Reconstruction of Two-Dimensional Signals from the Fourier Transform Magnitude

RLE Technical Report No. 517

May 1986

David Izraelevitz

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

This work has been supported in part by the Advanced Research Projects Agency monitored by the Office of Naval Research under Contract No. N00014-81-K-0742 and in part by the National Science Foundation under Grant ECS-84-07285.

Reconstruction of Two-Dimensional Signals from the Fourier Transform Magnitude

by

David Izraelevitz

Submitted in partial fulfillment of the requirements for the degree of Doctor of Science at the Massachusetts Institute of Technology.
May 8, 1986

Abstract

This thesis is concerned with the problem of reconstructing a discrete two-dimensional signal of known support from the Fourier transform magnitude only. This problem arises in many fields where imaging is desired, such as astronomy and wavefront sensing.

Since the autocorrelation function is easily calculated from the Fourier transform magnitude, we attack the equivalent problem of signal reconstruction from a known autocorrelation function. The main result of the thesis is a new algorithm for realizing this reconstruction. This algorithm is guaranteed to yield the correct solution given accurate measurements and is much more computationally attractive than previous reconstruction algorithms. The result is based on the detailed analysis of the zeros of a polynomial which is essentially the two-dimensional z-transform of the known autocorrelation signal. From this analysis, a large number of zeros of the z-transform of the unknown discrete signal are extracted. This set of zeros is then used to extract the signal values via the solution of a set of linear equations.

Examples of the application of this algorithm to several families of images is presented, along with a discussion of the accuracy and computational requirements of the new algorithm. We conclude with a discussion of the application of the ideas of this thesis to the area of two-dimensional filter design and stability testing.

Thesis Supervisor: Jae S. Lim
Title: Associate Professor of Electrical Engineering.

With Terry

Acknowledgements

Acknowledgements are always fun. I have heard of only one person who wrote the acknowledgements before writing the rest of the thesis and I believe he ended up never finishing the thesis. How could he, when the best part of the whole dissertation has been done and there is still left at least three years of hard work?

The rest of us, I suspect, wisely wait until the very end to slip in this page. By this time all the hard work is done; all the writing is complete, the equations and figures checked and the advisor and readers have approved of the contents. All the anxiety is over with and one can think back with fondness of all the help received on the way here.

My sincere thanks go to Professor Jae S. Lim for the advice and focus offered throughout the doctoral program. I have worked with Jae for the last five years, ever since I came to MIT, and during this time have learned from him a tremendous amount on the fine art of research. I also thank Professor Alan Oppenheim for his continuing interest and insight into the problem of signal reconstruction in general and phase retrieval in particular. Professors Michael Artin of the Mathematics Dept. at MIT and Eric Kaltofen of the Computer Science Dept. at RPI provided valuable help in the initial stages of my research. I also want to acknowledge the value of several rewarding conversations with Avideh Zakhor and Dr. Susan Curtis, who are involved in research related to the work presented here. In addition, the results of one of the appendices of this thesis is due to a collaboration between Ms. Zakhor and myself. Many, if not most of the members of the DSPG group engaged in the sometimes painful task of listening and discussing my half-baked ideas. Dr. Michael Wengrovitz, Cory Myers and Webster Dove were specially abused in this respect.

Of course, one does not complete a thesis by signal processing alone. Ever since I left home for college, I have always relied on the love, support and Sunday phone calls of my parents. I thank them now as I will thank them forever for this. Finally, at about the same time as I started my doctoral thesis, I fell in love with a cute girl from St. Louis. The deepest thanks go to my wife, Terry, for being with me. It made it all easy.

Contents

Chapter 1

Introduction

The magnitude and phase of the Fourier transform of an arbitrary multidimensional signal are independent functions of frequency. In many applications, however, there is additional information regarding the signal which provides a very strong connection between its Fourier transform magnitude and phase. One example of such additional information is the common condition that the signal is non-zero only over a specified region. In this case, it has been shown that almost all multidimensional signals which are non-zero only over a specified region are uniquely specified, in a sense, by knowledge of only its Fourier transform magnitude [1,2]. Hence, once the Fourier transform magnitude is known, the Fourier transform phase is determined as well. For this reason, the problem of reconstruction from Fourier transform magnitude is also called the phase retrieval problem.

The reconstruction of a two-dimensional signal from its Fourier transform magnitude has been the object of much study. This interest is guided by the wide range of applications of results in this area. One such application is in astronomy [3]. The effect limiting the resolution capabilities of the largest optical telescopes is not the diffraction limit of the lens, but rather the turbulence of the earth's atmosphere. During a short time period, the atmosphere can be considered to introduce on the incoming optical wave a spatially varying, time-invariant random phase delay due to inhomogeneities induced by thermal gradients. These inhomogeneities are slowly varying with respect to short exposure times. Thus, over such a short time period, the atmosphere can be modeled as a glass plate of spatially varying thickness over the telescope aperture.

This phase aberration, although it blurs each individual exposed image, does not affect the spatial autocorrelation function. A way of circumventing this blurring effect is to first measure an accurate estimate of the spatial autocorrelation function. This can be done via Labeyrie interferometry [4]. In this procedure an interferometer is used to image the spatial autocorrelation function over a small time period. This estimate will be very noisy because of the short exposure time. The signal-to-noise ratio however can be increased by averaging several short exposures. Thus a diffraction limited autocorrelation function can be measured which is not affected by atmospheric blurring. It is clear that a reliable method for extracting the image of the astronomical object from such interferometer data would in effect greatly increase the resolution capabilities of earth-based telescopes.

A possible application of phase retrieval to electron microscopy lies in the possibility of indirect phase measurement from magnitude measurement [5]. Photographic film can only record the intensity of the field impinging on it. However, the phase of the field provides important information on the object being viewed. For example, thin objects may be considered as modulating the phase of the electron wave while not affecting the magnitude. The actual phase delay introduced depends on the thickness and composition of the specimen. Retrieving the phase delays from the recorded field intensity would yield an indirect way of measuring the specimen properties.

X-ray crystallography is a third realm where reconstruction from Fourier transform magnitude may prove useful [6]. Physical arguments show that the angles at which the x-rays are diffracted from a crystal specimen and the intensity of the diffracted wave at each angle are related to the Fourier transform magnitude of the electron density of the crystal under study. An important part of crystallography is the task of deducing the arrangement of atoms in the crystal from knowledge of such diffraction data.

The importance of the phase retrieval problem has led several researchers to propose algorithms for reconstruction from Fourier transform magnitude. However, previously presented algorithms fall into either of two categories; they are heuristic algorithms which often do not converge to the true reconstruction, or they are computationally too expensive for even moderate size signals. The purpose of this thesis is to present a new algorithm for reconstruction of multidimensional discrete signals from Fourier transform magnitude which is a closed form solution to the problem and which has been used with success in reconstructing signals of moderate size.

The thesis is divided into eight chapters; Chapter 2 develops an appropriate notation, reviews basic properties of signals and introduces some mathematical concepts. Chapter 3 is concerned with the review of previous results in reconstruction of both one- and two-dimensional signals from the Fourier transform magnitude. The formulation of the phase retrieval problem as a bivariate polynomial factorization problem is given in Chapter 4. Previous algorithms for factoring polynomials in two variables and their connection and possible application to the phase retrieval problem are also discussed. Based on this framework, a new algorithm for factoring large polynomials in two variables is developed in Chapter 5. This leads to a new closed form algorithm for solving the phase retrieval problem. Examples of the application of the new phase retrieval algorithm is the subject of Chapter 6. Chapter 7 discusses the application of the ideas developed in this thesis to the areas of general bivariate polynomial factorization and filter stability testing. A summary of the work presented and suggestions for future research is the subject of Chapter 8.

Chapter 2

Notation and Basic Properties

In this chapter we review some concepts from digital signal processing and polynomials in one or several variables. The purpose of this review is primarily to develop a consistent notation and provide a base for our later development. The interested reader may consult [7] for a more detailed discussion of the signal processing topics described here. A development of the properties of polynomials reviewed here is contained in [8].

2.1 One- and Two-Dimensional Signals

A one-dimensional discrete signal $x[n]$ or a two-dimensional discrete signal $x[m, n]$ is a real or complex function of a single integer index n or two integer indices m and n respectively. The support of $x[n]$ is the set (or sometimes a superset) of all indices n such that $x[n]$ is non-zero. The support of $x[m, n]$ is also defined as the set of all index pairs such that $x[m, n]$ is non-zero.

In our discussion we will find special cases of support to be especially useful. A signal, either one- or two-dimensional, is said to be of finite extent if its support is bounded; in one dimension, this means that there is an index pair, (n_{min}, n_{max}) such that $x[n] = 0$ for $n > n_{max}$ and $n < n_{min}$. Similarly in two dimensions a signal is of finite extent if there is an index quadruple $(m_{max}, m_{min}, n_{max}, n_{min})$ such that $x[m, n] = 0$ whenever any of the four conditions below hold:

$$m > m_{max}, \; m < m_{min}, \; n > n_{max}, \; n < n_{min} \tag{2.1}$$

Pictorially, this means that the non-zero values of $x[m, n]$ can be enclosed in a box, Figure 2.1.

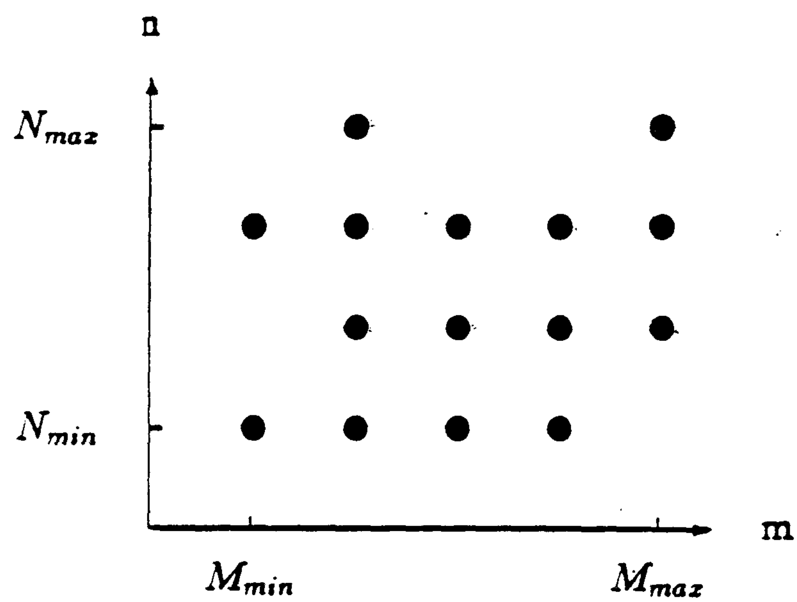

Figure 2.1: A two-dimensional signal with support $[M_{min}, M_{max}] \times [N_{min}, N_{max}]$.

We will denote a region of support consisting of all indices $a \leq n \leq b$ as $[a, b]$. In two dimensions a support comprising all index pairs (m, n) satisfying $a \leq n \leq b$, $c \leq m \leq d$ will be denoted by $[a, b] \times [c, d]$. We will abbreviate $[a, b] \times [a, b]$ by $[a, b]^2$.

Associated with any signal $x[n]$ is a Laurent series called its z-transform,

$$X(z) = \sum_n x[n] z^{-n} \tag{2.2}$$

where z is a complex number. The set of all z where the summation converges is called the region of convergence (ROC) of $X(z)$. In this thesis, we will be dealing primarily with signals of finite extent, in which case the ROC includes all of the complex z-plane with the possible exclusion of the origin or infinity. Evaluating the z-transform in (2.2) on the unit circle $|z| = 1$ yields the Fourier transform,

$$X(e^{jv}) = \sum_n x[n] e^{-jnv} \tag{2.3}$$

Even if $x[n]$ is real, its Fourier transform may be complex-valued, and can thus be represented in terms of its real and imaginary components,

$$X(e^{jv}) = X_r(e^{jv}) + jX_i(e^{jv}) \tag{2.4}$$

or in polar form

$$X(e^{jv}) = |X(e^{jv})| e^{j\Theta_x(v)} \tag{2.5}$$

Two-dimensional discrete signals also have a z-transform which is a complex-valued function of two complex numbers w and z,

$$X(w, z) = \sum_m \sum_n x[m, n] w^{-m} z^{-n} \tag{2.6}$$

The definition of an ROC also applies to two-dimensional z-transforms. Whenever the ROC includes the bicircle $|w| = 1$, $|z| = 1$, the Fourier transform of $x[m, n]$ is defined:

$$X(e^{ju}, e^{jv}) = \sum_m \sum_n x[m, n] e^{-jum} e^{-jvn} \tag{2.7}$$

An important signal which is generated from $x[n]$ is its autocorrelation function, defined via the summation below,

$$r[n] = \sum_l x[l]x^*[l+n] \tag{2.8}$$

It is straightforward to show that if $x[n]$ is a finite extent signal of support $[a, b]$ then $r[n]$ has support $[-(b-a), (b-a)]$. Note that the autocorrelation function is invariant to multiplication of $x[n]$ by a constant of unit magnitude or to a translation of $x[n]$.

From the convolution theorem [7], the following relationship is derived between $X(z)$ and the z-transform of $r[n]$, $R(z)$,

$$R(z) = X(z)X^*(\frac{1}{z^*}) \tag{2.9}$$

When evaluated on the unit circle, the above equation collapses to a relationship between the Fourier transforms of $x[n]$ and $r[n]$,

$$R(e^{j\upsilon}) = |X(e^{j\upsilon})|^2 \tag{2.10}$$

From (2.10) one can make an important observation that if two signals have the same autocorrelation function, they must have the same Fourier transform magnitude, and vice versa.

The autocorrelation function of $x[m, n]$ is similarly defined by

$$r[m,n] = \sum_k \sum_l x[k,l]x^*[k+m, l+n] \tag{2.11}$$

The support of $r[m, n]$ is given by $[-(b-a), (b-a)] \times [-(d-c), (d-c)]$ for $x[m, n]$ with support $[a, b] \times [c, d]$.

The convolution theorem also applies to the two-dimensional case; the z-transform of $r[m,n]$, $R(w,z)$, is given by

$$R(w,z) = X(w,z)X^*(\frac{1}{w^*}, \frac{1}{z^*}) \tag{2.12}$$

On the unit bicircle, (2.12) becomes

$$R(e^{ju}, e^{jv}) = |X(e^{ju}, e^{jv})|^2 \tag{2.13}$$

Thus we see that as in the one-dimensional case, if the Fourier transform magnitude of a signal is known then its autocorrelation function is known also and vice versa.

2.2 Polynomials in One and Two Variables

In the subsequent discussion, we will be dealing primarily with signals of finite extent. In this case, the corresponding z-transforms are essentially polynomials. Therefore, it is important to understand some properties of polynomials which will be used later on.

A polynomial in one variable z is a function of the form

$$p(z) = \sum_{n=0}^{N} p_n z^n \tag{2.14}$$

where the p_n are complex numbers and N is finite. The degree of a polynomial in one variable (or one-dimensional polynomials) is the largest power to which the indeterminate variable is raised. The degree of a polynomial $p(z)$ will be denoted by $deg(p)$. Thus the polynomial $p(z)$ in (2.14) above has degree $deg(p) = N$.

As a result of the Fundamental Theorem of Algebra [8], all one dimensional polynomials of degree N can always be expressed as a product of N polynomials of degree

1 and a constant,

$$p(z) = p_N \prod_{n=1}^{N} (z - z_n) \tag{2.15}$$

This product representation is unique up to a permutation of the product terms. Since $p(z_n) = 0$ for all n, z_n is called a zero of $p(z)$. We note that z_n will generally be complex, even if the p_n set is real.

A polynomial $p(z)$ with $deg(p) > 0$ is called reducible if it can be expressed as the product of two polynomials $p_1(z)$, $p_2(z)$ with $deg(p_1) > 0$ and $deg(p_2) > 0$, i.e.,

$$p(z) = p_1(z)p_2(z) \tag{2.16}$$

If no such decomposition is possible, then $p(z)$ is called irreducible. From the decomposition presented in (2.15), we see that the only irreducible polynomials in one variable are polynomials of degree 1 [1]. We will see, however, that the situation is quite different for polynomials in two (or more) variables.

Associated with any polynomial is a "mirror" polynomial consisting of coefficients in reversed order and conjugated. For example, for the polynomial $p(z)$ in (2.14), the mirror polynomial $\tilde{p}(z)$ is defined by

$$\tilde{p}(z) = \sum_{n=0}^{N} p^*_{N-n} z^n \tag{2.17}$$

There is a very simple relationship between the zeros of $p(z)$ and $\tilde{p}(z)$; namely, if z_0 is a zero of $p(z)$, then z_0^{*-1} is a zero of $\tilde{p}(z)$.

[1] Strictly speaking, irreducibility is defined with respect to a specific field. Whether a polynomial is irreducible or not may depend on the field of interest. However, in our discussion we will only be dealing with polynomials over the field of complex numbers.

One can also study polynomials in two variables, also called two-dimensional polynomials or bivariate polynomials. In this case there are two variables w, z,

$$p(w,z) = \sum_{m=0}^{M} \sum_{n=0}^{N} p_{m,n} w^m z^n \tag{2.18}$$

where again $p_{m,n}$ are allowed to be complex numbers. The degree in w, $deg_w(p)$, of $p(w,z)$ is the highest order to which the indeterminate w is raised. In the example above, $deg_w(p) = M$. Similarly, the degree in z of $p(w,z)$, $deg_z(p)$, is N. The degree of the polynomial $p(w,z)$ $deg(p)$, is defined by the pair of integers $(deg_w(p), deg_z(p))$; in this case, $deg(p) = (M,N)$. We will also define the total degree of $p(w,z)$, $totdeg(p)$, as the degree of the univariate polynomial $p(w,w)$. We note that in some areas of mathematics, e.g., algebraic geometry, the total degree is considered *the* degree of $p(w,z)$.

A decomposition of an arbitrary bivariate polynomial into a product of polynomials of a lower degree is not always possible. This is because, unlike the case for one-dimensional polynomials, there are irreducible polynomials in two variables for any degree, except of course polynomials which are of degree 0 in one of the variables.

Example: The polynomial of degree (M,N) below is easily checked to be irreducible for any $N > 0$ and $M > 0$ [9],

$$p(w,z) = p_0 + p_1 w + \cdots + p_1 w^M + z^N \tag{2.19}$$

If a product decomposition does exist, then it is essentially unique as expressed in the following theorem [10],

Theorem 2.1 *Every polynomial $f(w,z)$ is expressible as the product*

$$f = P_1 P_2 \cdots P_k \tag{2.20}$$