Charles L. Byrne
Department of Mathematical Sciences
University of Massachusetts Lowell
Book Introduction and Sample Chapter
ISBN 978-1-4822-4184-6 CRC Press
December 2014

Signal Processing: A Mathematical Approach

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 back-scattered 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 least-squares 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-

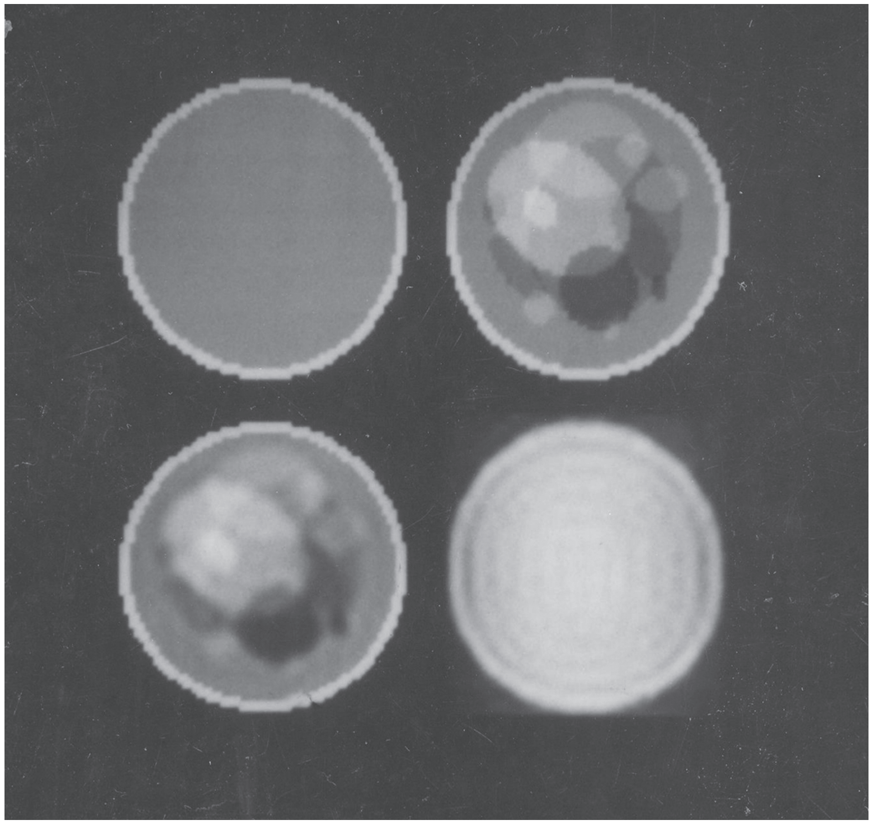

FIGURE 1.1: Extracting information in image reconstruction.

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 nth 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 **??** 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 jth 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 nth 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 jth urn was used; here, I want to estimate the frequency with which the jth voxel is the site of an emission, which is assumed to be equal to the proportion of the radionuclide within the jth 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_0th 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

Remote Sensing

2.1 Chapter Summary

A basic problem in remote sensing is to determine the nature of a distant object by measuring signals transmitted by or reflected from that object. If the object of interest is sufficiently remote, that is, is in the *far field*, the data we obtain by sampling the propagating spatio-temporal field is related, approximately, to what we want by *Fourier transformation*. In this chapter we present examples to illustrate the roles played by Fourier series and Fourier coefficients in the analysis of remote sensing and signal transmission. We use these examples to motivate several of the computational problems we shall consider in detail later in the text. We also discuss two inverse problems involving the Laplace transform.

We consider here a common problem of remote sensing of transmitted or reflected waves propagating from distant sources. Examples include optical imaging of planets and asteroids using reflected sunlight, radio-astronomy imaging of distant sources of radio waves, active and passive sonar, radar imaging using microwaves, and infrared (IR) imaging to monitor the ocean temperature. In such situations, as well as in transmission and emission tomography and magnetic-resonance imaging, what we measure are essentially the Fourier coefficients or values of the Fourier transform of the function we want to estimate. The image reconstruction problem then becomes one of estimating a function from finitely many noisy values of its Fourier transform.

2.2 Fourier Series and Fourier Coefficients

We suppose that $f : [-L, L] \to \mathbb{C}$, and that its Fourier series converges to $f(x)$ for all x in $[-L, L]$. In the examples in this chapter, we shall see how Fourier coefficients can arise as data obtained through measurements. However, we shall be able to measure only a finite number of the Fourier coefficients. One issue that will concern us is the effect on the estimation of $f(x)$ if we use some, but not all, of its Fourier coefficients.

Suppose that we have c_n, as defined by Equation (**??**), for $n = 0, 1, 2, ..., N$. It is not unreasonable to try to estimate the function $f(x)$ using the *discrete Fourier transform* (DFT) estimate, which is

$$f_{DFT}(x) = \sum_{n=0}^{N} c_n e^{i\frac{n\pi}{L}x}.$$

When we know that $f(x)$ is real-valued, and so $c_{-n} = \overline{c_n}$, we naturally assume that we have the values of c_n for $|n| \leq N$.

2.3 The Unknown Strength Problem

In this example, we imagine that each point x in the interval $[-L, L]$ is sending out a signal that is a complex-exponential-function signal, also called a *sinusoid*, at the frequency ω, each with its own strength $f(x)$; that is, the signal sent by the point x is

$$f(x)e^{i\omega t}.$$

In our first example, we imagine that the strength function $f(x)$ is unknown and we want to determine it. It could be the case that the signals originate at the points x, as with light or radio waves from the sun, or are simply reflected from the points x, as is sunlight from the moon or radio waves in radar. Later in this chapter, we shall investigate a related example, in which the points x transmit known signals and we want to determine what is received elsewhere.

2.3.1 Measurement in the Far Field

Now let us consider what is received by a point P on the circumference of a circle centered at the origin and having large radius D. The point P corresponds to the angle θ as shown in Figure 2.1; we use θ in the interval $[0, \pi]$. It takes a finite time for the signal sent from x at time t to reach P, so there is a delay.

We assume that c is the speed at which the signal propagates. Because D is large relative to L, we make the *far-field assumption*, which allows us to approximate the distance from x to P by $D - x\cos\theta$. Therefore, what P receives at time t from x is approximately what was sent from x at time $t - \frac{1}{c}(D - x\cos\theta)$.

Ex. 2.1 *Show that, for any point P on the circle of radius D and any $x \neq 0$, the distance from x to P is always greater than or equal to the far-field approximation $D - x\cos\theta$, with equality if and only if $\theta = 0$ or $\theta = \pi$.*

At time t, the point P receives from x the signal

$$f(x)e^{i\omega(t - \frac{1}{c}(D - x\cos\theta))} = e^{i\omega(t - \frac{1}{c}D)} f(x)e^{i\frac{\omega\cos\theta}{c}x}.$$

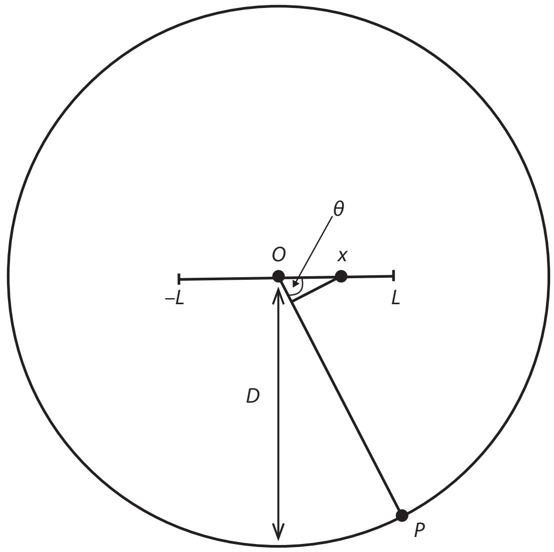

FIGURE 2.1: Far-field measurements.

Because the point P receives signals from all x in $[-L, L]$, the signal that P receives at time t is

$$e^{i\omega(t-\frac{1}{c}D)}\int_{-L}^{L} f(x)e^{i\frac{\omega\cos\theta}{c}x}dx.$$

Therefore, from measurements in the far field, we obtain the values

$$\int_{-L}^{L} f(x)e^{i\frac{\omega\cos\theta}{c}x}dx.$$

When θ is chosen so that

$$\frac{\omega\cos\theta}{c} = \frac{-n\pi}{L} \tag{2.1}$$

we have c_n.

2.3.2 Limited Data

Note that we will be able to solve Equation (2.1) for θ if and only if we have

$$|n| \le \frac{L\omega}{\pi c}.$$

This tells us that we can measure only finitely many of the Fourier coefficients of $f(x)$. It is common in signal processing to speak of the *wavelength* of a sinusoidal signal; the wavelength associated with a given ω and c is

$$\lambda = \frac{2\pi c}{\omega}.$$

Therefore we can measure $2N+1$ Fourier coefficients, where N is the largest integer not greater than $\frac{2L}{\lambda}$, which is the length of the interval $[-L, L]$, measured in units of wavelength λ. We get more Fourier coefficients when the product $L\omega$ is larger; this means that when L is small, we want ω to be large, so that λ is small and N is large. As we saw previously, using these finitely many Fourier coefficients to calculate the DFT reconstruction of $f(x)$ can lead to a poor estimate of $f(x)$, particularly when N is small.

Consider the situation in which the points x are reflecting signals that are sent to probe the structure of an object described by the function f, as in radar. This relationship between the number $L\omega$ and the number of Fourier coefficients we can measure amounts to a connection between the frequency of the probing signal and the resolution attainable; finer detail is available only if the frequency is high enough.

The wavelengths used in primitive early radar at the start of World War II were several meters long. Since resolution is proportional to aperture, that is, the length of the array measured in units of wavelength, antennas for such radar needed to be quite large. As Körner notes in [102], the general feeling at the time was that the side with the shortest wavelength would win the war. The cavity magnetron, invented during the war by British scientists, made possible microwave radar having a wavelength of 10 cm, which could then be mounted easily on planes.

2.3.3 Can We Get More Data?

As we just saw, we can make measurements at any points P in the far field; perhaps we do not need to limit ourselves to just those angles that lead to the c_n. It may come as somewhat of a surprise, but from the theory of complex analytic functions we can prove that there is enough data available to us here to reconstruct $f(x)$ perfectly, at least in principle. The drawback, in practice, is that the measurements would have to be free of noise and impossibly accurate. All is not lost, however.

2.3.4 Measuring the Fourier Transform

If θ is chosen so that

$$\frac{\omega \cos\theta}{c} = \frac{-n\pi}{L},$$

then our measurement gives us the Fourier coefficients c_n. But we can select any angle θ and use any P we want. In other words, we can obtain the values

$$\int_{-L}^{L} f(x)e^{i\frac{\omega\cos\theta}{c}x}dx,$$

for any angle θ. With the change of variable

$$\gamma = \frac{\omega\cos\theta}{c},$$

we can obtain the value of the Fourier transform,

$$F(\gamma) = \int_{-L}^{L} f(x)e^{i\gamma x}dx,$$

for any γ in the interval $[-\frac{\omega}{c}, \frac{\omega}{c}]$.

We are free to measure at any P and therefore to obtain values of $F(\gamma)$ for any value of γ in the interval $[-\frac{\omega}{c}, \frac{\omega}{c}]$. We need to be careful how we process the resulting data, however.

2.3.5 Over-Sampling

Suppose, for the sake of illustration, that we measure the far-field signals at points P corresponding to angles θ that satisfy

$$\frac{\omega\cos\theta}{c} = \frac{-n\pi}{2L},$$

instead of

$$\frac{\omega\cos\theta}{c} = \frac{-n\pi}{L}.$$

Now we have twice as many data points and from these new measurements we can obtain

$$d_n = \int_{-L}^{L} f(x)e^{-i\frac{n\pi}{2L}x}dx,$$

for $|n| \leq 2N$. We say now that our data is *twice over-sampled*. Note that we call it *over-sampled* because the rate at which we are sampling is higher, even though the distance between samples is shorter. The values d_n are not simply more of the Fourier coeffcients of f. The question now is: What are we to do with these extra data values?

The values d_n are, in fact, Fourier coefficients, but not of f; they are Fourier coefficients of the function $g : [-2L, 2L] \to \mathbb{C}$, where $g(x) = f(x)$ for $|x| \leq L$, and $g(x) = 0$, otherwise. If we simply use the d_n as Fourier coefficients of the function $g(x)$ and compute the resulting DFT estimate of $g(x)$,

$$g_{DFT}(x) = \sum_{n=-2N}^{2N} d_n e^{i\frac{n\pi}{2L}x},$$

this function estimates $f(x)$ for $|x| \leq L$, but it also estimates $g(x) = 0$ for the other values of x in $[-2L, 2L]$. When we graph $g_{DFT}(x)$ for $|x| \leq L$ we find that we have no improvement over what we got with the previous estimate f_{DFT}. The problem is that we have wasted the extra data by estimating $g(x) = 0$ where we already knew that it was zero. To make good use of the extra data we need to incorporate this prior information about the function g. The MDFT and PDFT algorithms provide estimates of $f(x)$ that incorporate prior information.

2.3.6 The Modified DFT

The modified DFT (MDFT) estimate was first presented in [22]. For our example of twice over-sampled data, the MDFT is defined for $|x| \leq L$ and has the algebraic form

$$f_{MDFT}(x) = \sum_{n=-2N}^{2N} a_n e^{i\frac{n\pi}{2L}x}, \tag{2.2}$$

for $|x| \leq L$. The coefficients a_n are not the d_n. The a_n are determined by requiring that the function f_{MDFT} be consistent with the measured data, the d_n. In other words, we must have

$$d_n = \int_{-L}^{L} f_{MDFT}(x) e^{-i\frac{n\pi}{2L}x} dx. \tag{2.3}$$

When we insert $f_{MDFT}(x)$ as given in Equation (2.2) into Equation (2.3) we get a system of $2N+1$ linear equations in $2N+1$ unknowns, the a_n. We then solve this system for the a_n and use them in Equation (2.2). Figure **??** shows the improvement we can achieve using the MDFT. The data used to construct the graphs in that figure was thirty times over-sampled. We note here that, had we extended f initially as a $2L$-periodic function, it would be difficult to imagine the function $g(x)$ and we would have a hard time figuring out what to do with the d_n.

In this example we measured twice as much data as previously. We can, of course, measure even more data, and it need not correspond to the Fourier coefficients of any function. The potential drawback is that, as we use more data, the system of linear equations that we must solve to obtain the MDFT estimate becomes increasingly sensitive to noise and round-off error in the data. It is possible to lessen this effect by *regularization*, but not to eliminate it entirely. Regularization can be introduced here simply by multiplying by, say, 1.01, the entries of the main diagonal of the matrix of the linear system. This makes the matrix less *ill-conditioned*.

In our example, we used the prior knowledge that $f(x) = 0$ for $|x| > L$. Now, we shall describe in detail the use of other forms of prior knowledge about $f(x)$ to obtain reconstructions that are better than the DFT.

2.3.7 Other Forms of Prior Knowledge

As we just showed, knowing that we have over-sampled in our measurements can help us improve the resolution in our estimate of $f(x)$. We may have other forms of prior knowledge about $f(x)$ that we can use. If we know something about large-scale features of $f(x)$, but not about finer details, we can use the PDFT estimate, which is a generalization of the MDFT. In Chapter 1 the PDFT was compared to the DFT in a two-dimensional example of simulated head slices.

The MDFT estimator can be written as

$$f_{MDFT}(x) = \chi_L(x) \sum_{n=-2N}^{2N} a_n e^{i\frac{n\pi}{2L}x}.$$

We include the prior information that $f(x)$ is supported on the interval $[-L, L]$ through the factor $\chi_L(x)$. If we select a function $p(x) \geq 0$ that describes our prior estimate of the shape of $|f(x)|$, we can then estimate $f(x)$ using the PDFT estimator, which, in this case of twice over-sampled data, takes the form

$$f_{PDFT}(x) = p(x) \sum_{n=-2N}^{2N} b_n e^{i\frac{n\pi}{2L}x}.$$

As with the MDFT estimator, we determine the coefficients b_n by requiring that $f_{PDFT}(x)$ be consistent with the measured data.

There are other things we may know about $f(x)$. We may know that $f(x)$ is nonnegative, or we may know that $f(x)$ is approximately zero for most x, but contains very sharp peaks at a few places. In more formal language, we may be willing to assume that $f(x)$ contains a few Dirac delta functions in a flat background. There are nonlinear methods, such as the maximum entropy method, the indirect PDFT (IPDFT), and eigenvector methods, that can be used to advantage in such cases; these methods are often called *high-resolution methods*.

2.4 Generalizing the MDFT and PDFT

In our discussion so far the data we have obtained are values of the Fourier transform of the support-limited function $f(x)$. The MDFT and PDFT can be extended to handle those cases in which the data we have are more general linear-functional values pertaining to $f(x)$.

Suppose that our data values are finitely many linear-functional values,

$$d_n = \int_{-L}^{L} f(x)\overline{g_n(x)}dx,$$

for $n = 1, ..., N$, where the $g_n(x)$ are known functions. The extended MDFT estimate of $f(x)$ is

$$f_{MDFT}(x) = \chi_L(x) \sum_{m=1}^{N} a_m g_m(x),$$

where the coefficients a_m are chosen so that f_{MDFT} is consistent with the measured data; that is,

$$d_n = \int_{-L}^{L} f_{MDFT}(x)\overline{g_n(x)}dx,$$

for each n. To find the a_m we need to solve a system of N equations in N unknowns.

The PDFT can be extended in a similar way. The extended PDFT estimate of $f(x)$ is

$$f_{PDFT}(x) = p(x) \sum_{m=1}^{N} b_m g_m(x),$$

where, as previously, the coefficients b_m are chosen by forcing the estimate of $f(x)$ to be consistent with the measured data. Again, we need to solve a system of N equations in N unknowns to find the coefficients.

For large values of N, setting up and solving the required systems of linear equations can involve considerable effort. If we discretize the functions $f(x)$ and $g_n(x)$, we can obtain good approximations of the extended MDFT and PDFT using the iterative ART algorithm [142, 143].

2.5 One-Dimensional Arrays

In this section we consider the reversed situation in which the sources of the signals are the points on the circumference of the large circle and we are measuring the received signals at points of the x-axis. The objective is to determine the relative strengths of the signals coming to us from various angles.

People with sight in only one eye have a difficult time perceiving depth in their visual field, unless they move their heads. Having two functioning ears helps us determine the direction from which sound is coming; blind people, who are more than usually dependent on their hearing, often move their heads to get a better sense of where the source of sound is. Snakes who smell with their tongues often have forked tongues, the better to detect the direction of the sources of different smells. In certain remote-sensing situations the sensors respond equally to arrivals from all directions. One then obtains the needed directionality by using multiple sensors, laid out in some spatial configuration called the sensor *array.* The simplest configuration is to have the sensors placed in a straight line, as in a sonar towed array.

Now we imagine that the points $P = P(\theta)$ in the far field are the sources of the signals and we are able to measure the transmissions received at points x on the x-axis; we no longer assume that these points are confined to the interval $[-L, L]$. The P corresponding to the angle θ sends $f(\theta)e^{i\omega t}$, where the absolute value of $f(\theta)$ is the strength of the signal coming from P. We allow $f(\theta)$ to be complex, so that it has both magnitude and phase, which means that we do not assume that the signals from the different angles are in phase with one another; that is, we do not assume that they all begin at the same time.

In narrow-band passive sonar, for example, we may have hydrophone sensors placed at various points x and our goal is to determine how much acoustic energy at a specified frequency is coming from different directions. There may be only a few directions contributing significant energy at the frequency of interest, in which case $f(\theta)$ is nearly zero for all but a few values of θ.

2.5.1 Measuring Fourier Coefficients

At time t the point x on the x-axis receives from $P = P(\theta)$ what P sent at time $t - (D - x\cos\theta)/c$; so, at time t, x receives from P

$$e^{i\omega(t-D/c)}f(\theta)e^{i\frac{\omega x}{c}\cos\theta}.$$

Since x receives signals from all the angles, what x receives at time t is

$$e^{i\omega(t-D/c)}\int_0^{\pi} f(\theta)e^{i\frac{\omega x}{c}\cos\theta}d\theta.$$

We limit the angle θ to the interval $[0,\pi]$ because, in this sensing model, we cannot distinguish receptions from θ and from $2\pi-\theta$.

To simplify notation, we shall introduce the variable $u=\cos\theta$. We then have

$$\frac{du}{d\theta}=-\sin(\theta)=-\sqrt{1-u^2},$$

so that

$$d\theta=-\frac{1}{\sqrt{1-u^2}}du.$$

Now let $g(u)$ be the function

$$g(u)=\frac{f(\arccos(u))}{\sqrt{1-u^2}},$$

defined for u in the interval $(-1,1)$. Since

$$\int_0^{\pi} f(\theta)e^{i\frac{\omega x}{c}\cos\theta}d\theta=\int_{-1}^{1} g(u)e^{i\frac{\omega x}{c}u}du,$$

we find that, from our measurement at x, we obtain $G(\gamma)$, the value of the Fourier transform of $g(u)$ at γ, for

$$\gamma=\frac{\omega x}{c}.$$

Since $g(u)$ is limited to the interval $(-1,1)$, its Fourier coefficients are

$$a_n=\frac{1}{2}\int_{-1}^{1} g(u)e^{-in\pi u}du.$$

Therefore, if we select x so that

$$\gamma=\frac{\omega x}{c}=-n\pi,$$

we have a_n. Consequently, we want to measure at the points x such that

$$x=-n\frac{\pi c}{\omega}=-n\frac{\lambda}{2}=-n\Delta, \tag{2.4}$$

where $\lambda=\frac{2\pi c}{\omega}$ is the wavelength and $\Delta=\frac{\lambda}{2}$ is the *Nyquist spacing*.

A one-dimensional array consists of measuring devices placed along a straight line (the x-axis here). Obviously, there must be some smallest bounded interval, say $[A, B]$, that contains all these measuring devices. The *aperture* of the array is $\frac{B-A}{\lambda}$, the length of the interval $[A, B]$, in units of wavelength. As we just saw, the aperture is directly related to the number of Fourier coefficients of the function $g(u)$ that we are measuring, and therefore, to the accuracy of the DFT reconstruction of $g(u)$. This is usually described by saying that aperture determines resolution. As we saw, a one-dimensional array involves an inherent ambiguity, in that we cannot distinguish a signal from the angle θ from one from the angle $2\pi - \theta$. In practice a two-dimensional configuration of sensors is sometimes used to eliminate this ambiguity.

In numerous applications, such as astronomy, it is more realistic to assume that the sources of the signals are on the surface of a large sphere, rather than on the circumference of a large circle. In such cases, a one-dimensional array of sensors does not provide sufficient information and two- or three-dimensional sensor configurations are used.

The number of Fourier coefficients of $g(u)$ that we can measure, and therefore the resolution of the resulting reconstruction of $f(\theta)$, is limited by the aperture. One way to improve resolution is to make the array of sensors longer, which is more easily said than done. However, *synthetic-aperture radar* (SAR) effectively does this. The idea of SAR is to employ the array of sensors on a moving airplane. As the plane moves, it effectively creates a longer array of sensors, a *virtual array* if you will. The one drawback is that the sensors in this virtual array are not all present at the same time, as in a normal array. Consequently, the data must be modified to approximate what would have been received at other times.

The far-field approximation tells us that, at time t, every point x receives from $P(\frac{\pi}{2})$ the same signal

$$e^{i\omega(t-D/c)} f\left(\frac{\pi}{2}\right).$$

Since there is nothing special about the angle $\frac{\pi}{2}$, we can say that the signal arriving from any angle θ, which originally spread out as concentric circles of constant value, has flattened out to the extent that, by the time it reaches our line of sensors, it is essentially constant on straight lines. This suggests the *plane-wave approximation* for signals propagating in three-dimensional space. As we shall see in Chapter **??**, these plane-wave approximations are solutions to the three-dimensional wave equation. Much of array processing is based on such models of far-field propagation.

As in the examples discussed previously, we do have more measurements we can take, if we use values of x other than those described by Equation (2.4). The issue will be what to do with these *over-sampled* measurements.

2.5.2 Over-Sampling

One situation in which over-sampling arises naturally occurs in sonar array processing. Suppose that an array of sensors has been built to operate at a *design frequency* of ω_0, which means that we have placed sensors a distance of Δ_0 apart in $[A, B]$, where λ_0 is the wavelength corresponding to the frequency ω_0 and $\Delta_0 = \frac{\lambda_0}{2}$ is the Nyquist spacing for frequency ω_0. For simplicity, we assume that the sensors are placed at points x that satisfy the equation

$$x = -n\frac{\pi c}{\omega_0} = -n\frac{\lambda_0}{2} = -n\Delta_0,$$

for $|n| \leq N$. Now suppose that we want to operate the sensing at another frequency, say ω. The sensors cannot be moved, so we must make do with sensors at the points x determined by the design frequency.

Consider, first, the case in which the second frequency ω is less than the design frequency ω_0. Then its wavelength λ is larger than λ_0, and the Nyquist spacing $\Delta = \frac{\lambda}{2}$ for ω is larger than Δ_0. So we have over-sampled.

The measurements taken at the sensors provide us with the integrals

$$\int_{-1}^{1} g(u)e^{i\frac{n\pi}{K}u}du,$$

where $K = \frac{\omega_0}{\omega} > 1$. These are Fourier coefficients of the function $g(u)$, viewed as defined on the interval $[-K, K]$, which is larger than $[-1, 1]$, and taking the value zero outside $[-1, 1]$. If we then use the DFT estimate of $g(u)$, it will estimate $g(u)$ for the values of u within $[-1, 1]$, which is what we want, as well as for the values of u outside $[-1, 1]$, where we already know $g(u)$ to be zero. Once again, we can use the MDFT, the modified DFT, to include the prior knowledge that $g(u) = 0$ for u outside $[-1, 1]$ to improve our reconstruction of $g(u)$ and $f(\theta)$. In sonar, for the over-sampled case, the interval $[-1, 1]$ is called *the visible region* (although *audible region* seems more appropriate for sonar), since it contains all the values of u that can correspond to actual angles of plane-wave arrivals of acoustic energy. In practice, of course, the measured data may well contain components that are not plane-wave arrivals, such as localized noises near individual sensors, or near-field sounds, so our estimate of the function $g(u)$ should be regularized to allow for these non-plane-wave components.

2.5.3 Under-Sampling

Now suppose that the frequency ω that we want to consider is greater than the design frequency ω_0. This means that the spacing between the sensors is too large; we have *under-sampled.* Once again, however, we cannot move the sensors and must make do with what we have.

Now the measurements at the sensors provide us with the integrals

$$\int_{-1}^{1} g(u) e^{i \frac{n\pi}{K} u} du,$$

where $K = \frac{\omega_0}{\omega} < 1$. These are Fourier coefficients of the function $g(u)$, viewed as defined on the interval $[-K, K]$, which is smaller than $[-1, 1]$, and taking the value zero outside $[-K, K]$. Since $g(u)$ is not necessarily zero outside $[-K, K]$, treating it as if it were zero there results in a type of error known as *aliasing*, in which energy corresponding to angles whose u lies outside $[-K, K]$ is mistakenly assigned to values of u that lie within $[-K, K]$. Aliasing is a common phenomenon; the strobe-light effect is aliasing, as is the apparent backward motion of the wheels of stagecoaches in cowboy movies. In the case of the strobe light, we are permitted to view the scene at times too far apart for us to sense continuous, smooth motion. In the case of the wagon wheels, the frames of the film capture instants of time too far apart for us to see the true rotation of the wheels.

2.6 Resolution Limitations

As we have seen, in the unknown-strength problem the number of Fourier coefficients we can measure is limited by the ratio $\frac{L}{\lambda}$. Additional measurements in the far field can provide additional information about the function $f(x)$, but extracting that information becomes an increasingly ill-conditioned problem, one more sensitive to noise the more data we gather.

In the line-array problem just considered, there is, in principle, no limit to the number of Fourier coefficients we can obtain by measuring at the points $n\Delta$ for integer values of n; the limitation here is of a more practical nature.

In sonar, the speed of sound in the ocean is about 1500 meters per second, so the wavelength associated with 50 Hz is $\lambda = 30$ meters. The Nyquist spacing is then 15 meters. A towed array is a line array of sensors towed behind a ship. The length of the array, and therefore the number of Nyquist-spaced sensors for passive sensing at 50 Hz, is, in principle, unlimited. In practice, however, cost is always a factor. In addition, when the array becomes too long, it is difficult to maintain it in a straight-line position.

Radar imaging uses microwaves with a wavelength of about one inch, which is not a problem; synthetic-aperture radar can also be used to simulate a longer array. In radio astronomy, however, the wavelengths can be more than a kilometer, which is why radio-astronomy arrays have to be enormous. For radio-wave imaging at very low frequencies, a sort of synthetic-aperture approach has been taken, with individual antennas located in different parts of the globe.

2.7 Using Matched Filtering

We saw previously that the signal that x receives from $P(\frac{\pi}{2})$ at time t is the same for all x. If we could turn the x-axis counter-clockwise through an angle of ϕ, then the signals received from $P(\frac{\pi}{2}+\phi)$ at time t would be the same for all x. Of course, we usually cannot turn the array physically in this way; however, we can *steer* the array mathematically. This mathematical steering makes use of *matched filtering*. In certain applications it is reasonable to assume that only relatively few values of the function $f(\theta)$ are significantly nonzero. Matched filtering is a commonly used method for dealing with such cases.

2.7.1 A Single Source

To take an extreme case, suppose that $f(\theta_0) > 0$ and $f(\theta) = 0$, for all $\theta \neq \theta_0$. The signal received at time t at x is then

$$s(x,t) = e^{i\omega(t-D/c)} f(\theta_0) e^{i\frac{\omega x}{c}\cos\theta_0}.$$

Our objective is to determine θ_0.

Suppose that we multiply $s(x,t)$ by $e^{-i\frac{\omega x}{c}\cos\theta}$, for arbitrary values of θ. When one of the arbitrary values is $\theta = \theta_0$, the product is no longer dependent on the value of x; that is, the resulting product is the same for all x. In practice, we can place sensors at some finite number of points x, and then sum the resulting products over the x. When the arbitrary θ is not θ_0, we are adding up complex exponentials with distinct phase angles, so destructive interference takes place and the magnitude of the sum is not large. In contrast, when $\theta = \theta_0$, all the products are the same and the sum is relatively large. This is *matched filtering*, which is commonly used to determine the true value of θ_0.

2.7.2 Multiple Sources

Having only one signal source is the extreme case; having two or more signal sources, perhaps not far apart in angle, is an important situation, as well. Then resolution becomes a problem. When we calculate the matched filter in the single-source case, the largest magnitude will occur when $\theta =$

θ_0, but the magnitudes at other nearby values of θ will not be zero. How quickly the values fall off as we move away from θ_0 will depend on the aperture of the array; the larger the aperture, the faster the fall-off. When we have two signal sources near to one another, say θ_1 and θ_2, the matched-filter output can have its largest magnitude at a value of θ between the two angles θ_1 and θ_2, causing a loss of resolution. Again, having a larger aperture will improve the resolution.

2.8 An Example: The Solar-Emission Problem

In [15] Bracewell discusses the *solar-emission* problem. In 1942, it was observed that radio-wave emissions in the one-meter wavelength range were arriving from the sun. Were they coming from the entire disk of the sun or were the sources more localized, in sunspots, for example? The problem then was to view each location on the sun's surface as a potential source of these radio waves and to determine the intensity of emission corresponding to each location.

For electromagnetic waves the propagation speed is the speed of light in a vacuum, which we shall take here to be $c = 3 \times 10^8$ meters per second. The wavelength λ for gamma rays is around one Angstrom, that is, 10^{-10} meters, which is about the diameter of an atom; for x-rays it is about one millimicron, or 10^{-9} meters. The visible spectrum has wavelengths that are a little less than one micron, that is, 10^{-6} meters, while infrared radiation (IR), predominantly associated with heat, has a wavelength somewhat longer. Infrared radiation with a wavelength around 6 or 7 microns can be used to detect water vapor; we use near IR, with a wavelength near that of visible light, to change the channels on our TV sets. Shortwave radio has a wavelength around one millimeter. Microwaves have wavelengths between one centimeter and one meter; those used in radar imaging have a wavelength about one inch and can penetrate clouds and thin layers of leaves. Broadcast radio has a λ running from about 10 meters to 1000 meters. The so-called long radio waves can have wavelengths several thousand meters long, necessitating clever methods of large-antenna design for radio astronomy.

The sun has an angular diameter of 30 min. of arc, or one-half of a degree, when viewed from earth, but the needed resolution was more like 3 min. of arc. Such resolution requires a larger aperture, a radio telescope 1000 wavelengths across, which means a diameter of 1km at a wavelength of 1 meter; in 1942 the largest military radar antennas were less than 5 meters across. A solution was found, using the method of reconstructing an object from line-integral data, a technique that surfaced again in tomography.

2.9 Estimating the Size of Distant Objects

Suppose, in the previous example of the unknown strength problem, we assume that $f(x) = B$, for all x in the interval $[-L, L]$, where $B > 0$ is the unknown *brightness* constant, and we don't know L. More realistic, two-dimensional versions of this problem arise in astronomy, when we want to estimate the diameter of a distant star.

In this case, the measurement of the signal at the point P gives us

$$\int_{-L}^{L} f(x) \cos\left(\frac{\omega \cos\theta}{c} x\right) dx$$

$$= B \int_{-L}^{L} \cos\left(\frac{\omega \cos\theta}{c} x\right) dx = \frac{2Bc}{\omega \cos\theta} \sin\left(\frac{L\omega \cos\theta}{c}\right),$$

when $\cos\theta \neq 0$, whose absolute value is then the strength of the signal at P. Notice that we have zero signal strength at P when the angle θ associated with P satisfies the equation

$$\sin\left(\frac{L\omega \cos\theta}{c}\right) = 0,$$

without

$$\cos\theta = 0.$$

But we know that the first positive zero of the sine function is at π, so the signal strength at P is zero when θ is such that

$$\frac{L\omega \cos\theta}{c} = \pi.$$

If

$$\frac{L\omega}{c} \geq \pi,$$

then we can solve for L and get

$$L = \frac{\pi c}{\omega \cos\theta}.$$

When $L\omega$ is too small, there will be no angle θ for which the received signal strength at P is zero. If the signals being sent are actually *broadband*, meaning that the signals are made up of components at many different frequencies, not just one ω, which is usually the case, then we might be able to filter our measured data, keep only the component at a sufficiently high frequency, and then proceed as before.

But even when we have only a single frequency ω and $L\omega$ is too small, there is something we can do. The received strength at $\theta = \frac{\pi}{2}$ is

$$F_c(0) = B\int_{-L}^{L} dx = 2BL.$$

If we knew B, this measurement alone would give us L, but we do not assume that we know B. At any other angle, the received strength is

$$F_c(\gamma) = \frac{2Bc}{\omega\cos\theta}\sin\left(\frac{L\omega\cos\theta}{c}\right).$$

Therefore,

$$F_c(\gamma)/F_c(0) = \frac{\sin(H(\theta))}{H(\theta)},$$

where

$$H(\theta) = \frac{L\omega\cos\theta}{c}.$$

From the measured value $F_c(\gamma)/F_c(0)$ we can solve for $H(\theta)$ and then for L. In actual optical astronomy, atmospheric distortions make these measurements noisy and the estimates have to be performed more carefully. This issue is discussed in more detail in Chapter **??**, in Section **??** on Two-Dimensional Fourier Transforms.

There is a simple relationship involving the intrinsic luminosity of a star, its distance from earth, and its apparent brightness; knowing any two of these, we can calculate the third. Once we know these values, we can figure out how large the visible universe is. Unfortunately, only the apparent brightness is easily determined. As Alan Lightman relates in [111], it was Henrietta Leavitt's ground-breaking discovery, in 1912, of the "period-luminosity" law of variable Cepheid stars that eventually revealed just how enormous the universe really is. Cepheid stars are found in many parts of the sky. Their apparent brightness varies periodically. As Leavitt, working at the Harvard College Observatory, discovered, the greater the intrinsic luminosity of the star, the longer the period of variable brightness. The final step of calibration was achieved in 1913 by the Danish astronomer Ejnar Hertzsprung, when he was able to establish the actual distance to a relatively nearby Cepheid star, essentially by parallax methods.

There is a wonderful article by Eddington [69], in which he discusses the use of signal processing methods to discover the properties of the star Algol. This star, formally Algol (Beta Persei) in the constellation Perseus, turns out to be three stars, two revolving around the third, with both of the first two taking turns eclipsing the other. The stars rotate around their own axes, as our star, the sun, does, and the speed of rotation can be estimated by calculating the Doppler shift in frequency, as one side of the star comes toward us and the other side moves away. It is possible to measure one side at a time only because of the eclipse caused by the other revolving star.

2.10 The Transmission Problem

Now we change the situation and suppose that we are designing a broadcasting system, using transmitters at each x in the interval $[-L, L]$.

2.10.1 Directionality

At each x we will transmit $f(x)e^{i\omega t}$, where both $f(x)$ and ω are chosen by us. We now want to calculate what will be received at each point P in the far field. We may wish to design the system so that the strengths of the signals received at the various P are not all the same. For example, if we are broadcasting from Los Angeles, we may well want a strong signal in the north and south directions, but weak signals east and west, where there are fewer people to receive the signal. Clearly, our model of a single-frequency signal is too simple, but it does allow us to illustrate several important points about directionality in array processing.

2.10.2 The Case of Uniform Strength

For concreteness, we investigate the case in which $f(x) = 1$ for $|x| \leq L$. In this case, the measurement of the signal at the point P gives us

$$\begin{aligned} F(P) &= \int_{-L}^{L} f(x) \cos\Big(\frac{\omega \cos\theta}{c} x\Big) dx \\ &= \int_{-L}^{L} \cos\Big(\frac{\omega \cos\theta}{c} x\Big) dx \\ &= \frac{2c}{\omega \cos\theta} \sin\Big(\frac{L\omega \cos\theta}{c}\Big), \end{aligned}$$

when $\cos\theta \neq 0$. The absolute value of $F(P)$ is then the strength of the signal at P.

In Figures 2.2 through 2.7 we see the plots of the function $\frac{1}{2L}F(P)$, for various values of the aperture

$$A = \frac{L\omega}{\pi c} = \frac{2L}{\lambda}.$$

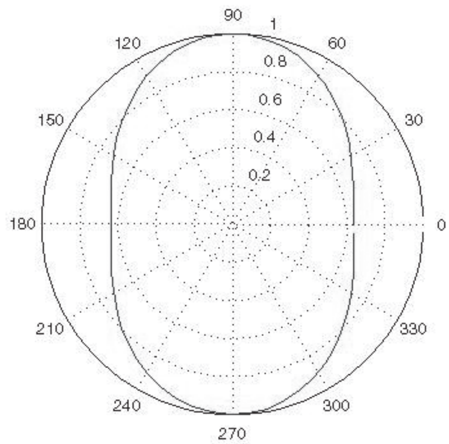

FIGURE 2.2: Relative strength at P for $A = 0.5$.

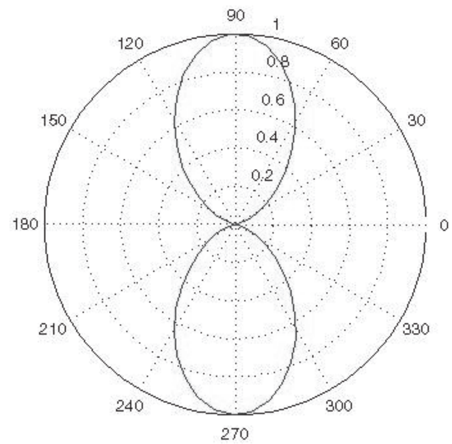

FIGURE 2.3: Relative strength at P for $A = 1.0$.

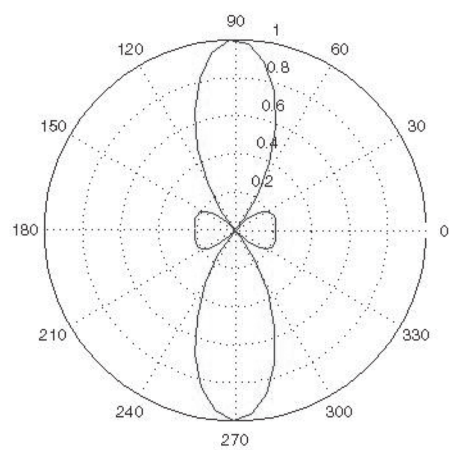

FIGURE 2.4: Relative strength at P for $A = 1.5$.

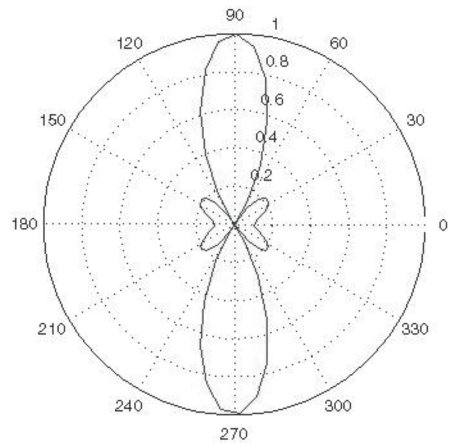

FIGURE 2.5: Relative strength at P for $A = 1.8$.

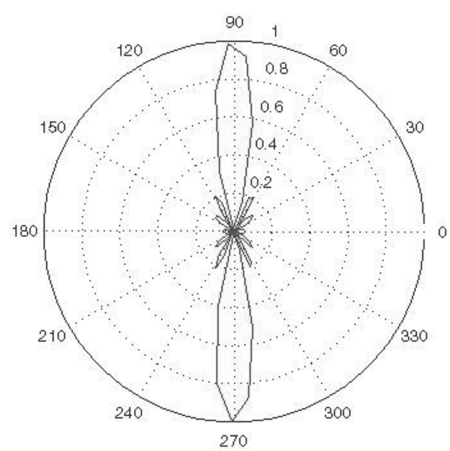

FIGURE 2.6: Relative strength at P for $A = 3.2$.

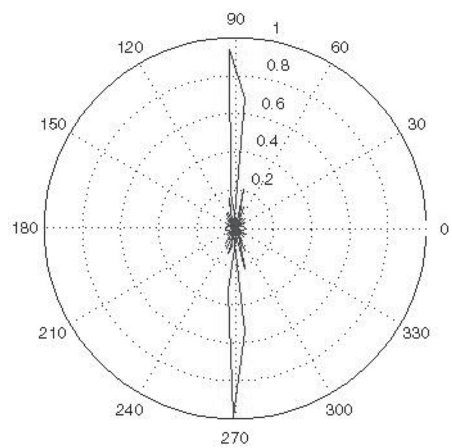

FIGURE 2.7: Relative strength at P for $A = 6.5$.

2.10.2.1 Beam-Pattern Nulls

Is it possible for the strength of the signal received at some P to be zero? As we saw in the previous section, to have zero signal strength, that is, to have $F(P) = 0$, we need

$$\sin\left(\frac{L\omega\cos\theta}{c}\right) = 0,$$

without

$$\cos\theta = 0.$$

Therefore, we need

$$\frac{L\omega\cos\theta}{c} = n\pi,$$

for some positive integers $n \geq 1$. Notice that this can happen only if

$$n \leq \frac{L\omega\pi}{c} = \frac{2L}{\lambda}.$$

Therefore, if $2L < \lambda$, there can be no P with signal strength zero. The larger $2L$ is, with respect to the wavelength λ, the more angles at which the signal strength is zero.

2.10.2.2 Local Maxima

Is it possible for the strength of the signal received at some P to be a local maximum, relative to nearby points in the far field? We write

$$F(P) = \frac{2c}{\omega\cos\theta}\sin\left(\frac{L\omega\cos\theta}{c}\right) = 2L\text{sinc}\,(H(\theta)),$$

where

$$H(\theta) = \frac{L\omega\cos\theta}{c}$$

and

$$\text{sinc}\,(H(\theta)) = \frac{\sin H(\theta)}{H(\theta)},$$

for $H(\theta) \neq 0$, and equals one for $H(\theta) = 0$. The value of A used previously is then $A = H(0)$.

Local maxima or minima of $F(P)$ occur when the derivative of $\text{sinc}\,(H(\theta))$ equals zero, which means that

$$H(\theta)\cos H(\theta) - \sin H(\theta) = 0,$$

or

$$\tan H(\theta) = H(\theta).$$

If we can solve this equation for $H(\theta)$ and then for θ, we will have found angles corresponding to local maxima of the received signal strength. The largest value of $F(P)$ occurs when $\theta = \frac{\pi}{2}$, and the peak in the plot of $F(P)$ centered at $\theta = \frac{\pi}{2}$ is called the *main lobe*. The smaller peaks on either side are called the *grating lobes*. We can see grating lobes in some of the polar plots.

2.11 The Laplace Transform and the Ozone Layer

We have seen how values of the Fourier transform can arise as measured data. The following examples, the first taken from Twomey's book [156], show that values of the Laplace transform can arise in this way as well.

2.11.1 The Laplace Transform

The Laplace transform of the function $f(x)$, defined for $0 \leq x < +\infty$, is the function

$$\mathcal{F}(s) = \int_0^{+\infty} f(x)e^{-sx}dx.$$

2.11.2 Scattering of Ultraviolet Radiation

The sun emits ultraviolet (UV) radiation that enters the earth's atmosphere at an angle θ_0 that depends on the sun's position, and with intensity $I(0)$. Let the x-axis be vertical, with $x = 0$ at the top of the atmosphere and x increasing as we move down to the earth's surface, at $x = X$. The intensity at x is given by

$$I(x) = I(0)e^{-kx/\cos\theta_0}.$$

Within the ozone layer, the amount of UV radiation scattered in the direction θ is given by

$$S(\theta, \theta_0)I(0)e^{-kx/\cos\theta_0}\Delta p,$$

where $S(\theta, \theta_0)$ is a known parameter, and Δp is the change in the pressure of the ozone within the infinitesimal layer $[x, x+\Delta x]$, and so is proportional to the concentration of ozone within that layer.

2.11.3 Measuring the Scattered Intensity

The radiation scattered at the angle θ then travels to the ground, a distance of $X - x$, weakened along the way, and reaches the ground with intensity

$$S(\theta, \theta_0)I(0)e^{-kx/\cos\theta_0}e^{-k(X-x)/\cos\theta}\Delta p.$$

The total scattered intensity at angle θ is then a superposition of the intensities due to scattering at each of the thin layers, and is then

$$S(\theta, \theta_0)I(0)e^{-kX/\cos\theta_0}\int_0^X e^{-x\beta}dp,$$

where

$$\beta = k\left(\frac{1}{\cos\theta_0} - \frac{1}{\cos\theta}\right).$$

This superposition of intensity can then be written as

$$S(\theta, \theta_0)I(0)e^{-kX/\cos\theta_0}\int_0^X e^{-x\beta}p'(x)dx.$$

2.11.4 The Laplace Transform Data

Using integration by parts, we get

$$\int_0^X e^{-x\beta}p'(x)dx = p(X)e^{-\beta X} - p(0) + \beta\int_0^X e^{-\beta x}p(x)dx.$$

Since $p(0) = 0$ and $p(X)$ can be measured, our data is then the Laplace transform value

$$\int_0^{+\infty} e^{-\beta x}p(x)dx;$$

note that we can replace the upper limit X with $+\infty$ if we extend $p(x)$ as zero beyond $x = X$.

The variable β depends on the two angles θ and θ_0. We can alter θ as we measure and θ_0 changes as the sun moves relative to the earth. In this way we get values of the Laplace transform of $p(x)$ for various values of β. The problem then is to recover $p(x)$ from these values. Because the Laplace transform involves a smoothing of the function $p(x)$, recovering $p(x)$ from its Laplace transform is more ill-conditioned than is the Fourier transform inversion problem.

2.12 The Laplace Transform and Energy Spectral Estimation

In x-ray transmission tomography, x-ray beams are sent through the object and the drop in intensity is measured. These measurements are then used to estimate the distribution of attenuating material within the object. A typical x-ray beam contains components with different energy levels. Because components at different energy levels will be attenuated differently, it is important to know the relative contribution of each energy level to the entering beam. The energy spectrum is the function $f(E)$ that describes the intensity of the components at each energy level $E > 0$.

2.12.1 The Attenuation Coefficient Function

Each specific material, say aluminum, for example, is associated with attenuation coefficients, which is a function of energy, which we shall denote by $\mu(E)$. A beam with the single energy E passing through a thickness x of the material will be weakened by the factor $e^{-\mu(E)x}$. By passing the beam through various thicknesses x of aluminum and registering the intensity drops, one obtains values of the absorption function

$$R(x) = \int_0^\infty f(E)e^{-\mu(E)x}dE. \tag{2.5}$$

Using a change of variable, we can write $R(x)$ as a Laplace transform.

2.12.2 The Absorption Function as a Laplace Transform

For each material, the attenuation function $\mu(E)$ is a strictly decreasing function of E, so $\mu(E)$ has an inverse, which we denote by g; that is, $g(t) = E$, for $t = \mu(E)$. Equation (2.5) can then be rewritten as

$$R(x) = \int_0^\infty f(g(t))e^{-tx}g'(t)dt.$$

We see then that $R(x)$ is the Laplace transform of the function $r(t) = f(g(t))g'(t)$. Our measurements of the intensity drops provide values of $R(x)$, for various values of x, from which we must estimate the functions $r(t)$, and, ultimately, $f(E)$.

Bibliography

[1] Anderson, T. (1972) "Efficient estimation of regression coefficients in time series." In *Proc. of Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: The Theory of Statistics.* Berkeley, CA:University of California Press, pp. 471–482.

[2] Ash, R. and Gardner, M. (1975) *Topics in Stochastic Processes.* Boston: Academic Press.

[3] Auslander, L., Kailath, T., and Mitter, S., eds. (1990) *Signal Processing Part I: Signal Processing Theory*, IMA Volumes in Mathematics and Its Applications, Volume 22. New York: Springer-Verlag.

[4] Baggeroer, A., Kuperman, W., and Schmidt, H. (1988) "Matched field processing: source localization in correlated noise as optimum parameter estimation." *Journal of the Acoustical Society of America* **83**, pp. 571–587.

[5] Benson, M. (2003) "What *Galileo* Saw." In *The New Yorker*; reprinted in [57].

[6] Bertero, M. (1992) "Sampling theory, resolution limits and inversion methods." In [8], pp. 71–94.

[7] Bertero, M. and Boccacci, P. (1998) *Introduction to Inverse Problems in Imaging.* Bristol, UK: Institute of Physics Publishing.

[8] Bertero, M. and Pike, E.R., eds. (1992) *Inverse Problems in Scattering and Imaging.* Malvern Physics Series, Adam Hilger, IOP Publishing, London.

[9] Bertsekas, D.P. (1997) "A new class of incremental gradient methods for least squares problems." *SIAM J. Optim.* **7**, pp. 913–926.

[10] Blackman, R. and Tukey, J. (1959) *The Measurement of Power Spectra.* New York: Dover Publications.

[11] Boggess, A. and Narcowich, F. (2001) *A First Course in Wavelets, with Fourier Analysis.* Englewood Cliffs, NJ: Prentice-Hall.

[12] Born, M. and Wolf, E. (1999) *Principles of Optics: 7th edition*. Cambridge, UK: Cambridge University Press.

[13] Bochner, S. and Chandrasekharan, K. (1949) *Fourier Transforms*, Annals of Mathematical Studies, No. 19. Princeton, NJ: Princeton University Press.

[14] Bolles, E.B. (1997) *Galileo's Commandment: 2,500 Years of Great Science Writing*. New York: W.H. Freeman.

[15] Bracewell, R.C. (1979) "Image reconstruction in radio astronomy." In [90], pp. 81–104.

[16] Bruckstein, A., Donoho, D., and Elad, M. (2009) "From sparse solutions of systems of equations to sparse modeling of signals and images." *SIAM Review* **51(1)**, pp. 34–81.

[17] Bruyant, P., Sau, J., and Mallet, J.J. (1999) "Noise removal using factor analysis of dynamic structures: application to cardiac gated studies." *Journal of Nuclear Medicine* **40(10)**, pp. 1676–1682.

[18] Bucker, H. (1976) "Use of calculated sound fields and matched field detection to locate sound sources in shallow water." *Journal of the Acoustical Society of America* **59**, pp. 368–373.

[19] Burg, J. (1967) "Maximum entropy spectral analysis." Paper presented at the 37th Annual SEG meeting, Oklahoma City, OK.

[20] Burg, J. (1972) "The relationship between maximum entropy spectra and maximum likelihood spectra." *Geophysics* **37**, pp. 375–376.

[21] Burg, J. (1975) *Maximum Entropy Spectral Analysis*. Ph.D. dissertation, Stanford University.

[22] Byrne, C. and Fitzgerald, R. (1979) "A unifying model for spectrum estimation." In *Proceedings of the RADC Workshop on Spectrum Estimation*. October 1979. Rome, NY: Griffiss AFB.

[23] Byrne, C. and Fitzgerald, R. (1982) "Reconstruction from partial information, with applications to tomography." *SIAM J. Applied Math.* **42(4)**, pp. 933–940.

[24] Byrne, C., Fitzgerald, R., Fiddy, M., Hall, T. and Darling, A. (1983) "Image restoration and resolution enhancement." *J. Opt. Soc. Amer.* **73**, pp. 1481–1487.

[25] Byrne, C. and Wells, D. (1983) "Limit of continuous and discrete finite-band Gerchberg iterative spectrum extrapolation." *Optics Letters* **8(10)**, pp. 526–527.

[26] Byrne, C. and Fitzgerald, R. (1984) "Spectral estimators that extend the maximum entropy and maximum likelihood methods." *SIAM J. Applied Math.* **44(2)**, pp. 425–442.

[27] Byrne, C., Levine, B.M., and Dainty, J.C. (1984) "Stable estimation of the probability density function of intensity from photon frequency counts." *JOSA Communications* **1(11)**, pp. 1132–1135.

[28] Byrne, C. and Steele, A. (1985) "Stable nonlinear methods for sensor array processing." *IEEE Transactions on Oceanic Engineering* **OE-10(3)**, pp. 255–259.

[29] Byrne, C. and Wells, D. (1985) "Optimality of certain iterative and non-iterative data extrapolation procedures." *Journal of Mathematical Analysis and Applications* **111(1)**, pp. 26–34.

[30] Byrne, C. and Fiddy, M. (1987) "Estimation of continuous object distributions from Fourier magnitude measurements." *JOSA A* **4**, pp. 412–417.

[31] Byrne, C. and Fiddy, M. (1988) "Images as power spectra; reconstruction as Wiener filter approximation." *Inverse Problems* **4**, pp. 399–409.

[32] Byrne, C., Brent, R., Feuillade, C., and DelBalzo, D (1990) "A stable data-adaptive method for matched-field array processing in acoustic waveguides." *Journal of the Acoustical Society of America* **87(6)**, pp. 2493–2502.

[33] Byrne, C., Frichter, G., and Feuillade, C. (1990) "Sector-focused stability methods for robust source localization in matched-field processing." *Journal of the Acoustical Society of America* **88(6)**, pp. 2843–2851.

[34] Byrne, C. (1993) "Iterative image reconstruction algorithms based on cross-entropy minimization." *IEEE Transactions on Image Processing* **IP-2**, pp. 96–103.

[35] Byrne, C., Haughton, D., and Jiang, T. (1993) "High-resolution inversion of the discrete Poisson and binomial transformations." *Inverse Problems* **9**, pp. 39–56.

[36] Byrne, C. (1995) "Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization'." *IEEE Transactions on Image Processing* **IP-4**, pp. 225–226.

[37] Byrne, C. (1996) "Block-iterative methods for image reconstruction from projections." *IEEE Transactions on Image Processing* **IP-5**, pp. 792–794.

[38] Byrne, C. (1997) "Convergent block-iterative algorithms for image reconstruction from inconsistent data." *IEEE Transactions on Image Processing* **IP-6**, pp. 1296–1304.

[39] Byrne, C. (1998) "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative (RBI) methods." *IEEE Transactions on Image Processing* **IP-7**, pp. 100–109.

[40] Byrne, C. (2002) "Iterative oblique projection onto convex sets and the split feasibility problem." *Inverse Problems* **18**, pp. 441–453.

[41] Byrne, C. (2008) *Applied Iterative Methods*. Wellesley, MA: A K Peters, Publ.

[42] Byrne, C. (2009) *Applied and Computational Linear Algebra: A First Course*. Unpublished text available at my web site.

[43] Byrne, C. (2014) *A First Course in Optimization*. Boca Raton: CRC Press.

[44] Byrne, C. (2014) *Iterative Optimization in Inverse Problems*. Boca Raton: CRC Press.

[45] Candès, E., Romberg, J., and Tao, T. (2006) "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." *IEEE Transactions on Information Theory* **52(2)**, pp. 489–509.

[46] Candès, E. and Romberg, J. (2007) "Sparsity and incoherence in compressive sampling." *Inverse Problems* **23(3)**, pp. 969–985.

[47] Candès, E., Wakin, M., and Boyd, S. (2007) "Enhancing sparsity by reweighted l_1 minimization." Preprint available at http://www.acm.caltech.edu/ emmanuel/publications.html.

[48] Candy, J. (1988) *Signal Processing: The Modern Approach*. New York: McGraw-Hill Publ.

[49] Capon, J. (1969) "High-resolution frequency-wavenumber spectrum analysis." *Proc. of the IEEE* **57**, pp. 1408–1418.

[50] Cederquist, J., Fienup, J., Wackerman, C., Robinson, S., and Kryskowski, D. (1989) "Wave-front phase estimation from Fourier intensity measurements." *Journal of the Optical Society of America A* **6(7)**, pp. 1020–1026.

[51] Champeney, D.C. (1987) *A Handbook of Fourier Theorems*. Cambridge, UK: Cambridge University Press.

[52] Chang, J.-H., Anderson, J.M.M., and Votaw, J.R. (2004) "Regularized image reconstruction algorithms for positron emission tomography." *IEEE Transactions on Medical Imaging* **23(9)**, pp. 1165–1175.

[53] Childers, D., ed. (1978) *Modern Spectral Analysis*. New York: IEEE Press.

[54] Christensen, O. (2003) *An Introduction to Frames and Riesz Bases*. Boston: Birkhäuser.

[55] Chui, C. (1992) *An Introduction to Wavelets*. Boston: Academic Press.

[56] Chui, C. and Chen, G. (1991) *Kalman Filtering, 2nd edition*. Berlin: Springer-Verlag.

[57] Cohen, J., ed. (2010) *The Best of The Best American Science Writing*. New York: Harper-Collins Publ.

[58] Cooley, J. and Tukey, J. (1965) "An algorithm for the machine calculation of complex Fourier series." *Math. Comp.* **19**, pp. 297–301.

[59] Cox, H. (1973) "Resolving power and sensitivity to mismatch of optimum array processors." *Journal of the Acoustical Society of America* **54**, pp. 771–785.

[60] Csiszár, I. (1991) "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems." *The Annals of Statistics* **19(4)**, pp. 2032–2066.

[61] Csiszár, I. and Tusnády, G. (1984) "Information geometry and alternating minimization procedures." *Statistics and Decisions* **Supp. 1**, pp. 205–237.

[62] Dainty, J. C. and Fiddy, M. (1984) "The essential role of prior knowledge in phase retrieval." *Optica Acta* **31**, pp. 325–330.

[63] Daubechies, I. (1988) "Orthogonal bases of compactly supported wavelets." *Commun. Pure Appl. Math.* **41**, pp. 909–996.

[64] Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics.

[65] De Bruijn, N. (1967) "Uncertainty principles in Fourier analysis." In *Inequalties*, O. Shisha, editor, pp. 57–71. Boston: Academic Press.

[66] Dhanantwari, A., Stergiopoulos, S., and Iakovidis, I. (2001) "Correcting organ motion artifacts in x-ray CT medical imaging systems by adaptive processing. I. Theory." *Med. Phys.* **28(8)**, pp. 1562–1576.

[67] Donoho, D. (2006) "Compressed sensing." *IEEE Transactions on Information Theory* **52(4)**, pp. 1289–1306.

[68] Duda, R., Hart, P., and Stork, D. (2001) *Pattern Classification.* New York: John Wiley and Sons, Inc.

[69] Eddington, A. (1927) "The story of Algol." *Stars and Atoms*; reprinted in [14].

[70] Feuillade, C., DelBalzo, D., and Rowe, M. (1989) "Environmental mismatch in shallow-water matched-field processing: geoacoustic parameter variability." *Journal of the Acoustical Society of America* **85**, pp. 2354–2364.

[71] Feynman, R. (1985) *QED: The Strange Theory of Light and Matter.* Princeton, NJ: Princeton University Press.

[72] Feynman, R., Leighton, R., and Sands, M. (1963) *The Feynman Lectures on Physics, Vol. 1.* Boston: Addison-Wesley.

[73] Fiddy, M. (1983) "The phase retrieval problem." In *Inverse Optics*, SPIE Proceedings 413 (A.J. Devaney, editor), pp. 176–181.

[74] Fiddy, M. (2008) *Private communication.*

[75] Fienup, J. (1979) "Space object imaging through the turbulent atmosphere." *Optical Engineering* **18**, pp. 529–534.

[76] Fienup, J. (1982) "Phase retrieval algorithms: a comparison." *Applied Optics* **21**, pp. 2758–2769.

[77] Fienup, J. (1987) "Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint." *Journal of the Optical Society of America A* **4(1)**, pp. 118–123.

[78] Frieden, B. R. (1982) *Probability, Statistical Optics and Data Testing.* Berlin: Springer-Verlag.

[79] Gabor, D. (1946) "Theory of communication." *Journal of the IEE (London)* **93**, pp. 429–457.

[80] Gasquet, C. and Witomski, F. (1998) *Fourier Analysis and Applications.* Berlin: Springer-Verlag.

[81] Gelb, A., ed. (1974) *Applied Optimal Estimation*, written by the technical staff of The Analytic Sciences Corporation. Cambridge, MA: MIT Press.

[82] Geman, S. and Geman, D. (1984) "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, pp. 721–741.

[83] Gerchberg, R. W. (1974) "Super-restoration through error energy reduction." *Optica Acta* **21**, pp. 709–720.

[84] Gordon, R., Bender, R., and Herman, G.T. (1970) "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography." *J. Theoret. Biol.* **29**, pp. 471–481.

[85] Green, P. (1990) "Bayesian reconstructions from emission tomography data using a modified EM algorithm." *IEEE Transactions on Medical Imaging* **9**, pp. 84–93.

[86] Groetsch, C. (1999) *Inverse Problems: Activities for Undergraduates.* Washington, DC: The Mathematical Association of America.

[87] Grünbaum, F.A., Helton, J.W., and Khargonekar, P., eds. (1990) *Signal Processing Part II: Control Theory and Applications*, IMA Volumes in Mathematics and Its Applications, Volume 23. New York: Springer-Verlag.

[88] Haykin, S. (1985) *Array Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall.

[89] Hebert, T. and Leahy, R. (1989) "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors." *IEEE Transactions on Medical Imaging* **8**, pp. 194–202.

[90] Herman, G.T., ed. (1979) *Image Reconstruction from Projections: Topics in Applied Physics, Vol. 32*, Berlin: Springer-Verlag.

[91] Hinich, M. (1973) "Maximum likelihood signal processing for a vertical array." *Journal of the Acoustical Society of America* **54**, pp. 499–503.

[92] Hinich, M. (1979) "Maximum likelihood estimation of the position of a radiating source in a waveguide." *Journal of the Acoustical Society of America* **66**, pp. 480–483.

[93] Hoffman, K. (1962) *Banach Spaces of Analytic Functions.* Englewood Cliffs, NJ: Prentice-Hall.

[94] Hogg, R. and Craig, A. (1978) *Introduction to Mathematical Statistics.* New York: MacMillan.

[95] Hubbard, B. (1998) *The World According to Wavelets.* Natick, MA: A K Peters, Inc.

[96] Johnson, C., Hendriks, E., Berezhnoy, I., Brevdo, E., Hughes, S., Daubechies, I., Li, J., Postma, E., and Wang, J. (2008) "Image Processing for Artist Identification." *IEEE Signal Processing Magazine* **25(4)**, pp. 37–48.

[97] Kaiser, G. (1994) *A Friendly Guide to Wavelets*. Boston: Birkhäuser.

[98] Kalman, R. (1960) "A new approach to linear filtering and prediction problems." *Trans. ASME, J. Basic Eng.* **82**, pp. 35–45.

[99] Katznelson, Y. (1983) *An Introduction to Harmonic Analysis*. New York: John Wiley and Sons, Inc.

[100] Kheifets, A. (2004) *Private communication.*

[101] Körner, T. (1988) *Fourier Analysis*. Cambridge, UK: Cambridge University Press.

[102] Körner, T. (1996) *The Pleasures of Counting*. Cambridge, UK: Cambridge University Press.

[103] Lane, R. (1987) "Recovery of complex images from Fourier magnitude." *Optics Communications* **63(1)**, pp. 6–10.

[104] Lange, K. and Carson, R. (1984) "EM reconstruction algorithms for emission and transmission tomography." *Journal of Computer Assisted Tomography* **8**, pp. 306–316.

[105] Lange, K., Bahn, M. and Little, R. (1987) "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography." *IEEE Trans. Med. Imag.* **MI-6(2)**, pp. 106–114.

[106] Leahy, R., Hebert, T., and Lee, R. (1989) "Applications of Markov random field models in medical imaging." In *Proceedings of the Conference on Information Processing in Medical Imaging*. Berkeley, CA: Lawrence-Berkeley Laboratory.

[107] Lent, A. (1998) *Private communication.*

[108] Levi, A. and Stark, H. (1984) "Image restoration by the method of generalized projections, with application to restoration from magnitude." *J. Opt. Soc. Am. A* **1**, pp. 932–943.

[109] Levitan, E. and Herman, G.T. (1987) "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography." *IEEE Transactions on Medical Imaging* **6**, pp. 185–192.

[110] Liao, C.-W., Fiddy, M., and Byrne, C. (1997) "Imaging from the zero locations of far-field intensity data." *Journal of the Optical Society of America: A* **14(12)**, pp. 3155–3161.

[111] Lightman, A. (2005) *Discoveries: Great Breakthroughs in 20th Century Science*. New York: Vintage Books.

[112] Lindberg, D. (1992) *The Beginnings of Western Science*. Chicago: University of Chicago Press.

[113] Luenberger, D. (1969) *Optimization by Vector Space Methods*. New York: John Wiley and Sons, Inc.

[114] Magness, T. and McQuire, J. (1962) "Comparison of least squares and minimum variance estimates of regression parameters." *Annals of Mathematical Statistics* **33**, pp. 462–470.

[115] Mallat, S.G. (1989) "A theory of multiresolution signal decomposition: The wavelet representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-11**, pp. 674–693.

[116] McClellan, J., Schafer, R., and Yoder, M. (2003) *Signal Processing First*. Upper Saddle River, NJ: Prentice Hall, Inc.

[117] Meidunas, E. (2001) *Re-scaled Block Iterative Expectation Maximization Maximum Likelihood (RBI-EMML) Abundance Estimation and Sub-pixel Material Identification in Hyperspectral Imagery*. MS thesis, Department of Electrical Engineering, University of Massachusetts Lowell.

[118] Meyer, Y. (1993) *Wavelets: Algorithms and Applications*. Philadelphia, PA: SIAM Publ.

[119] Muller, R. (2008) *Physics for Future Presidents: The Science Behind the Headlines*. New York: W.W. Norton and Company.

[120] Oppenheim, A. and Schafer, R. (1975) *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.

[121] Papoulis, A. (1975) "A new algorithm in spectral analysis and band-limited extrapolation." *IEEE Transactions on Circuits and Systems* **22**, pp. 735–742.

[122] Papoulis, A. (1977) *Signal Analysis*. New York: McGraw-Hill.

[123] Paulraj, A., Roy, R., and Kailath, T. (1986) "A subspace rotation approach to signal parameter estimation." *Proceedings of the IEEE* **74**, pp. 1044–1045.

[124] Pelagotti, A., Del Mastio, A., De Rosa, A., Piva, A. (2008) "Multispectral imaging of paintings." *IEEE Signal Processing Magazine* **25(4)**, pp. 27–36.

[125] Penrose, R. (2007) *The Road to Reality: A Complete Guide to the Laws of the Universe*. New York: Vintage Books.

[126] Pisarenko, V. (1973) "The retrieval of harmonics from a covariance function." *Geoph. J. R. Astrom. Soc.* **30**, pp. 347–366 .

[127] Pižurica, A., Philips, W., Lemahieu, I., and Acheroy, M. (2003) "A versatile wavelet domain noise filtration technique for medical imaging." *IEEE Transactions on Medical Imaging: Special Issue on Wavelets in Medical Imaging* **22**, pp. 323–331.

[128] Poggio, T. and Smale, S. (2003) "The mathematics of learning: dealing with data." *Notices of the American Mathematical Society* **50(5)**, pp. 537–544.

[129] Priestley, M. B. (1981) *Spectral Analysis and Time Series*. Boston: Academic Press.

[130] Prony, G.R.B. (1795) "Essai expérimental et analytique sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansion de la vapeur de l'alcool, à différentes températures." *Journal de l'Ecole Polytechnique* (Paris) **1(2)**, pp. 24–76.

[131] Qian, H. (1990) "Inverse Poisson transformation and shot noise filtering." *Rev. Sci. Instrum.* **61**, pp. 2088–2091.

[132] Ribés, A., Pillay, R., Schmitt, F., and Lahanier, C. (2008) "Studying that smile." *IEEE Signal Processing Magazine* **25(4)**, pp. 14–26.

[133] Romberg, J. (2007) "Compressed sensing creates a buzz at ICIAM '07." *SIAM NEWS*, October 2007, p. 7.

[134] Rudin, W. (1953) *Principles of Mathematical Analysis*. New York: McGraw-Hill.

[135] Sato, M. (1958) "On the generalization of the concept of a function." *Proc. Japan Acad.* **34**, pp. 126–130.

[136] Schmidt, R. (1981) *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. PhD thesis, Stanford University.

[137] Schultz, L., Blanpied, G., Borozdin, K., *et al.* (2007) "Statistical reconstruction for cosmic ray muon tomography." *IEEE Transactions on Image Processing* **16(8)**, pp. 1985–1993.

[138] Schuster, A. (1898) "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena." *Terrestrial Magnetism* **3**, pp. 13–41.

[139] Shang, E. (1985) "Source depth estimation in waveguides." *Journal of the Acoustical Society of America* **77**, pp. 1413–1418.

[140] Shang, E. (1985) "Passive harmonic source ranging in waveguides by using mode filter." *Journal of the Acoustical Society of America* **78**, pp. 172–175.

[141] Shang, E., Wang, H., and Huang, Z. (1988) "Waveguide characterization and source localization in shallow water waveguides using Prony's method." *Journal of the Acoustical Society of America* **83**, pp. 103–106.

[142] Shieh, M., Byrne, C., and Fiddy, M. (2006) "Image reconstruction: a unifying model for resolution enhancement and data extrapolation: Tutorial." *Journal of the Optical Society of America A* **23(2)**, pp. 258–266.

[143] Shieh, M., Byrne, C., Testorf, M., and Fiddy, M. (2006) "Iterative image reconstruction using prior knowledge." *Journal of the Optical Society of America: A* **23(6)**, pp. 1292–1300.

[144] Shieh, M. and Byrne, C. (2006) "Image reconstruction from limited Fourier data." *Journal of the Optical Society of America: A* **23(11)**, pp. 2732–2736.

[145] Smith, C. Ray and Grandy, W.T., eds. (1985) *Maximum-Entropy and Bayesian Methods in Inverse Problems*. Dordrecht: Reidel Publ.

[146] Smith, C. Ray and Erickson, G., eds. (1987) *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*. Dordrecht: Reidel Publ.

[147] Sondhi, M. (2006) "The History of Echo Cancellation." *IEEE Signal Processing Magazine*, September 2006, pp. 95–102.

[148] Sondhi, M., Morgan, D., and Hall, J. (1995) "Stereophonic acoustic echo cancellation- an overview of the fundamental problem." *IEEE Signal Processing Letters* **2(8)**, pp. 148–151.

[149] Stark, H. and Woods, J. (2002) *Probability and Random Processes, with Applications to Signal Processing*. Upper Saddle River, NJ: Prentice-Hall.

[150] Stark, H. and Yang, Y. (1998) *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets and Optics*. New York: John Wiley and Sons, Inc.

[151] Strang, G. (1980) *Linear Algebra and Its Applications*. New York: Academic Press.

[152] Strang, G. and Nguyen, T. (1997) *Wavelets and Filter Banks*. Wellesley, MA: Wellesley-Cambridge Press.

[153] Therrien, C. (1992) *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.

[154] Tindle, C., Guthrie, K., Bold, G., Johns, M., Jones, D., Dixon, K., and Birdsall, T. (1978) "Measurements of the frequency dependence of normal modes." *Journal of the Acoustical Society of America* **64**, pp. 1178–1185.

[155] Tolstoy, A. (1993) *Matched Field Processing for Underwater Acoustics*. Singapore: World Scientific.

[156] Twomey, S. (1996) *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurement*. New York: Dover Publ.

[157] Unser, M. (1999) "Splines: A perfect fit for signal and image processing." *IEEE Signal Processing Magazine* **16**, pp. 22–38.

[158] Van Trees, H. (1968) *Detection, Estimation and Modulation Theory*. New York: John Wiley and Sons, Inc.

[159] Walnut, D. (2002) *An Introduction to Wavelets*. Boston: Birkhäuser.

[160] Watson, J. (1968) *The Double Helix*. New York: Atheneum.

[161] Widrow, B. and Stearns, S. (1985) *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.

[162] Wiener, N. (1949) *Time Series*. Cambridge, MA: MIT Press.

[163] Wiener, N. (1964) *Generalized Harmonic Analysis and Tauberian Theorems*. Cambridge, MA: MIT Press.

[164] Wright, W., Pridham, R., and Kay, S. (1981) "Digital signal processing for sonar." *Proc. IEEE* **69**, pp. 1451–1506.

[165] Yang, T.C. (1987) "A method of range and depth estimation by modal decomposition." *Journal of the Acoustical Society of America* **82**, pp. 1736–1745.

[166] Yin, W. and Zhang, Y. (2008) "Extracting salient features from less data via l_1-minimization." *SIAG/OPT Views-and-News* **19(1)**, pp. 11–19.

[167] Young, R. (1980) *An Introduction to Nonharmonic Fourier Analysis*. Boston: Academic Press.