaged by mapping the variations in the pull of gravity, as measured by satellites.

Gravity, or better, changes in the pull of gravity from one location to another, was used in the discovery of the crater left behind by the asteroid strike in the Yucatan that led to the extinction of the dinosaurs. The rocks and other debris that eventually filled the crater differ in density from the surrounding material, thereby exerting a slightly different gravitational pull on other masses. This slight change in pull can be detected by sensitive instruments placed in satellites in earth orbit. When the intensity of the pull, as a function of position on the earth's surface, is displayed as a two-dimensional image, the presence of the crater is evident.

Studies of the changes in gravitational pull of the Antarctic ice between 2002 and 2005 revealed that Antarctica is losing 36 cubic miles of ice each year. By way of comparison, the city of Los Angeles uses one cubic mile of water each year. While this finding is often cited as clear evidence of global warming, it contradicts some models of climate change that indicate that global warming may lead to an increase of snowfall, and therefore more ice, in the polar regions. This does not show that global warming is not taking place, but only the inadequacies of some models [119].

### 1.3.13 Echo Cancellation

In a conference call between locations A and B, what is transmitted from A to B can get picked up by microphones in B, transmitted back to speakers in A and then retransmitted to B, producing an echo of the original transmission. Signal processing performed at the transmitter in A can reduce the strength of the second version of the transmission and decrease the echo effect.

### 1.3.14 Hearing Aids

Makers of digital hearing aids include signal processing to enhance the quality of the received sounds, as well as to improve localization, that is, the ability of the hearer to tell where the sound is coming from. When a hearing aid is used, sounds reach the ear in two ways: first, the usual route directly into the ear, and second, through the hearing aid. Because that part that passes through the hearing aid is processed, there is a slight delay. In order for the delay to go unnoticed, the processing must be very fast. When hearing aids are used in both ears, more sophisticated processing can be used.

### 1.3.15 Near-Earth Asteroids

An area of growing importance is the search for potentially damaging near-earth asteroids. These objects are initially detected by passive optical observation, as small dots of reflected sunlight; once detected, they are then imaged by active radar to determine their size, shape, rotation, path, and other important parameters. Satellite-based infrared detectors are being developed to find dark asteroids by the heat they give off. Such satellites, placed in orbit between the sun and the earth, will be able to detect asteroids hidden from earth-based telescopes by the sunlight.

### 1.3.16 Mapping the Ozone Layer

Ultraviolet light from the sun is scattered by ozone. By measuring the amount of scattered UV at various locations on the earth's surface, and with the sun in various positions, we obtain values of the Laplace transform of the function describing the density of ozone, as a function of elevation.

### 1.3.17 Ultrasound Imaging

While x-ray tomography is a powerful method for producing images of the interior of patients' bodies, the radiation involved and the expense make it unsuitable in some cases. Ultrasound imaging, making use of backscattered sound waves, is a popular method of inexpensive preliminary screening for medical diagnostics, and for examining a developing fetus.

### 1.3.18 X-ray Vision?

The MIT computer scientist and electrical engineer Dina Katabi and her students are currently exploring new uses of wireless technologies. By combining *Wi-Fi* and *vision* into what she calls *Wi-Vi*, she has discovered a way to detect the number and approximate location of persons within a closed room and to recognize simple gestures. The scattering of reflected low-bandwidth wireless signals as they pass through the walls is processed to eliminate motionless sources of reflection from the much weaker reflections from moving objects, presumably people.

---

## 1.4 The Common Core

The examples just presented look quite different from one another, but the differences are often more superficial than real. As we begin to use mathematics to model these various situations we often discover a common core of mathematical tools and ideas at the heart of each of these applications. For example, the Fourier transform and its variants play an important role in many areas of signal and image processing, as do the language and theory of matrix analysis, iterative optimization and approximation techniques, and the basics of probability and statistics. This common core provides the subject matter for this book. Applications of the core material to tomographic medical imaging, optical imaging, and acoustic signal processing are among the topics to be discussed in some detail.

Although the applications of interest to us vary in their details, they have common aspects that can be summarized as follows: the data has been obtained through some form of sensing; physical models, often simplified, describe how the data we have obtained relates to the information we seek; there usually isn't enough data and what we have is corrupted by noise and other distortions.

## 1.5 Active and Passive Sensing

In some signal and image processing applications the sensing is *active*, meaning that we have initiated the process, by, say, sending an x-ray through the body of a patient, injecting a patient with a radionuclide, transmitting an acoustic signal through the ocean, as in sonar, or transmitting a radio wave, as in radar. In such cases, we are interested in measuring how the system, the patient, the quiet submarine, the ocean floor, the rain cloud, will respond to our probing. In many other applications, the sensing is *passive*, which means that the object of interest to us provides its own signal of some sort, which we then detect, analyze, image, or process in some way. Certain sonar systems operate passively, listening for sounds made by the object of interest. Optical and radio telescopes are passive, relying on the object of interest to emit or reflect light, or other electromagnetic radiation. Night-vision instruments are sensitive to lower-frequency, infrared radiation.

From the time of Aristotle and Euclid until the middle ages there was an ongoing debate concerning the active or passive nature of human sight [112]. Those like Euclid, whose interests were largely mathematical, believed that the eye emitted rays, the *extramission theory*. Aristotle and others, more interested in the physiology and anatomy of the eye than in mathematics, believed that the eye received rays from observed objects outside the body, the *intromission theory*. Finally, around 1000 AD, the Arabic mathematician and natural philosopher Alhazen demolished the extramission theory by noting the potential for bright light to hurt the eye, and combined the mathematics of the extramission theorists with a refined theory of intromission. The extramission theory has not gone away completely, however, as anyone familiar with Superman's x-ray vision knows.

## 1.6 Using Prior Knowledge

An important point to keep in mind when doing signal processing is that, while the data is usually limited, the information we seek may not be lost. Although processing the data in a reasonable way may suggest otherwise, other processing methods may reveal that the desired information is still available in the data. Figure 1.1 illustrates this point.

The original image on the upper right of Figure 1.1 is a discrete rectangular array of intensity values simulating the distribution of the x-ray-attenuating material in a slice of a head. The data was obtained by taking the two-dimensional discrete Fourier transform of the original image, and then discarding, that is, setting to zero, all these spatial frequency values, except for those in a smaller rectangular region around the origin. Reconstructing the image from this limited data amounts to solving a large system of linear equations. The problem is under-determined, so a minimum-norm solution would seem to be a reasonable reconstruction method. For now, "norm" means the Euclidean norm.

The minimum-norm solution is shown on the lower right. It is calculated simply by performing an inverse discrete Fourier transform on the array of modified discrete Fourier transform values. The original image has relatively large values where the skull is located, but the least-squares reconstruction does not want such high values; the norm involves the sum of squares of intensities, and high values contribute disproportionately to the norm. Consequently, the minimum-norm reconstruction chooses instead to conform to the measured data by spreading what should be the skull intensities throughout the interior of the skull. The minimum-norm reconstruction does tell us something about the original; it tells us about the existence of the skull itself, which, of course, is indeed a prominent feature of the original. However, in all likelihood, we would already know about the skull; it would be the interior that we want to know about.

Using our knowledge of the presence of a skull, which we might have obtained from the minimum-norm reconstruction itself, we construct the prior estimate shown in the upper left. Now we use the same data as before, and calculate a minimum-weighted-norm reconstruction, using as the weight vector the reciprocals of the values of the prior image. This minimum-

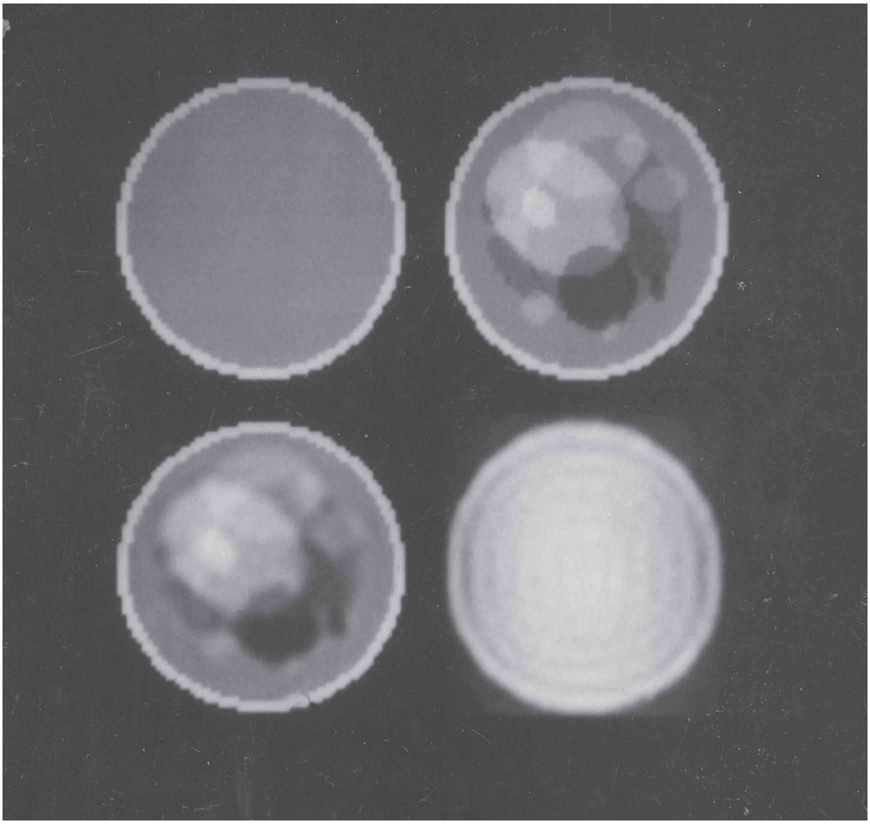


**FIGURE 1.1**: Extracting information in image reconstruction.

weighted-norm reconstruction, also called the PDFT estimator, is shown on the lower left; it is clearly almost the same as the original image. The calculation of the minimum-weighted-norm solution can be done iteratively using the ART algorithm [143].

When we weight the skull area with the inverse of the prior image, we allow the reconstruction to place higher values there without having much of an effect on the overall weighted norm. In addition, the reciprocal weighting in the interior makes spreading intensity into that region costly, so the interior remains relatively clear, allowing us to see what is really present there.

When we try to reconstruct an image from limited data, it is easy to assume that the information we seek has been lost, particularly when a reasonable reconstruction method fails to reveal what we want to know. As this example, and many others, show, the information we seek is often still in the data, but needs to be brought out in a more subtle way.

## 1.7 An Urn Model of Remote Sensing

Most of the signal processing that we shall discuss in this book is related to the problem of *remote sensing*, which we might also call *indirect measurement*. In such problems we do not have direct access to what we are really interested in, and must be content to measure something else that is related to, but not the same as, what interests us. For example, we want to know what is in the suitcases of airline passengers, but, for practical reasons, we cannot open every suitcase. Instead, we x-ray the suitcases. A recent paper [137] describes progress in detecting nuclear material in cargo containers by measuring the scattering, by the shielding, of cosmic rays; you can't get much more *remote* than that. Before we get into the mathematics of signal processing, it is probably a good idea to consider a model that, although quite simple, manages to capture many of the important features of remote-sensing applications. To convince the reader that this is indeed a useful model, we relate it to the problem of image reconstruction in *single-photon emission computed tomography* (SPECT). There seems to be a tradition in physics of using simple models or examples involving urns and marbles to illustrate important principles. In keeping with that tradition, we have here two examples, both involving urns of marbles, to illustrate various aspects of remote sensing.

### 1.7.1 An Urn Model

Suppose that there is a box containing a large number of small pieces of paper, and on each piece is written one of the numbers from $j = 1$ to $j = J$. I want to determine, for each $j = 1, ..., J$, the probability of selecting a piece of paper with the number $j$ written on it. Unfortunately, I am not allowed to examine the box. I am allowed, however, to set up a remote-sensing experiment to help solve my problem.

My assistant sets up $J$ urns, numbered $j = 1, ..., J$, each containing marbles of various colors. Suppose that there are $I$ colors, numbered $i = 1, ..., I$. I am allowed to examine each urn, so I know precisely the probability that a marble of color $i$ will be drawn from urn $j$. Out of my view, my assistant removes one piece of paper from the box, takes one marble from the indicated urn, announces to me the color of the marble, and then replaces both the piece of paper and the marble. This action is repeated $N$ times, at the end of which I have a long list of colors, $\mathbf{i} = \{i_1, i_2, ..., i_N\}$, where $i_n$ denotes the color of the $n$th marble drawn. This list $\mathbf{i}$ is my data, from which I must determine the contents of the box.

This is a form of remote sensing; what we have access to is related to, but not equal to, what we are interested in. What I wish I had is the list of urns used, $\mathbf{j} = \{j_1, j_2, ..., j_N\}$; instead I have $\mathbf{i}$, the list of colors. Sometimes data such as the list of colors is called "incomplete data," in contrast to the "complete data," which would be the list $\mathbf{j}$ of the actual urn numbers drawn from the box.

Using our urn model, we can begin to get a feel for the *resolution problem*. If all the marbles of one color are in a single urn, all the black marbles in urn $j = 1$, all the green in urn $j = 2$, and so on, the problem is trivial; when I hear a color, I know immediately which urn contained that marble. My list of colors is then a list of urn numbers; $\mathbf{i} = \mathbf{j}$. I have the complete data now. My estimate of the number of pieces of paper containing the urn number $j$ is then simply the proportion of draws that resulted in urn $j$ being selected.

At the other extreme, suppose two urns have identical contents. Then I cannot distinguish one urn from the other and I am unable to estimate more than the total number of pieces of paper containing either of the two urn numbers. If the two urns have nearly the same contents, we can distinguish them only by using a very large $N$. This is the resolution problem.

Generally, the more the contents of the urns differ, the easier the task of estimating the contents of the box. In remote-sensing applications, these issues affect our ability to resolve individual components contributing to the data.

### 1.7.2 Some Mathematical Notation

To introduce some mathematical notation, let us denote by $x_j$ the proportion of the pieces of paper that have the number $j$ written on them. Let $P_{ij}$ be the proportion of the marbles in urn $j$ that have the color $i$. Let $y_i$ be the proportion of times the color $i$ occurs in the list of colors. The expected proportion of times $i$ occurs in the list is $E(y_i) = \sum_{j=1}^{J} P_{ij} x_j = (Px)_i$, where $P$ is the $I$ by $J$ matrix with entries $P_{ij}$ and $x$ is the $J$ by 1 column vector with entries $x_j$. A reasonable way to estimate $x$ is to replace $E(y_i)$ with the actual $y_i$ and solve the system of linear equations $y_i = \sum_{j=1}^{J} P_{ij} x_j$, $i = 1, ..., I$. Of course, we require that the $x_j$ be nonnegative and sum to one, so special algorithms may be needed to find such solutions. In a number of applications that fit this model, such as medical tomography, the values $x_j$ are taken to be parameters, the data $y_i$ are statistics, and the $x_j$ are estimated by adopting a probabilistic model and maximizing the likelihood function. Iterative algorithms, such as the expectation maximization maximum likelihood (EMML) algorithm, are often used for such problems; see Chapter 14 for details.

### 1.7.3 An Application to SPECT Imaging

In *single-photon emission computed tomography* (SPECT) the patient is injected with a chemical to which a radioactive tracer has been attached. Once the chemical reaches its destination within the body the photons emitted by the radioactive tracer are detected by gamma cameras outside the body. The objective is to use the information from the detected photons to infer the relative concentrations of the radioactivity within the patient.

We discretize the problem and assume that the body of the patient consists of $J$ small volume elements, called *voxels*, analogous to *pixels* in digitized images. We let $x_j \geq 0$ be the unknown proportion of the radioactivity that is present in the $j$th voxel, for $j = 1, ..., J$. There are $I$ detectors, denoted $\{i = 1, 2, ..., I\}$. For each $i$ and $j$ we let $P_{ij}$ be the known probability that a photon that is emitted from voxel $j$ is detected at detector $i$; these probabilities are usually determined by examining the relative positions in space of voxel $j$ and detector $i$. We denote by $i_n$ the detector at which the $n$th emitted photon is detected. This photon was emitted at some voxel, denoted $j_n$; we wish that we had some way of learning what each $j_n$ is, but we must be content with knowing only the $i_n$. After $N$ photons have been emitted, we have as our data the list $\mathbf{i} = \{i_1, i_2, ..., i_N\}$; this is our *incomplete data*. We wish we had the *complete data*, that is, the list $\mathbf{j} = \{j_1, j_2, ..., j_N\}$, but we do not. Our goal is to estimate the frequency with which each voxel emitted a photon, which we assume, reasonably, to be proportional to the unknown proportions $x_j$, for $j = 1, ..., J$.

This problem is completely analogous to the urn problem previously discussed. Any mathematical method that solves one of these problems will solve the other one. In the urn problem, the colors were announced; here the detector numbers are announced. There, I wanted to know the urn numbers; here I want to know the voxel numbers. There, I wanted to estimate the frequency with which the $j$th urn was used; here, I want to estimate the frequency with which the $j$th voxel is the site of an emission, which is assumed to be equal to the proportion of the radionuclide within the $j$th voxel. In the urn model, two urns with nearly the same contents are hard to distinguish unless $N$ is very large; here, two neighboring voxels will be very hard to distinguish (i.e., to resolve) unless $N$ is very large. But in the SPECT case, a large $N$ means a high dosage, which will be prohibited by safety considerations. Therefore, we have a built-in resolution problem in the SPECT case.

Both problems are examples of probabilistic mixtures, in which the mixing probabilities are the $x_j$ that we seek. The *maximum likelihood* (ML) method of statistical parameter estimation can be used to solve such problems. The interested reader should consult the text [42].

## 1.8 Hidden Markov Models

In the urn model we just discussed, the order of the colors in the list is unimportant; we could randomly rearrange the colors on the list without affecting the nature of the problem. The probability that a green marble will be chosen next is the same, whether a blue or a red marble was just chosen the previous time. This independence from one selection to another is fine for modeling certain physical situations, such as emission tomography. However, there are other situations in which this independence does not conform to reality.

In written English, for example, knowing the current letter helps us, sometimes more, sometimes less, to predict what the next letter will be. We know that, if the current letter is a "q", then there is a high probability that the next one will be a "u". So what the current letter is affects the probabilities associated with the selection of the next one.

Spoken English is even tougher. There are many examples in which the pronunciation of a certain sound is affected, not only by the sound or sounds that preceded it, but by the sound or sounds that will follow. For example, the sound of the "e" in the word "bellow" is different from the sound of the "e" in the word "below"; the sound changes, depending on whether there is a double "l" or a single "l" following the "e". Here the entire context of the letter affects its sound.

Hidden Markov models (HMM) are increasingly important in speech processing, optical character recognition, and DNA sequence analysis. They allow us to incorporate dependence on the context into our model. In this section we illustrate HMM using a modification of the urn model.

Suppose, once again, that we have $J$ urns, indexed by $j = 1, ..., J$ and $I$ colors of marbles, indexed by $i = 1, ..., I$. Associated with each of the $J$ urns is a box, containing a large number of pieces of paper, with the number of one urn written on each piece. My assistant selects one box, say the $j_0$th box, to start the experiment. He draws a piece of paper from that box, reads the number written on it, call it $j_1$, goes to the urn with the number $j_1$ and draws out a marble. He then announces the color. He then draws a piece of paper from box number $j_1$, reads the next number, say $j_2$, proceeds to urn number $j_2$, etc. After $N$ marbles have been drawn, the only data I have is a list of colors, $\mathbf{i} = \{i_1, i_2, ..., i_N\}$.

The *transition probability* that my assistant will proceed from the urn numbered $k$ to the urn numbered $j$ is $b_{jk}$, with $\sum_{j=1}^{J} b_{jk} = 1$. The number of the current urn is the current *state*. In an ordinary *Markov chain* model, we observe directly a sequence of states governed by the transition probabilities. The Markov chain model provides a simple formalism for describing a system that moves from one state into another, as time goes on. In the hidden Markov model we are not able to observe the states directly; they are hidden from us. Instead, we have indirect observations, the colors of the marbles in our urn example.

The probability that the color numbered $i$ will be drawn from the urn numbered $j$ is $a_{ij}$, with $\sum_{i=1}^{I} a_{ij} = 1$, for all $j$. The colors announced are the *visible states*, while the unannounced urn numbers are the *hidden states*.

There are several distinct objectives one can have, when using HMM. We assume that the data is the list of colors, $\mathbf{i}$.

- **Evaluation:** For given probabilities $a_{ij}$ and $b_{jk}$, what is the probability that the list $\mathbf{i}$ was generated according to the HMM? Here, the objective is to see if the model is a good description of the data.

- **Decoding:** Given the model, the probabilities, and the list $\mathbf{i}$, what list $\mathbf{j} = \{j_1, j_2, ..., j_N\}$ of urns is most likely to be the list of urns actually visited? Now, we want to infer the hidden states from the visible ones.

- **Learning:** We are told that there are $J$ urns and $I$ colors, but are not told the probabilities $a_{ij}$ and $b_{jk}$. We are given several data vectors $\mathbf{i}$ generated by the HMM; these are the *training sets*. The objective is to learn the probabilities.

Once again, the ML approach can play a role in solving these problems [68]. The *Viterbi algorithm* is an important tool used for the decoding phase (see [149]).

# Chapter 2

## Fourier Series and Fourier Transforms



### 2.1 Chapter Summary

We begin with Fourier series and Fourier transforms, which are essential tools in signal processing. In this chapter we give the formulas for